UMassAmherst

Donahue Institute

Economic and
Public Policy Research

UMDI Interim 2020 Population Estimates by Age, Sex, Race, and Municipality:
Statement of Methods, Assumptions, and Limitations

March 1, 2022

Context

The Massachusetts Department of Public Health (MA DPH) publishes disease rates and other health-related incidence rates for specific populations at small-level geographies, including the city, town, and even neighborhood level. To calculate these rates, small-level geography population data (including population by age, sex, race, ethnicity, and geography) are needed on an annual basis. These population data serve as the denominators when calculating health-related incidence rates. While the Census Bureau enumerates population by age, sex, race, and ethnicity at a very fine level of geography—the census block—it does so only every ten years. In the years following the census, or "post-censual" years, the Census Bureau's Population Estimates Program (PEP) produces annual estimates of the population down to the town level; however, they break out age, sex, race, and ethnicity only at the *county level* and higher.

In Fiscal Year 2016, MA DPH contracted with UMDI to produce small-area population estimates through 2015 at the city, town, and Census-tract level by age, sex, race, and ethnicity for: single age cohorts 0-20, 5-year age cohorts 0-4 through 80-84, and an 85+ cohort and for bridged race groups including: White, Black, Asian, American Indian and Alaska Native, Hawaiian Native or Pacific Islander, Two or More Races, and All Races (total races.) UMDI also provided MA DPH with uncontrolled, estimated population shares by age, race, sex, and ethnicity through 2020 along with STATA code and instructions on how to independently update the estimates dataset each year through 2019 using the Census Bureau's annual county-level population estimates by age, sex, race, and ethnicity.

MA DPH currently has a need for updated 2020 population by age, sex, race, and ethnicity in order to continue publishing health and disease rates past the year 2019. As of the date of this methods statement, March 1, 2022, the U.S. Census Bureau has tabulated and released some, but not all, of the decennial Census products and variables that will eventually be released using the 2020 Census count results. To date, the Bureau has released state-level apportionment counts and the PL-94.171 Redistricting dataset (PL-94). The PL-94 dataset is the more detailed of the two, and it includes 63 race groups alone or in combination, ethnicity by Hispanic and non-Hispanic, and age groups by under-18 and 18-plus, for 252 possible combinations of race, ethnicity, and age (63 x 2 x 2 = 252). PL-94 data is released for all standard Census geographies including state, county, tract, Minor Civil Division (or

"MCD", the Census equivalent of cities and towns in Massachusetts), block-groups, and blocks – all expressed according to 2020 statistical geographic boundaries.

While the Census 2020 PL-94 data is the preferred source for the most accurate and up-to-date population counts for Massachusetts geographies, it does not provide the single- and five-year age and sex detail required by MA DPH for its work. Also, the race categories in the PL-94 dataset do not match the MA DPH race categories. Specifically, the Census 2020 PL-94 dataset includes counts for "Some Other Race" responses, which have not yet been assigned to specific race categories (Table 1.)

Table 1. "Race Alone" Groups by Product

Census 2020 Count (PL-94)	MA DPH (current)
White	White
Black	Black
Asian	Asian
American Indian or Alaska Native	American Indian or Alaska Native
Hawaiian Native or Pacific Islander	Hawaiian Native or Pacific Islander
Two or More Races	Two or More Races
Some Other Race	
x Hispanic or Non-Hispanic	

To meet MA DPH program needs while we await the release of more detailed Census 2020 data, UMDI has produced "Interim Population Estimates" for MA DPH for the year 2020 by Massachusetts municipality for the same age/sex/race/ethnicity cohorts produced by UMDI for MA DPH in the FY 2016 series, namely:

- Age by single years 0-20 and by 5-year cohorts 0-4 through 80-84 plus 85+
- Sex by Male, Female, and Total
- By race including: White, Black, Asian, American Indian or Alaska Native, Hawaiian Native or Pacific Islander, Two or More Races, and All Races
- Ethnicity by Hispanic, Non-Hispanic, and Total

The methods and related assumptions and limitations used to develop the UMDI Interim 2020 Population Estimates by Age, Sex, Race, and Municipality are described in the sections below.

