Online Appendix for: Estimating the Civilian Noninstitutional Population for Small Areas

A Modified Cohort Component Approach Using Public Use Data

Andrew C. Forrester*

U.S. Bureau of
Labor Statistics

June 2023

*Forrester: forrester.andrew@bls.gov, Division of Local Area Unemployment Statistics, Office of Employment and Unemployment Statistics, U.S. Bureau of Labor Statistics. All views are my own and do not necessarily reflect those of the U.S. Bureau of Labor Statistics, the Department of Labor, or the United States.

A Comparison with the American Community Survey

One incomplete benchmark for the modified cohort component method is the American Community Survey (ACS), which provides annual, county-level estimates of the CNP16. These data tabulate the CNP16 for a subset of areas through a disability status table, Table S1811. Due to sample size constraints on the annual ACS sample, the 1-year estimates only provide a subset of counties each year, of which a smaller subset of counties have published CNP16 estimates. The ACS 1-year estimates are also produced by weighting respondents using the current year vintage population controls from PEP and are not revised using intercensal adjustments. Consequently, there are a few details to consider when comparing the ACS-derived CNP16 estimates with those produced using cohort components.

A.1 *Caveat Emptor*

Conceptually, each method estimates the same population universe; however, the ACS design and publishing standards complicate any comparison between the two data series. For one, the ACS estimates, by design, conform to independent population control totals developed by the *Population Estimates Program* (PEP) that represent the resident population as of July 1st each year. For years in between the decennial enumerations, the ACS is controlled to the current year postcensal population estimates and only use intercensal estimates in decennial years (U.S. Census Bureau, 2021). For example, the population bases will match for the 2010 census and diverge each year, as the ACS uses Vintage 2011 estimates for the 2011 ACS, Vintage 2012 for the 2012 ACS, and so on. This feature of the ACS design means the ACS estimates will reflect different population controls each year compared to the population estimates used in the modified cohort component method between census years.

The ACS also limits the number of areas published in the 1-year estimates to a population threshold of 65,000 and above. Since the ACS 1-year data only contain CNP16 estimates for around 100 large counties (around 44% of the national CNP16 in the 2021 ACS), county-to-county comparisons will capture the efficacy of the modified cohort component method for only a small fraction of more populous areas where vital statistics are more likely to be complete.

A final caveat to this benchmark is that the 2020 ACS estimates were not published for counties. Due to data collection errors resulting from the COVID-19 pandemic, the U.S. Census Bureau was unable to collect a robust sample for the ACS over the course of the year and instead published only a limited set of state-level estimates, developed using an experimental weighting methodology. I therefore leave out a comparison for 2020 to maintain consistency between the ACS 1-year estimates over time.

A.2 Direct Comparisons

I make two types of comparisons with the ACS — between the percentage and level differences across each method. Specifically, I compare the ACS estimates with July 1st cohort

component estimates. Table 1 summarizes each set of comparisons, while Figure 1 and Figure 2 display each comparison visually.

First, I compare the percentage differences (MAPE and MALPE) between the modified cohort component method and the ACS 1-year estimates over time in Figure 1. In each calculation I assume the ACS 1-year estimate is the "true" CNP16 value. Panel A shows that the the percentage differences between the ACS and the modified cohort component method increase almost monotonically over 2010 to 2019, increasing from a low of 0.4% in 2010 to 1.7% in 2019. It is unclear whether the discrepancy between the two series reflects the difference in ACS survey weights or the increasing number of published counties in the ACS data. Supporting the population controls hypothesis, the differences between the ACS and modified cohort component method subsequently drop to a level of 0.4% in 2021, similar in magnitude to the 2010 estimates closer to the 2010 census. Looking at the direction of differences between the ACS and modified cohort component method in Panel B, I find that the modified cohort component method tends to overshoot the ACS estimates in most years. Again, it is unclear whether this results from the aforementioned difference in population controls using in the underlying ACS estimates or the number of published areas represented in the ACS 1-year data.

