Civilian Non-Institutional Population Estimates for Counties and Equivalents

Andrew C. Forrester

Division of Local Area Unemployment Statistics
U.S. Bureau of Labor Statistics

FCSM 2024
Oct. 22, 2024

Disclaimer

The views expressed here do not necessarily reflect the positions or policies of the U.S. Bureau of Labor Statistics, the Department of Labor, or the United States.

Background (1 of 2)

The *civilian non-institutional population ages 16 and over* (CNP16) is the base population for labor force statistics.

- Describes the population likely engaged in the civilian labor force and forms basis for employment-population ratios and labor force participation rates.
- Defines the sampling frame for the Current Population Survey (CPS).

U.S. Census Bureau produces monthly *statewide* CNP16 estimates, although substate geographies not available.

- **This work: produce CNP16 estimates for counties and county equivalents in the U.S. and Puerto Rico.**

Background (2 of 2)

Objective: Develop monthly estimates of the CNP16 by county/equivalent from April 2000 to the present.

- Utilize standard demographic cohort component modelling and public use data which accounts for population change through births, deaths, and net migration (Preston et al., 2000; Bryan, 2004).
- Ensure CNP16 series are linked *temporally* and *geographically*.
- Develop CNP16 projections for monthly data production.
- Evaluate potential for labor force rates and ratios research series.

A quick note on terminology:

- *Postcensal estimates* are extrapolations/projections from a prior census.
- *Intercensal estimates* are estimates linked between two censuses.

Current Methods

Applied demographic methods to produce monthly estimates of the CNP16 by county/equivalent in the U.S. and Puerto Rico.

- Forrester (2024b) applied a modified cohort component method to estimate CNP16 using publicly available data from the census, official population estimates, and vital statistics.
  - Data sourced directly from U.S. Census Bureau Population Estimates Program (PEP) and National Center for Health Statistics (NCHS).
  - Account for changes to geographic definitions over time, i.e., changes to county boundaries.
  - Develop monthly CNP16 projections through a production year.
- Separate method developed for Puerto Rico, adapting to Puerto Rico's vital statistics data (Forrester, 2024a).

Civilian Non-Institutional Population

The *civilian non-institutional population 16 plus* (CNP16) is defined by the resident population less the institutional group quarters population and active-duty armed forces:

$$\text{CNP16}_{i,t} = \text{RES16}_{i,t} - \text{GQINS}_{i,t} - \text{AFMIL}_{i,t}. \tag{1}$$

Where:

- $\text{RES16}_{i,t}$: resident population ages 16 and over
- $\text{GQINS}_{i,t}$: institutionalized group quarters population 16 and over
- $\text{AFMIL}_{i,t}$: active-duty armed forces

To estimate the CNP16, first need estimates of the *resident population*, then subtract out GQ and military.

Cohort Component Method

Estimates for the *resident population* use the cohort component method for area i in month t, adapted for aging into the reference population (ages 16 and over).

$$\text{RES16}_{i,t+1} = \text{RES16}_{i,t} + \text{AGING}_{i,t} - \text{DEATHS}_{i,t} + \text{NETMIG}_{i,t} \quad (2)$$

Where:

- $\text{RES16}_{i,t}$: resident population ages 16 and over
- $\text{AGING}_{i,t}$: Population aging from 15 to 16
- $\text{DEATHS}_{i,t}$: Deaths ages 16 and over
- $\text{NETMIG}_{i,t}$: Net migration ages 16 and over

The following discussion explains each input process.

Aging

Approximate month-to-month aging estimated by taking the lagged distribution of births in an area i, lagged 16 years (proxy for birth months).

- Let y indicate an observation year ($t \in y$), aging from 15 to 16 over the year is approximated by

$$\text{AGING}_{i,t} = \text{AGE15}_{i,t} \times \left(\frac{\text{BIRTHS}_{i,(y-16),t}}{\sum_{t \in (y-16)} \text{BIRTHS}_{i,y,t}} \right), \tag{3}$$

where:

- $\text{AGE15}_{i,t}$: Resident population age 15
- $\text{BIRTHS}_{i,y,t}$: Births in month t of year y
- $\text{AGING}_{i,t}$: Estimated aging from 15 to 16

In the projection model the population age 15 is moved into age 16 group until exhausted, then switch to age 14.

Mortality

Decedents ages 16 and over available by county/equivalent and month from NCHS.

- Data disclosure rules limit tabular data to county/month counts of either > 10 or zero.
- Impute missing months by allocating difference between annual totals and published months.

For projections, use combination of provisional data and time series forecasts.

- Provisional mortality data usually available through the previous calendar month.
- Forecast remaining mortality counts using county-specific seasonal ARIMA models in line with Hauer (2019).

