Kem C. Gardner
POLICY INSTITUTE
THE UNIVERSITY OF UTAH
DAVID ECCLES SCHOOL OF BUSINESS

Experimental Age, Sex, Race and Ethnicity Projections for Every State for 2030, with Scenarios

Overview

The Kem C. Gardner Policy Institute produced new experimental population projections for 2030 by sex, age, race and ethnicity, and state as part of a contract with the Sorenson Impact Center (SIC).[1] Demographic work at the Gardner Institute does not typically extend beyond Utah. However, due to the need expressed by SIC for data that did not exist in the necessary detail, we produced these experimental population projections by combining less detailed publicly available projections using statistical models.

Experimental projections are available for all 50 states and Washington, D.C., and include low, medium, and high scenarios, providing a range of potential futures for planning.

These projections should be used to make state-level comparisons on the metrics provided. If seeking projections for an individual state, please seek out independent projections produced by that state.

Table 1: Experimental Projection Output Categories

GEOGRAPHIES
- All 50 states
- Washington, D.C., United States

AGE (5-YEAR AGE GROUPS)
- 0-4,
- 5-9,

  …
- 80-84,
- 85 and older

SEX
- Male
- Female

SCENARIOS
- Low
- Medium
- High

RACE AND ETHNICITY
- White, not Hispanic
- Black or African American, not Hispanic
- American Indian or Alaska Native, not Hispanic
- Asian, not Hispanic
- Native Hawaiian or Pacific Islander, not Hispanic
- Two or more races, not Hispanic
- Hispanic (any race)

YEARS
- 2019 (past estimate)
- 2030 (future projection)

Data Products

Excel Data Workbooks

The Policy Institute has produced two Excel data workbooks. The first is the "Totals" workbook, which includes:

- The total population and total population by each race and ethnicity group for each state, Washington D.C., and the nation
- 2019 (estimate) and 2030 (projection) for each group and state
- One worksheet for each scenario - low, medium, and high

The second workbook contains detailed data designed for more experienced users, and includes:

- Experimental 2030 projections for each scenario-geography-race-age-sex combination.
- Vintage 2019 Census Bureau population estimates (for comparison).[2]
- One worksheet for each scenario – low, medium, and high
- Details on variable names and coding
- 51,408 data points (3 scenarios and 1 historical timepoint, 51 geographies, 2 sexes, 18 age groups, and 7 race data points and ethnic groups, the table includes 4*51*2*18*7 data points)

Interactive visuals

Several interactive dashboards generated in Tableau are available online here. These provide an easy way to visually explore potential demographic trends between 2019 and 2030.

Current and Potential Uses

These experimental projections were requested and funded by SIC, which is using the data as inputs into their MAPS project. The project uses models to plan for future higher educational demand; equity, diversity, and inclusion; and finance. The finer demographic detail and alternative scenarios provide a fuller picture of the nation's future higher education needs, especially in terms of the increasingly diverse racial, ethnic, and age makeup of students.

The primary purpose of these experimental projections is to inform planning or research, with the ability to compare across the 50 states and the District of Columbia. These can fill the need for age, sex, race and ethnicity projections for multiple states. If you are looking for data for a single state, and do not need cross-state comparisons, many states provide their own projections that use local data sources and expert knowledge.

The high and low experimental projections can help inform planning for a range of future scenarios. A different set of age, sex, and race and ethnicity projections for Utah produced by the Gardner Institute in 2018 illustrated how Utah's overall higher education system and workforce would be affected if current educational attainment patterns continue.[3]

Limitations

The most important limitation is the lack of the detailed 2020 decennial census data needed to make more accurate projections. Data and models can incorporate this new information once the Census Bureau releases the detailed datasets in 2023. Additionally, the Weldon Cooper Center and U.S. Census Bureau plan to update their projections once the 2020 census data are released.[4] For this reason, we have only produced a single year of projections (2030). The new census data will make it feasible to produce projections for each year up until 2030.

Additionally, the race-ethnic groups are restricted to the standard seven OMB categories. This makes them consistent with the Integrated Postsecondary Education Data System (IPEDS) data used for higher education research and planning. While this race-ethnicity grouping procedure is standard practice, aggregating race-ethnic groups into these categories masks important variation. Using different categories produces different results, which can influence planning, media reports, and public perception of the future.[5] Race and ethnicity are social concepts that change with history, so it is possible these standard planning categories might change by 2030.[6]

We have included different scenarios to help communicate the future's uncertainty. There is no way to predict our demographic future with absolute certainty or perfectly accurate statistical confidence bounds.

