Technical Notes:
July 1, 2000 Town-level Bridged Race Estimates for Connecticut

The July 1, 2000 town-level ASRH (vintage 2004) inter-censal population file contains estimates of the resident population of Connecticut as of July 1, 2000 by town, five-year age group (0-4, 5-9, 10-14 … 85 years and over), sex, and race/ethnicity category (White non-Hispanic, Black or African American non-Hispanic, American Indian or Alaska Native non-Hispanic, Asian or Pacific Islander non-Hispanic, and Hispanic or Latino of any race). These estimates were produced by the Health Information Systems and Reporting section of the Department of Public Health and were made available in 2007.

Purpose

Population estimates by age, sex, race and Hispanic ethnicity (ASRH) are invaluable for the estimation of rates of health and illness in Connecticut's towns and counties. Each year, the population as of July 1 for the state, counties, and towns is estimated through a collaboration between the U.S. Census Bureau (USCB) and the CT Department of Public Health. These figures provide a mid-year population estimate that can be utilized for year-to-year comparisons. Unfortunately, the annual town population estimates do not include the ASRH components.

The decennial Census in 2000 estimated the population of Connecticut's counties and towns as of April 1, 2000 with the desired ASRH demographic components. The population change between April 1 and July 1 is substantial enough to prohibit the direct substitution of demographic counts; however, the proximity of the population estimates makes a derivation of July 1, 2000 demographic estimates feasible – if only for the year 2000. To address this issue, we reviewed the data sources available and determined that July 1, 2000 town population estimates by ASRH could be reasonably derived with the use of several USCB datasets, a process called 'bridging', and the statistical method of iterative proportional fitting. 'Bridging' is the term used for the process of making data collected using one set of race categories consistent with data collected using a different set of race categories.[1] More specifically, race bridging is a method used to make multiple-race and single-race data collection systems sufficiently comparable to permit estimation and analysis of race-specific statistics. Iterative proportional fitting (IPF) is the process of adjusting a table of values (in this case, the April 1 population by ASRH) so that the rows and columns sum to specific totals (in this case, total town population and county population by ASRH). For our 2000 data, the relative proportions of each demographic variable to one another in the April 1, 2000 estimates will be applied to the larger population totals for July 1, 2000.

The development of this 'bridged' July 1, 2000 dataset allows rates of health and illness to be calculated at the town and county level using the mid-year population rather than the lower population estimate from the Census in April. Also, by using the mid-year estimate, rates calculated using the population denominators will be consistent with the official state and town estimates of the population for 2000 and provide a consistent reference point with the annual inter-censal population estimates.

Methods

The first step in creating a July 1, 2000 town-level dataset by ASRH is to prepare each of the source datasets for use in the IPF procedure. This requires making the ASRH variables in each of the source datasets consistent with each other as well as making the variables consistent with the needs of Connecticut towns. In Connecticut, race is often collected and tabulated based on four race groups defined by the U.S. Office of Management and Budget in 1977 (OMB77): White, Black or African American, American Indian or Alaskan Native, and Asian.[2] Meanwhile, the Census 2000 data uses five race groups defined by U.S. Office of Management and Budget in 1997 (OMB97; White, Black or African American, American Indian or Alaskan Native, Asian, Native Hawaiian or Other Pacific Islander)[3] and a sixth category for 'some other race'. In order to proportionally fit the April 1, 2000 Census data to the July 1, 2000 data needed for Connecticut, the Census 2000 data needs to be reclassified into the same four race categories that Connecticut utilizes. No reclassification was necessary for age, sex, or Hispanic ethnicity.

July 1, 2000 Town totals: Each year, the U.S. Census Bureau (USCB) publishes total resident July 1 population estimates for Connecticut's 169 towns. These annual estimates are referred to as inter-censal estimates as they provide population counts following the decennial census.[4] For this project, we used the July 1, 2000 town population estimates from the vintage 2004 annual estimates of the population for Minor Civil Divisions in Connecticut.[5] The vintage 2004 estimates incorporated Count Question Resolution (CQR) [6] corrections through May 2004 into the July 1, 2000 town-level estimates.

July 1, 2000 County-level ASRH totals: The National Center for Health Statistics (NCHS) releases bridged-race inter-censal population estimates of the resident population for each county in the United States by single year of age, sex, bridged race, and Hispanic origin. These estimates result from bridging the 31 race categories used in Census 2000 to only four race categories. A detailed explanation of how NCHS bridges the Census 2000 race categories is available online through the NCHS.[7] For this project, we used the NCHS bridged-race population estimates for July 1, 2000 vintage 2004.[8] The vintage 2004 estimates incorporated Count Question Resolution (CQR) corrections through May 2004 into the July 1, 2000 county-level estimates. Population estimates by age, sex, four OMB77 race categories (White, Black or African American, American Indian or Alaskan Native, and Asian or Pacific Islander), and Hispanic origin for Connecticut's eight counties were extracted from this dataset.