Net Migration

Net migration more difficult to measure at the county level, apply PEP net migration rates prorated to months and applied to aged population.

$$\text{NETMIG}_{i,t} = (\text{RES16}_{i,t} + \text{AGING}_{i,t}) \times \left(\frac{\text{NMR}_{i,y}}{k} \right), \qquad (4)$$

$$k \in \{3, 12\}.$$

Where:

- $\text{NETMIG}_{i,t}$: Level of net migration ages 16 and over
- $\text{RES16}_{i,t}$: Resident population 16 and over
- $\text{AGING}_{i,t}$: Aging into age 16 and over age group
- $\text{NMR}_{i,y}$: Total net migration rate, either annual or quarterly
- k: pro-rata factor for quarterly (3) or annual (12) PEP

Projected net migration carries most recent PEP NMR forward.

Group Quarters Population (1 of 4)

Need to remove the *institutionalized GQ* and *military* population from the resident population each month.

- GQ populations evolve differently than household population (Bryan, 2004).
- Annual GQ data from PEP only cover total GQ; decennial data cover age distribution and facility type.
- GQ totals adjusted for error of closure using Das Gupta (1981) method.

Solution: apply a *prevalence rate* to allocate GQ total by age and facility type (Land and Hough, 1986).

- Represents *share* of the total GQ ages 16 and over who reside in institutional or military GQ.

Group Quarters Population (2 of 4)

Using decennial data by area i and month t, compute prevalence rates as:

$$\text{INSRATE}_{i,t} = \frac{\text{GQINS16}_{i,t} + \text{GQMIL}_{i,t}}{\text{GQTOTAL}_{i,t}}, \quad t \in \{Apr.\ 2000, Apr.\ 2010,\ Apr.2020\}. \tag{5}$$

Where:

- $\text{INSRATE}_{i,t}$: Share of institutional and military ages 16 and over in the total GQ population
- $\text{GQTOTAL}_{i,t}$: Total GQ population
- $\text{GQINS16}_{i,t}$: Institutional GQ population ages 16 and over
- $\text{GQMIL}_{i,t}$: Military GQ population

Group Quarters Population (3 of 4)

There are a few caveats to note with the GQ population:

- Single years of age interpolated for institutionalized GQ using the Beers (1945) osculatory formula.
  - Public county-level GQ data only available in 5-year age ranges.
- Military GQ used as a proxy for active-duty armed forces, research shows data line up well with military base locations (Forrester, 2024b).
- Prevalence rates linearly interpolated between censuses and extrapolated after (Forrester, 2024b; Bryan, 2004).
  - PEP uses annual GQ report to obtain time series variation, otherwise carry forward census GQ population.

Group Quarters Population (4 of 4)

Monthly GQ adjustment obtained by multiplying prevalence rate by PEP total GQ population.

$$\text{GQADJ16}_{i,t} = \text{INSRATE}_{i,t} \times \text{GQESTIMATE}_{i,t}, \tag{6}$$

where:

- $\text{GQADJ16}_{i,t}$: Institutional and military GQ ages 16 and over
- $\text{INSRATE}_{i,t}$: Institutional and military ages 16 and over share of total GQ population
- $\text{GQESTIMATE}_{i,t}$: Total annual GQ population

Statewide Population Controls (1 of 2)

Next, obtain an initial area-level CNP16 estimate by subtracting the GQ adjustment from the resident population 16 and over.

$$\text{CNP16PRE}_{i,t} = \text{RES16}_{i,t} - \text{GQADJ16}_{i,t}, \tag{7}$$

where:

- $\text{CNP16PRE}_{i,t}$: Initial, uncontrolled CNP16
- $\text{RES16}_{i,t}$: Resident population ages 16 and over
- $\text{GQADJ16}_{i,t}$: Institutional and military GQ ages 16 and over

To the extent that components of change or GQ contains measurement error, can apply independent population controls...

Statewide Population Controls (1 of 2)

Official CNP16 from the BLS provide an independent population control and allow raking to the monthly statewide (or Commonwealth) population control for each state s:

$$\text{CNP16FIN}_{i,t} = \text{CNP16PRE}_{i,t} \times \left(\frac{\text{CNP16CON}_{s,t}}{\sum_{i \in s} \text{CNP16PRE}_{i,t}} \right), \quad \forall\, i, t, \quad (8)$$

where:

- $\text{CNP16FIN}_{i,t}$: Final, controlled CNP16
- $\text{CNP16PRE}_{i,t}$: Initial, controlled CNP16
- $\text{CNP16CON}_{i,t}$: Statewide CNP16 control total

Two desirable features:

- Ensures additivity to the statewide (or Commonwealth) estimates.
- Reduces measurement error in estimates due to errors in data inputs.

Data Overview

Utilize official, public-use **demographic data** and **vital statistics** for the U.S. and Puerto Rico.

