EXTENDED ABSTRACT

Estimating the Population of Military Counties in the United States, 2000-2020

Alexa Jones-Puthoff
Population Division
U.S. Census Bureau

Irene Dokko
Population Division
U.S. Census Bureau

For submission as a paper presentation to the annual meeting of the Population Association of America, Boston, MA May 1-3, 2014

BACKGROUND

In the years between decennial censuses, the U.S. Census Bureau produces population estimates for the nation, states, and counties by demographic characteristics. They are considered "estimates" because they are not based solely on a census; they are derived by using available data (for example, vital statistics data on births and deaths) in conjunction with census counts to measure change in the population of the United States since the last census. The population estimates serve as the official estimates between censuses, and as such are used for many important purposes, including informing the distribution of federal funds, serving as population controls for several major surveys (including the American Community Survey and the Current Population Survey), providing denominators for statistical rates, and helping federal, state, and local leaders plan for their communities' changing needs.

The Population Estimates Program utilizes a cohort component method to create population estimates for the nation, states, and counties. Fundamentally, we begin with the most recent census and estimate forward year by year, adding births, subtracting deaths, and adding in estimates of net international migration (at the national, state, and county levels) and domestic migration (at the state and county levels). Estimates of the Armed Forces population—both stock (the number, location, and characteristics of the Armed Forces population), and flow (the movement of the Armed Forces population between the 50 states and the District of Columbia and overseas)—are part of the calculation of the estimates at the national, state, and county levels.

While vital statistics records on births and deaths are considered to be complete and accurate, estimating the size and movement of populations with irregular migratory patterns (such as the Armed Forces population) is a more challenging proposition. Several aspects of the 2000-2010 time period complicated efforts to accurately estimate the size and characteristics of counties with substantial

military populations, including the wars in Iraq and Afghanistan and implementation of the 2005 Base Realignment and Closure Commission's recommendations.

This paper focuses on the county-level resident population estimates for two groups of counties: "military" counties, with substantial Armed Forces populations, and "non-military" counties, with smaller (or negligible) Armed Forces populations. This work explores the Census Bureau's population estimates for these two types of counties, and considers the estimates' accuracy relative to 2010 Census counts for these groups of counties.

ESTIMATES EVALUATION METHODOLOGY

This work is part of a broader project designed to evaluate the performance of the methods used to produce population estimates for the 2000-2010 time period. Given that the Population Estimates Program's ultimate goal is to accurately estimate the population in the period between censuses, our primary measure of success in this endeavor is accuracy of the estimate for a census date relative to the census counts on that same date. In the context of evaluating the 2000-2010 postcensal population estimates, we define accuracy as the degree of closeness of the (2000-based) population estimates for April 2010 to the 2010 Census counts (which also have a reference date of April 2010).

We have developed a series of "measures of accuracy," the results of which provide an accuracy profile for a set of population estimates, in this case "military" and "non-military" counties. The measures of accuracy utilized in this work are described below. Each measure was selected prior to beginning this analysis, and they provide information on four key criteria identified in a Committee on National Statistics report. In this paper, the "evaluation geographies" are counties, and the analysis is done separately for the two groups of counties (military and non-military).

Mean Absolute Percent Error (MAPE) = ((∑ ((| Estimate – Census |)/ Census))/N)*100

This measure takes the absolute value of the difference between the estimate and the census value for each evaluation geography, divides that by each respective census value, sums them, divides by the number of evaluation geographies, and multiplies the result by 100. The goal is to provide a relative measure of error. It ranges from zero to positive infinity and represents the average error across cases, regardless of sign. This is one of the most commonly used measures for assessing the accuracy of a series of estimates.

Mean Algebraic Percent Error (MALPE) = ((∑((Estimate – Census)/Census))/N)*100

Similar to the MAPE, this measure takes the difference between the estimate and the census value for each evaluation geography, divides that by each respective census value, sums them, divides by the number of evaluation geographies, and multiplies the result by 100. The goal is to identify systematic bias and provide an additional relative measure of error. Its main value is that it preserves the sign of the error, allowing us to assess whether the estimates were generally higher or lower than the census count.

Root Mean Squared Error (RMSE) = SQRT(∑ ((Estimate – Census)^2)/N)

This measure squares the difference between the estimate and the census value for each evaluation geography, sums these values across evaluation geographies, divides by the number of evaluation geographies, and finds the square root of this value. It presents an alternative measure that places greater emphasis on large numeric errors versus mean absolute errors.

Percent Difference Threshold = Number of areas with differences above/below a certain threshold

Unlike the other measures, this is a numeric value that relies upon an arbitrarily set threshold (e.g., 5 and 10 percent). The percent difference is computed by dividing the difference between the estimate and census value for a given area by the census value for that area and multiplying by 100. The end measure is reported for each summary area (in this case, military and non-military county groups), and is a count of how many evaluation geographies had absolute percent differences that exceeded the specified percent difference threshold.

Total Absolute Error of Shares = $\sum \left| ((\text{Estimate}/\sum\text{Estimate}) - (\text{Census}/\sum\text{Census})) \right|$

This measure finds the proportion of each estimate to the total estimate for the summary geography (in this case, groups of counties [military/non-military]) and subtracts the proportion of the census value to the total census value for the summary geography. The absolute value of these proportional differences across evaluation geographies is then summed to the summary geography level. The goal is to provide a measure of the distributional error in the estimated shares.

RESULTS

The results of the estimates evaluation of "military" and "non-military" counties are shown in terms of the individual measures of accuracy detailed above. Collectively, they provide a profile of accuracy for these two sets of counties. Preliminary results indicate reduced accuracy of the 2010 population estimates for counties with substantial Armed Forces populations relative to counties with smaller (or negligible) Armed Forces populations, when compared to 2010 Census counts.

Following the discussion of the accuracy of the population estimates for these two groups of counties, we explore potential methodological improvements to the estimation of stateside military populations and the movement of the Armed Forces between the United States and overseas. Several recommendations are made which, if implemented, may assist the Census Bureau in more accurately estimating military counties in the 2010-2020 time period.