This is a unique experimental product for the nation, 50 states, and the District of Columbia. If you are looking for data focused within Utah, please use our previously published Utah Race/Ethnicity Projections.

Data and Methodology

Data Sources

Table 2. Data sources

Product	Main and alternative population projections, vintage 2017 [a]	Population estimates by state, age, sex, and race and ethnicity, vintage 2019 [b]	Population projections by state, age, and sex, vintage 2018 [c]
Producer	U.S. Census Bureau	U.S. Census Bureau	Weldon Cooper Center for Public Service
Application	Total population, and marginal race-ethnic distributions	Association between geography and race-ethnicity.	Marginal state-level distributions.

a. [U.S. Census Bureau. 2017 National Population Projections Datasets.]

b. [U.S. Census Bureau. 2020. Annual State Resident Population Estimates for 6 Race Groups (5 Race Alone Groups and Two or More Races) by Age, Sex, and Hispanic Origin: April 1, 2010 to Jul 1, 2019.]

c. [University of Virginia Weldon Cooper Center, Demographics Research Group. (2018). National Population Projections.]

Notes: None of these data sources has incorporated the detailed 2020 decennial census data. Those data are not yet available in the necessary detail. We hope to update when the full data become available. For that reason, we have labeled these projections "experimental."

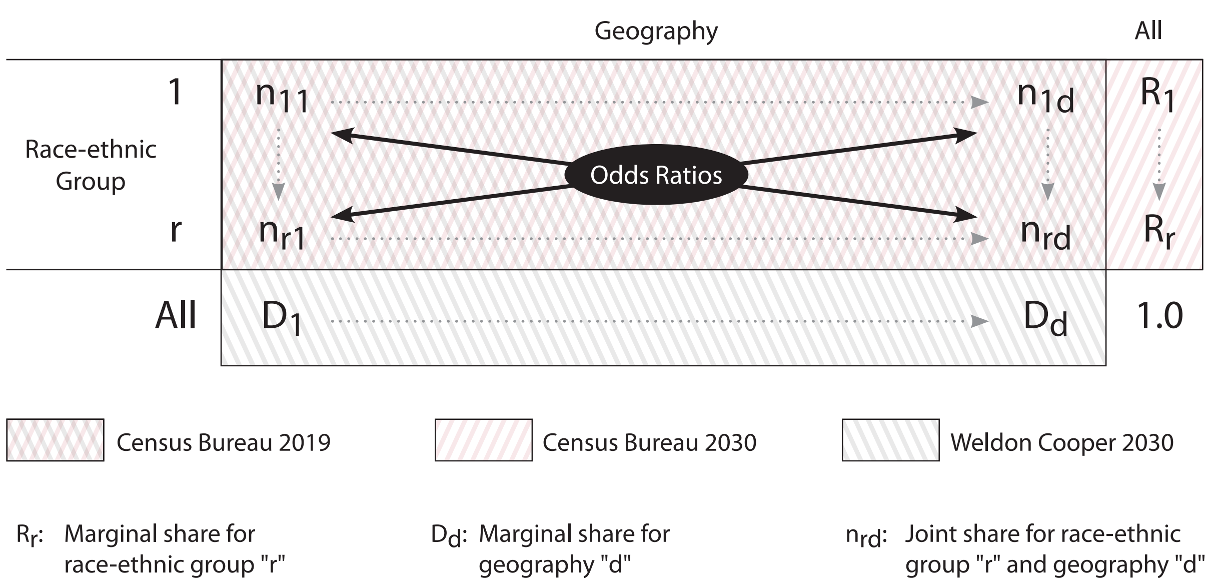
Summary Methods

A more detailed summary is available in Appendix A. Figure 1 diagrams a summary methodology.

Steps:

1. We start with the Census Bureau's projected population total and the projected share in each of the seven OMB race and ethnicity categories.
2. We combine this with the Weldon Cooper Center's projected share in each geography.
3. We model the race-by-geography association
   a. If race and geography were statistically independent (i.e., assumed race-ethnic groups do not cluster in different regions of the country), then the two marginals could simply be multiplied together. But, they are not independent.
   b. We modeled the race-geography dependence by estimating interaction terms from the 2019 census population estimates using log-linear models and applied them to the 2030 marginal shares using iterative proportionate fitting (IPF), summarized in Figure 1.
4. We implemented this procedure *(total * race share * geography share * interaction, followed by IPF)* separately for each of the 108 age-sex-scenario groups.
5. To facilitate quality control, we first modeled the nine census divisions, then modeled each division's states.

