MAPPING AMERICA'S FUTURES, BRIEF 5

Methodology and Assumptions for the Mapping America's Futures Project

Kaitlin Franks Hildner, Austin Nichols, and Steven Martin

January 2015

The Mapping America's Futures project has developed multiple series of population projections for 740 commuting zones in the United States by age and race and ethnicity. This brief explains the assumptions and methodology of our population projections.

Project Overview

The "Mapping America's Futures: Population" tool produces a series of population projections that provide considerable local geographic and demographic detail for the entire United States. To do this, we project populations by age and race and ethnicity at the commuting zone level. Commuting zones are a spatial measure of a local labor market consisting of one or more counties or county equivalents. The tool allows users to explore both the population size and demographic make-up of each commuting zone. Our objective is to help the user explore what America's future *could be*, considering alternative scenarios. The possible scenarios we project are nevertheless constrained by the inertia of demographic processes that range from unlikely to extremely unlikely to change their momentum.

Detailed Population Projections

To provide population projections at a local level across the United States, we use the commuting zone. As a collection of counties or county equivalents that often cross state lines, a commuting zone is not a political entity. Instead, it is a measure of a local labor market, identifying an area of shared economic activity and social activity. Though the counties within a commuting zone might not follow identical trends in birth, mortality, or migration rates, the commuting zone's demographic trajectory can plausibly be projected as the joint development of a single local economy (see Tolbert and Sizer 1996). Hence, use of the commuting zone as a unit of analysis makes sense for population projections. Further,

the projections generated for these commuting zones can then be applied to the discussion of economic outcomes (such as housing markets) as well as social or other outcomes, where the commuting zone also serves as a natural unit of social analysis.

For each of the 740 commuting zones in our analysis, we develop projections for the population in the years 2020 and 2030. In addition to overall population counts, the projection series provides access to separate counts for Hispanics, non-Hispanic blacks, non-Hispanic whites, and other non-Hispanic races, and for five-year age categories ranging from birth to 4 years through 85 years and older.

The Cohort-Component Method

Similar to the method the US Census Bureau uses to create national projections, these projections were created using a cohort-component method, which applies three components of demographic change—birth, survival, and migration rates—to a population and then ages that population (see US Census Bureau 2013). The idea is that a 27-year-old today, who continues to live in the same place, should be counted as a 37-year-old 10 years from now. Because our model uses five-year age groups, we apply five-year survival and migration rates before aging the group. Thus, to get population counts for 25- to 30-year-olds in 2015, we apply geography-, age-, race and ethnicity–, and sex-specific five-year survival and migration rates to the 20- to 25-year-old population in 2010. Additionally, to get those ages 85 and older in 2015, we apply this same method to 80- to 85-year-olds in 2010 and add to it the 85-and-older population in 2010 projected to survive and remain in the same location for the next five years. We also project a new birth cohort, from birth to age 4, which consists of births projected to occur between 2010 and 2015 by applying five-year birthrates to the 2010 population. This method continues in five-year increments until 2040.

Investigating Alternative Futures

Since the future of the United States could follow many paths, the Mapping America's Futures project includes projections under various scenarios for each commuting zone, decade, and demographic group. The current projections include three possibilities each for future birth, mortality, and migration rates, which allow the user to explore the combined effects of choosing different scenarios for the three components of demographic change on future populations to 2030.

For example, a user can explore the effects of the advancing age of the baby boom generation on his or her local population. What would that distribution look like if medical advances lower mortality rates? What if retirees decide to migrate out to warmer climates? The series of projections allows users to explore these possible futures.

Detailed Methodology

Our projections begin with county-level population, birth, survival, and migration data by age, race and ethnicity, and sex from 1990 to 2010. We then age the group's 2010 population five years and apply

geography-, age-, race and ethnicity–, and sex-specific birth, survival, and migration rates to arrive at "precalibrated" populations for each cohort in 2015. Finally, we adjust these projected numbers so they sum to the US Census Bureau's national population projection by age, race and ethnicity, and sex for that year.[1] This process is repeated until we reach the 2030 population. Although we project populations in five-year intervals, we provide only decennial data. Below, we describe the details of the data used; the assumptions we made about how these birth, survival, and migration rates affect population; and the methods used to adjust our projections to match the US Census Bureau's national numbers.