April 1, 2000 Town-level ASRH totals: For the Census 2000, the USCB collected race using the five OMB97 race categories with two major allowances: 1) respondents were permitted to select more than one race, and 2) respondents were provided an additional "some other race" category with a write-in option if he or she did not identify with any of the five pre-defined races. These two allowances prevent the direct use of the Census 2000 race data with other datasets that use the OMB97 standards. To address the second allowance, the USCB developed a procedure to re-assign those who wrote in a response for "some other race" to one of the five OMB97 races.[9] This *Modified Race* dataset contains the April 1, 2000 population data using 31 single or multiple race combinations of the five OMB97 race

categories. For this project, we used the vintage 2004 version of the Modified Race estimates which incorporated Count Question Resolution (CQR) corrections through May 2004.[10]

Bridging:
We still needed to bridge the Census 2000 Modified Race data to the four OMB77 race categories NCHS uses for their July 1, 2000 county-level estimates. We contacted NCHS for the county-level allocation factors that they used to convert the multiple races to single races for July 1, 2000.[11] We then proportionally reallocated (i.e., bridged) each multiple race category into its component single race groups for each town within each county. The end result was a dataset containing the population estimate for April 1, 2000 for each of the 169 towns by age, sex, 4 bridged-race OMB77 categories, and Hispanic ethnicity.

Iterative Proportional Fitting:
The next step in creating a July 1, 2000 town-level dataset by ASRH is to apply the April 1, 2000 bridged ASRH proportions to the July 1, 2000 town and ASRH totals. Iterative Proportional Fitting (IPF), sometimes referred to as 'raking', is a procedure for adjusting a two-dimensional table of data cells such that the new values in each column and in each row sum to pre-determined totals.[12] The unadjusted data cells may be referred to as 'seed' cells and the pre-determined totals may be referred to as the 'marginal' totals. IPF acts as a weighting system whereby the original table values are gradually adjusted through repeated calculations to fit the row and column constraints. IPF is employed in various disciplines but has been particularly useful in census-related analyses to provide updated population statistics and to estimate individual-level attribute characteristics.[13]

IPF is feasible because of the three datasets we have available. The April 1, 2000 dataset provides the 'seed' proportions that we want expanded to the larger July 1, 2000 population estimate. The July 1, 2000 town dataset provides the town totals (i.e., rows) to which we want all of the April 1, 2000 town-level demographic groups to conform. It is conceivable that we could use just these two datasets to inflate one dataset into the other; however it is unlikely that the population increases occurred equally amongst each of the demographic groups. The inclusion of a third dataset, the July 1, 2000 county population by ASRH, provides an added control (i.e., columns) by providing an estimate of the demographic distribution for July 1, 2000 even though the estimates are aggregated to the county level. This third dataset will help control for the unequal change between demographic groups although it will not control for unequal change between towns within each county. In sum, the seed demographic distribution for each town will be proportionally fitted to the town population total and to the demographic distribution of the county.

The Enhanced Raking Macro (eRake) for the statistical software SAS was developed by David Izrael, Abt Associates, in June 1999 to compute weights for a sample to make it agree with independent control totals on a number of characteristics.[14] To reduce the number of 'seed' cells with a null value (i.e., zero population), we collapsed individual years of age into 18 five-year age groups and all Hispanic cases into a single "race/ethnicity" group. The final aggregated 'seed' dataset contained 18 five-year age groups, 2 sexes, and 5 race/ethnicity groups. To keep the marginal totals consistent, the July 1, 2000 NCHS county-level ASRH control totals were also collapsed into the same age, sex, and race/ethnicity groups prior to

raking. Since the ASRH totals and the town totals are specific to each county, we ran each county separately through the eRake macro. We set the convergence threshold for the raking to 0.01 which means the row and column totals must be within 0.01 of the control totals. One of the eight counties failed to converge at 0.01, so for this county, a convergence of 0.05 was used.

The final raked datasets contain fractional values for each of the ASRH cells. Although these fractional ASRH values sum to an integer for town totals (i.e., by row) and county ASRH totals (i.e., by column), population estimates require integers (i.e., whole person units) for each of the ASRH cells. Simple rounding of the ASRH values results in 1) town totals that do not sum to the county population and 2) ASRH group totals that do not sum to the county population. To address the need for integer ASRH cell values, we created a program in SAS IML to round the ASRH cells in such a way that single cell changes are balanced within the rows and columns. This results in a process that retains the correct row and column totals while converting all fractional cell values to whole person units.