- Intercensal and postcensal resident population by age from Population Estimates Program (PEP).
- Vital events (births and deaths) from NCHS or Puerto Rico *Dpto. de Salud*.
- Net migration levels and rates from PEP.
- Postcensal GQ population total by year from PEP.
- Census GQ population by age and facility type from decennial censuses.
- Population controls from BLS or *Dpto. de Trabajo y Recursos Humanos*

Before entering data inputs into cohort component framework, some pre-processing steps needed...

Data Processing

Recall: need consistent series over *time* and *space*.

- *Time Series*. Use the Das Gupta (1981) method to link postcensal resident and GQ population to next census.
  - Needed to reconcile Vintage 2020 postcensal estimate to Vintage 2023 blended base (error of closure).
- *Geographic*. Incorporate county boundary changes over the 2000s.
  - Extensive changes to Alaska census areas, Connecticut replaced counties with planning regions.
  - Use census disaggregations to split geographies.

Resulting data are consistent population and components of change data for $N = 3,222$ counties and equivalents in the U.S. and Puerto Rico spanning April 2000 through December 2023.

Data and Variable Lists (1 of 2)

Table: Input Data

Variable ID	Description	Source	Frequency	Availability
RES16	Resident population ages 16 and over	PEP	A	2000-2023
AGE15	Resident population age 15	PEP	A	2000-2023
DEATHS	Deaths ages 16 and over	NCHS	M	2000-2024
BIRTHS	Births (lagged)	NCHS	M	1984-2023
NMR	Net Migration Rate per 1,000	PEP	A	2000-2023
GQTOTAL	Total GQ	PEP	A	2000-2023
GQINS	Institutional GQ by age	Census	D	2000-2020
GQMIL	Military GQ	Census	D	2000-2020

- PEP: Population Estimates Program, U.S. Census Bureau
- NCHS: National Center for Health Statistics
- A: Annual; M: Monthly; D: Decennial

Data and Variable Lists (2 of 2)

Table: Intermediate and Output Data

ID	Description	Frequency
AGING	Aging into 16+ Age Group	M
NETMIG	Net Migration Ages 16+	M
INSRATE	Institutional and military GQ Prev. Rate	M
GQADJ16	Institutional and Military GQ Level	M
CNP16CON	Control Civilian Non-Institutional pop. 16+	M
CNP16PRE	Prelim Civilian Non-Institutional pop. 16+	M
CNP16FIN	Final Civilian Non-Institutional pop. 16+	M

- A: Annual; M: Monthly; D: Decennial

Discussion

Combining public use data across the federal statistical system enables reliable estimates of the *civilian non-institutional population ages 16 and over* for substate geographies.

- Cohort component method using public data produces reliable results of official/equivalent county resident population and statewide CNP16.
- Able to produce geographically consistent estimates back to April 2000.

Timeliness of data availability permits data production in real-time.

- Vital statistics and forecasting allows current-year production.
- Official population controls from BLS and Puerto Rico Dept. of Labor available for raking current month estimates.

Further Work

Ongoing research to evaluate various refinements to data inputs for the cohort component method...

- Impact of differentially private census data on GQ prevalence rates and geographic adjustments.
- Examining the GQ military assumption and incorporating data from the Dept. of Defense and the American Community Survey.
- Testing more complex forecasting models for area-level mortality beyond seasonal ARIMA models.

...and developing labor force rates and ratios as research series using official LAUS data

- Developing employment-population ratios and labor force participation rates on a research basis.

Thank you!

References I

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). Population Estimates. In Siegel, J. S. and Swanson, D. A., editors, *The Materials and Methods of Demography*. Elsevier, 2 edition.

Das Gupta, P. (1981). Intercensal Estimates for the States [Memorandum]. Technical report, U.S. Census Bureau, Washington, DC: U.S. Department of Commerce.

Forrester, A. C. (2024a). Civilian Non-Institutional Population Estimates for Puerto Rico Municipios Using A Modified Cohort Component Method. Working Paper.

Forrester, A. C. (2024b). Estimating Monthly Area Population with Public Use Data: A Cohort Component Approach. *Journal of Population Research*, 41(5).

References II

Hauer, M. E. (2019). Population Projections for U.S. Counties by Age, Sex, and Race Controlled to Shared Socioeconomic Pathway. *Scientific Data*, 6:190005.

Land, K. C. and Hough, G. C. (1986). Improving the Accuracy of Intercensal Estimates and Postcensal Projections of the Civilian Noninstitutional Population. *Journal of the American Statistical Association*, 81(393):62–74.

Preston, S., Heuveline, P., and Guillot, M. (2000). *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell.

Contact Information

Andrew C. Forrester
Economist and Lead Demographer

Division of Local Area Unemployment Statistics

U.S. Bureau of Labor Statistics
forrester.andrew@bls.gov