



Section IV. Technical Discussion of Methods and Assumptions

excerpt from:

Long-term Population Projections for Massachusetts Regions and Municipalities

Prepared for the
Office of the Secretary of the Commonwealth
of Massachusetts



Henry Renski
University of Massachusetts, Amherst
Department of Landscape Architecture and Regional Planning

Susan Strate
Population Estimates Program Manager, UMass Donahue Institute

UMass Donahue Institute Contributors:

Daniel Hodge, Director of Economic and Public Policy Research

William Proulx, Senior Research Analyst

Katherine Paik, Research Analyst

Steffen Herter, Research Assistant

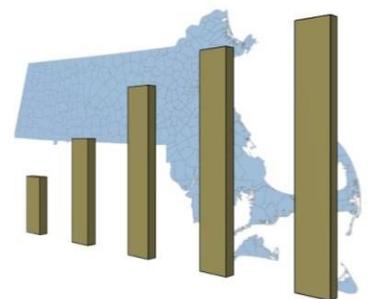
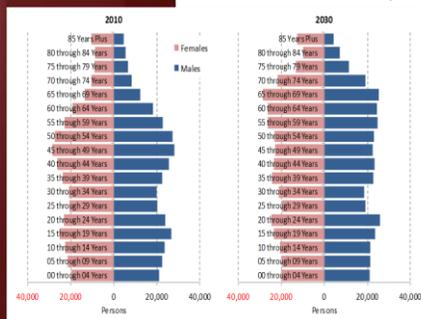


Table of Contents

IV.	Technical Discussion of Methods and Assumptions	61
	A. Regional-Level Methods and Assumptions	61
	Summary	61
	Regional definitions	62
	Estimating the components of change	63
	<i>Determining the launch year and cohort classes</i>	63
	<i>Deaths and Survival</i>	63
	<i>Domestic Migration</i>	63
	<i>College Migration</i>	66
	<i>International Migration (immigration and emigration)</i>	67
	<i>Births and Fertility</i>	69
	<i>Aging the population and generating projections for later years</i>	69
	<i>Reconciliation to Current Population Estimates</i>	70
	B. Municipal-Level Methods and Assumptions	71
	MCD-Level Model Overview	71
	Data Sources	72
	MCD Projections Launch Population	72
	<i>Initial Launch Population</i>	72
	MCD Projections: Mortality	72
	<i>Forward Cohort Survival Method</i>	72
	<i>Five-Year Survival Rates by Age/Sex</i>	72
	<i>Survived Population for MCDs</i>	73
	<i>Key Assumptions</i>	73
	MCD Projections: Migration	73
	<i>Residual Net Migration from Vital Statistics</i>	73
	<i>Determination of Net Migration Rates</i>	73
	<i>Key Assumptions</i>	74
	Fertility	74
	<i>Vital Statistics Method</i>	74
	<i>Fertility by Age of Mother</i>	74
	<i>Fertility Rates</i>	74
	<i>Key Assumptions</i>	75
	Controlling to the Regional-level Projections	75

List of Tables and Figure

Figure 4.1:	Massachusetts Regions for Population Forecasts	62
--------------------	--	-----------



IV. Technical Discussion of Methods and Assumptions

This section provides a technical description of the process used to develop the 1) regional and 2) municipal-level population projections using a cohort-component approach. While both levels of projections are prepared using a cohort-component method, the major methodological difference is in the way migration is modeled: the municipal-level estimates (also referred to as Minor Civil Divisions, or MCDs) rely on residual net migration rates computed from vital statistics, while the sub-state regional projections use gross domestic migration rates based on the American Community Survey Public Use Microdata (ACS PUMS). MCD projections are controlled to projections developed for eight sub-state regions in order to smooth out variations due to data quality issues at the MCD level and ensure more consistent and accurate projections at higher-level geographies. These controlled MCD projections can then be re-aggregated to other areas of interest, such as counties or regional planning areas.

A. Regional-Level Methods and Assumptions

Summary

This section describes the process and data used to develop the regional population projections. These projections were developed separately for eight Massachusetts regions, although each region was produced following a generally similar framework. The methodology describing how the regional projections were used to estimate municipal population projections follows in Part B of this section.

Our regional projections are based on a demographic accounting framework for modeling population change, commonly referred to as a cohort-component model. The cohort-component method recognizes that there are only four ways that a region's population can change from one time period to the next. It can add residents through either births or in-migration, or it can lose residents through deaths or out-migration. We further divide migration by whether domestic or international, and use separate estimation methods for each.

