

# Vintage 2013 State Single-Year-of-Age Projection Methodology

*Prepared by:  
David Egan-Robertson  
UW-Madison Applied Population Laboratory  
April 2014*

The single year of age and sex projections for the State of Wisconsin involve the disaggregation of the final state-level age-sex projections that Demographic Services Center released in January 2014. The state-level projections, which are the summation of county age-sex projections, are calculated in five-year age groups and at five-year intervals, from the base year 2010 through 2040. The single-year methodology separates these five-year results into individual years of age for single calendar years from 2010 through 2025.<sup>1</sup>

The methodology for single-year projections differs from the cohort-component method, which is used for the county projections. For the latter, separate calculations are made for each five-year age-sex cohort to obtain the expected population and net migrants, then these two numbers are summed to produce the projected population. In the single year of age series, a survival methodology is applied to each one-year age-sex cohort, but net migration rates and net migrants are not calculated. Instead, a “residual” population—the difference between the projected and the surviving population—is distributed within each five-year band.

## **Data Inputs**

For this vintage of single-year-of-age (SYA) projections, the following data inputs were employed:

- The base-year age distribution by single year of age and sex, as tabulated in the 2010 U.S. Census.
- The unabridged period life table for the United States, prepared for calendar year 2009 by the National Center for Health Statistics (NCHS). This NCHS life table was the most recent unabridged (i.e., single-year) public-source life table available. (NCHS prepares state-specific unabridged life tables centered on the census years, but they are not available until many years after a census.) The life table is needed to compute survival rates at each age.
- Demographic Services Center’s five-year age-sex projections through 2025. The single-year age projections are controlled to these projections in 5-year age groups and 5-year time steps.
- Births by single year and sex were generated by disaggregating each five years of births, as projected in the five-year age-sex calculations. The overall pattern displayed an upward annual trend in births, although in reality much greater year-to-year variation (including declines) will likely occur. Births were divided between the sexes based on the proportions of actual male and female births in Wisconsin from April 1990 through March 2010.

---

<sup>1</sup> Based on the recommendation of Demographic Services’ former demographer, Balkrishna Kale, single-year projections beyond a 15-year horizon are not reliable enough to warrant continuation of the calculations farther into the future.

### Three Five-Year Cycles

The SYA calculations are carried out in three five-year cycles. The 2010 Census values serve as the base for the first five years of projections, 2011 through 2015. The 2015 projections serve as the base for 2016 through 2020 projections and, similarly, the 2020 projections are the base for the 2021 through 2025 projections.

Within this documentation, the algebraic subscripts used in formulas are:

$x$  = year of age

$z$  = year in time

The value of  $z$  will only vary from 0 through 5. For the base years 2010, 2015 and 2020,  $z = 0$ ; for 2011, 2016 and 2021,  $z = 1$ ; and so on. The  $x$  values proceed from age 0 through 100 and over.

### Calculating the Expected Population

Using the NCHS unabridged life tables, SYA survival rates are calculated for each sex, as illustrated below:

$$S_0 = L_0 \div 100,000 \text{ (the NCHS radix, also notated } l_0)$$

$$S_1 = L_1 \div L_0$$

$$S_2 = L_2 \div L_1$$

$$S_3 = L_3 \div L_2$$

...

$$S_{98} = L_{98} \div L_{97}$$

$$S_{99} = L_{99} \div L_{98}$$

$$S_{100+} = L_{100+} \div (L_{99} + L_{100+})$$

The survival rates are multiplied by the base-year population to yield an expected population one year hence (i.e., April 1 of the following year). The  $S_0$  value is multiplied by the number of births (B) projected for the 12-month period. For example, to obtain the expected population of age 0 at April 1, 2011:

$$E_{0, 2011} = B_{2010-11} \times S_0$$

This calculation produces the number of births that survive to age 0 (that is, up to but not including exactly 1 year of age). Similarly, survival rates calculated for a given age  $x$  are applied to the base-year population at age  $x-1$  to produce the expected population at age  $x$ .

