Methodology for Preparing Demographic Estimates
Data release: 2011
Data pertaining to: 2005, 2006, 2007, 2008, 2009, 2010, 2011
IECAM demographic methodology page:
http://iecam.illinois.edu/data/methodology/

# Methodology for IECAM Demographic Estimates

by Katherine Nesse & Andrew M. Isserman

The method of creating demographic estimates for IECAM balances the advantages of multiple estimates from the US Census Bureau, resulting in estimates that are both current and precise. The two primary sources that we use are the American Community Survey and the Population Estimates Program. In addition, we employ the decennial census. The population surveys conducted by the US Census Bureau, in particular the "long form" survey that has accompanied the decennial census, have been the preferred sources for population estimates by policy analysts. The difficulty with this source is that it was out of date by the middle of the decade. The Census Bureau has been transitioning to a new method of surveying the population which was fully realized in 2010 in the 2005-2009 American Community Survey (ACS). The ACS provides rolling updates that are more current but less precise in some cases than the decennial "long form." In addition, the Census Bureau creates annual estimates for incorporated areas through the Population Estimates Program. The method of creating estimates of population and demographic characteristics for the state, counties, townships, municipalities, state legislative districts, and federal congressional districts in Illinois relies on the combination of these estimates.

We give preference to the estimates created in the Population Estimates Program. The estimates for counties and states are broken down by age and race but in most cases the estimate is limited to the total population. If the population estimates program does not create an estimate for the place or characteristic of interest, we rely on the ACS 1-year release. If the place or characteristic is not estimated in the ACS 1-year release either, we apply a ratio of the place to its parent area for the characteristic of interest derived from the most current source. The preferred source for the ratio is the ACS 3-year release, however if the characteristic or geographic area is not available in the ACS 3-year release, we use the ACS 5-year release. In rare cases we will rely on the decennial census long form (figure 1). With data from so many different sources one of the challenges for creating estimates of multiple areas is for them to add up to the total. In some cases we adjust the estimates uniformly. In other cases we rely on a method of adjusting matrices described by Miller and Blair (1985).

A. Sources for creating estimates

Until recently, the Census Bureau has published a limited number of intercensal estimates of the population and its characteristics. They produced estimates of total population for states, counties, municipalities, and sub-county regions such as townships. In addition they produced estimates of a few characteristics for some geographic areas. For example, they produced estimates of the number of children under 18 in poverty in counties and school districts. Except for those few statistics, analysts would have to rely on the long form survey that accompanied

the decennial census. Early in the decade the long form statistics were useful but their usefulness declined as time marched away from the turn of the decade. The long form survey has been replaced by the American Community Survey (ACS), a rolling sample of the population. The advantage of the ACS is that estimates are released every year. However, it has some disadvantages including a lack of precision in some cases.

The primary data source is the estimates made by the US Census Bureau's Population Estimates Program. It collaborates with states through the Federal-State Cooperative for Population Estimates to create estimates for all incorporated areas: states, counties, townships, and municipalities. These estimates are created for counties using primarily the component of change method (US Census Bureau 2009a). For sub-county areas such as townships and places, the estimates are created using primarily the housing unit change method. The estimates are controlled by the county estimates so that the sum of townships is equal to the county they are in (US Census Bureau 2009b). In addition to these estimates of total population, the Population Estimates Program works with the National Center for Health Statistics to create estimates of the population in states and counties by sex, age, race, and Hispanic status (NCHS 2009).

The annual estimates of total population created by the Population Estimates Program are used as benchmarks for the American Community Survey (Scardamalia 2006). In some respects the ACS is a more accurate survey than the long form (Scardamalia 2006). However, it sends surveys out throughout the year and samples at a lower rate than the long form did. Because of this, responses are pooled across 12 months for very large places (over 65,000 people), across 36 months for mid-sized places (over 20,000 people) and across 60 months for all other places, including very small places like census tracts. The types of data are referred to by the size of the response pool, as 1-, 3-, or 5-year releases.

