COUNTY-SPECIFIC LIFE TABLES

Balkrishna Kale, David Egan-Robertson
Wisconsin Department of Administration
and
Charles Palit, Paul Voss
University of Wisconsin-Madison

KEYWORDS: Survival Probability, Small Area Estimates and Projections, Life Expectancy

The Problem

County-specific life tables are usually difficult to create. Detailed mortality data by age and sex are often not available by county; only the total number of deaths of males and females may be known. Moreover, the problem of small numbers arises even when detailed data are available. Many counties have small populations: for example, 24 of 72 counties in Wisconsin—one out of three—had less than 25,000 persons in 2000; four of them had populations under 5,000. Furthermore, the number of deaths in many age-sex groups may be minimal—even for a decade, not just in a year—and therefore subject to a large random variation. The values would be even smaller if disaggregated by race.

A state life table is not as difficult to construct. However, it would not accurately reflect the life expectancy in a particular county. Survival chances would be expected to vary among counties as a result of environmental, health care, socio-economic and compositional differences. In particular, a county's racial composition is an important element because life expectancy differences by race are well known.

The Solution

The solution presented in this paper enables the estimation of county-specific life tables for the midpoint of a decade. It begins with a calculation of age-sex specific state-level decennial survival rates for whites and non-whites separately. (For this purpose, it is possible to use life tables that might be readily available for the midpoint of the decade. Note that life tables from other years are also required in this case for surviving the births of the inter-censal period.) The state-level survival rates used in this exercise were based on an extension of the vital statistics method (Kale, et. al., 1993).

Next, state-level rates by race are combined into a single set of age-specific preliminary county rates using white/non-white population shares as weights within each age group in the county. Carried out for males and females separately, these preliminary rates are then applied to the county's population at the first census and the county's births in the first and second half of the inter-censal period. This step enables the calculation of a preliminary estimate of the county's resident deaths by age for males and females. This estimate includes deaths attributable to migrants that are derived using a procedure detailed in a previous paper (Kale, et. al., 1995).

The preliminary number of resident deaths by age as calculated in the above step is iteratively adjusted within each sex using (a) the ratio of actual total deaths to total calculated deaths in that county and (b) the ratio of actual total deaths in the state of a specific age cohort to the calculated total deaths of that age across all counties. This step provides the final estimate of resident deaths by age and sex in each county.

The proportions as given by the preliminary calculations of deaths among the county's residents at the beginning of the decade (and births of the inter-censal period) on the one hand and those attributable to the migrants on the other are used to disaggregate the final estimate of resident deaths between the two categories. A final estimate of the expected population at the end of the decade in the absence of any migration is then derived by subtracting from the residents at the decade's beginning (and births of the inter-censal period) the final estimate of deaths among them. The final decennial survival rates are then given by dividing the final estimate of the expected population by the population at the first census (or births of the inter-censal period).

County-specific mid-decade stationary population is calculated on the basis of these final decennial survival rates. This enables the computation of the county's male and female life expectancy for the midpoint of the decade.

The Data

Data for Wisconsin and its counties are used here for illustration. Population counts by race groupings (i.e., white and non-white), age and sex are drawn from the 1990 and 2000 Census files. We have used the 1990 census data by race and age as modified by the Census Bureau (U.S.

Bureau of the Census, 1990 CPH-L-74). These 1990 modified data (the MARS file) provide more accurate classifications.

Presumably, age misreporting was less serious in 2000. However, race classification changed significantly in the 2000 census. Unlike in 1990, when only a single race could be chosen or assigned, people were able to report their race, if they so chose, as a combination of two or more.

No adjustment is made here to make the 2000 data comparable to the 1990 MARS file. Those reporting "white alone" as their race are considered as the 2000 counts of the white race. All others, including those who gave "other" as their race, are classified as non-white. This dissimilarity on race between 1990 and 2000 introduces errors in the survival rates calculated for whites and non-whites. However, these errors are of a counterbalancing nature when the two sets are combined into one with appropriate weights based on their shares in the total population in each cohort.

White and non-white male and female births of the two five-year periods in the 1990s by county form the second set of data and the 1990-2000 male and female resident deaths by five-year birth cohorts constitute the final set. The proposed solution to the stated problem uses total male and female resident deaths only at the county level and a detailed age distribution of deaths just at the state level. The detailed age distribution of deaths at the county level is utilized for evaluating the suggested model.