The cohort-component approach also accounts for population change associated with the aging of the population. The current age profile is a strong predictor of future population levels, growth and decline. The age profile of the population can differ greatly from one region to another. For example, the Greater Boston region has a high concentration of residents in their twenties and early thirties, while the Cape and Islands have large shares of near and post-retirement age residents. Furthermore, the likelihood of birth, death, and in- and out-migration all vary by age. Because fertility rates are highest among women in their twenties and early thirties, a place that is anticipating a large number of women coming into their twenties and thirties in the next decade will likely experience more births. Similarly, mortality rates are notably higher for persons 70-years

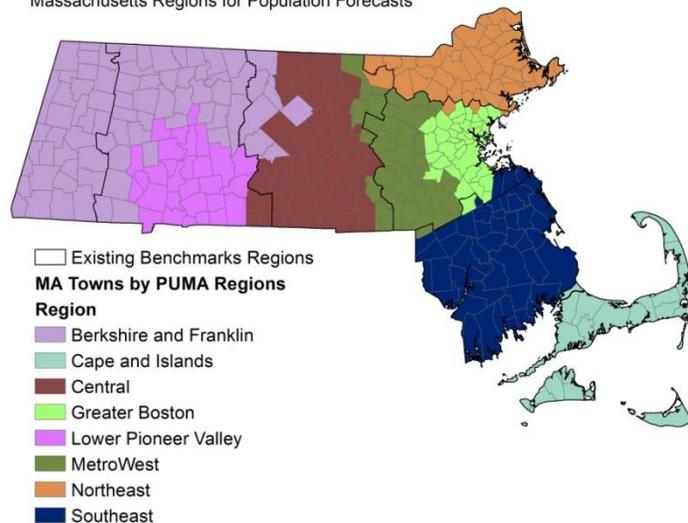
and older, such that an area with a large concentration of elderly residents will experience more deaths in decades to come.

Developing a cohort-component model involves estimating rates of change for each separate component and age-sex cohort (i.e. age-specific fertility rates, survival rates, and in- and out-migration rates) - typically based on recent trends. It then applies these rates to the current age profile in order to predict the likely number of births, deaths, and migrants in the coming years. The changes are added to or subtracted from the current population, with the resulting population aged forward by a set number of years (five years, in our case). The result is a prediction of the anticipated number of people in each cohort X years in the future. This prediction becomes the new starting baseline for estimating change due to each component an additional X years in the future. The process is repeated through several iterations until the final target projection year has been reached.

Regional definitions

A preliminary step in generating our regional projections was to determine the boundaries for each of our study areas. We use the definitions for the MassBenchmarks regions as a starting point. The Benchmarks regions were designed by the UMASS Donahue Institute to approximate functional regional economies (sets of communities with roughly similar characteristics in terms of overall demographic characteristics, industry structure, and commuting patterns). These Benchmarks regions constitute a widely accepted standard among policy officials and analysts statewide that meet common perceptions of distinct regional economies in Massachusetts.

Figure 4.1:
Massachusetts Regions for Population Forecasts



We then compared the Benchmarks regions to the boundaries of Public Use Micro-Sample Areas, also known as PUMAs. PUMAs are the smallest geographic units used by the U.S. Census Bureau for reporting data taken from the detailed (micro) records of the American Community Survey (ACS) – our primary source of migration data. PUMA boundaries are defined so that they include no fewer than 100,000 persons, and thus their physical size varies greatly between densely settled urban and sparsely settled rural areas. And although PUMAs do not typically match county boundaries, in Massachusetts individual PUMAs can be grouped together to form regions whose outer boundaries match aggregated groups of individual municipalities. This critically important feature allows us to match Census micro-data with other Census data and State vital statistics estimates we obtained at

the municipal level (i.e. births and deaths). We performed our regional grouping using Geographic Information System mapping software. The resulting study regions are presented in Figure 4.1.

Estimating the components of change

Determining the launch year and cohort classes

We begin by classifying the composition of resident population into discrete cohorts by age and sex. Following standard practice, we use five year age cohorts (e.g. 0- 4 years old, 5- 9,... 80-84, and 85- and older) and develop separate profiles for males and females, based on information provided in the 100% Count (SF 1) file of the 2010 Decennial Census of Population. This will also serve as the starting point (i.e. launch year) for generating forecasts.

Deaths and Survival

The first component of change is survival. Our projections require an estimate of the number of people in the current population who are expected to live an additional five years into the future. Estimating the survival rate of each cohort is fairly straightforward. The Massachusetts Department of Public Health provided us with a detailed dataset that included all known deaths in the Commonwealth that occurred between 2000 to the end of calendar year 2009. This database includes information on the sex, age, and place of residence of the deceased, which we aggregated into our study regions by age/sex cohort. We estimate the five year survival rate for each cohort (j) in study region (i) as one minus the average number of deaths over the past five years (2005 to 2009) divided by the base population in 2005 and then raised to the fifth power, or:

$$Survival Rate_{i,j} = \left[1 - \left(\frac{Deaths_{i,j}}{Population_{i,j}} \right) \right]^5. \quad (1)$$

Following the recommendations of Isserman (1993), we calculate an operational survival rate as the average of the five year survival rates across successive age cohorts. The operational rate recognizes that, over the next five years, the average person will spend half their time in their current age cohort and half their time in the next cohort. We estimate the number of eventual survivors in each cohort by 2015 by multiplying the operational survival rate against the cohort population count as reported by the 2010 Census.

Domestic Migration

Migration is the most dynamic component of change, and often makes the difference between whether a region shows swift growth, relative stability, or gradual decline. Migration is also the most difficult component to estimate and is the most likely source of uncertainty and error in population projections. Whereas fertility and mortality follow fairly regular age-related patterns, the migration behavior of similar age groups is influenced by regional and national differences in socio-economic conditions. Furthermore, the data needed to estimate migration is often restricted or limited; especially for many small areas. Even when it is available, it is based on statistical samples and not actual population counts, and thus is prone to sampling error – which will be larger for smaller regions.

