Methodology for Preparing Demographic Estimates
Data release: 2012
Data pertaining to: 2008, 2009, 2010, 2011, 2012
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 (PEP). 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 The difficulty with this source is that it was out of date by the middle of the decade.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 5-year release. If the place or characteristic is not estimated in the ACS 5-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. See Figure 1 for using and adapting the sources.

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 informational details (for example, some income and poverty questions are asked in a slightly different way on the ACS); 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 5-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.

C. Adapting estimates from other sources

The population characteristics and the geographic areas estimated by the Population Estimates Program are limited. For areas or characteristics not found in the PEP, 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.

$$POPsub,t = (POPsub,t,5 / POPsup,t,5) * POPsup,t \qquad (1)$$

In equation (1) *POPsub,t* is the population in sub-area *sub* at time *t*, *POPsub,t,5* is the population estimate in sub-area *sub* at time *t* in the 5-year ACS release, *POPsup,t,5* is the population estimate in super-area *sup* containing sub-area *sub* at time *t* in the 5-year ACS release, and *POPsup,t* is the population estimate in super-area *sup* at time *t* that has already been estimated. 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, suppose we are making estimates for year 2008, using 3-year ACS where 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" (i.e., most recent) available at the time we are making the estimates. [1]

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 could constrain 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 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 all age cohorts will add up to the total number of children in a given geographic area, but the number of children in a given age cohort does not add up to the state total. To resolve this, one could consider 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, one may want to constrain the model to make it so. If the estimate of the number of children living below 100 percent of poverty in township 1 is 52 and the estimate of the number of children living below 185 percent of poverty is 50, we reduced the number of children living below 100 percent of poverty to 50. This means that the sum of the townships is now 2 people less than the estimate for the county. Then, the other townships in the county could be adjusted to make up for this. This process is repeated until all of the townships in the county obey the constraint and sum to the county estimate.

This release is more concerned on horizontal add up. So, in a given area, the sum of estimates in each applicable age cohort should be equal to the estimate for the age range of which each age cohort belongs to. The estimation process for this release does not do vertical adjustment.