The final source is the decennial census and the accompanying long form survey. The census is the decennial count of people in the US as of April 1 and the answers on the long form are pegged to the same date. These two sources (often referred to as just "the decennial census") have been the primary source for demographic information about US residents for decades. The ACS has replaced the long form survey so the ACS and the long form have very similar questions and are tabulated in similar ways. However, the ACS is slightly different in geographic coverage (for example, the ACS does not have estimates for zip code tabulation areas) and informational details (for example, some income and poverty questions are asked in a slightly different way on the ACS) and so we still occasionally rely on the decennial census.

B. Using existing estimates

It was impractical to create estimates from scratch because we need estimates for many geographic areas and specific characteristics of the population; therefore we rely on other estimates. We use estimates created by the Census Bureau as a starting point for the IECAM estimates. If estimates exist from a reliable source it makes sense to use those. We use estimates of total population created annually by the Census Bureau's Population Estimates Program through the Federal-State Cooperative for Population Estimates. The estimates are made for all incorporated areas in the US. The Population Estimates program also creates annual estimates of the population by age and race/ethnicity for states and counties. For places and characteristics for which the Population Estimates program does not create estimates, we turn to the ACS 1-year release.

Within each set of estimates, the estimates are consistent: the sum of counties equals the state and the sum of townships within a county is equal to that county. However, in combining the two sets, the estimates are not always consistent. Most of the time adjustments do not need to be made. The ACS 1-year release only has estimates for some cities, townships and counties so the case where all of the estimates of the lower geographic level are from the ACS and the upper geographic level is from the Population Estimates Program is rare. When this does occur, it is necessary to adjust one or the other to maintain consistent estimates. We prefer the Population Estimates Program estimates so when the two are in conflict, we adjust the ACS 1-year release.

In creating estimates for IECAM, we ran into this difficulty with federal congressional districts. In theory, the sum of estimates for federal congressional districts in Illinois should add up to the estimate for the state. The Population Estimates Program publishes estimates for total population and population by age for the state. However, it does not publish those estimates for federal congressional districts. Estimates for those areas come from the ACS 1-year release. However, the sum of the estimates for federal congressional districts from the ACS 1-year release does not add up to the state total from the Population Estimates Program. Therefore we reduced the estimates of federal legislative districts uniformly to equal the state total.

C. Adapting estimates from other sources

The population characteristics estimated by the Population Estimates Program and the geographic areas estimated in the ACS 1-year release are limited. For areas or characteristics not found in these two sources we found the ratio of sub-area (e.g., township) to super-area (e.g., county) for the characteristic of interest in another source and applied that ratio to the estimate of the super-area.

$$POP_{g,t} = POP_{g,t,3} / POP_{s,t,3} * POP_{s,t} \qquad (1)$$

In equation (1) $POP_{g,t}$ is the population in geographic area g at time t, $POP_{g,t,3}$ is the population estimate in geographic area g at time t in the 3-year ACS release, $POP_{s,t,3}$ is the population estimate in super-area s containing geographic area g at time t in the 3-year ACS release, and $POP_{s,t}$ is the population estimate in super-area s at time t that has already been estimated either through the ACS 1-year estimate or with this process. We use the ACS 3-year release if an estimate for the geographic area and characteristic can be found in it. If not, we draw the ratio from the ACS 5-year release. In rare cases where this ratio cannot be found in either source, we use the 2010 decennial census.

There are potentially four ACS releases that include the target year. For a place over 65,000 in 2007, in addition to the 1-year release, there are three 3-year releases that cover 2007 (2005-2007, 2006-2008, and 2007-2009), and there are potentially five 5-year releases, although the 5-year ACS has only been released twice so far (2005-2009 and 2006-2010). Even for smaller places, there are multiple releases to choose between. The Census Bureau cautions users not to peg the estimates to a specific point in time (US Census Bureau 2009c) yet offers little advice on how to choose among the multiple estimates.