Due to data limitations and the other methodological challenges, applied demographers have developed a variety of alternate models and methods to estimate migration rates. No single method works best in all circumstances, and we evaluated numerous approaches in the development of our projections. Those presented in this report are based on a particularly novel approach known as a multi-region gross migration model as discussed by Isserman (1993); Smith, Tayman and Swanson (2001); and Renski and Strate (2013). Most analysts use a net migration approach, where a single net migration rate is calculated as the number of net new migrants per cohort (in-migrants minus out-migrants) divided by the baseline cohort population of the study region. Although common, the net migration approach suffers from several conceptual and empirical flaws. A major problem is that denominator of the net migration rate is based purely on the number of residents in the study region. However, none of the existing residents are at risk of migrating into the region – they already live there. While this may seem trivial, it has been shown to lead to erroneous and biased projections especially for fast growing and declining regions.

A gross-migration approach calculates separate rates for in- and out-migrants. Beyond generating more accurate forecasts in most cases, it has an added benefit in that it connects regional population change to broader regional and national forces – rather than simply treating any one region as an isolated area. This type of model is made possible by utilizing the rich detail of information available through the newly released Public Use Micro-Samples (PUMS) of the American Community Survey (ACS). The ACS is a relatively new data product of the U.S. Census Bureau that replaced the detailed information collected on the long-form of the decennial census (STF 3). It asks residents questions about where they lived one year prior, which can be used to estimate the number of domestic in- and out-migrants. Unfortunately, the ACS does not report enough detail to estimate migration rates by detailed age-sex cohorts in its standard products. This information can be tabulated from the ACS PUMS – which is 5% random sample of individual records taken drawn the ACS surveys¹. Each record in the PUMS is given a survey weight, which we use to estimate the total number of migrants by detailed age and sex cohorts. It is very important to realize that the PUMS records are based on small, although representative, samples – and that the smaller the sample the greater the margin of error². Sample sizes can be particularly small when distributed by age and sex cohorts for different types of migrants, especially in small regions. For this reason, the Berkshire/Franklin and Cape & Islands are two regions that can be treated with more skepticism in our projections results and which lend themselves to greater cross-examination

¹ To account for small or missing samples in some cohorts in some regions, we make some limited adjustments to the ACS PUMS data before calculating migration rates based on the data. In the Cape and Berkshire/Franklin regions, male and female migrants under the age of 15 are assigned the male/female average number of migrants before a rate is calculated in order to smooth out male/female ratios resulting from small sample sizes. In other regions, cohorts under age 75 with a sample size of zero in the ACS data are assigned values from the opposite gender when it is available to reduce instances of rates calculated from a null value.

² While we are aware of the potential for sampling error in using ACS PUMS data for these small regions, it is the only direct source of gross migration by age available to us at this time. IRS data on migration does include gross migration data for tax-filers at the county level; however the released data does not include age detail. The Current Population Survey, another sample survey product from the U.S. Census Bureau, provides migration data by age, but only down to the U.S. regional level of geography. Other methods commonly used to estimate migration do so using an indirect method of calculating net migration by age as a residual of a cohort-survival method

by alternative methods³. These two regions were counted at fewer than 250,000 persons each in the 2010 Census and are subject to larger sampling error than the other six sub-state regions which all number more than 600,000 persons, and sometimes over 1 million. In our model, we develop migration rates using data from the 2005 to 2009 ACS PUMS as well as the 2007 to 2011 ACS PUMS, the most recent five-year dataset available at the PUMA level of geography.⁴

Estimating domestic out-migration is largely similar to estimating net-migration. Because current residents of the study region (*i*) are those who are ‘at risk’ of moving out, so the appropriate cohort (*j*) migration rate is:

$$\text{Out Migration Rate}_{i,j} = \left(\frac{\text{OutMigrants}_{i,j}}{\text{Population}_{i,j}} \right). \quad (2)$$

Because migration in the ACS is based on place of residence one year prior, the out-migration rate reported in equation (2) is the equivalent of a single year rate. We multiply this by five to estimate the five-year equivalent rate, and, as we did with survival rates, average the five year rates across succeeding cohorts to craft an operational five year rate.⁵ The operational rate for each cohort is then multiplied against the number of eventual survivors in 2015 to estimate the number of likely out-migrants from the surviving population.

In-migration is more challenging. The candidate pool of potential domestic in-migrants is not those currently living in the region, but people living elsewhere in the U.S. Modeling in-migration thus requires collecting data on the age-sex profile of not only the study region, but for other regions as well. We model two separate regions as possible sources of incoming migrants in the multi-regional framework - those originating in neighboring regions and states (New York, Connecticut, Rhode Island, New Hampshire, and other Massachusetts regions) and those coming from elsewhere in the U.S. By doing so, we recognize that most inter-regional migration is fairly local and that the migration behavior of the Northeast is likely to differ considerably from that of the rest of the nation – in part due to our older and less racially diverse demographic profile.