To produce the expected population for the second year, the first-year expected populations become the "base" populations and the same survival rates are multiplied by them. This approach is continued through the fifth year. Representing the pattern generically, the calculations are:

$$\text{Year 1: } E_x = P_{x-1} \times S_x$$

$$\text{Year 2: } E_{x+1} = E_x \times S_{x+1}$$

$$\text{Year 3: } E_{x+2} = E_{x+1} \times S_{x+2}$$

$$\text{Year 4: } E_{x+3} = E_{x+2} \times S_{x+3}$$

$$\text{Year 5: } E_{x+4} = E_{x+3} \times S_{x+4}$$

### Calculating the Residual Population in Five-Year Groups

Once the expected populations have been calculated, the results must be brought into agreement with those from the state final age-sex projections. The difference between total projected and expected population is analogous to net migrants, but is not exactly the same as computed in the cohort-component model used in the state and county projections. To truly represent net migrants, each year's population at age  $x$  to age  $x+1$  would need to be survived, then net migrants calculated by multiplying a net migration rate by the survivors for each age. The base data as assembled for the state and county projections does not permit the calculation of single year net migration rates.<sup>2</sup>

The first step is to sum the expected populations in five-year age bands (e.g., 0-4, 5-9, etc.) at year five and subtract this result from our total projections for the corresponding five-year groups. This gives us the sum of the residual populations for the five years in the age band. Formulaically:

$$\sum_{z=1}^5 \sum_{x=k}^{k+4} R_{x,z} = P_{x-x+4, z=5} - \sum_{x=k}^{k+4} E_{x,z=5}$$

where  $\sum_{z=1}^5 \sum_{x=k}^{k+4} R_{x,z}$  is the sum of residual population for a five-year age group;

$P_{x-x+4, z=5}$  is the 5-year age group population at year 5 in the cycle;

and  $\sum_{x=k}^{k+4} E_{x,z=5}$  is the sum of the expected populations in five-year bands at year 5 in the cycle.

In this particular equation, the values of  $x$  are limited to years of age that are divisible by 5 with no remainder (i.e., 0, 5, 10, 15 and so on).

### Calculating the Preliminary Residual Population in One-Year Groups

The calculated residual population must be distributed along two dimensions: the five ages in the age group, and the five years in the projection period.

For ages 5-9 through 95-99, the first cut at distributing the residual population is to divide it equally among the 5-by-5 matrix formed by each five year cohort ( $x = k$  to  $k+4$ ) and five single years of projection ( $z = 1$  to 5). In simple terms, the residual population for each cell ( $r_{x,z}$ ) is the total residual population, as calculated above, divided by 25.

For ages 0-4, such a distribution does not make sense: For those born at 2014, for example, they could only be "exposed" to one year of residual population before the 2015 endpoint. For this five-year cohort, the pattern of exposure to residual population in the 5-by-5 matrix may be represented binarily as:

---

<sup>2</sup> In the state and county age-sex projections, the net migration rates are calculated based on the ten-year experience of the prior intercensal period (e.g., from 2000 to 2010), then adjusted to five-year increments based on a target number of net migrants.

	z	1	2	3	4	5
x						
0		0	0	0	0	1
1		0	0	0	1	1
2		0	0	1	1	1
3		0	1	1	1	1
4		1	1	1	1	1

Note: x values are at time z = 5.

Thus, of the 25 cells, 15 are exposed to residual population. Therefore, the residual when k = 0 is divided by 15, and each 1/15<sup>th</sup> is entered in the cells where the value is one; all remaining cells are null.

As the very elderly age toward and then into the age group 100 and over, the appearance of the matrix is reversed:

	z	1	2	3	4	5
x						
95		1	1	1	1	1
96		1	1	1	1	0
97		1	1	1	0	0
98		1	1	0	0	0
99		1	0	0	0	0

Note: x values are at time z = 0.

Of the 25 cells, 15 are exposed to residual population.<sup>3</sup> Therefore, the residual when k = 95 is divided by 15, and each 1/15<sup>th</sup> is entered in the cells where the value is one; all remaining cells are null.

### **Adjusting the Preliminary One-Year Residual Populations in the College-Aged Cohorts**

While the equal-distribution methodology for apportioning the residual population produced generally reasonable results, problems arise among the cohorts containing the college-aged populations. It is recognized that Wisconsin is a net “importer” of college students (at approximately ages 17-19) and a net “exporter” of college graduates (at approximately ages 21-23). At the five-year cohort level, these demographic events are apparent: for the three five-year periods from 2010 through 2025, both males and females show a net in-migration at ages 15-19 (i.e., age 10-14 at the base year aging to 15-19 in the fifth projection year) and a net out-migration at ages 20-24 and 25-29.