Data Sources

All our projections begin with the US Census county population estimates by age, race and ethnicity, and sex for April 1, 2010, for the base population.[2] Subsequent steps in the projection process use Census and other data as described below.

Geographies

Geographers, planners, and journalists have long recognized a mismatch between state boundaries and "meaningful" geographies in the United States. To account for this mismatch, they have produced a wide assortment of geographies to better reflect common patterns, histories, and economic factors. (See, e.g., Garreau 1981; Gottman 1961; and Nelson and Lang 2011.)[3] Both the commuting zones and 24 larger regions that we defined for this project are such geographies.

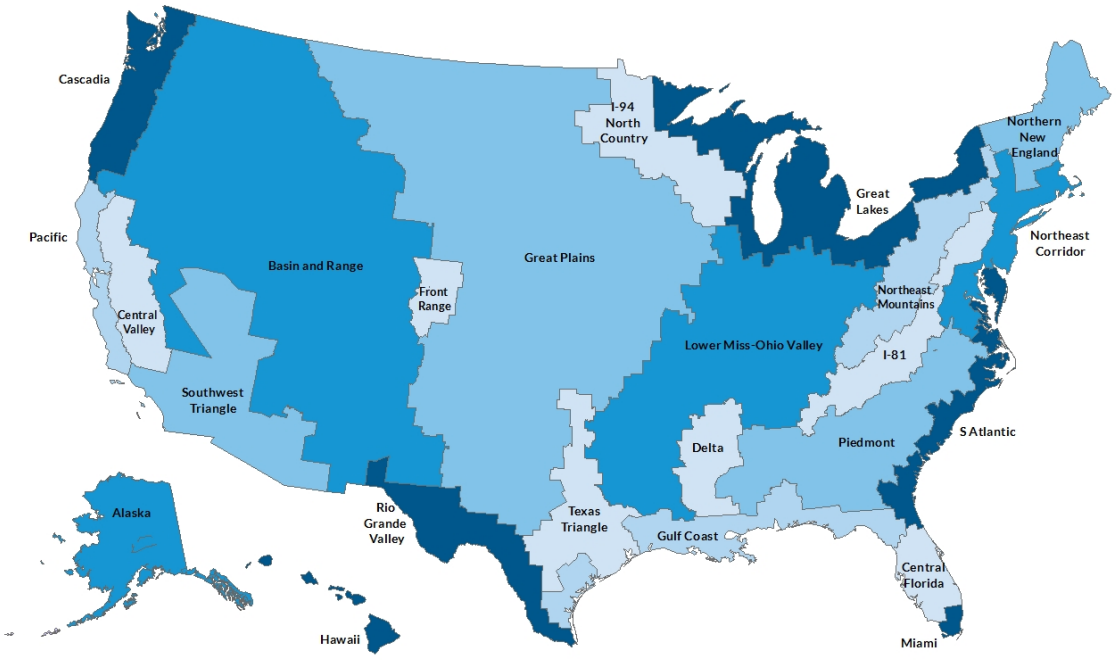
The base of our projection series consists of 740 commuting zones derived from the 1990 Census's 741 commuting zones (Tolbert and Sizer 1996).[4] In 2000, this number was reduced to 709 after the commuting-zone criteria for local labor markets were reestimated for the 2000 Census. But we retain the 1990 designations because components of our projection series use local demographic data extending back to 1990. Our projection series used a county-to–commuting zone crosswalk to maintain 740 of these commuting zones over time (Autor and Dorn 2013).[5] Further adjustments were made to accommodate new county designations.