Thus the in-migration rates characterizing migration behavior from neighboring regions (*NE*) to study region (*i*) and from the rest of the United States (U.S.) are calculated as:

$$\text{In Migration Rate}_{NE \text{ to } i,j} = \left(\frac{\text{InMigrants}_{NE \text{ to } i,j}}{\text{Population}_{NE,j} - \text{Population}_{i,j}} \right) \quad (3)$$

³ For information on alternative projections methods and results for the Berkshire/Franklin and Cape & Islands regions, researchers may contact the Population Estimates Program of the UMass Donahue Institute.

⁴ To account for small or missing samples in some cohorts in some regions, we make some limited adjustments to the ACS PUMS data before calculating migration rates based on the data. In the Cape and Berkshire/Franklin regions, male and female migrants under the age of 15 are assigned the male/female average number of migrants before a rate is calculated in order to smooth out male/female ratios resulting from small sample sizes. In other regions, cohorts under age 75 with a sample size of zero in the ACS data are assigned values from the opposite gender when it is available to reduce instances of rates calculated from a null value.

⁵ This differs from calculating the five-year survival rate, where the one-year rate was taken to the fifth power. Survival is modeled as a non-recurring probability, since you can only die one. However, we assume that any individual migrant could move more than once during the study period, and multiply the single year rate by five to estimate a five-year equivalent.

$$\text{In Migration Rate}_{US \text{ to } i,j} = \left(\frac{\text{InMigrants}_{US \text{ to } i,j} - \text{InMigrants}_{NE \text{ to } i,j}}{\text{Population}_{US,j} - \text{Population}_{NE,j}} \right). \quad (4)$$

As with the out-migration, each single-year in-migration rate is converted into a five-year operational migration rate. Unlike out-migration, these in-migration rates are not multiplied against the surviving regional population for the study region but instead the cohort population for the region of origin (neighboring regions for equation 3 or the rest of the U.S. for equation 4) to reflect the true population at risk of in-migration. The data for estimating the launch year cohort size for other regions is aggregated from the 2010 Census of Population (SF 1), with the study region cohort population subtracted from the base of neighbor regions and neighbor populations subtracted from the United States cohort population.

College Migration

Tracking the migration of college students is often problematic for researchers, as neither the ACS nor conventional tax-return migration data seems to capture their movement comprehensively or accurately. For this reason, the U.S. Census Bureau applies a “college fix” in their annual county-level population estimates to areas that meet their criteria for percent of population enrolled in college and other population thresholds⁶. In the basic application of the “college fix”, the college-enrolled population in a region is held back from aging and migration experienced by the non-college population over the specified time period, and is then restored to the region at the end of the period. In this way, the college-enrolled population remains more or less fixed for a region while other cohorts migrate and age over time.

In the UMDI Vintage 2015 projections model, we apply a “college fix” method to the 15-19, 20-24, and 25-29 age cohorts in three regions: Greater Boston, Lower Pioneer Valley, and the Central Region. According to ACS 20072011 data, these regions all show significant percentages of college enrollment as follows:

ACS 2007-2011 Population Enrolled in College or Graduate School by Region						
UMDI Region	Greater Boston		Lower Pioneer Valley		Central Region	
Age cohort	# enrolled	% of cohort	# enrolled	% of cohort	# enrolled	% of cohort
15-19	55,018	39%	19,565	36%	14,207	27%
20-24	97,496	54%	30,255	57%	22,624	49%
25-29	44,479	24%	5,557	15%	5,613	14%

The UMDI college fix method, like the Census Bureau’s, holds out the college enrolled portion of these three cohorts from aging and migration and then adds it back into its original cohort five years later. For each of the “College Fix” regions, we use 2007-2011 ACS data to determine the share of population enrolled in college or graduate school in each of the age cohorts. The share is based on the region’s enrolled cohort as a percent of the total U.S. cohort. We apply this share by

⁶ *The “College Fix”: Overcoming Issues in the Age Distribution of Population in College Counties*. Ortman, Sink, King. Population Division, U.S. Census Bureau. October 2014.

age and sex to the base year population in order to estimate the regional college population and then subtract this from the total regional population. The difference is the estimated “non-college” population. This non-college population is subject to the same migration method described in the domestic migration section above, except that the migration rates are based solely on the non-college population and migrants in the ACS data. The resulting net number of non-college domestic migrants is added to each non-college cohort, which is then aged forward by five years. Finally, the enrollment share for each cohort is applied to the latest U.S. cohort total to determine a new estimate of the college-enrolled population for the region. This updated college estimate is added to the projected population. Below is an example for the 2010 to 2015 period.