Estimates Methodology

Overview

To prepare the *Interim 2020 Population Estimates by age sex, race, and ethnicity*, UMDI uses inputs including: Census 2010 Summary File 1 population counts by age, sex, race, ethnicity, and MCD; Census

2020 PL-94.171 population counts by age and MCD; Census Annual County-Level Population Estimates by Age, Sex, Race, and Ethnicity (2019 estimates); and the Census 2010 Modified Race Data Summary File. UMDI first produces population estimates by age, sex, and MCD and then distributes these using race and ethnicity shares to produce estimates by age, sex, race, ethnicity, and MCD.

Age/Sex Estimates

To produce the age/sex/MCD estimates, UMDI takes the Census 2020 PL-94.171 population count for each Massachusetts municipality and distributes it among five-year-age/sex cohorts according to estimated proportions developed by UMDI.

To estimate the age/sex distributions, UMDI starts with a modified *Hamilton-Perry* or *cohort-change-ratio* (CCR) model using 2000 and 2010 decennial Census data by age, sex, and MCD. The CCR method accounts for the aging of each individual cohort from one census to the next and creates a ratio between a specific cohort population (by age, sex, and geography) age $_{a}$ in year $_{y}$ to its corresponding cohort ten years younger, aged $_{a\text{-}10}$, and ten year earlier, in year $_{y\text{-}10}$. The single ratio captures the combined effect of death and net migration for each age cohort in a specified geography as it ages forward from one Census to the next.[1]

For each age/sex/MCD cohort, the resulting cohort-specific ratio is then applied to the corresponding base population (the Census 2010 population in our model) in order to estimate the population 10 years later (2020 in this case). As a modification to the standard Hamilton-Perry model, before we integrate the resulting CCRs into our model, we cap them at "1" for cohort groups including fewer than 25 people and "2" for groups under 100 people. Population estimates for years 2011 through 2019 are developed by interpolating populations for each single year in the time series between the Census 2010 populations and the 2020 estimates.[2]

In the UMDI Interim Estimates, the resulting population estimates by age/sex/MCD in 2019 are next controlled to the U.S. Census Bureau's V2019 county-level population estimates by age and sex.[3] This control measure leverages the post-2010 updates that the Census Bureau makes to each county's population based on actual and estimated county-level births, deaths, domestic migration, immigration, and reported changes in the group quarters populations since 2010. The resulting controlled estimates are, in turn, controlled to the Census Bureau's Census 2020 PL-94 population counts of under-18 and 18-plus by MCD. By this method, the age/sex estimates in the UMDI Interim estimates will sum to the Census 2020 PL-94 totals for each city and town.

---

[1] For cohorts aged 0-9, Child-to-Women (CTW) ratios are used instead of Cohort-Change Ratios. children aged 0-4 are calculated as a ratio of the female population aged 20-44 and children aged 5-9 are calculated as a ratio of the population aged 30-49.

[2] For additional details on UMDI's modified cohort-change ratio method, see the Methodology section of: *Small Area Population Estimates for 2011 through 2020*, UMass Donahue Institute. October 2016.

[3] CC-EST2019-AGESEX-[ST-FIPS]: Annual County and Puerto Rico Municipio Resident Population Estimates by Selected Age Groups and Sex: April 1, 2010 to July 1, 2019, Source: U.S. Census Bureau, Population Division. Release Date: June 2020.