Second, Figure 2 shows the level differences between the ACS and modified cohort component method across all matched counties with the associated data in Table 1. The modified cohort component method tracks the ACS well in level terms, both for the unadjusted and raked series. In every year, the differences between the total CNP16 were less than 1%, comparing both the adjusted and unadjusted series.

Across all matched counties from the ACS, the modified cohort component method produces estimates that are nearly identical to those produced by the ACS with the appropriate caveats. While percentage errors between the ACS and the cohort component estimates increased each year between 2010 and 2019, the disparity likely arose from the population controls used to weight ACS respondents. Aggregating all published counties each year, the level and percentage differences are minimal and amount to less than 1% each year. Taken together — and with appropriate caution — the ACS and proposed cohort component method produce virtually indistinguishable results.

B Updating Institutional Prevalence Rates

The most recent group quarters data by facility type, age, and sex were released on May 25, 2023 from the 2020 census as part of the Demographic and Housing Characteristics (DHC) file. These data include the enumeration of the population residing in group quarters by facility type, sex, and age group. To compute institutional prevalence rates using the new 2020 data, I first derive the the institutional and resident military populations by 5-year age groups using the 2020 census data in Tables PC1 (totals), PC2 (institutional), and PC9 (military). I recode the under 20 years old age group to reflect the 15 to 19 age range before computing the prevalence rates, since new military recruits must be at least 17 years old by

law.[1] Using the standard 5-year age-intervalled data, I apply the Beers (1945) formula to compute the 16 to 19 age group.

The next consideration is how to reconcile the difference between the 2010 and 2020 institutional prevalence rates. Since the group quarters population does not change according to the standard demographic components of change, I linearly interpolate the institutional prevalence rates between April 2010 and April 2020 to prevent a discontinuous break in the series when incorporating the new 2020 data. After April 2020 I follow the literature by carrying the institutional prevalence rate forward through the projection horizon, as the standard is to assume no change in the group quarters population unless other data are received (Bryan, 2004).

B.1 Results

Figure 3 compares the age distribution of the total group quarters and institutional and resident military populations between the 2010 and 2020 enumerations. Each chart compares the share of each age group relative to the respective total, i.e., the 20 to 24 year old share of the institutional and military group quarters population. Moving from 2010 to 2020 I find the largest shifts in the institutional and resident military populations was into the ages 65 plus age group. This trend likely reflects general aging in the U.S. population, as institutional group quarters include nursing facilities and hospices.

Next, I examine the effect of integrating the 2020 enumeration-based institutional prevalence rates into the overall CNP16 estimates. Figure 4 compares three series: the CNP16 series produced with the 2010-based institutional prevalence rates (blue), integrating the 2020 rates as described above (red), and with the final raked estimates to the independent CNP16 series from the U.S. Bureau of Labor Statistics. Both over the historical and projected periods, both methods produce nearly identical estimates, with the 2010-based institutional prevalence rates producing slightly larger estimates at the national level.

[1] 10 U.S.C. 505.(a) requires that new enlistments must be at least 17 years of age.

