The Census Bureau's Use of the Vital Statistics

Ben Bolender – Assistant Division Chief: Population Estimates and Projections, Population Division

Overview

The U.S. Census Bureau's Population Estimates and Projections program (PEP) creates annual population estimates for the nation, states, counties, cities, and towns and include varying levels of detail by age, sex, race, and Hispanic origin. The population estimates are used to allocate hundreds of billions of federal dollars annually both directly or indirectly through statistics that use them as controls or denominators.

For example, the estimates are used as controls to generate the premier socioeconomic indicators from the American Community Survey (ACS), the unemployment rates from the Current Population Survey (CPS), the population covered by health insurance, rates of cancer and disease by race and ethnicity from the National Center for Health Statistics (NCHS), crime and victimization rates from the National Crime Victimization Survey (NCVS), and many others. Further, the estimates of the population age 18 or over are delivered to the federal organizations such as the Federal Election Commission, the Department of Defense, and the Department of Justice to facilitate the enforcement of laws and regulations. Finally, the population estimates are released publicly. They inform researchers, policy makers, and private industry about the growth, demographic characteristics, and the geographic distribution of the nation's population allowing for more accurate decision-making.

To generate population estimates, we rely primarily on administrative records to measure population change between decennial censuses. We obtain data on births, deaths, and migration then use a cohort component method to calculate how the population grows over time. The national vital statistics system forms a core component of our population estimates. While the movement of population from one place to another is inherently difficult to estimate, the vital statistics are generally seen as a reliable component. Our data on vital statistics is derived from three sources: the National Center for Health Statistics (NCHS); the Federal-State Cooperative for Population Estimates (FSCPE); and data internal to our program.

We obtain the bulk of the data annually from the NCHS. These data have been cleaned and processed to account for residence/occurrence issues, cross-state events, and other quality concerns. The NCHS data represent full demographic and geographic detail for two years prior to the final year of our series (or "vintage") of estimates. We also use NCHS national-level preliminary or provisional totals that to help control our short range projections.

The second major source is our FSCPE members. These are generally state demographers or other researchers designated by the governor to partner with us to provide data for, review of, and input on our population estimates. Data from the FSCPE mostly takes the form of county level distributions of births and deaths within their state. Often the data come indirectly from state health departments or other similar organizations and are able to provide a more nuanced distribution for the year prior to the vintage year.

The final source of vital statistics data is internal to our program. We generate additional demographic detail and the short range projections that allow us to estimate the vintage year itself and a short distance into the future (for particular surveys like CPS or for the Population Clock). To generate race detail for deaths, we build upon research by NCHS around 2000 to convert the four races reported by

many states until recently into the 31 races used in the estimates. We use a similar process for parents, then utilize decennial census data to generate probability distributions for the resulting race of the births. To project births and deaths into the vintage year and beyond, we calculate internal vital rates based on the most recent NCHS data and the latest population estimates.

In addition to the population estimates program, the Census Bureau also uses vital statistics for a number of other projects. Within PEP, we perform Demographic Analysis (DA) every 10 years as one of only two major coverage measures for the decennial census (the other being the post-enumeration survey). Essentially, DA starts with a population of zero and adds historic records of births, deaths, and migration to reach a population estimate that is almost entirely independent of the decennial census. Across the Census, there are other uses of vital records as well, although they tend to be primarily for benchmarking.

Complications and Opportunities for Improvement

The vital statistics system is central to the Census Bureau's production of high quality population estimates and projections. The previous discussion highlights two areas of difficulty that could be addressed with future improvements. The first deals with race coding. We spend a lot of time and effort converting vital statistics data from the Office of Management and Budget (OMB) 1977 race categories into the 1997 categories, then reviewing the results to make sure the distributions are demographically reasonable. We also convert the resulting estimates back into the 1977 race categories for use as controls and denominators for health related research. Only in 2016 did all states report in the 1997 categories. We are currently looking at ways to incorporate these new data into our production process. The second major concern revolves around the timeliness of the data collection. The primary reason we use inputs provided by our state partners is to get more recent geographic distributions for use in our projections.

Given these concerns, we would argue that moving to a more centralized, online or electronic reporting system could have several advantages. Currently, it is a monumental and costly undertaking to make a change to the vital reporting system. This includes changes not just to state forms, reporting, and record keeping systems, but also to private solutions developed for hospitals and other health workers. With OMB again examining the standards on race and ethnicity, moving toward a more centralized system could offer significant cost savings. The initial costs of a centralized system may be high, but the costs of trying to organize and consolidate a plethora of smaller systems as we navigate ever-changing definitions of racial and ethnic identity could be much higher in the end, both in economic and social terms.

In addition, electronic entry and storage may allow for improved timeliness in processing. Reducing the two-year time lag on final data would allow us to reduce the amount of projection needed to estimate the current population and may eliminate the need to request data from our state partners. Not only could the Census Bureau and other federal agencies benefit from more recent data, but it is also likely that states and reporting districts could see benefits. Automated data entry and access to the system may allow them to more quickly and easily clean their own data, sort out cross-district events, and compare their trends to other areas.