Age/Sex Estimates Assumptions

While the *level* of combined net migration[4], births, and deaths for each city and town is updated to reflect the Census 2020 count totals – or the change from 2010 to 2020 - in the UMDI Interim Estimates, our method assumed that:

- the *distribution* of combined net migration and deaths by age and sex in each city and town *relative to its county* is the same as was experienced between 2000 and 2010, and that
- the *distribution* of births, deaths, and net migration by five-year age group by sex within each county is aligned with Census Bureau estimates by age, sex, and county through 2019.[5]

Race Distributions

Once age/sex/MCD estimates are developed, UMDI distributes the age/sex/MCD-cohort estimates to race/ethnicity cohorts. To perform this allocation, once again UMDI leverages estimates developed from a cohort-change-ratio method that projects 2000-to-2010 cohort-population change forward to 2019, controls the results to the county-level age/sex/race/ethnicity estimates published by the U.S. Census Bureau, and, lastly, controls the resulting age/sex/race/ethnicity/MCD estimates back to the interim 2020 age/sex/MCD estimates developed in the earlier step.[6]

In order to ensure that the resulting cohort estimates reflect the most recent Census 2020 count information available, the estimates are finally controlled to the Census 2020 PL-94 counts by race, ethnicity, and MCD for populations aged under 18 and 18-and-over. In order to execute this last control measure, first the Census 2020 data must be "bridged" from Census-count race categories to MA-DPH public-use categories, as described earlier in this report (Table 1). UMDI bridges the 2020 PL-94 population counts by race and ethnicity to MA-DPH race categories using the *Census 2010 Modified Race Summary File,* which is the most recent bridging file currently available from the U.S. Census Bureau. The Modified Race Summary File "provides data from Census 2010 that have been modified to assign each of the persons who selected the Some Other Race category to an OMB race category" at the county level and by sex and five-year age group.[7]

To apply this control, UMDI first translates, or "bridges" each MCD's Census 2020 PL-94 population by age group (under-18 and 18+) by race and ethnicity to the MA-DPH race categories.[8] Next, UMDI takes each single-year-age/sex/race/ethnicity cohort's share of the MCD's population under 18 or 18-and-over by race and ethnicity, as generated in the CCR model described earlier, and applies these shares to the corresponding "bridged" PL-94 populations under 18 and 18 and over by race, ethnicity and MCD. In a

---

[4] "Net migration" in this summary refers to combined Net Domestic Migration and Immigration.

[5] Additional model assumptions inherent in the UMDI cohort-change-ratio model are described in UMDI's *Small Area Population Estimates for 2011 through 2020* October 2016 report.

[6] For additional details on UMDI's modified cohort-change ratio method, see the Methodology section of: *Small Area Population Estimates for 2011 through 2020*, UMass Donahue Institute. October 2016.

[7] Modified Race Summary File Methodology U.S. Census Bureau, Population Division Updated: 07/05/2012. https://www.census.gov/data/datasets/2010/demo/popest/modified-race-data-2010.html

[8] Note that this bridging step assumes that MCD-level age/race/ethnicity groups follow the same bridging pattern – or distribution of "some other race" to DPH categories – as the parent county

last step, the resulting age/sex/race/ethnicity/MCD cohorts are controlled back to the *Interim 2020 age/sex/MCD* estimates developed in an earlier step.

After final controls, the interim 2020 estimates of population by *single-year-of-age* by sex by MCD sum to the corresponding interim 2020 estimates by *5-year-age* by sex by MCD; however, after the single-year age cohorts are distributed to the various race and ethnicity groups, there are small variations, averaging 1.2 persons, between the *sum of the single-year-of-age* by sex, race, ethnicity, and MCD cohorts and the corresponding *5-year age* cohorts by sex, race, ethnicity, and MCD.[9] This variation occurs because race bridging files are produced for 5-year cohorts and PL-94 data cuts across a 5-year cohort (with populations under 18 and 18+) such that, after final controls, the summed single year cohorts by race, ethnicity, and MCD are not expected to sum exactly to their respective 5-year age/sex/race/ethnicity/MCD cohorts. In the UMDI interim 2020 estimates model, age/sex/MCD control totals are prioritized – meaning that age/sex/MCDs were distributed *down to* race/ethnicity categories (or, conversely, that race/ethnicity cohorts were controlled *up to* age/sex/MCD totals). By this method, by-age rates developed for analyses for single-year age/sex/MCD cohorts will be consistent with corresponding 5-year age/sex/MCD rates.