Our goal is to use the best available data and adapt it to create the estimate of the population characteristic we desire. The graph of the estimates of total population by different sources (figure 2) shows that generally, the ACS estimate where the year of interest is the center year in the release (e.g., the ACS 3-year release covering 2006-2008 when 2007 is the year of interest) is probably the most accurate. Using the middle year also makes sense when the ACS is thought of as an average across 3 years. The Census Bureau advises against pegging the 3-year estimates to any given year (US Census Bureau 2009c). However, we are looking for the ratio that will most likely reflect the ratio in a given year. That ratio is most likely to be drawn from an ACS release where the year of interest is the middle year. On the other hand, this release is not always available. For example, the first 3-year release that includes 2008 is the 2006-2008. To create estimates for 2008 before the ideal 3-year release is published (2007-2009), we use the 2006-2008 since it is the best available at the time we are making the estimates. [1]

The estimates derived from the 3-year ACS release, the 5-year ACS release and the 2000 decennial census are not consistent. The sum of townships estimated this way will not necessarily add up to the county. One township may be covered by the ACS 1-year release and therefore fixed, a few other townships may be covered by the ACS 3-year release and the rest covered by the decennial census. It would be unusual for the sum of the sub-areas to be exactly the same as the estimate for the super-area since the ratio of sub-area to super-area is not the same across the releases. To adjust the estimates to add up to the super-area, we uniformly reduce (or enlarge) all of the estimates of the sub-area that are based on the ACS 3-year release or the decennial census. Those estimates based on the ACS 1-year release or on the Population Estimates Program estimates are not adjusted.

Occasionally there are estimates that cannot be made by comparing a place to its super-area. This is because the estimate does not exist at the higher level. In this case we compared the estimate of the characteristic in the place to a similar characteristic of the place. For example, one of the estimates we created was of the number of children in each age cohort living below 185 percent of poverty. The American Community Survey does not publish such an estimate at any geographic level. However, we had estimated the number of all children in each age cohort. Instead of comparing the number of children aged 1 year living below 185 percent of the poverty line in the place to the estimate of that characteristic in its super-area, we apply the ratio of the number of children aged 1 year to all children to the estimate of all children living below 185 percent of the poverty line:

$$CHAR_{g,t} = REL1_{g,t} / REL2_{g,t,3} * POP_{g,t} \quad (2)$$

In equation (2) $CHAR_{g,t}$ is the population characteristic of interest in geographic area g at time t (e.g., the number of people aged 1 year living below 185 percent of the poverty line), $REL1_{g,t}$ is the first related population characteristic estimate in geographic area g at time t (e.g., the number of people aged 1 year), $REL2_{g,t}$ is the second related population characteristic estimate in geographic area g at time t (e.g., the total number of children), $POP_{g,t}$ is the population estimate in geographic area g at time t (e.g., the total number of children living below 185 percent of the poverty line). We used this method to estimate the number of children in each age cohort for all levels of poverty.

D. Constraining and adjusting the estimates

There are many constraints that can be placed on the estimates to ensure that the result is logical and internally consistent. We constrained the estimates of the number of children in poverty so that the estimate of each level of poverty was greater than or equal to the estimate of the level below (e.g., the number of children living below 185 percent of the poverty line is greater than or equal to the number of children living below 100 percent of the poverty line). In addition, although the estimates are adjusted to add up to the parent area as they are made, there is a special circumstance when the estimates need to be adjusted after all of the estimates are made. In some cases the estimates need to add up horizontally: the sum of children age 0 to age 5 should add up to the estimate for the number of children age 5 and under; and add up vertically: the number of children age 0 in townships should equal the estimate of the number of children age 0 in the county. There are situations where this is not the case, for example, the number of children in a given age cohort will add up to the state total or the number of children in all age cohorts will add up to the total number of children in a given geographic area but not both. To resolve this, we employed a method similar to the RAS method described by Miller and Blair to adjust input-output matrices (Miller and Blair 1985).

To ensure that the estimate of the number of children living below 100 percent of the federal poverty line was not larger than the estimate of the number of children living below 185 percent of the poverty line, we constrained the model to make it so. If the estimate of the number of children living below 100 percent of poverty in township 1 is 50 and the estimate of the number of children living below 185 percent of poverty is 52, we reduced the number of children living below 185 percent of poverty to 50. This means that the sum of the townships is now 2 people less than the estimate for the county. We adjusted the other townships in the county to make up for this. We repeated this process until all of the townships in the county obey the constraint and sum to the county estimate.

