

## A Note on Demographic Benchmarks

Howard Hogan

July 24, 2021

Besides the operational indicators, the Census Bureau produces two demographic benchmarks to which the Census can be compared. These are known as the Demographic Analysis (DA) estimates and the Population Estimates. Both sets rely on measures of births, deaths and migration to construct alternative population estimates. They differ principally in their base-line starting points. The Population Estimates start with the previous census and project that forward ten years to the time of the current census. The Demographic Analysis estimates use as their base the past 65 years of birth records plus information from administrative records to construct a set of estimates of the population at the time of the current census<sup>1</sup>. Because of this long data series, Demographic Analysis only produces estimates at the national level. The Population Estimates track internal migration over the 10 years and thus can produce sub-national estimates.

These measures are rightly termed benchmarks. Although both of these are constructed independently of the current census, the reverse cannot be said. An important and explicit part of the census post-collection processing consists of comparing the preliminary census results with Population Estimates in order to identify and correct census errors, and the staff doing the post-processing would be aware if their results fell outside the limits estimated by Demographic Analysis.

Although the Census Bureau has produced Demographic Analysis estimates since the 1960 Census, comparisons are difficult as these earlier estimates were based, in part, on the Census results and thus produced only after the Census results were known. The implementation of the American Community Survey provided timely data on immigration which made it possible to prepare the estimates in advance. Previously, the Demographic Analysis acquired this data series from the census 'long form.' Both in 2010 and in 2020, before the final census results were known, the Census Bureau published a set of independent Demographic Analysis estimates, shown here as net percent coverage error, measured as  $\text{Census count} - \text{Estimated True population size}$ .

---

<sup>1</sup> In fact, the DA estimates have been built up over several censuses with each decades estimates using the results from the previous decade's estimates.

	<b>Low</b>	<b>Low Middle</b>	<b>Middle</b>	<b>High Middle</b>	<b>High</b>
2010	1.0	0.4	0.1	-0.4	-1.3
2020		0.2	-0.3	-1.2	

Source: *A Preliminary Analysis of U.S. and State-Level Results From the 2020 Census*, Table 8.

If one accepts, as the Census Bureau argues, that the 2010 Low Middle and High Middle are the most comparable to the 2020 estimates, some conclusions are clear. A plausible estimate is that both censuses produced a small overcount. It is equally plausible that the 2010 Census produced a small undercount, while the 2020 Census produced a somewhat larger undercount.

A note of caution is in order. The Census Bureau now argues:

For the 2020 Census, the count falls between the low and middle series, with the estimate of net coverage for the low series (0.22) suggesting an overcount, and the estimate for the middle series (-0.35) suggesting an undercount. (Hartley et al. page 10)

This is circular logic. The two lower Demographic Analysis estimates are chosen because they are those closest to the census. If one assumes that the census is accurate, then these are the most accurate of the Demographic Analysis estimates. One cannot then conclude that the census must be accurate because it agrees with these two Demographic Analysis estimates. A similar ‘logic’ was used by the Census Bureau to highlight only the one, out of five, Demographic Analysis estimates that most closely agreed with the 2010 Census. There is no point in producing multiple independent estimates only to choose from among them the one or two which make the census look best.

The second benchmark comes from the Population Estimates Program and is properly termed ‘postcensal’ estimates, as it tracks change from the previous census, using measures of births, deaths, international migration and internal movement.

At the end of each decade the postcensal estimates are compared with the next census. The difference is termed the ‘Error of Closure’ (EoC) and has traditionally been used to assess the quality of the postcensal estimates. The difference between the two sets of population measures is comprised of:

- Errors in the previous census
- Errors in the current census

Errors in measuring intercensal change  
Definitional changes between censuses<sup>2</sup>.

If there are no errors in the population estimates, the EoC measures the relative coverage between the two censuses. If the two censuses have equal coverage, the EoC measures errors in estimating population change. Obviously, in almost all cases, the EoC results from both changes in census coverage and errors in measuring population change.

National Errors of Closure: 1990 to 2020

Year	EoC
2020	0.62
2010	0.10
2000	2.48
1990	-0.60

Source: Tables compiled by J. Gregory Robinson.