State-Level Survival Rates

The vital statistics method was extended to derive the decennial state-level survival rates for this exercise. (For a detailed discussion of this procedure see Kale, et. al., 1993).

The familiar demographic equation $\mathbf{P_2 = P_1 - D + M}$ may be rewritten for any age cohort as follows:

$$\mathbf{M = P_2 - P_1 + D},$$

where $\mathbf{P_1}$ is the population of age x at the first census, $\mathbf{P_2}$ is the population of age x+10 at the second census, $\mathbf{D}$ is the number of resident deaths of persons (which includes the deaths of non-migrants as well as in-migrants to the state during the decade) who were of age x at the time of the first census, and $\mathbf{M}$ is net migration during the decade among persons of age x at the first census.

For children under 10 at the second census, $\mathbf{P_1}$ in this equation would refer to the inter-censal births. Net migration ($\mathbf{M}$) given by this equation includes migrants who were alive at the time of the second census as well as those who had died after their migration.

If all migration occurs at the decade's beginning, then the inter-censal survival rate ($\mathbf{s_1}$) is given by:

$$\mathbf{s_1 = P_2 / (P_1 + M)}$$

If all migration occurred at the decade's end, the survival rate ($\mathbf{s_2}$) is

$$\mathbf{s_2 = (P_2 - M) / P_1}$$

Since the timing of migration is unknown, a reasonable approximation of the inter-censal survival rate ($\mathbf{S}$) would be:

$$\mathbf{S = (s_1 + s_2) / 2}$$

This procedure is used to produce three sets of state-level decennial survival rates for males and females separately: (i) Survival rates for the whites $\mathbf{S_w}$, (ii) Survival rates for others $\mathbf{S_o}$, and (iii) Survival rates for all races together $\mathbf{S}$. Note that

$$\mathbf{S = (S_w * p_w) + (S_o * p_o)},$$

where $\mathbf{p_w}$ is proportion of whites in an age cohort and $\mathbf{p_o}$ is proportion of non-whites in that age cohort. These proportions are calculated as follows:

$$\mathbf{p_w = ((P_{1w,x} / P_{1,x}) + (P_{2w,x+10} / P_{2,\,x+10})) / 2}, \text{ and}$$
$$\mathbf{p_o = ((P_{1o,x} / P_{1,x}) + (P_{2o,x+10} / P_{2,\,x+10})) / 2},$$

where $\mathbf{P_{1w,x}}$, $\mathbf{P_{1o,x}}$, and $\mathbf{P_{1,x}}$ represent the white, non-white and total population of age x at the first census, and $\mathbf{P_{2w,x}}$, $\mathbf{P_{2o,x}}$ and $\mathbf{P_{2,x}}$ represent the white, non-white and total population of age x+10 at the second census. Births of the first half of the decade are considered along with 5-9 year olds at the second census for this purpose and births of the second half are considered along with 0-4 year olds at the second census.

The age cohort of persons who were 90 and over in 1990 becoming 100 and over in 2000 is sufficiently large at the state level but extremely small in many counties. The decennial survival probability for this group is extremely low. Therefore, it was decided to have persons 80 and over in 1990 becoming 90 and over in 2000 as the oldest age group for this study.

Preliminary County-Specific Estimates

(1) All calculations at the county level, here and elsewhere, are done for males and females separately. Decennial survival rates ($\mathbf{S'_c}$) for each age group are obtained as follows:

$$\mathbf{S'_c = (S_w * p_{w,c}) + (S_o * p_{o,c})}$$

where $\mathbf{p_{w,c}}$ and $\mathbf{p_{o,c}}$ are proportions of whites and non-whites in a county's total population in an age group. These proportions are calculated in the same way as explained at the state level.

(2) Survivors in 2000 if no migration occurred during the decade are given by:

$$\mathbf{P_{e}'_{,c} = P_{1,c} - (P_{1,c} * S'_{c})}$$

where $\mathbf{P_{e}'_{,c}}$ is the county's surviving (or expected) population in 2000 of age x+10 (in the absence of migration) and $\mathbf{P_{1,c}}$ is the county's population in 1990 of age x. For 0-4 and 5-9 year olds in 2000, $\mathbf{P_{1,c}}$ represents county births in the 1995-2000 and 1990-1995 periods respectively.