We employed a method for adjusting matrices described in Miller and Blair (1985) to guarantee that the sum of the estimates in the child areas is equal to the parent area and that the estimates are consistent within a geographic area as well. In the example below, County A has an estimated 10,000 children age 5 and under living in it. An estimated 15 percent (1,500) live in Township 1, 77 percent (7,700) in Township 2, 3 percent (300) in township 3, and 5 percent (500) in Township 4 (table 1). While the individual age cohorts in each township add up to the total for children under age 6, the sum of the township estimates for a given age cohort do not add up to the county estimate for that age cohort. We resolve this in a two-step iterative process.

The first step is to divide the township estimate by the total for all the township estimates and multiply it by the estimate for the county for each age cohort:

$T1,0 = BT / JT * AT$ (3)

Where T1,0 is the revised estimate for Township 1 for age 0, BT is the current estimate for Township 1 for age 0 found on line B (column T) in table 1, JT is the current sum of the township estimates for age 0 found on line J (column T) in table 1, and AT is the estimate for

County A for age 0 found on line A (column T) in table 1. The results are shown in table 2. You will notice that now the sum of the townships in each age cohort is equal to the county estimate for that age cohort (line A equals line J) but the sum of age 0 to age 5 is not equal to the estimate for the number of children under age 6 in each township (column S is not equal to column Z).

The second step reconciles column S with column Z. For each township, divide each age cohort by the sum of the age cohorts and multiply it by the estimate of the number of children under age 6 in the township:

$$T1,0 = TB / ZB * SB \quad (5)$$

Where T1,0 is the revised estimate for Township 1 for age 0, TB is the current estimate for Township 1 for age 0 found in column T (line B) in table 2, ZB is the current sum of all age cohorts for Township 1 found in column Z (line B) in table 2, and S is the current estimate for Township 1 for age 0 found in column S (line B) in table 2. The results are shown in table 3. Now the sum of age 0 to age 5 is equal to the estimate of the number of children age 6 and under in each township (column Z is equal to column S) and for each age cohort the sum of townships is almost equal to the estimate of the age cohort for the county (line J is almost equal to line A). In this example, the discrepancy was resolved in one round but it sometimes takes more iteration to reduce the discrepancy to less than one. We repeat steps one and two until a solution is found where both the rows and columns summed correctly (a discrepancy of less than one). Then we round those numbers so that the estimates reflect whole people. This may slightly change the estimates (for example, the final estimate for the number of children under age 0 in Township 1 is 250, not very different from the original estimate of 253) but the estimates are now more logical.