2010		2015	
non college pop 10-14	<i>age 5 years and add net migrants 2010-2015→</i>	non-college pop 15-19	
college pop 15-19	<i>not aged; apply % enrolled to 2015 U.S. population 15-19→</i>	college pop 15-19	
non college pop 15-19	<i>age 5 years and add net migrants 2010-2015→</i>	non-college pop 20-24	
college pop 20-24	<i>not aged; apply % enrolled to 2015 U.S. population 20-24→</i>	college pop 20-24	
non college pop 20-24	<i>age 5 years and add net migrants 2010-2015→</i>	non college pop 25-29	
college pop 25-29	<i>not aged; apply % enrolled to 2015 U.S. population 25-29→</i>	college pop 25-29	
non college pop 25-29	<i>age 5 years and add net migrants 2010-2015→</i>	non college pop 30-34	

Because the college population is held out of the aging process, and because migration is only captured for the non-college population, we had to make two additional adjustments to our model. First, we allow portions of the college-enrolled population aged 20-24 and 25-29 to age forward into the non-college population⁷. This accounts for the college-enrolled population that ages in place into the non-college population (i.e. those that come for college or graduate and stay). Additionally, we account for the region’s non-college population that joins the college population upon migrating out of the region (i.e. those who leave their homes in Massachusetts to attend college elsewhere in the U.S.) by capturing them as out-migrants⁸.

International Migration (immigration and emigration)

International immigration in our model is estimated according to the number of international migrants, by age and sex, indicated for each region by the ACS 2007-2011 PUMS dataset. Unlike domestic migration in our model, however, the estimates of international immigrants from the ACS are not then converted to rates. With domestic migration, we can more comfortably make the assumption that there is a relationship between the number of migrants (our numerator) and another region (our denominator) that might be expected to remain relatively constant over time - for example the number of out-migrants relative to the region’s population or the number of in-migrants relative to the U.S. population. In the case of international migration, it is harder to make an assumption that, for example, as the world population by age increases, the region’s immigrants will increase at the same rate. In reality, a great number of factors not related to any particular

⁷ To determine this proportion we applied a residual survival method using estimates of the college-enrolled and total populations by age in 2005 and 2010, based on enrollment levels by age indicated in the ACS 2005-2009 PUMS data.

⁸ Out-migrants that are enrolled in college in regions outside of the study area, as captured in the ACS PUMS datasets.

region's current population will influence future immigration levels, including federal immigration policy change, college recruitment policies, and labor needs, to name just a few. Instead of trying to guess at which way these changes will affect immigration to each region, we assume that the levels experienced in recent history, in this case the 2007 to 2011 period, will be sustained, and in our Vintage 2015 model the number of immigrants by cohort remain constant over the time period.

There is no consensus on how best to deal with emigration in a gross-migration context. One quirk of the ACS is that while it does contain information on the residence of recent international immigrants, it contains no information that might be used to estimate emigration. This is because the ACS only surveys people currently living in the U.S. This includes recent immigrants, but not people that moved out of the nation during the last year.

But, while we cannot directly estimate the number of emigrants in a five-year period using regional level ACS data, there are alternative methods that can be borrowed to at least approximate the a number for each region. The U.S. Census Bureau developed emigration rates for the foreign born population -- the population most prone to emigration -- for a demographic analysis of net international migration. The rates were developed using a residual method and data from Census 2000, the American Community Survey, and life tables from the National Center for Health Statistics⁹. They estimated emigration rates ranging from of 12.8 to 15.5 per 1,000 among the population of recently arrived foreign born (those entering the U.S. within 10 years prior to the survey) and rates of just 1.7 to 3.5 per 1,000 for the foreign born population with longer residency -- (those arriving more than ten years prior to the survey).

To estimate emigration in our model, we first use ACS 2007-2011 information on the foreign born population by age and by decade of entry to create two estimates of the foreign born population for each state region: one recent-arrival group and one longer-residency group. Using a simplified survival method, we age these two populations forward every five years, decreasing them by letting the 85-and older population fall out (a rough proxy for mortality) and increasing them by the addition of new immigrants (using ACS 2007-2011 levels). After 10 years, new immigrants are moved into the longer-residency group. We apply the Census Bureau's middle-range rates for recently-arrived and longer-residency distinctly to each group in order to estimate the total number of emigrants by cohort in each time period.

It should be noted that in the Greater Boston, Central, and Lower Pioneer Valley regions, emigrating international students are already accounted for by the "revolving-door" approach of the college-fix method. In these three regions, we calculate international immigration and emigration only for the non-college population. College students in our model are withheld from the population at-risk for migration and aging. As such, they are not being counted as "immigrants" in the conventional sense, but instead are lumped in with all other college students, as a constant relative to the entire national population. In the Greater Boston region, college-enrolled immigrants ages 15-29 account for 30% of all international immigrants in the 2007-2011 ACS period, while in the Lower Pioneer Valley, they account for about 36%. These proportions can be thought of in our model as now

⁹ Source: Population Division Working Paper No. 97: *Estimating Net International Migration for 2010 Demographic Analysis: An Overview of Methods and Results*, U.S. Census Bureau, February 2013.

removed from the foreign born population that would typically drive both immigration and emigration numbers, and so reduces the effect of any error in estimating emigration based on foreign born population estimates.

Finally, international immigrants who become part of the resident population are then subject to the same out-migration rates as the general population. If they move on to other parts of the U.S., they are captured as out-migrants in the next five-year period.

The final step of the migration model adds the estimated net number of domestic migrations (in-migrants minus out-migrants) and the estimated international migrants to the expected surviving population in order to estimate the expected number of “surviving stayers.” This is an estimate of the number of current residents who neither die nor move out of the region in the coming five years, plus any new migrants to the region. These surviving stayers are then used as the basis for estimating anticipated births.