C Additional Figures

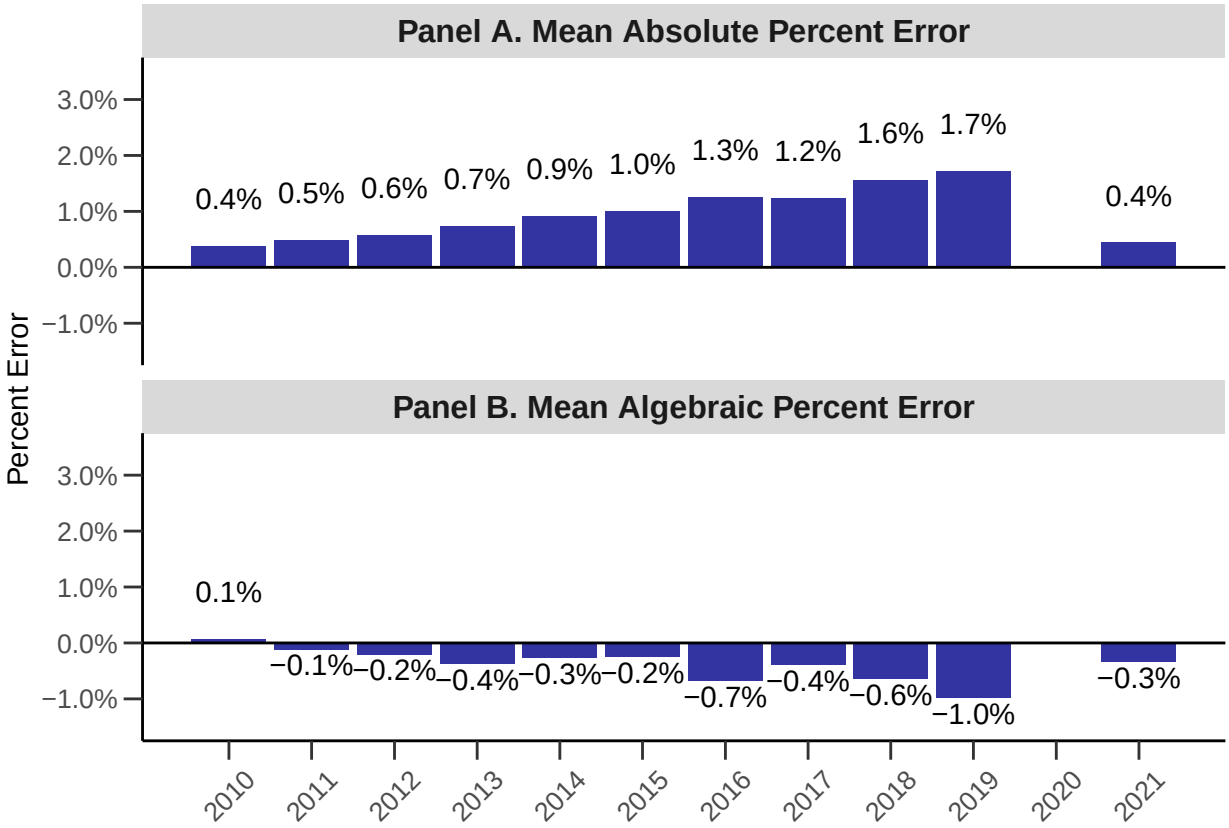

Figure A 1: Percent Errors Between CNP16 from the ACS and Cohort Components

Notes: Figure shows the MAPE and MALPE between the CNP16 from the ACS 1-year Table S1811 and the modified cohort component estimates, raked to the state CNP16 totals from the Current Population Survey (CPS), as of July 1st each year. ACS data are weighted based on the current year vintage population estimates and do not include an intercensal adjustment. 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