Figure 1. Decision tree for using and adapting sources

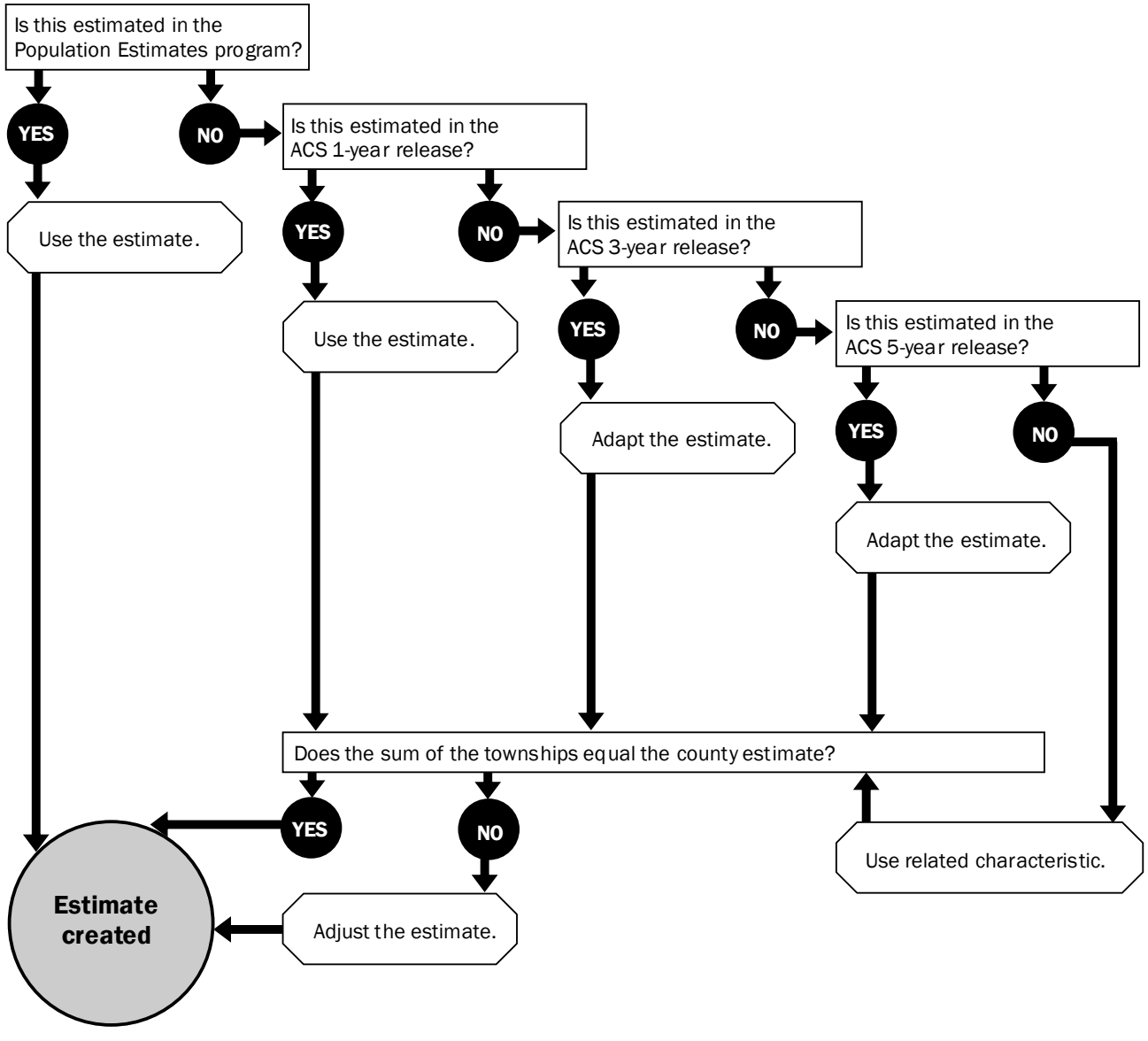

Figure 2. Estimates of total population in Champaign County (The ACS 1-year release estimates did not include the group-quarters population until 2006)