Age/Sex/Race/Ethnicity Assumptions

Like all estimates, the UMDI Interim 2020 estimates rely on certain assumptions which will affect the accuracy of the estimates. In addition to assumptions inherited from the Interim Age/Sex/MCD estimates, the Interim Age/Sex/Race/Ethnicity estimates also assume:

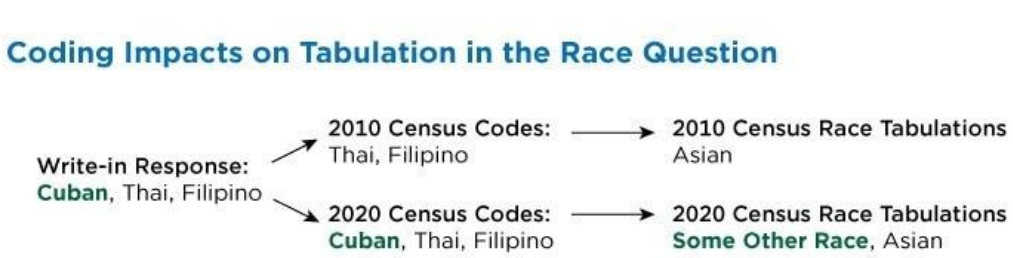
- that the change in sub-county race distribution by age and by MCD experienced by each age/sex/race cohort between Census 2000 and 2010 *relative to its the corresponding county* was carried forward through 2020.
- that the Census Bureau's annual county-level estimates of population by age, sex, race, and ethnicity are accurate
- that the modified race file assumptions subcounty (MCD-level) re-assignment of "some other Race" is the same as the county-level re-assignment.[10]

Additionally, a major limitation of the UMDI Interim 2020 estimates by race and ethnicity is that they rely on the *Census 2010 Modified Race Summary File*, as an updated 2020 bridging file is not yet available. This assumes that the allocation of the "some other race" cohorts counted in the decennial Census to both MA-DPH and Census annual estimates race categories developed in 2010 is the same in 2020. UMDI wishes to clarify that use of this file will undoubtedly lead to unquantified error in the estimates. Not only has self-reporting on race likely changed over the past decade, but the Census collection survey and race coding have also changed significantly from 2010 to 2020, resulting in increased multi-racial responses, among other changes. The U.S. Census Bureau explains that, among other changes in the 2020 Census race question, "we increased the number of characters captured from 30 to 200, which allowed us to capture and fully recognize longer write-in responses (and) instead of prioritizing multiple responses into only two codes, we coded up to six detailed codes for each write-in

[9] The resulting absolute differences in the *sum of single year age/race/ethnicity/MCD cohorts* compared to correspond*ing 5-year cohorts* are summarized in the *Technical Notes* appended to this report.

[10] Additional model assumptions inherent in the UMDI cohort-change-ratio model are described in UMDI's *Small Area Population Estimates for 2011 through 2020* October 2016 report.

area."[11] An example of changes in race designations resulting from coding changes, as provided by the Census Bureau, appears below: [12]



According to the U.S. Census Bureau, "It is important to note that … data comparisons between the 2020 Census and 2010 Census race data should be made with caution, taking into account the improvements we have made to the Hispanic origin and race questions and the ways we code what people tell us."[13]

Limitations and Disclaimer

The *UMDI Interim 2020 Population Estimates by Age, Sex, Race, Ethnicity and Municipality* are ***estimates only***, for UMDI and MA-DPH program use in the interim period between the Census Bureau's release of *Census 2020 PL-94.171* data and the forthcoming *Census 2020 Demographic and Housing Characteristics File* (DHC). According to the Census Bureau's most recently proposed release plan, the DHC will include populations by detailed age and sex by Minor Civil Division (or "MCD" - the Census geographic equivalent of municipality in Massachusetts).