Additionally, we defined a set of 24 regions (figure 1). The definitions of these regions build from the many regional divisions defined in earlier work linked with our own observations on how neighboring commuting zones experience similar trends in population change. Some of these regions are geographically distinct areas, such as Alaska and Hawaii. Some have been the subject of scholarship and political organization, such as the Central Valley, Gulf Coast, and Rio Grande Valley. Others are connected through transit, economies, and cultural identity, like the I-81 Corridor, Front Range, the Great Lakes, and the Mississippi Delta. Still others were adapted from earlier work, including the Texas Triangle, which was adopted from Nelson and Lang's *Megapolitan America* definition (2011); the Northeast Corridor, adopted from Gottman's *Megalopolis* (1961); and the Great Plains, adopted from the Interior Plains Province set out by the US Geological Survey.[6] Our purpose in defining these regions is to present a clearer story for the entire country at a level well above that of the 740 commuting

zones. These regions are not intended as an argument about policy or cause and effect relationships, nor are they a firm statement about the permanence of these boundaries and divisions.

FIGURE 1

Map of 24 Regions



Race and Ethnicity

We disaggregate the population into four distinct racial and ethnic categories: Hispanics (of any race), non-Hispanic blacks, non-Hispanic whites, and non-Hispanic others. The non-Hispanic other population includes American Indians and Alaska Natives, Asians, Native Hawaiians, Pacific Islanders, and those who are of two or more races who are not Hispanic. We use self-reported race and ethnicity from the 2010 Census population estimates. From earlier birth data, used to calculate birthrates as described below, we assign the same race and ethnicity as that reported from the mothers. Our current model similarly assigns the mother's race to children born after 2010.

Births

Birthrates are calculated using data from the averages of the National Vital Statistics System's birth records for 2000 and 2010.[7] These rates are calculated by age and race at the state level with county-specific adjustments based on a regression model with county-fixed effects. However, these fixed

effects (both negative and positive) were shrunk so that state-level estimates are still the dominant source of variation. For the scenario using average birthrates, the model holds birthrates to continue at 2000–10 levels. The high birthrate is 1.2 times the 2000–10 levels, and the low birthrate is 0.8 times the 2000–10 levels. We apply those birthrates to the population in five-year increments.

Deaths

Mortality rates are calculated by dividing the 1990–2000 counts of deaths from the National Vital Statistics System's data on mortality, by age, race, ethnicity, sex, and state by the 1990–2000 Census populations by age, race, ethnicity, sex, and state.[8] The survival rates are the share of the population who would survive after applying the mortality rates (i.e., 1-mortality rate). The high- and low-survival rate scenarios transform the 1990–2000 survival rates onto a real line, adjust them up by 0.5 and down by 0.5, respectively, and then use a logistic function to keep survival rates between 0 and 1. We apply those survival rates to the population in five-year increments.

Migration

The source data for migration rates is a file of 2000–10 county-level estimates of population change, net of reported births and deaths in each county (Winkler et al. 2013). However, we further tested our models using a similar file of 1990–2000 county-level migration (Voss et al. 2004). Because we project out 5-year age groups in 5-year increments, we take the square root of the 10-year migration rates to estimate a 5-year migration rate. Although this method works well for most age groups in most counties, it does not capture age groups that make significant moves within a 10-year period, including college students moving in and out of college towns.

We predict two migration rates, an exponential- and a constrained-migration rate, and then calculate a series of weighted averages of these two projections to create high-, low-, and average-migration scenarios. All migration rates are calculated for each cohort at the commuting-zone level using population-weighted averages of county-level migration rates from 2000 to 2010 for each age, race, and sex group.

TWO CALCULATIONS FOR MIGRATION RATES

The exponential-migration rates are calculated for each county using population-weighted averages of the ratio of the cohort's (*i*) actual 2010 population (*Pop*) over the cohort's expected 2010 population (*Exp*) for every county in the commuting zone:

$$\widehat{E_{i,c}} = \sqrt{Avg_{cz,pop}\left(\frac{Pop_{2010,i}}{Exp_{2010,i}}\right)},$$

where $\hat{E}_{i,c}$ = the expected exponential-migration ratio for cohort *i* in county *c*;

$Avg_{cz,pop}$ = the population-weighted average of the migration ratios of all counties in the commuting zone in which county *c* belongs;

$Pop_{2010,i}$ = the county-level population for cohort *i* in 2010; and

$Exp_{2010,i}$ = the expected population is the population that would occur in 2010 if the only changes to the cohort's 2000 population were from deaths that occurred between 2000 and 2010.