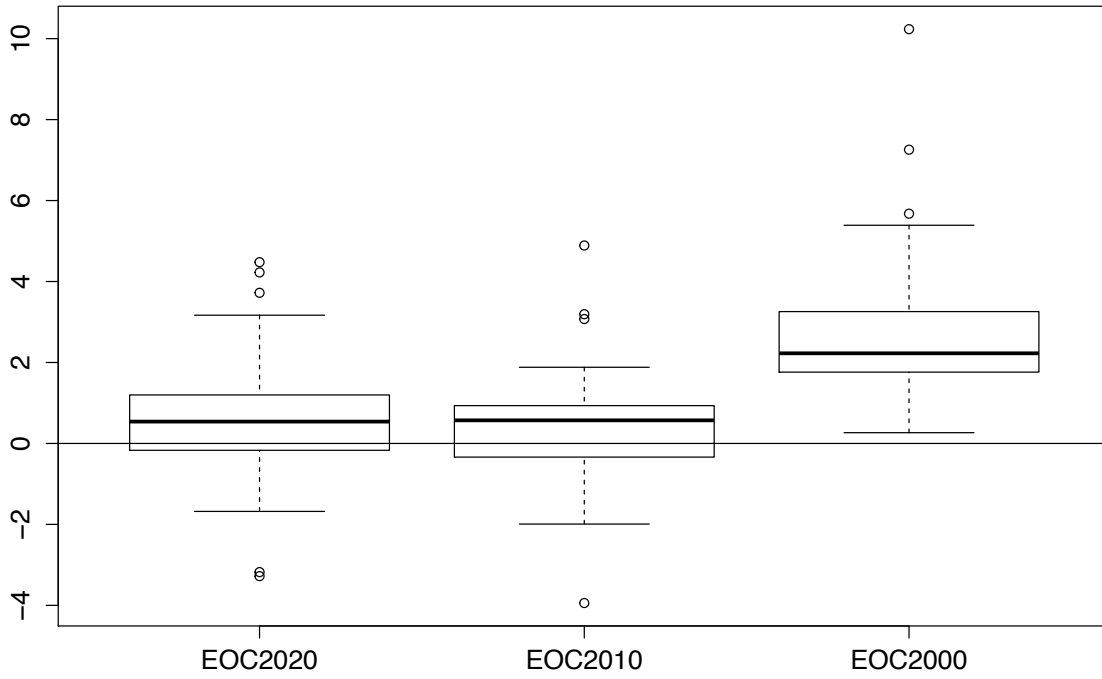
As one can see, the 2020 Census came in slightly higher than the population estimates, although well within the range observed in recent censuses. The outlier here is the 2000 Census, where the population estimates badly underestimated international migration. Again, with the American Community Survey providing timely information on the foreign-born population, the EoC has been small in the two most recent censuses.

Since the postcensal estimates are available subnationally, one can compute state EoC. Figure 1 compares the distribution of the state EoC for 2000, 2010, and 2020. Figure 2 shows the distribution of state EoC by census division for 2020, while Figure 3 shows the same for 2010. The 2020 distribution differs but is comparable to that of 2010. (Source: Tables compiled by J. Gregory Robinson)