After the population estimates for each of the eight counties had been raked and rounded, the estimates for all eight counties were appended together to create a single "bridged 2000" dataset with ASRH estimates for each of Connecticut's 169 towns.[15]

Suggested Citation:
Mueller LM, Backus KM, Stone CL. (2007) Town-level bridged race estimates for Connecticut, July 1 2000, Connecticut Department of Public Health, Health Care Quality Statistics Analysis & Reporting, Hartford, CT.

---

[1] National Center for Health Statistics. Documentation for Bridged-Race Vintage 2004 (July 1, 2000 - July 1, 2004) Postcensal Population Estimates for Calculating Vital Rates. Available on the Internet on September 9, 2005.

[2] Office of Management and Budget. Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting. Released May 12, 1977.

[3] Office of Management and Budget. Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Federal Register Notice, Vol. 62, No. 210, Thursday, October 30, 1997.

[4] Population Division, US Census Bureau. Methodology for the Subcounty Total Resident Population Estimates (Vintage 2004): April 1, 2000 to July 1, 2004. Released June 30, 2005.

[5] Population Division, US Census Bureau. Table 5: Annual Estimates of the Population for Minor Civil Divisions in Connecticut, Listed Alphabetically Within County: April 1, 2000 to July 1, 2004 (SUB-EST2004-05-09). Released June 30, 2005.

[6] The Count Question Resolution (CQR) Program was a planned administrative review program that handled external challenges to particular official Census 2000 counts of housing units and group quarters population received from state, local or tribal officials of governmental entities or their designated representatives in the U.S. and Puerto Rico. Census 2000 CQR corrections were published in the Census 2000 Notes and Errata. CQR corrections resulted in a net increase of only 37 in the Census 2000 base for Connecticut; however, the population change for some towns was substantial enough to require that corrected counts be used to minimize

error. The vintage 2004 estimates were the most current estimates available at the time this dataset was developed.

Date of official USCB letter	From	To	Count Changes	Estimate year correction is expected to be reflected in USCB Estimates
11/30/01	W Hartford	Hartford	2,543	2002
09/23/02	East Hampton	Middletown	2,396	2003
09/23/02	Coventry	Mansfield	36	2003
09/23/02	Tolland	Mansfield	60	2003
12/16/02	Waterford	New London	514	2003
01/09/03		Roxbury	1	2003
01/09/03	New Milford	Washington	23	2003
01/09/03	Woodbury	Washington	2	2003
01/09/03		Washington	18	2003
10/01/03		Groton	18	2004
12/29/03	Hamden	New Haven	150	2004
06/30/05	Plainville	Bristol	125	2005
Highlighted corrections reflect increases in town populations with no accompanying decrease in another town; the result is an increase in the total state population.				
Note: USCB CQR Resolutions report a shift of 722 from Groton City to Groton Town. This does not affect the total town population.				

[7] Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. National Center for Health Statistics. Vital Health Stat 2(135). 2003.

[8] National Center for Health Statistics. Estimates of the July 1, 2000-July 1, 2004 , United States resident population from the Vintage 2004 postcensal series by year, county, age, sex, race, and Hispanic origin, prepared under a collaborative arrangement with the U.S. Census Bureau. Available on the Internet on September 9, 2005.

[9] US Census Bureau. Census 2000 Modified Race Data (MR(31)-CO.txt):Technical Documentation. November 27, 2002.
A detailed explanation of race modifications can be found in the Modified Race Data Summary File 2000 Census of Population and Housing Technical Documentation.

[10] For a detailed review, see: Stone C, Mueller L. (2002) Effect of USCB-Generated Race Modifications on 2000 Census Race Allocations in Connecticut. Health Care Quality Statistics Analysis and Reporting, HISR, Planning Branch, Connecticut Department of Public Health.

[11] Personal communication with NCHS resulted in bridging constants for each of the multiple race categories by ASRH to be provided for Connecticut counties. The allocation factors were by county, age, sex, and Hispanic ethnicity for each of the 31 single and multiple race groups.

[12] Deming W, Stephan F. (1940) On least square adjustment of sampled frequency tables when the expected marginal totals are known. Annals of Mathematics and Statistics, 6, 427–444.

[13] Norman, P. (1999) Putting iterative proportional fitting on the researcher's desk. Working Paper 99/3. School of Geography, University of Leeds, Leeds.

14 Izrael D, Hoaglin DC, and Battaglia MP. (2000) A SAS Macro for Balancing a Weighted Sample. Proceedings of the Twenty-Fifth Annual SAS® Users Group International Conference (SUGI 25), Paper 258-25. Cary, NC: SAS Institute Inc.

15 Although efforts were made to use the best available data and methods to produce the bridged estimates, the modeling process introduces error into the estimates. The potential for error will be greatest for the smallest population groups, particularly the smaller race groups and town level estimates.