(3) Surviving migrants in 2000 of any age group are obtained as follows:

$$\mathbf{ML'_{c} = P_{2,c} - P_{e}'_{,c}}$$

where $\mathbf{ML'_{c}}$ represents a county's surviving migrants and $\mathbf{P_{2,c}}$ is the county's population in 2000, both of the same age as of the expected population ($\mathbf{P_{e}'_{,c}}$).

Preliminary Estimate of County Resident Deaths

(1) The preliminary estimate of a county's resident deaths comprises two components. These calculations are done by age. The first component $\mathbf{D_{e}'_{,c}}$ gives the preliminary estimate of deaths among 1990 residents and among births of the inter-censal period:

$$\mathbf{D_{e}'_{,c} = P_{1,c} - P_{e}'_{,c}}$$

where $\mathbf{P_{1,c}}$ is the county's population of age x in 1990, $\mathbf{P_{e}'_{,c}}$ is the county's expected population of age x+10 in 2000 in the absence of migration and $\mathbf{D_{e}'_{,c}}$ is the number of inter-censal deaths among the county's residents of age x in 1990. For age groups 0-4 and 5-9 in 2000, $\mathbf{P_{1,c}}$ would represent the births of the 1995-2000 and 1990-1995 periods, $\mathbf{P_{e}'_{,c}}$ the expected number of 0-4 and 5-9 year olds, and $\mathbf{D_{e}'_{,c}}$ would be the deaths among children born in the two 5-year periods.

(2) Then there is $\mathbf{D_{m}'_{,c}}$, the preliminary estimate of deaths among the migrants:

$$\mathbf{D_{m}'_{,c} = M'_{c} - ML'_{c}}$$

where $\mathbf{M'_{c}}$ represents for an age cohort the migrants including those who died, $\mathbf{ML'_{c}}$ the surviving migrants and $\mathbf{D_{m}'_{,c}}$ the number of deaths among those migrants. Now $\mathbf{M'_{c}}$ is given by:

$$\mathbf{M'_{c} = (ML'_{c} / ((S'_{c} + 1)/2)) - ML'_{c}}$$

where $\mathbf{(S'_{c} + 1)/2}$ is the migrant survival rate. Mortality risks are assumed to be the same for migrants as for others. If they came at the very beginning of the period, their survival rate is S'_{C}. If they came at the very end of the period, their survival rate would be 1. Thus, similar to the calculation of **S** at the state level, the migrant survival rate is the average of the starting and finishing survival rates.

(3) We get all resident deaths ($\mathbf{D'_{c}}$) for any age group by adding the two components:

$$\mathbf{D'_{c} = D_{e}'_{,c} + D_{m}'_{,c}}$$

Final Estimate of County Deaths and Survival Rates

The preliminary estimate of resident deaths by age given by the above procedure is iteratively adjusted (within each sex) using the ratio of (a) actual total deaths to the total estimated deaths of all ages in that county, and (b) actual total deaths in the state of a specific age cohort to the total estimated deaths of that age in all counties. Two iterations were sufficient to make the calculated totals equal to the actual county totals across all age groups on the one hand and the actual state totals for any age group across all counties on the other. This step gives the final estimate of county resident deaths by age ($\mathbf{Dx\wedge_{,c}}$).

The proportion based on the preliminary estimate of deaths among the county's 1990 residents and births of the inter-censal period out of total resident deaths is used to obtain the final estimate of deaths among the 1990 residents and births of the inter-censal period ($\mathbf{Dx_{e}\wedge_{,c}}$). This is done within each group separately as follows:

$$\mathbf{Dx_{e}\wedge_{,c} = Dx_{c} * (Dx_{e}'_{,c} / Dx'_{c})}$$

The final estimate of survivors in 2000 of the 1990 county residents and births of the inter-censal period by age ($\mathbf{Px_{e}\wedge_{,c}}$) is then given by:

$$\mathbf{Px_{e}\wedge_{,c} = P_{1}x_{,c} - Dx_{e}\wedge_{,c}}$$

The final estimates of age-specific survival rates by county ($\mathbf{Sx\wedge_{c}}$) can now be calculated:

$$\mathbf{Sx\wedge_{c} = (Px_{e}\wedge_{,c})/(P_{1}x_{,c})}$$

County Life Tables

Life tables representing the 1990-2000 period can be calculated using the survival rates derived in the manner discussed above. These life tables may be considered as life tables for the midpoint of the ten-year period.