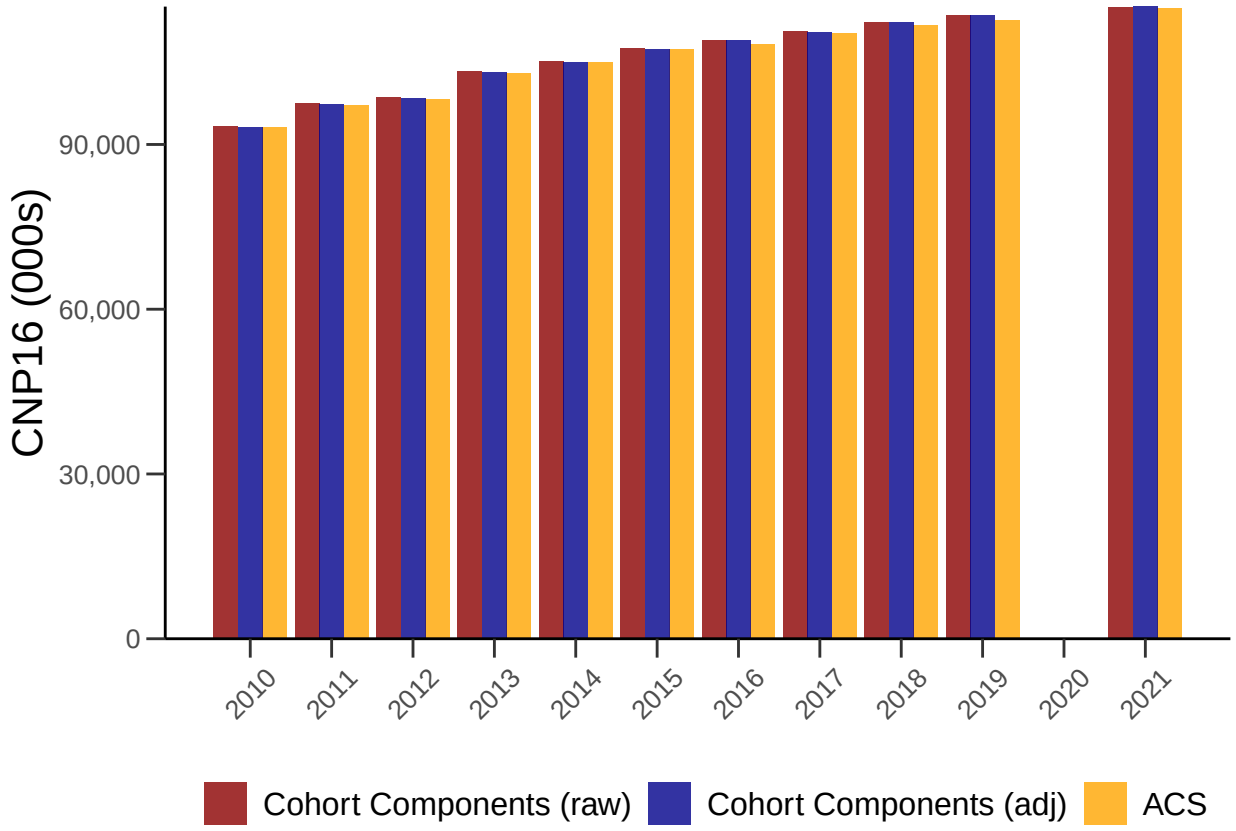

Figure A 2: Level Differences Between CNP16 from the ACS and Cohort Components

Notes: Figure shows the level differences between the CNP16 from the ACS 1-year Table S1811 and the modified cohort component estimates, raked to the state CNP16 totals from the Current Population Survey (CPS), as of July 1st each year. ACS data are weighted based on the current year vintage population estimates and do not include an intercensal adjustment. 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