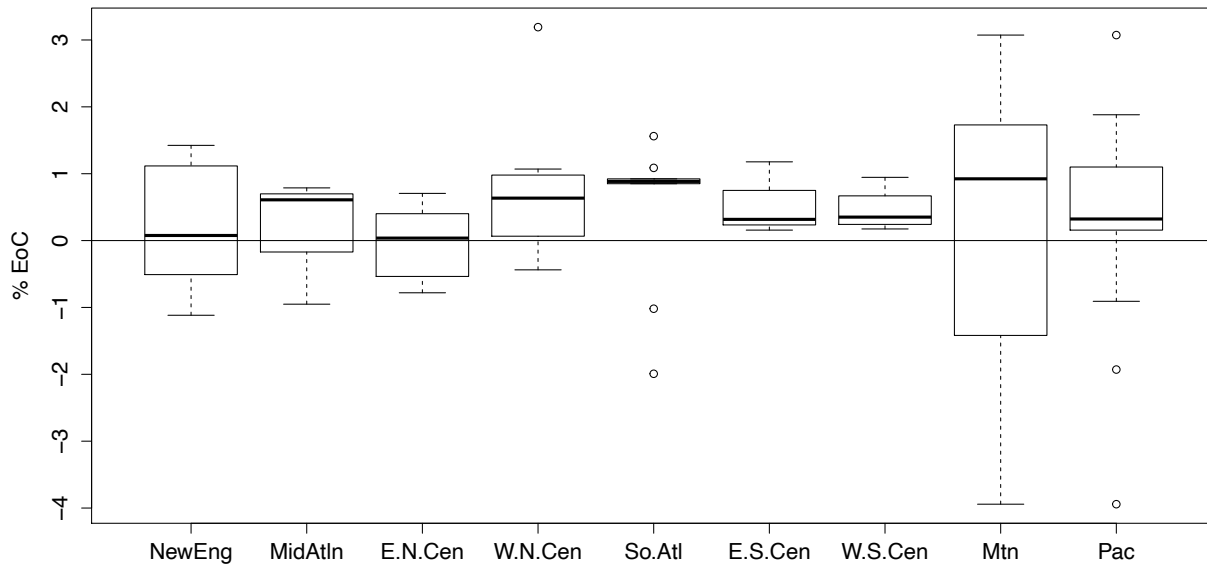
---

<sup>2</sup> The Overseas Federally Affiliated population measured by the census fell from 1,042,523 in 2010 to 350,686 in 2020, a decline of 691,837. It is unclear how much of this change represents net real movement of population, and how much is due to changes in the residence rules.

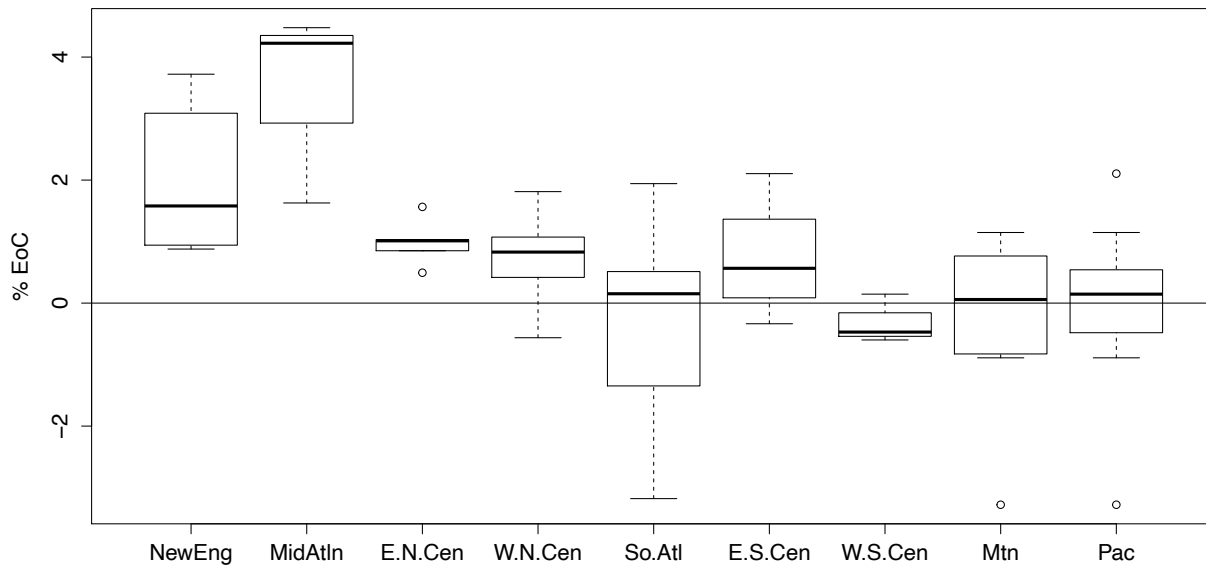
**Figure 1: Distribution of State Error of Closure: 2000 – 2020**



**Figure : Distribution of State Error of Closure by Division 2010**



**Figure :Distribution of State Error of Closure by Division 2020**



Turning attention now to 2020, three aspects stand out. First is the relatively high EoCs for New England and Mid-Atlantic. The next two are the outliers of Arizona and Alabama. It is not so much that these states are extreme, but that their EoCs differ so much from the surrounding states.

Table 2: Two Outlier States

AREA	EoC
Mountain	-0.90
Arizona	-3.28
Rest	0.09
East South Central	0.81
Alabama	2.10
REST	0.37

RECOMMENDATION: That the Census Bureau investigate whether there were special conditions in producing the 2020 population estimates that might have affected these states (Arizona and Alabama) differently from the surrounding states.

Both the Demographic Analysis estimates and the Postcensal estimates produced results that are reasonably near the census results. This should not come as a surprise as postcensal estimates were explicitly used in post-collection census processing to identify and correct errors in the census, while any census result outside the DA range would have been a warning flag. It would be quite alarming if, after this effort, the census still was far from one or both of these external measures. However, at this point, little can be concluded from this general agreement. When census age, race/ethnicity and sex data become available, a more meaningful analysis can take place.