This initial dependence on the five-year cohort results creates unusual transitions when viewed in one-year age progressions. Below is a segment of the worksheet for females from 2010 through 2015, illustrating the change in residual population from one five-year band to the next:

<sup>3</sup> Actually, all twenty-five cells are exposed: 100 and over at year x becomes 101 and over at year x+1, 102 and over at year x+2, and so on. However, because the values at 100 and over are summarized, each individual year beyond that point does not require separate calculation.

AgeYr0	AgeYr1	AgeYr2	AgeYr3	AgeYr4	AgeYr5	RPopYr1	RPopYr2	RPopYr3	RPopYr4	RPopYr5
10	11	12	13	14	15	70.23	70.23	70.23	70.23	70.23
11	12	13	14	15	16	70.23	70.23	70.23	70.23	70.23
12	13	14	15	16	17	70.23	70.23	70.23	70.23	70.23
13	14	15	16	17	18	70.23	70.23	70.23	70.23	70.23
14	15	16	17	18	19	70.23	70.23	70.23	70.23	<b>70.23</b>
15	16	17	18	19	20	-310.42	-310.42	-310.42	<b>-310.42</b>	-310.42
16	17	18	19	20	21	-310.42	-310.42	-310.42	-310.42	-310.42
17	18	19	20	21	22	-310.42	-310.42	-310.42	-310.42	-310.42
18	19	20	21	22	23	-310.42	-310.42	-310.42	-310.42	-310.42
19	20	21	22	23	24	-310.42	-310.42	-310.42	-310.42	-310.42

N.B. The bolded values are discussed in the following text.

The thicker black lines in the table are the “fences” within which the final population results, at  $z = 5$  and  $x = k$  through  $x = k+4$ , must sum to match the value for the corresponding five-year age-sex cohort. Note that, when crossing certain fences, the residual population changes direction, from positive to negative or vice versa. This phenomenon creates the unusual result that, for example, the age 18 cohort at year 4 aging to 19 at year 5 gains 70 in residual population, but the age 18 cohort at year 3 aging to 19 at year 4 loses 310 in residual population.

To correct these inconsistencies the following adjustments are applied, as illustrated in the table below (the bolded and italicized values are the ones that have been altered):

- 1) The residual change for the “late teen” years is carried down into the subsequent five-year band.
- 2) The residual populations of those aging from 19 to 20 and 20 to 21 remain as calculated originally.
- 3) In order to balance the gain to the five-year cohort created in the first step, the residual population loss in cohorts aging from 21 to 22 and 22 to 23 has to be increased. While the residual loss appears relatively large, it is not unreasonable, given the exodus of young college graduates.
- 4) This latter adjustment is carried down to the next five-year cohort for the same two single-year cohorts.
- 5) The residual populations of those aging from 23 to 24 and 24 to 25 remain as calculated originally.
- 6) In order to balance the loss to the five-year cohort created in the fourth step, the residual population change in cohorts aging from 25 to 26 and higher is increased.

AgeYr0	AgeYr1	AgeYr2	AgeYr3	AgeYr4	AgeYr5	RPopYr1	RPopYr2	RPopYr3	RPopYr4	RPopYr5
14	15	16	17	18	19	70.23	70.23	70.23	70.23	70.23
15	16	17	18	19	20	<b>70.23</b>	<b>70.23</b>	<b>70.23</b>	<b>70.23</b>	-310.42
16	17	18	19	20	21	<b>70.23</b>	<b>70.23</b>	<b>70.23</b>	-310.42	-310.42
17	18	19	20	21	22	<b>70.23</b>	<b>70.23</b>	-310.42	-310.42	<b>-1071.73</b>
18	19	20	21	22	23	<b>70.23</b>	-310.42	-310.42	<b>-1071.73</b>	<b>-1071.73</b>
19	20	21	22	23	24	-310.42	-310.42	<b>-1071.73</b>	<b>-1071.73</b>	-310.42
20	21	22	23	24	25	-282.24	<b>-1071.73</b>	<b>-1071.73</b>	-282.24	-282.24
21	22	23	24	25	26	<b>-1071.73</b>	<b>-1071.73</b>	-282.24	-282.24	<b>112.50</b>
22	23	24	25	26	27	<b>-1071.73</b>	-282.24	-282.24	<b>112.50</b>	<b>112.50</b>
23	24	25	26	27	28	-282.24	-282.24	<b>112.50</b>	<b>112.50</b>	<b>112.50</b>
24	25	26	27	28	29	-282.24	<b>112.50</b>	<b>112.50</b>	<b>112.50</b>	<b>112.50</b>

Note that this adjustment process could be carried into the next five-year cohort, and *ad infinitum*. However, the principal goal of this adjustment is to correct for *sizeable* directional changes in the residual population in adjoining five-year cohorts.