As estimates, the populations by age/sex/MCD published in the UMDI Interim set are subject to inaccuracies. The *Methodology* section in this document describes the methods, data sources, and assumptions used to develop the UMDI Interim Estimates. While UMDI is able to share these interim estimates with data users to support their data needs, UMDI will not be held liable for uses and decisions based upon the *UMDI 2020 Interim Estimates*. The population counts by age and sex to be released in the Census Bureau's planned Census 2020 Demographic and Housing Characteristics File may supersede and replace the UMDI Interim Estimates as the recommended dataset once they are published.

---

[11] *Improvements to the 2020 Census Race and Hispanic Origin Question Designs, Data Processing, and Coding Procedures.* U.S. Census Bureau Population Division. August 3, 2021. https://www.census.gov/newsroom/blogs/random-samplings/2021/08/improvements-to-2020-census-race-hispanic-origin-question-designs.html

[12] *Improvements to the 2020 Census Race and Hispanic Origin Question Designs* ,Ibid

[13] *2020 Census Illuminates Racial and Ethnic Composition of the Country. U.S. Census Bureau*. August 12, 2021 https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html

UMassAmherst

Donahue Institute

Economic and
Public Policy Research

March 1, 2022

Technical Notes to UMDI Interim 2020 Population Estimates by Age, Sex, Race, and Municipality

As noted in the *UMDI Interim 2020 Population Estimates Methods Statement*, the estimated populations by *single-year-of-age* by sex by MCD sum to the corresponding estimates by *5-year-age* by sex by MCD. However, after the single-year and 5-year groups are distributed to the various race and ethnicity groups, there is some small variation between the *sum of the single-year-of-age/*sex/race/ethnicity/MCD cohorts and their corresponding *5-year age*/sex/race/ethnicity/MCD cohorts. The resulting absolute differences in the *sum of single year age/race/ethnicity/MCD cohorts* compared to correspond*ing 5-year cohorts* range from 0 to 158 maximum and average 1.2 persons per cohort, or 8.3%. Note that the maximum error of 158 represents a percent difference of only 1.6% of the corresponding 5-year cohort population. As summarized below, 99.9% of the 5-year cohorts show differences of less than 25 compared to the summed 1-year cohorts. The vast majority, 98.4%, show differences of less than 10; 94.8% differ by less than 5, and 74.3% differ by less than 1 person.

Percent of all *non-zero* cohorts[14]	Absolute difference less than:
99.9%	25
98.4%	10
94.8%	5
74.3%	1

From largest observed absolute differences to smallest,

- A total of 7 observations out of 37,499 non-zero cohorts (rounding to 0.0% of all) had ***differences over 50*** with an average absolute difference of 104, a maximum absolute difference of 158, an average absolute percent difference of 1.4%, and a maximum absolute percent difference of 2.4%. All of the cohorts with differences greater than 50 are large cohorts in the 15-19 year-old age range in the City of Boston.

[14] Note that this summary analysis was conducted for 37,499 "non-zero" cohort observations only, defined for this purpose as all age/sex/race/ethnicity/MCD cohorts *except* those that showed zero population in *both* the 5-year-age cohort and the corresponding sum of single-year-age cohorts.[14]

- 49 out of 37,499 non-zero cohort observations (0.1% of all) had absolute ***differences ranging from 25 to <50*** with an average absolute difference of 33 and average absolute percent difference of 40.6%.

- 9,588 (or 25.6%) of non-zero cohorts showed absolute ***differences ranging between 1 and <25*** when comparing single-year sums to 5-year ag/sex/race/ethnicity/MCD cohorts, with an average absolute difference of 3.7 and an average absolute percent difference of 14.7%.

- 27,855 cohorts (74.3%) showed ***differences of less than 1*** person with an average absolute percent difference of 0.0%.