The methods for calculating the L_{x} values for the various age groups are as follows:

- Ages 0-4: The survival rate that relates the county's births of the 1995-2000 period to age 0-4 population in 2000 refers to the midpoint of the 1995-2000 period. The state-level survival rate for this age group in 1995 can be derived by interpolating between the 1990 and 1997 rates. The life table developed by NCHS for 1990 may be used for this purpose. The county-specific survival rate for the 0-4 age group in 1995 is now obtained based on the state-level rate of change in the survival rate of this age group between 1995 and 1997. Finally, the L_x value for the 0-4 age group is estimated by multiplying the survival rate for the 0-4 age group by 500,000.

- Ages 5-9: The survival rate, denoted here as *c*, that relates the births of the 1990-1995 period to age 5-9 population in 2000 may be viewed as a product of two ratios: (i) a ratio *a* by which births of the first half of the decade survive to age 0-4 at mid-decade and (ii) a ratio *b* by which children of age 0-4 at mid-decade survive to age 5-9 at the end of the decade. Ratio *a* may be placed at the midpoint of the first half of the 1990s (representing the average experience of the five-year period) and the second ratio *b* may be placed at the midpoint of the second half. Ratio *a* may be estimated on the basis of the state-level rate of change in the survival rate of this age group between 1992 and 1997. (The state-level survival rate for 1992 is obtained by interpolating between 1990 and 1997.) Ratio *b* is then obtained by dividing the survival rate *c* by *a*. Finally, the county-specific survival rate for the age group in 1995 is based on the state-level rate of change in the survival rate of this age group between 1995 and 1997. (The state-level survival rate for this age group in 1997 is obtained on the basis of the procedure just described and the state-level survival rate for 1995 is given by interpolating between 1990 and 1997.) Finally, the county-level L_x value for the 5-9 age group in 1995 is estimated by multiplying the county's L_x value for the 0-4 age group calculated above by the county's survival rate for the 5-9 age group calculated here.
- Age groups 10-14 through 85-89: The calculated survival rates for these age groups are 10-year survival rates and refer to the midpoint of the decade (representing the entire decade). The L_x value for 0-4 can be multiplied by the survival rate for the 10-14 age group to obtain the L_x value for the latter age group. The L_x value for the 15-19 age group can be obtained similarly by multiplying the survival rate for that age group by the L_x value for the 5-9 age group. This process can then continue up to ages 85-89.
- The open-ended (age 90 and over at 2000) age group: This age group's survival rate can be described as:

  $c = b / (a + b)$

  c is the ratio that survives persons 80 years and over in 1990 to 90 years and over in 2000, the same as relating the L_x values for the two age groups; *a* is the sum of L_x values for the 80-84 and 85-89 age groups; *b* is the L_x value for the 90 and over group. Recasting the equation we have:

  $b = (ac) / (1 - c)$

  Thus, the L_x value for this age group is obtained by solving for $((L_{x,80\text{-}84} + L_{x,85\text{-}89}) * S_{c,90+}) / (1\text{-}S_{c,90+})$.

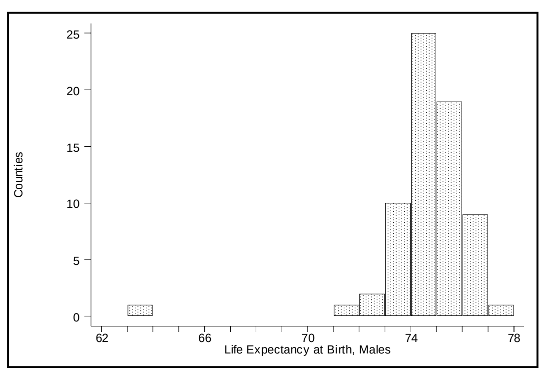
The L_x column is now complete and enables the calculation of the T_x column and life expectancy at birth for the midpoint of the decade. The calculation of county life tables for small populations is thus accomplished even in the absence of detailed mortality data.

The Results

Life tables were developed for 68 counties out of a total of 72 in Wisconsin. Four counties had misallocations of sizeable populations across county boundaries in the 2000 Census. The distribution by race and age of the misallocated persons could not be determined, so these counties were excluded. The state-level survival rates were calculated for the entire state. However, for the iterative adjustment process described above, the actual number of resident deaths by age at the state level included data for 68 counties only.