Births and Fertility

The last component in our regional cohort-component model requires estimating fertility rates using past data on the number of live births by the age of the mother. Like survival, information on births comes from the Massachusetts Department of Public Health which was aggregated, by region, into our five-year age cohorts according to the mother’s age, and averaged over five years (2005 to 2009). The number of births is then divided by the corresponding number of women in 2005 for each cohort to generate an approximate age-specific fertility rate. The births of males and females are modeled separately in our approach, however, in both cases it is only the number of women in each cohort that represents the population ‘at risk’ and appears in the denominator of the fertility rate. This single year fertility rate is multiplied by five to estimate a five-year equivalent, or:

$$Fertility\ Rate_{i,j} = 5 \left[\left(\frac{Births_{i,j}}{Number\ of\ Women_{i,j}} \right) \right]. \quad (7)$$

Next, the estimated fertility rates are multiplied against the number of females in the child-bearing age cohorts among the number of ‘surviving stayers’ as estimated in the previous step. This provides an estimate of the number of babies that are anticipated within the next five years, and this number is summed across all maternal age cohorts.

Aging the population and generating projections for later years

The next step in generating our first set of five year forecasts (for year 2015) is to age the surviving stayers in all cohorts by five years. The first (0- 4) and final (85+) cohorts are treated differently. The number of anticipated babies estimated in the previous step becomes the number of 0- 4 year olds in 2015. The number of persons in the 85+ cohort in 2015 is the number of surviving stayers in the 80- 84 age cohort (in 2010) added to the number of surviving stayers in the 85 and older cohort. As we made separate estimates for males and females, the two populations are added and summed across all cohorts to determine the projected number of residents in 2015.

This process is essentially repeated for all future year projections, except that the rates developed from historic data remain the same throughout the forecast horizon. Our 2015 projection becomes our launch year population for estimating the 2020 population, which in turn is used to seed the 2025 population and so-forth. The only notable difference in the process used to generate the later year forecasts is the need to have outside projections of future population levels for the nation as a whole and for neighboring states. This is necessary for estimating population ‘at-risk’ of domestic in-migration. The U.S. Census Bureau regularly generates highly detailed national population forecasts.¹⁰ We use the latest release of national forecasts (release date December 2014) which are based on information from the 2010 Decennial Census. Unfortunately, the Census Bureau no longer generates detailed state-level long-term projections; their last state-level projections were developed in 2005. So for estimating future in-migrants from neighboring Northeast states, we use the state-level age/sex projections developed by the University of Virginia’s Weldon Cooper Center for Public Service¹¹ (release 2013).

Reconciliation to Current Population Estimates

As a final step in the regional model, we align our projections to the most current population estimates from the U.S. Census Bureau at the state and regional levels. We aggregate the vintage 2013 sub-county estimates¹² to the UMDI regions and then calculate the annual percent change in population from 2010 to 2013 for each region. This annual percent change is applied to the 2013 population to create a 2014 estimate for each region. The 2014 regional totals are then controlled to the Census Bureau’s vintage 2014 state-level population estimate¹³ to create updated regional totals to 2014. For each region, the resulting annual percent growth from 2010 to 2014 is calculated and then applied to the 2014 total to create a 2015 “target” population.

In the first five-year period of our projection series, 2010 to 2015, migration rates are adjusted across all age/sex cohorts by a fixed percentage so that the 2015 projection now matches this 2015 target. In regions where our unadjusted 2015 projection is less than the 2015 target, in-migration was adjusted upward and out-migration downward. In regions that were over-projected, in-migration was adjusted downward and out-migration upward. Adjustment factors varied by region from 0.00 to 0.13 (where adjustment = original rate x [1 + adjustment factor]). Because the adjustment is applied as a percentage of the original cohort rate, the effect is that high-migratory age groups are affected to a greater degree than the groups with less migration activity, in terms of resulting number of migrants. These final migration rates for the 2010 to 2015 period are essentially “synthesized” age/sex rates that capture the 2010 to 2014 population change trend while conforming to the to the age/sex distribution of migration found in the 2007-2011 ACS, the latest five-year set of age/sex migration data available at the PUMA level.

¹⁰ Source: <http://www.census.gov/population/projections/>

¹¹ Source: *Population Projections by Age for the U.S. and States*. Updated August 9, 2013. Weldon Cooper Center for Public Service, University of Virginia. <http://www.coopercenter.org/demographics/national-population-projections>

¹² Source: Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2013, U.S. Census Bureau, Population Division, May 2014.

¹³ Source: Annual Estimates of the Resident Population: April 1, 2010 to July 1, 2013, U.S. Census Bureau, Population Division, December 2014.

Rates for subsequent projection periods – 2015 to 2020, 2020 to 2025, and so on – use an average of rates calculated from the 2005-2009 and 2007-2011 ACS datasets. The two sets are averaged in order to capture the longest recent time-span available in the ACS PUMS five-year datasets. This averaging also helps to reduce sample error for age/sex migration rates that occurs with sample survey data. While averaging these two overlapping periods effectively centers the migration rates on the 2007-2009 period, according to Census Bureau state-level component estimates¹⁴, the centered average of these two overlapping periods is nearly identical to the average net migration estimated by Census for the most recent ten-year period, 2005 to 2014.