We then cap the migration ratio at three to one per decade. We assume these exponential rates from 2000 and 2010 for each age, race, and sex group for a commuting zone continue into the future. These rates are applied to the group's population in five-year increments from 2015 to 2040, and, consequently, have a multiplicative effect on the number of migrants.

For the constrained-migration rates, we weigh down the exponential-migration rates, as described above, by the cohort's (*i*) share of the total county (*c*) population (*Pop*) in 2010. This holds a cohort's shares of the commuting zones' populations to a consistent rate, pushing migration toward zero.

$$\widehat{O_{i,c}} = \left(\widehat{E_{i,c}} - 1\right) \mathrm{X} \left(\frac{Pop_{2010,i,c}}{Pop_{2010,c}} + 1\right),$$

where $\hat{O}_{i,c}$ = the expected constrained-migration ratio for cohort *i* in county *c*;

$\hat{E}_{i\,c}$ = the expected exponential-migration ratio for cohort *i* in county *c*;

$Pop_{2010,i,c}$ = the county-level population for cohort *i* in 2010 in county *c*; and

$Pop_{2010,c}$ = the county-level total population in 2010 in county *c*.

Again, these rates are applied to the group's population in five-year increments from 2015 to 2040, further pushing the group's migration rate toward zero.

AVERAGE-, LOW-, AND HIGH-MIGRATION SCENARIOS

The average-migration scenario takes the mathematical mean of the populations projected using the exponential- and constrained-migration rates. The low-migration scenario takes a weighted average of these two projected populations, giving the projection using a constrained-migration rate a weight of three times that of the projection using the exponential-migration rate. Similarly, the high-migration scenario is a weighted average, but it gives the projection using the exponential-migration rate a weight of three times that of the projection using the constrained-migration rate.

Calibration

After using the cohort-component method to create population projections based on historical rates of births, deaths, and migration and creating different projections for high and low birth, mortality, and migration rates, the final step in our model is to calibrate these series of projections to national projections from the US Census Bureau.[9] The projections for the separate age-, race- and ethnicity-, sex-specific group populations within the separate commuting zones can be aggregated to produce projections for the separate age, racial and ethnic, and sex group populations for the nation as a whole. These national group population projections from our model can be compared with the 2012 National Population Projections Middle Series by the US Census Bureau, which projects national totals for each age, race, sex, and ethnicity group in five-year intervals through 2060.[10] To make our aggregate projections compatible with these official national projections, we rescale our age-, race- and ethnicity-,

sex- specific population projections aggregated across all commuting zones to match the national totals from the official Census national projections.

Our current model calibrates each age, racial and ethnic, and sex group's (*i*) population projections to the corresponding national projections in each year (*y*) for each migration scenario (*m*) by applying the following ratio to the future population projections:

$$R_{i,y,m} = \frac{CPop_{i,y,m}}{\sum_{n=1}^{740} CZPop_{i,y,m}},$$

where $R_{i,y,m}$ = the calibration ratio for group *i* in year *y* for migration scenario *m*;

$CPop_{i,y,m}$ = the national population for group *i* in year *y* for migration scenario *m* projected in the Population Projection Middle Series; and

$CZPop_{i,y,m}$ = the commuting zone-level population for group *i* in year *y* for migration scenario m projected by the model pre-calibration, assuming average birth, mortality, and migration rates.

The calibration ratio is calculated such that when the ratio is applied to the precalibrated, model-projected populations using the average birth and mortality scenarios, the sum of a group's commuting zone populations equals the Population Projections Middle Series for that same group and year, regardless of the migration scenario chosen.