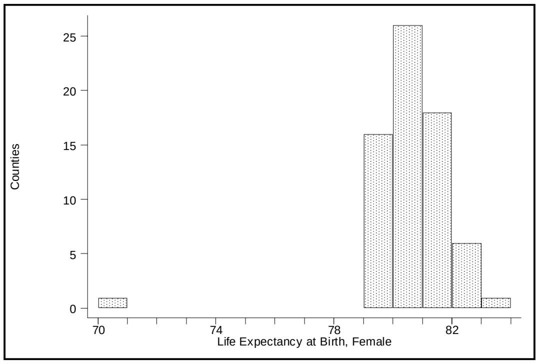
Male life expectancies in 1995 ranged from 74.0 to 74.9 in 24 out of 68 counties and from 75.0 to 75.9 in 21 counties. They ranged from 76.0 to 76.8 in nine counties and was as high as 77.4 in one county. On the low side life expectancies ranged from 73.1 to 73.9 in nine counties, from 71.7 to 72.4 in three counties, and as low as 63.9 in Menominee County, which has a predominantly Native American population (87 percent reported their race as American Indian alone in the 2000 Census).

Male Life Expectancies at Birth, Wisconsin Counties, 1995



Female life expectancy in Menominee County was higher at 70.8, but far below the next county with a female life expectancy of 79.1. Female life expectancies ranged from 79.1 to 79.9 in 15, from 80.0 to 80.9 in 24, 81.0 to 81.9 in 21 and from 82.2 to 83.1 in seven counties.

Female Life Expectancies at Birth, Wisconsin Counties, 1995

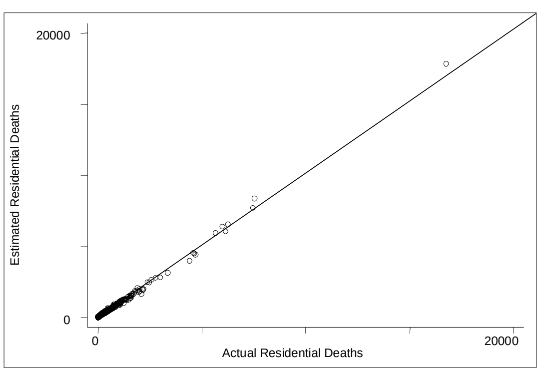


Clearly, survival chances do vary in Wisconsin counties, and in Menominee County life expectancy is substantially lower than in other counties. It is worthwhile to know that these differences exist from county to county.

Evaluation

In order to evaluate the efficacy of our proposed model, a comparison to detailed mortality data by age cohort and sex—which are available in Wisconsin for all counties—was carried out. (These detailed county-level data were not used in developing the life tables themselves.)

Estimated Residential Deaths versus Actual Residential Deaths, All Age Cohorts and Both Sexes (line has slope of 1)



The array of estimated number of resident deaths by age for the 68 counties was compared with the array of actual number of deaths, using regression and chi-square tests. The regression of estimated deaths on actual deaths, based on 2,584 observations (38 age-sex groups—19 male, 19 female—in 68 counties), resulted in an R-square of .996 and a slope of 1.02; the chi-square value was 4954.33. The high R-square value and slope close to one indicates that the proposed solution for developing life tables for small areas, especially in the absence of detailed mortality data, appears to work very well.

For comparison, we also developed a simplified model. It did not use data by race; in short, the state-level survival rates were for all races together and county-level calculations did not consider racial composition. All other steps were the same as the other model. The goal here was to determine if the additional attention to racial composition improved the results enough to warrant the additional data extraction and calculations. In this case, the R-square worked out to .993 and the slope 1.04; the chi-square value was 6119.29.

The refined model that used data by race certainly proved better than the simplified model. The regression slope and R-square values are closer to one, and the lower chi-square statistic indicates a better fit to the actual data.

Regression and Chi-Square Results, Simplified and Refined Models

Model	Slope	R-Square	Chi-Square
Simplified (no race inputs)	1.038456	.9933	6119.29
Refined (race inputs)	1.023336	.9959	4954.33

Nonetheless, the simpler model also gives satisfactory results. A lack of data by race may justify using the simpler model and yet produce reasonably good results.

Citations

Kale, Balkrishna, John Besl, Charles Palit, Paul Voss, Frederick Krantz and Henry Krebs, "An Extension of the Vital Statistics Method to Derive Survival Rates," 1993 Proceedings of the Social Statistics Section, American Statistical Association, pp. 791-796.

Kale, Balkrishna, John Besl, and Charles Palit, "A Procedure to Estimate Mortality Among Migrants," 1995 Proceedings of the Social Statistics Section, American Statistical Association, pp. 139-144.