Figure 1: Experimental Projection Modeling Approach



Note: Shares are then multiplied by total Census Bureau 2030 population for the age-sex-scenario group
Source: Kem C. Gardner Policy Institute

Appendix A. Detailed Methodology

The objective is to project population (*N*) for the year 2030 by age (*A*), sex (*S*), and race-ethnic (*R*) categories for U.S. States (*D*). There are $a=18$ age groups, $s=2$ sex groups, $r=7$ race-ethnic groups, and $d=51$ geographic subdivisions (See Table 1). This yields a total of $a*s*r*d=12,852$ data points. For three scenarios, this becomes *51,408* data points.

At a given time point, *t*, each of these data points is represented by the notation $N_{ASRD}(t)$. Each is the product of the grand total population (*GTOT*) and the share in that group (lower-case *n*):

$$N_{ASRD}(t) = GTOT(t) \times n_{ASRD}(t). \quad (1)$$

We already have good external projections of the age-sex populations (to be explained below). While they unfortunately do not incorporate the latest decennial census data, due to delays in the Census Bureau production schedule, they are among the most reliable 2030 population projections currently available. Taking those projections as fixed, we can make Eq. 1 specific to each age-sex group, and it simplifies to

$$N_{RD}(t) = TOT(t) \times n_{RD}(t), \quad (2)$$

where *TOT* is simply the total population for that age-sex group, and n_{RD} the share that is race *R* and state *D*. This is the model we need to specify for each of the 36 age-sex groups. It can be formulated as a multiplicative hierarchical log-linear model[7]:

$$N_{RD}(t) = TOT(t) \times R_i(t) \times D_j(t) \times RD_{ij}(t). \quad (3)$$

TOT determines the level; the terms R_i and D_j are the marginal race and geographic distributions, respectively; and RD_{ij} is the "interaction" effect[8]. The marginal effects can be viewed as the "allocation structure," and the interaction effects as the "association structure"[9]. The allocation and association (expressed as odds ratios) structures can vary independently of each other[10]. The interaction term could be disregarded if race and geography were uncorrelated; but that is not the case.

We can statistically combine datasets by taking the different terms from different sources. In our application, we already had two independent population projections for 2030, though not in the full demographic detail required.

1. The Census Bureau published national projections by age, sex, and race[11].
2. The Weldon Cooper Center for Public Service published state projections by age, sex, and geography[12].

The two sources each provide their own population total and 36 age-sex distributions for 2030. We use the Census Bureau's for three key reasons.

1. They have incorporated detailed data and assumptions about international migration that are tied to projections on international populations. They have also incorporated detailed data and assumptions regarding fertility and mortality rates by age, sex, race, ethnicity, and nativity.
2. They have already produced low and high scenarios. We can easily apply our model to these scenarios to get a range of possible projections. The scenarios are based upon different international migration assumptions. Migration is the most unpredictable demographic component, and will powerfully impact future race-ethnic demographics.
3. The Census Bureau projections are more well-known and utilized.

For each age-sex group, the *TOT* term came from the Census Bureau data, as did the marginal race distribution. The Weldon Cooper Center data formed the marginal distribution for geography. Only the interaction term remained. Following the approach of Rogers et al.[13], we took this "auxiliary information" or "offset" term from a previous time point—2019, the last published Census Bureau data with the necessary detail[14].

Adding a superscript to denote the data source (1=Census Bureau, 2=Weldon Cooper) and specifying the time points (2=2030, and 1=2019), our estimate for Eq. 3 becomes

$$N_{RD}(2) = TOT^1(2) \times R_i^1(2) \times D_j^2(2) \times RD_{ij}^1(1). \quad (4)$$

This is diagrammed in Figure 1 above. First, the procedure involves solving for the interaction term in 2019 by dividing the left-hand side by the non-interaction terms on the right. That interaction term is then multiplied by the totals and marginal distributions from the various 2030 data sources to produce the projections. Iterative proportional fitting is used to ensure the final numbers are consistent. For more details on calculation, refer to the several readings cited in the endnotes.

The method is very adaptable. For the high/low scenarios, we utilized the *TOT* and R_i terms from the respective Census Bureau high/low projections scenarios. It is easy to incorporate other projections, if desired. All the input data we used are publicly available online.