ALLOWING VARIATIONS IN BIRTH AND MORTALITY RATES TO AFFECT NATIONAL POPULATION COUNTS

These calibration ratios are then applied to the populations projected using high and low birth and mortality rates. This effectively allows the groups' national population counts to vary based on the birth and mortality scenarios chosen, but it still constrains the projections to the same ratio-based adjustments made for the population projected using the average birth- and mortality-rates scenario. This means a population projected using the low-birthrate scenario is going to have a lower national population count than a population projected using the high-birthrate scenario.

RESTRICTING THE EFFECT OF VARIATIONS IN MIGRATION RATES ON POPULATION COUNTS

Unlike the birth and mortality scenarios, all of the migration scenarios are calibrated to the US Census Bureau Population Projections Middle Series for the country separately, again holding birth and mortality rates constant at the average level. This limits the national number of net international migration for each age and racial and ethnic group to the Middle Series projection, regardless of the migration scenario chosen. However, it will allow significant variation in the pace of movement within the United States. This variation across migration scenarios affects both the level of domestic migration and the geographic destinations for immigrants. Future plans for scenario development include migration scenarios for variation in net immigration and emigration.

The calibration's effects on the commuting zone's group population in the low-migration scenario, in addition to keeping the commuting zone's current group populations where they are, will favor disbursement of any increases in the national population to commuting zones where that group already

has a substantial portion of that group's national population. Similarly, the effects of the calibration on the commuting zone's group population in the high-migration scenario, in addition to moving populations toward commuting zones to which they were moving in the past decade, will favor disbursement of any increases in the national population to commuting zones where that group already has experienced substantial growth in the past decade. For example, in a low-migration scenario, because Los Angeles already had a large Hispanic population in 2010, it would experience greater increases in its Hispanic population through internal processes (i.e., more births than deaths among an incumbent population with less propensity to migrate out) as well as external ones (i.e., increases from immigration). In a high-migration scenario, however, commuting zones like Charlotte, which did not have as large a Hispanic population in 2010 but had significant Hispanic in-migration rates in the 2000s, would experience greater increases in its Hispanic population. A corollary to this is that Charlotte would experience lower increases in its population because of migration in a low-migration scenario, and Los Angeles would experience lower increases in its Hispanic population because of migration in the high-migration scenario.

Future Plans

We see our model, and the data and tools that come from it, as part of an ongoing process. The current model is our initial step toward building the data and tools necessary to create useful, local-level projections that allow users to explore and plan for what their communities might look like in the future. Our future plans already include modeling immigration and emigration and capturing the nuances of five-year migration patterns to better model the migration of college students. We also hope to build additional tools that would give users more control over and potentially create their own scenarios.

Notes

1. "2012 National Population Projections," US Census Bureau Population Division, accessed December 1, 2014, http://www.census.gov/population/projections/data/national/2012.html.
2. "Annual Estimates of the Resident Population by Age, Sex, Race, and Hispanic Origin for Counties: April 1, 2010 to July 1, 2011," US Census Bureau Population Division, accessed December 1, 2014, http://www.census.gov/popest/data/counties/asrh/2011/CC-EST2011-alldata.html.
3. See also "MegaRegions," America 2050, accessed November 26, 2014, http://www.america2050.org/content/megaregions.html#more.
4. The one 1990 commuting zone not included in our projections is Adak, Alaska, because it was not included in the "County-Specific Net Migration by Five-Year Age Groups, Hispanic Origin, Race and Sex: 2000-2010 (ICPSR34638-v1)" Stata data files.
5. Commuting-zone crosswalks for the 1990 and 2000 Censuses were obtained from Autor and Dorn (2013) and are available at http://www.ddorn.net/data.htm.
6. "Geologic Provinces of the United States: Interior Plain Province," US Geological Survey, Last modified October 2, 2014, accessed December 1, 2014, http://geomaps.wr.usgs.gov/parks/province/intplain.html.
7. National Center for Health Statistics, "Vital Statistics Data Available Online," Birth data files from 2000 and 2010, last modified October 8, 2014, accessed November 26, 2014, www.cdc.gov/nchs/data_access/VitalStatsOnline.htm.