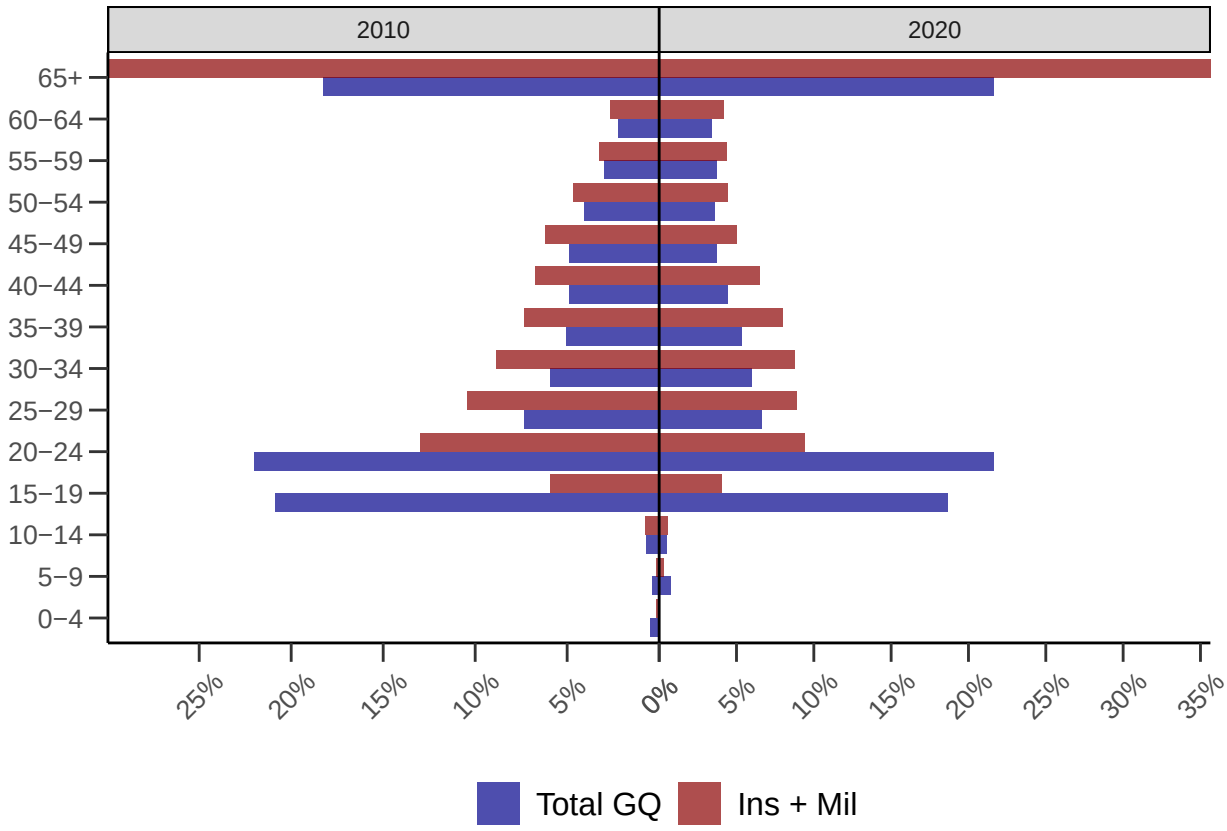

Figure A 3: Group Quarters Population Population Share for the Total and Institutional Plus Resident Military

Notes: Figure shows the share of the total group quarters and institutional and resident military population in each each group in the 2010 and 2020 censuses. Data from the 2010 census are from Summary File 1 and data from the 2020 census are from the Demographic and Housing Characteristics (DHC) file.