**Accumulating the Residual Population for Each One-Year Group**

The residual population values, as illustrated above, are the changes from year to year. However, the residual effect is actually cumulative within each five-year period.

For example, in the portion of the residual population table above, 15-year olds at  $z = 0$  age to 16 at  $z = 1$ , and gain approximately 70 residents. As those 16 year olds at  $z = 1$  age to 17 at  $z = 2$ , they gain an additional 70 17-year-old companions, and so on.

The generic formula for this process is:

$$R_{x,z} = \sum_{x=k}^{k+z} r_{x,z}$$

where  $R_{x,z}$  is the (cumulative) residual population at age  $x$  and year  $z$ ;  
and  $r_{x,z}$  is the value (in each cell) obtained as described in the previous two sections.

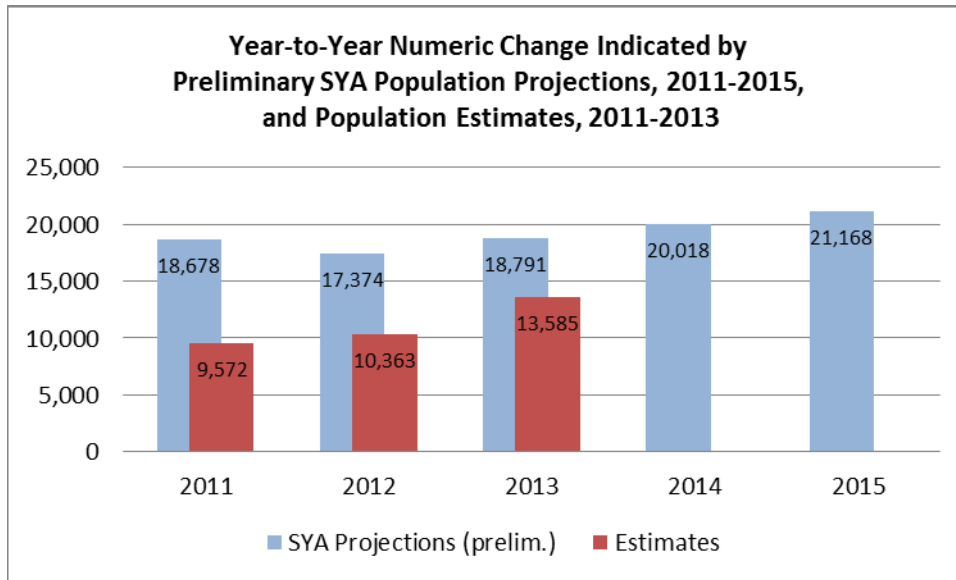
**Calculating the Preliminary Total Population in One-Year Age Groups**

With the expected population and residual population for each individual age and year (and sex) calculated, the total projected population can now be completed.

$$P_{x,z} = E_{x,z} + R_{x,z}$$

**Final Adjustments to Projections**

While the SYA projection methodology is constructed and calculations are carried out independently from Demographic Services Center’s population estimates, the first three years—2011, 2012 and 2013—are coincident. The graph below summarizes the year-to-year change in total population, as projected by the preliminary SYA series, and the first three estimates since the 2010 Census (interpolated and extrapolated to the April 1 reference date of the projections). The SYA projections tend to produce a relatively smooth annual change in projected population, but the annual estimates are much more variable.



The relatively large discrepancies between the SYA projections and the estimates necessitated an adjustment to the projection series; the total difference in the projections from 2010 to 2015 must remain the same, but the annual change needed to be modified. As a modification, the preliminary total population projection for 4/2014 was lowered slightly so that the year-to-year change produced a pattern that appeared reasonable, both within this five-year time frame and within the longer fifteen-year projection horizon.

Finally, there was discontinuity in the annual change from the 2015-2020 to the 2020-2025 period. The annual changes in the latter were modified to create a smoother and more reasonable pattern of change across time.