Even in parsimonious models, a lot can go wrong. Data visualization is critical to capture anomalies. Going straight from one region to 51 made this vetting and visualization process especially difficult. Therefore, we implemented the above model in two stages. First, we went from the U.S. to nine census divisions (one model run for each scenario). After quality control, we then went from each of the nine divisions to the constituent states (nine model runs for each scenario)[15].

Endnotes

1. Race and ethnic groups are the exhaustive and mutually exclusive categories used by the U.S. Office of Management and Budget (OMB). They include (1) non-Hispanic White alone, (2) non-Hispanic Black or African American alone, (3) non-Hispanic American Indian or Alaska Native alone, (4) non-Hispanic Asian alone, (5) non-Hispanic Native Hawaiian or Pacific Islander alone, (6) non-Hispanic Two or More Races, and (7) Hispanic, any Race. See U.S. Census Bureau. (2022). *About the Topic of Race.* Washington, D.C. https://www.census.gov/topics/population/race/about.html .
2. When these experimental projections were produced, we judged the 2019 estimates to be the most reliable data available. Our models are unlikely to be improved considerably by more recent data until the detailed 2020 census data tables are published in late 2023.
3. Curtin, J. A. (2019). *Utah's Growing Opportunity Gap: The Impact of Shifting Demographics on Utah's Postsecondary Educational Attainment.* Issue Brief No. 2019-4. Utah System of Higher Education. Salt Lake City, Utah. https://eric.ed.gov/?q=source%3A%22Utah+System+of+Higher+Education%22&id=ED601919. Also, Hollingshaus, M., Harris, E., & Perlich, P. S. (2019). *Utah's increasing diversity: population projections by race/ethnicity.* Kem C. Gardner Policy Institute. https://gardner.utah.edu/demographics/population-projections/raceethnicity-projections/ .
4. Interagency communications.
5. See Alba, R. (2018). *What Majority-minority Society? A Critical Analysis of the Census Bureau's Projections of America's Demographic Future.* Socius, 4. Also, Myers, D., & Levy, M. (2018). *Racial Population Projections and Reactions to Alternative News Accounts of Growing Diversity.* The ANNALS of the American Academy of Political and Social Science, 677(1), 215–228
6. Census Bureau National Advisory Committee on Racial, Ethnic, and Other Populations. (2022). *What 2020 Census Results Tell Us About Persisting Problems with Separate Questions on Race and Ethnicity in the Decennial Census.* Washington, D.C. https://www2.census.gov/about/partners/cac/nac/meetings/2022-05/presentation-what-2020-census-results-tell-us.pdf
7. The same procedure could be done for the full model in Eq. 1. But, there are so many interaction terms, including a four-way interaction, that the model is difficult to work with. Since we already had the age-sex control totals, we simply implemented 36 sub-models that were easier to manage and interpret. For an example full-model approach, see: Raymer, J., Bai, X., & Smith, P. W. F. (2020). *Forecasting Origin-Destination-Age-Sex Migration Flow Tables with Multiplicative Components.* In S. Mazzuco & N. Keilman (Eds.), *Developments in Demographic Forecasting*. Springer.
8. Taking the natural log, Eq. 3 can also be expressed as a generalized linear model with log(TOT) as the intercept; and hence the terminology "log-linear."
9. Koch, G. G., Freeman, D. H., Jr., & Tolley, D. (1975). *The Asymptotic Covariance Structure of Estimated Parameters from Contingency Table Log-Linear Models* (Institute of Statistics Memo Series No. 1046). University of North Carolina at Chapel Hill. p. 35.
10. Rudas, Tamás. (1998). *Odds Ratios in the Analysis of Contingency Tables*. Thousand Oaks, CA: SAGE Publications.
11. U.S. Census Bureau. *2017 National Population Projections Datasets*. https://www.census.gov/data/datasets/2017/demo/popproj/2017-popproj.html .
12. University of Virginia Weldon Cooper Center, Demographics Research Group. (2018). *National Population Projections.* https://demographics.coopercenter.org/national-population-projections .
13. Rogers, A., Little, J., & Raymer, J. (2010). *The indirect estimation of migration: Methods for dealing with irregular, inadequate, and missing data*. Springer Science+Business Media.
14. Tests on historical data showed that holding the 2010 interactions terms constant predicted 2019 better than more complicated models that included a linear or log trend from 2000-2010.
15. For a map of census divisions and states, see https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf .

Authored by: Mike Hollingshaus, Senior Demographer; Emily Harris, Senior Demographer

(DE) DemUserGuire May2022