8. National Center for Health Statistics, "Vital Statistics Data Available Online," Mortality multiple cause files for 1990 and 2000, accessed November 26, 2014, www.cdc.gov/nchs/data_access/VitalStatsOnline.htm.

9. Data from US Census Bureau Population Division, "2012 National Population Projections," Table 1 Projected Population by Single Year of Age, Sex, Race, and Hispanic Origin for the United States: 2012 to 2060: Middle Series, accessed November 6, 2013, http://www.census.gov/population/projections/data/national/2012/downloadablefiles.html.

10. In December 2014, the US Census Bureau released new 2014 US population projections to the year 2060. Our current projections tool remains calibrated to the 2012 Census projection series that was available during development of the tool. A planned update of our projection tool in 2015 will include calibration to the 2014 Census projections as well as an adjustment for upcoming Census projections for variation in migration that is expected in spring 2015.

References

Autor, David H., and David Dorn. 2013. "The Growth of Low Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (5): 1553–97.

Garreau, Joel. 1981. *The Nine Nations of North America*. Boston: Houghton Mifflin.

Gottman, Jean. 1961. *Megalopolis: The Urbanized Northern Seaboard of the United States.* New York: Twentieth Century Fund.

Nelson, Chris, and Robert Lang. 2011. *Megapolitan America: A New Vision for Understanding America's Metropolitan Geography*. Chicago: APA Planners Press.

Tolbert, Charles M., and Molly Sizer. 1996. *U.S. Commuting Zones and Labor Market Areas: A 1990 Update*. No. 9614. Washington, DC: Economic Research Service, US Department of Agriculture.

Voss, Paul R., Scott McNiven, Roger B. Hammer, Kenneth M. Johnson, and Glenn V. Fuguitt. 2004. *County-Specific Net Migration by Five-Year Age Groups, Hispanic Origin, Race and Sex 1990–2000*. CDE Working Paper No. 2004-24. Madison: Center for Demography and Ecology, University of Wisconsin—Madison.

US Census Bureau. 2013. *Methodology and Assumptions for the 2012 National Projections*. Washington, DC: US Census Bureau.

Winkler, Richelle L., Kenneth M. Johnson, Cheng Cheng, Paul R. Voss, and Katherine J. Curtis. 2013. *County-Specific Net Migration by Five-Year Age Groups, Hispanic Origin, Race and Sex 2000-2010*. ICPSR34638-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

About the Authors

Kaitlin Franks Hildner is a research associate in the Metropolitan Housing and Communities Policy Center. Her research focuses on using data and analysis to better inform decisionmakers at both the policy and the programmatic levels. Her current work includes providing technical assistance to US Department of Education Promise Neighborhoods grantees on how to use and develop longitudinal data systems and assisting with the National Neighborhood Indicators Partnership.

Austin Nichols is an Urban Institute affiliated scholar who specializes in applied econometrics, labor economics, and public finance. His research focuses on the well-being of families and social insurance programs, including work on child poverty, disability insurance, income volatility, and economic mobility (within and across generations). He also studies education, health, and labor market interventions, and determinants of poverty and economic inequality.

Steven Martin is a senior research associate in the Center on Labor, Human Services, and Population at the Urban Institute, having joined in 2013. He works on various topics in social demography; his particular area of interest has been modeling demographic events across the life course. His recent work has covered a range of demographic topics across the life course, such as nonmarital childbearing, fertility timing, childlessness, union formation and dissolution, and age at entry into sexual activity as well as topics in time use, well-being, the "digital divide" (the unequal diffusion of Internet and computer use in the United States), and the quality of data from event-history surveys.

URBAN INSTITUTE

2100 M Street NW
Washington, DC 20037

www.urban.org

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

Funders do not determine research findings or influence scholars' conclusions. Urban scholars and experts are independent and empowered to share their evidence-based views and recommendations shaped by research.