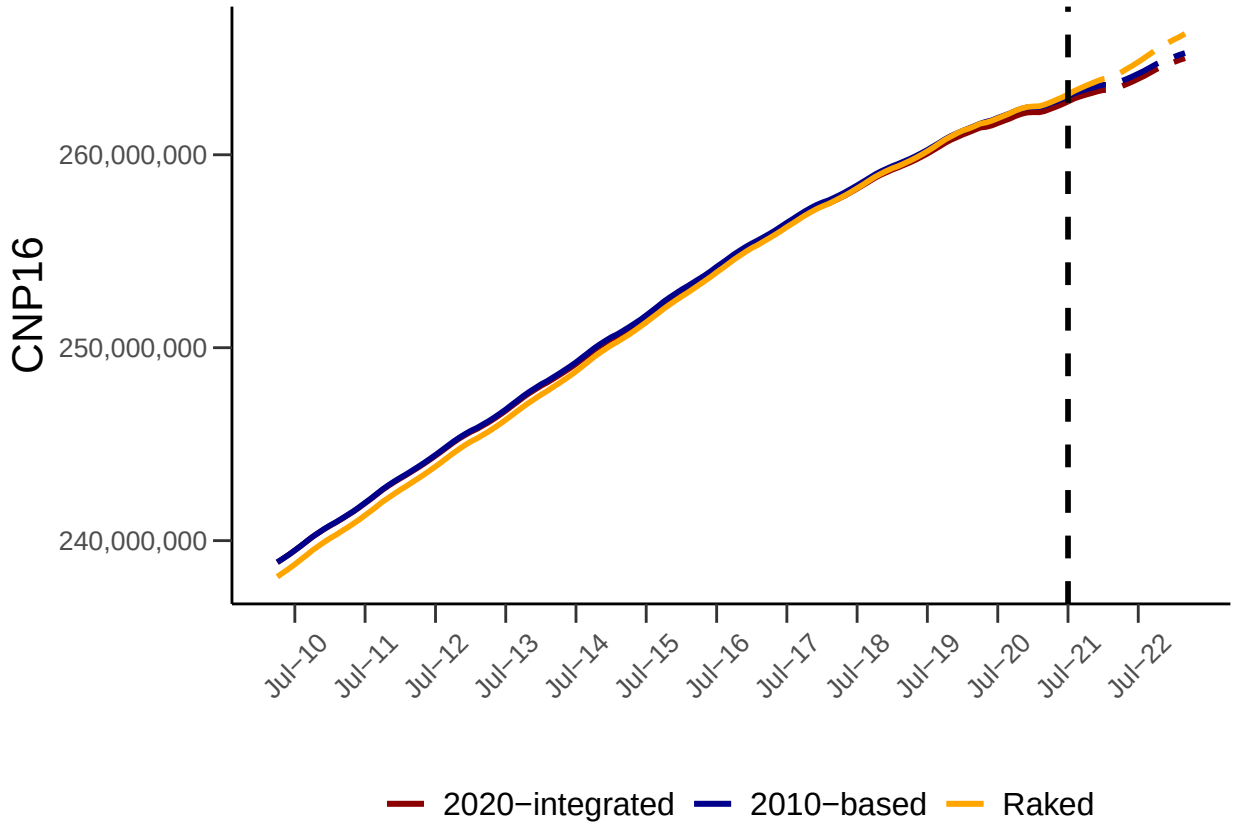

Figure A 4: Effect of Using 2010-Based Institutional Prevalence Rates v. 2010 and 2020 Rates

Notes: Figure shows the civilian noninstitutional population estimated using institutional-prevalence rates derived from the 2010 census alone and interpolating the 2010 and 2020 rates. Data from the 2010 census are from Summary File 1 and data from the 2020 census are from the Demographic and Housing Characteristics (DHC) file.

D Additional Tables

Table A 1: Comparing CNP16 with the ACS

Year	Counties	ACS (1yr)	CCM (raw)	CCM (adj)	MAPE (raw)	MAPE (adj)
2010	89	93,167,959	93,380,301	93,095,865	0.23%	0.08%
2011	93	97,197,163	97,509,156	97,252,892	0.32%	0.06%
2012	93	98,233,105	98,579,072	98,356,332	0.35%	0.13%
2013	99	102,968,137	103,444,327	103,238,132	0.46%	0.26%
2014	100	104,933,562	105,215,158	105,040,723	0.27%	0.10%
2015	102	107,256,213	107,439,775	107,323,367	0.17%	0.06%
2016	103	108,310,428	108,972,663	108,898,032	0.61%	0.54%
2017	104	110,266,617	110,533,402	110,489,834	0.24%	0.20%
2018	106	111,643,863	112,183,781	112,187,494	0.48%	0.48%
2019	107	112,659,748	113,522,413	113,575,043	0.76%	0.81%
2021	108	114,730,408	114,937,952	115,091,195	0.18%	0.31%

Notes: Table shows the differences in aggregate CNP16 from the ACS 1-year Table S1811 and the raw and adjusted cohort component estimates as of July 1st each year. The second column shows the number of published counties matched in the ACS 1-year data. Adjusted data are raked to the statewide CNP16 data from the Current Population Survey (CPS). 2020 ACS data for counties were not published due to data collection issues resulting from the COVID-19 pandemic.

E References

Beers, H. S. (1945). Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries*, *33*(2), 245–260.

Bryan, T. (2004). *The Materials and Methods of Demography* (J. S. Siegel & D. A. Swanson, Eds.; 2nd ed.). Elsevier.

U.S. Census Bureau. (2021). *Population controls for the 2020 ACS*. U.S. Census Bureau; 2010 Census of Population and Housing. https://www.census.gov/programs-surveys/acs/technical-documentation/user-notes/2021-01.html