B. Municipal-Level Methods and Assumptions

MCD-Level Model Overview

As described in the regional-level methods section of this report, separate projections are produced for the 351 MCDs and for the eight state sub-regions. The MCD results are then controlled to the corresponding projected regional cohorts to help smooth any inconsistencies in the MCD-level results and to reflect migration trends that may be more accurately reflected by the regional projection methodology.¹⁵ While both of the regional and MCD-level projections are prepared using a cohort-component method, the MCD estimates rely on residual net migration rates computed from vital statistics, while the sub-region projections use gross domestic migration rates based on the American Community Survey Public Use Microdata (ACS PUMS).

The population aged five and over is projected by the mortality and migration methods, while the population age 0-4 is projected by the fertility method. The initial launch year is 2010, with projections made in five-year intervals from 2015 to 2035 using the previous projection as the new launch population. Projections for eighteen five-year age groups (0-4, 5-9 ...80-84, and 85-and older) are reported for males and females. (Throughout this document, the term “age” refers to a five-year age cohort). The cohort-component method is used to account for the effects of mortality, migration, and fertility on population change.

Population projections for each age and sex cohort for each five-year period are created by applying a survival rate to the base population, adding net migration for each age/ sex/ MCD cohort, and finally adding births by sex and mother’s age, as shown in the table below.

Component	Projection
Mortality	Survived population by age/sex
Migration	Net migration by age/sex
Fertility	Births by sex and mother’s age
Launch	2010 Census count by age/sex for 2015 projection; Five-year projection thereafter

¹⁴ Source: ST-2000-7; CO-EST2010-ALLDATA; and NST-EST2013-ALLDATA, U.S. Census Bureau Population Division.

¹⁵ The regional projection methodology, discussed at length in Section IV.A. of this report, projects domestic migration using migration data from the American Community Survey, therefore explicitly accounting for recent domestic migration trends. As explained in this section, the MCD methodology uses a “residual” method based on vital statistics to project migration.

Data Sources

The launch populations by sex, age cohort, and MCD were obtained from U.S. Census 2010 data¹⁶. UMDI estimated population by age and sex for 2005 from the 2000 and 2010 U.S. Censuses using a simple linear interpolation by age and sex.

UMDI requested and received confidential vital statistics data for births and deaths from January 1, 2000 through December 31, 2009 from the Massachusetts Department of Public Health. From these, UMDI estimated survival, birth and residual net migration rates.

MCD Projections Launch Population

Initial Launch Population

The initial launch population for the 2015 projection is the 2010 Census population by age/sex for each MCD¹⁷. Corrected census counts from the Count Question Resolution (CQR) program are incorporated where applicable. Each projection thereafter uses the previous projection as the launch population (i.e. the 2020 projection uses the 2015 projection as the launch population).

MCD Projections: Mortality

Forward Cohort Survival Method

The forward cohort survival method is used to account for the mortality component of population change. This procedure applies five-year survival rates by age/sex to the launch population by age/sex for MCDs in order to survive their populations out five years, resulting in the expected population age five and over before accounting for migration.

Five-Year Survival Rates by Age/Sex

UMDI calculated five-year survival rates by age and sex using deaths by age, sex and MCD from 2000 to 2009 (January 1, 2000 through December 31, 2009). Survival rates by age, sex and MCD were assumed to be constant for the duration of the projections (from 2010 through 2035). Survival rates for each age cohort up to 80-84 were averaged with the next-older cohort to account for the fact that roughly half of each cohort would age into the next cohort over the course of each five-year period. The 85-and older cohort's survival rate was used as-is, since there was no older cohort to average.

¹⁶ An exception is made in our model for the town of Lincoln, Massachusetts. For the Lincoln base we have instead created 2010 age/sex estimates using cohort-change ratios observed in the 1990-2000 period applied to the Census 2000 age/sex base. We do this because Lincoln was counted in Census 2010 with a significantly reduced population. This happened because, at the time of the Census count, a large number of the housing units at a military base had been demolished, with their replacement happening only later in 2011. This gave the town a Census 2010 base count that was out of trend with its population in the years right before and again shortly after, with population reduced by as much as 21%. While the 2010 Census may be considered as a relatively accurate point-in-time count, using it as a point of reference in a residual net migration model will create drastically altered migration rates for the town, and using it as the population base for future years will also produce unreasonably low projections.

¹⁷ See footnote (above) on exception in the town of Lincoln.

MCDs with smaller populations demonstrated a degree of variability in survival rates that we considered too broad for optimal results. Therefore, for MCDs with populations lower than 10,000 as of the 2000 Census, we used regional survival rates by age and sex instead of MCD-specific rates to smooth the results. We calculated regional rates using the same MCD-based vital statistics data from 2000 to 2009 as we used in calculating the MCD rates.

Survived Population for MCDs

The base population by age/sex for MCDs is survived to the next five-year projection by applying the corresponding averaged five-year survival rates by age/sex.

Key Assumptions

The methodology assumes that survival rates vary most significantly by age and sex. To some extent, the use of MCD-specific rates will also indirectly account for varying socioeconomic factors, including race and ethnicity, which vary by MCD and may affect survival rates. The methodology assumes that survival rates by age, sex and MCD will stay constant over the next 25 years.