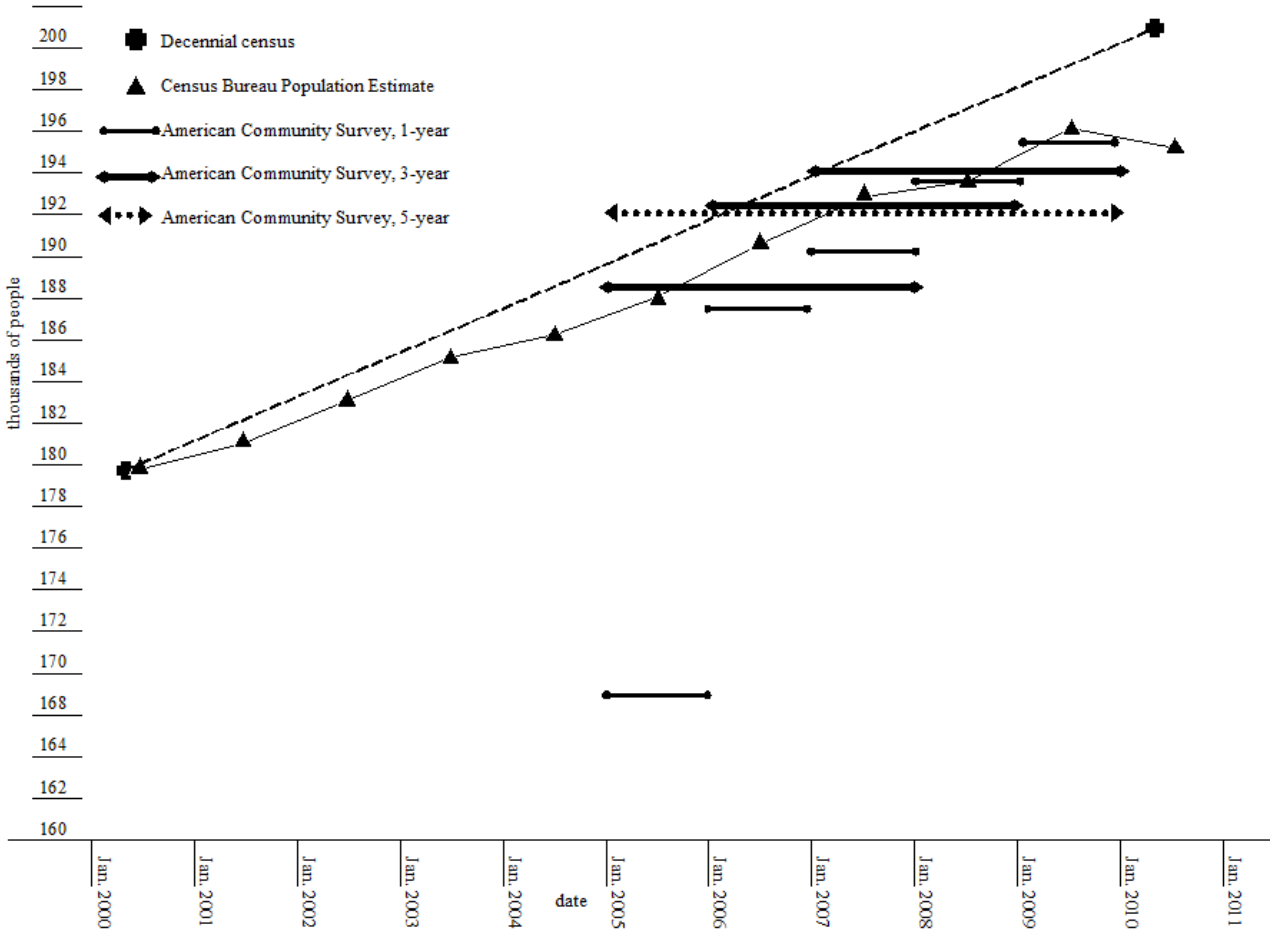

Table 1. Example of adjusting estimates, step 1

| | | under age 6 | age 0 | age 1 | age 2 | age 3 | age 4 | age 5 | sum of age 0 to age 5 |
|---|---|---|---|---|---|---|---|---|---|
| A | County A | 10000 | 1547 | 1821 | 1642 | 1689 | 1703 | 1598 | 10000 |
| B | Township 1 | 1500 | 253 | 222 | 236 | 281 | 243 | 265 | 1500 |
| C | | 15% | | | | | | | |
| D | Township 2 | 7700 | 1179 | 1288 | 1276 | 1354 | 1292 | 1311 | 7700 |
| E | | 77% | | | | | | | |
| F | Township 3 | 300 | 42 | 39 | 58 | 69 | 51 | 41 | 300 |
| G | | 3% | | | | | | | |
| H | Township 4 | 500 | 97 | 106 | 80 | 74 | 62 | 81 | 500 |
| I | | 5% | | | | | | | |
| J | Total of townships | | 1571 | 1655 | 1650 | 1778 | 1648 | 1698 | |
| | | S | T | U | V | W | X | Y | Z |

Table 2. Example of adjusting estimates, step 2

| | | under age 6 | age 0 | age 1 | age 2 | age 3 | age 4 | age 5 | sum of age 0 to age 5 |
|---|---|---|---|---|---|---|---|---|---|
| A | County A | 10000 | 1547.00 | 1821.00 | 1642.00 | 1689.00 | 1703.00 | 1598.00 | 10000.00 |
| B | Township 1 | 1500 | 249.13 | 244.27 | 234.86 | 266.93 | 251.11 | 249.39 | 1495.70 |
| C | | 15% | | | | | | | |
| D | Township 2 | 7700 | 1160.99 | 1417.19 | 1269.81 | 1286.22 | 1335.12 | 1233.79 | 7703.13 |
| E | | 77% | | | | | | | |
| F | Township 3 | 300 | 41.36 | 42.91 | 57.72 | 65.55 | 52.70 | 38.59 | 298.82 |
| G | | 3% | | | | | | | |
| H | Township 4 | 500 | 95.52 | 116.63 | 79.61 | 70.30 | 64.07 | 76.23 | 502.36 |
| I | | 5% | | | | | | | |
| J | Total of townships | | 1547.00 | 1821.00 | 1642.00 | 1689.00 | 1703.00 | 1598.00 | |
| | | S | T | U | V | W | X | Y | Z |