MCD Projections: Migration

Residual Net Migration from Vital Statistics

The residual net migration method is used to account for the migration component of population change. “Residual” refers to the fact that migration is assumed to be responsible for past population change after accounting for births and deaths. This residual net migration is then used to estimate past migration rates. The procedure applies the resulting net migration rates by age/sex estimated for each MCD to the MCD’s survived population by age/sex in order to project net migration by age/sex for the population ages five and older. For the population ages 0-4, it is assumed that residence of infants will be determined by the migration of their birth mothers. For MCDs with 2000 Census population below 10,000, a linear migration assumption (described below) is used to smooth migration.

Determination of Net Migration Rates

Vital statistics are used to infer net migration totals for 2000 to 2009. In order to calculate five-year net migration by age, sex and MCD, natural increase (births minus deaths) by age/sex for 2000 to 2005 is added to the 2000 population by age/sex for each MCD. The results are then subtracted from the interpolated 2005 population by age/sex for each MCD to estimate net migration by age/sex and MCD for 2000 to 2005. A similar process calculates migration between 2005 and 2010.

For MCDs with 2000 population equal to or greater 10,000, the two five-year net migration estimates are averaged and rates are then calculated for each age, sex and MCD. The resulting rates are applied to the base population to project five-year net migration. The resulting average five-year net migration rates by age/sex are held constant throughout the projection period.

For MCDs with 2000 population under 10,000, five-year net migration by age, sex and MCD is held constant, and population cohorts are never allowed to go below zero. This avoids applying

unrealistically high migration rates to small populations. For instance, if an MCD starts with four males aged 70-74 and net migration shows four more move in over five years, the result is a migration rate of 2. This results in highly variable and unrealistic results in some cases. In this example, holding migration linear means that in each five-year projection period, four males aged 70-74 will move into the MCD. UMDI conducted sensitivity testing for this method and found that the model with constant migration for small places in most cases resulted in more realistic, gradual population growth or decline, as well as more realistic sex and age profiles for these MCDs.

Key Assumptions

The use of a net migration rate relies on a base for migration that includes only current residents – in other words, only those at risk of out-migration. Nonresidents who are at risk of in-migration are not explicitly accounted for in the MCD method, and this results in some inaccuracy which is minimized by the process of controlling to regional total projections that are based on a gross migration model.

We assume that age, sex and MCD are the key factors by which migration rates vary. Other factors, including non-demographic factors such as macroeconomic factors or local policy changes, are not explicitly included in this model. Future projection models may incorporate these or other factors.

Fertility

Vital Statistics Method

We apply age-specific fertility rates to the migrated female population by age to project births by age of mother, followed by survival rates for the population aged 0-4. Total survived births are then derived by summing across all maternal age groups, and the results represent the projected population age 0-4. For each MCD, the number of males and females is assumed to be the same as the proportion of male or female births statewide.

Fertility by Age of Mother

Average births by age of mother for each MCD are calculated for two five-year periods (2000 to 2005 and 2005 to 2010) using nine maternal age groups, from 10-14...50-54.

Fertility Rates

Age-specific fertility rates are computed for each time period by dividing the average number of births by age of mother by the corresponding number of females of that age group. The average age-specific fertility rates are held constant throughout the projection period. The base population for launching a new five-year projection is the survived, post-migration projected female population by age.

MCDs with smaller populations demonstrated a degree of variability in fertility rates that we considered too broad for optimal results. Therefore, for MCDs with populations lower than 10,000 as of the 2000 Census, we used regional fertility rates by age and sex instead of MCD-specific rates

to smooth the results¹⁸. We calculated regional rates using the same MCD-based vital statistics data from 2000 to 2009 as we used in calculating the MCD rates.

Key Assumptions

We assume age, sex and MCD to be adequate indicators of fertility rates for MCD for the first vintage projections. We assume that the proportion of male to female births does not vary significantly by geography or maternal age. We assume that fertility rates by maternal age and MCD will not change significantly over time. Future iterations of the projections may amend these assumptions based on available data.

Controlling to the Regional-level Projections

The resulting MCD-level projected cohorts are finally controlled to the regional-level projected cohorts. To do this, we assume that each MCD's share of the region's population, for each age and sex cohort, is given by the MCD population projections. Those shares are then applied to the regional projections to arrive at adjusted age/ sex cohorts for each MCD.

¹⁸ While MCDs with populations less than 10,000 are given the regional rate in this model, we make exception for “college bedroom” towns. Because fertility rates are generally lower among females enrolled in college compared to the general population of the same age group, applying regional fertility rates to small towns with high percentages of college-enrolled population resulted in inflated births. We developed criteria for identifying “college bedroom” towns and applied town-specific fertility rates to these instead of the regional rates. Criteria is: population under 10,000 in 2010; >20% of 18 and over female population is enrolled in college or graduate school according to 2008-2012 ACS; and use of regional fertility rate resulted in a $\geq 25\%$ Increase in the 0-4 age group from 2010 to 2015. The three MCDs subject to the “college bedroom” exception include Wenham, Sunderland, and Williamstown.