Table 3. Example of adjusting estimates, results

| | | under age 6 | age 0 | age 1 | age 2 | age 3 | age 4 | age 5 | sum of age 0 to age 5 |
|---|---|---|---|---|---|---|---|---|---|
| A | County A | 10000 | 1547.00 | 1821.00 | 1642.00 | 1689.00 | 1703.00 | 1598.00 | 10000.00 |
| B | Township 1 | 1500 | 249.85 | 244.97 | 235.53 | 267.70 | 251.83 | 250.11 | 1500.00 |
| C | | 15% | | | | | | | |
| D | Township 2 | 7700 | 1160.52 | 1416.61 | 1269.30 | 1285.70 | 1334.58 | 1233.29 | 7700.00 |
| E | | 77% | | | | | | | |
| F | Township 3 | 300 | 41.52 | 43.08 | 57.95 | 65.80 | 52.91 | 38.74 | 300.00 |
| G | | 3% | | | | | | | |
| H | Township 4 | 500 | 95.07 | 116.08 | 79.24 | 69.97 | 63.77 | 75.87 | 500.00 |
| I | | 5% | | | | | | | |
| J | Total of townships | | 1546.96 | 1820.75 | 1642.01 | 1689.17 | 1703.09 | 1598.01 | |
| | | S | T | U | V | W | X | Y | Z |

**Works cited**

Miller, Ronald E. and Peter D. Blair. 1985. *Input-output analysis: Foundations and extensions.* Englewood Cliffs, NJ: Prentice Hall.

National Center for Health Statistics [NCHS]. 2009. Documentation for Vintage 2008 Bridged-Race Postcensal Population Estimates for Calculating Vital Rates. http://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm

Scardamalia, Robert. 2006. The American Community Survey: general commentary on the findings from external evaluations, *Population Research and Policy Review* 25: 293-303.

US Census Bureau. 2009a. Methodology for the State and County Total Resident Population Estimates (Vintage 2009): April 1, 2000 to July 1, 2009. http://www.census.gov/popest/methodology/2009-st-co-meth.pdf

US Census Bureau. 2009b. Methodology for the Subcounty Total Resident Population Estimates (Vintage 2009): April 1, 2000 to July 1, 2009. http://www.census.gov/popest/methodology/2009-su-meth.pdf

US Census Bureau. 2009c. *A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know.* Washington, DC: U.S. Government Printing Office, (May).

**Notes**

[1] The 2011 release IECAM estimates use these sources for each estimate year:

- 2005 (v2011 Population Estimates Program, 2005 ACS 1-year, 2005-2007 ACS 3-year, 2005-2009 ACS 5-year);
- 2006 (v2011 Population Estimates Program, 2006 ACS 1-year, 2005-2007 ACS 3-year, 2005-2009 ACS 5-year);
- 2007 (v2011 Population Estimates Program, 2007 ACS 1-year, 2006-2008 ACS 3-year, 2005-2009 ACS 5-year);
- 2008 (v2011 Population Estimates Program, 2008 ACS 1-year, 2007-2009 ACS 3-year, 2005-2009 ACS 5-year);
- 2009 (v2011 Population Estimates Program, 2009 ACS 1-year, 2008-2010 ACS 3-year, 2007-2011 ACS 5-year).
- 2010 (v2010 Decennial Census, 2010 ACS 1-year2009-2011 ACS 3-year, 2007-2011 ACS 5-year);
- 2011 (v2011 Population Estimates Program, 2011 ACS 1-year, 2009-2011 ACS 3-year, 2007-2011 ACS 5-year).