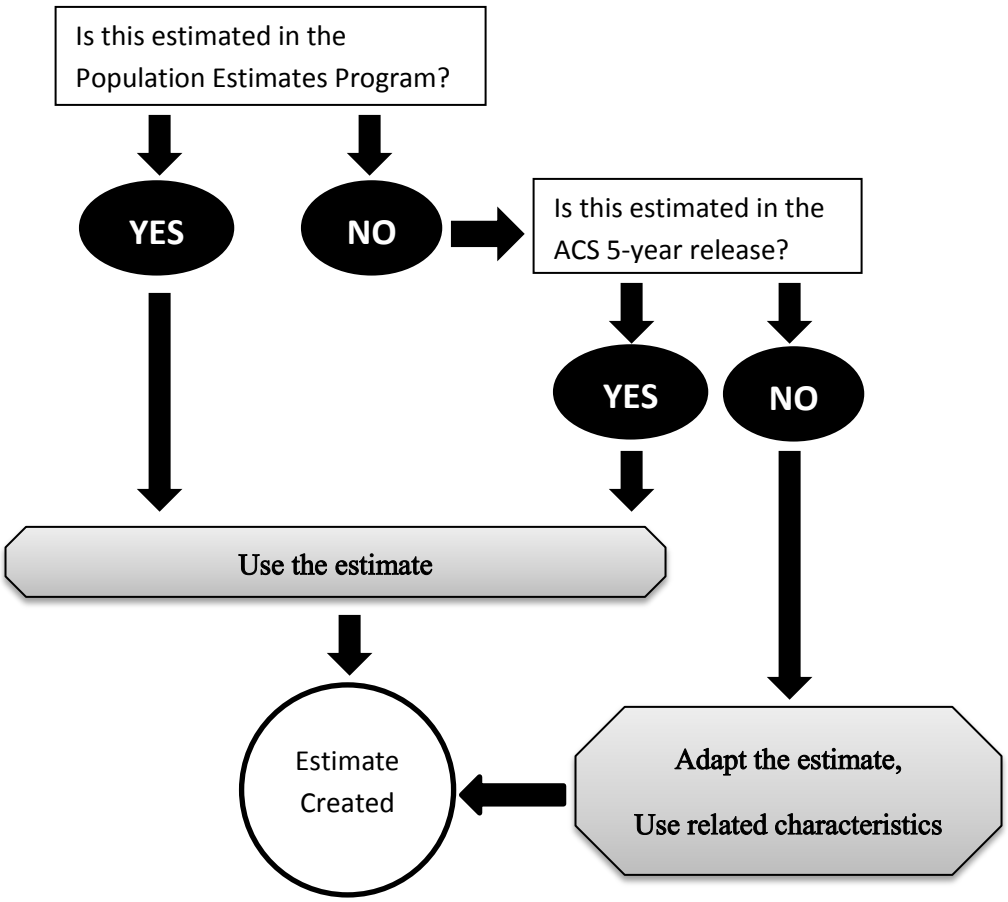

Figure1. Decision tree for using and adapting sources

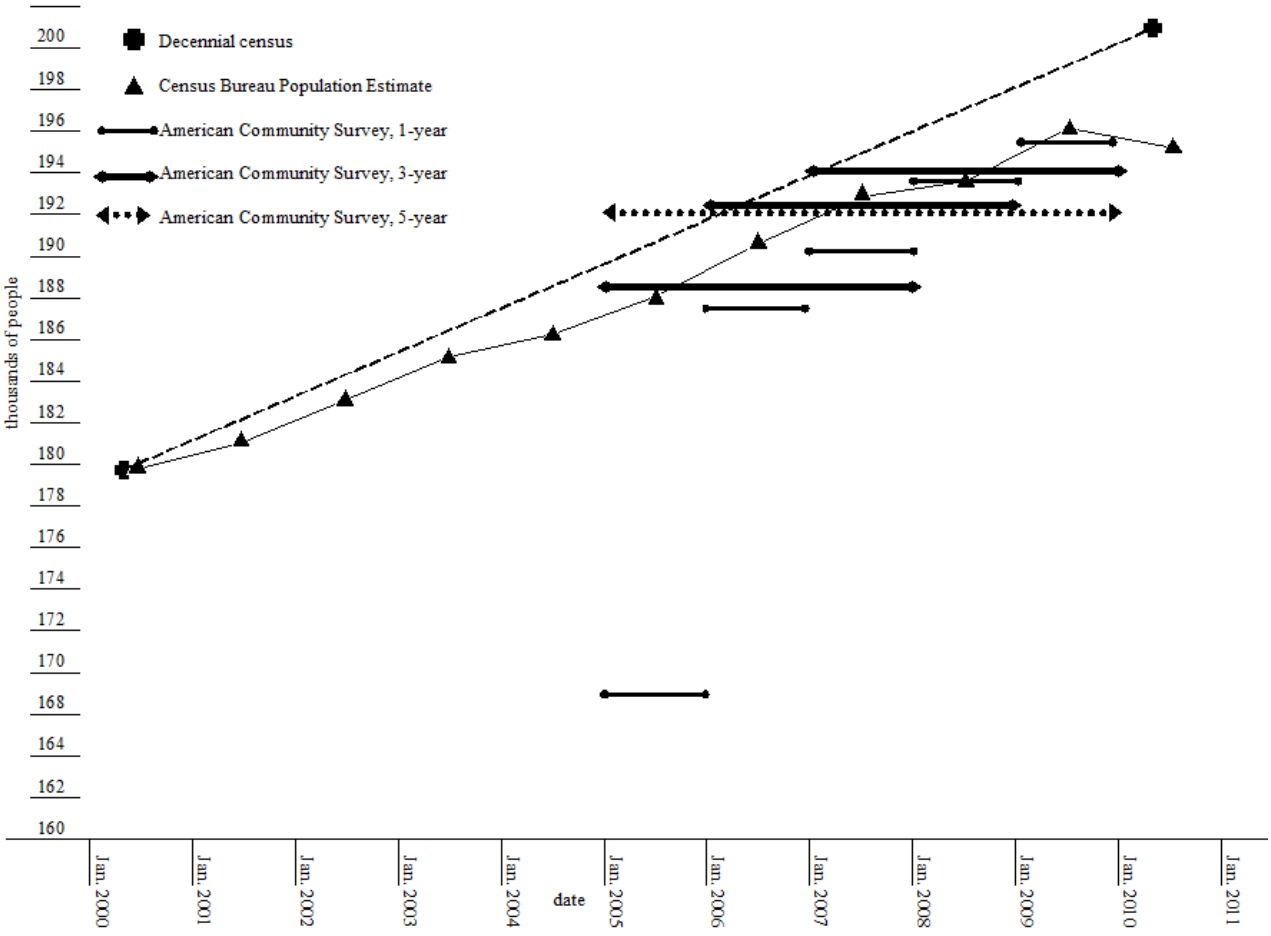

Figure2. Estimates of total population in Champaign County (The ACS 1-year release estimates did not include the group-quarters population until 2006)

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 2012 release IECAM estimates use these sources for each estimate year:

- 2008 (PEP v2012 year 2008, 2006-2010 ACS 5-year)
- 2009 (PEP v2012 year 2009, 2007-2011 ACS 5-year)
- 2010 (2010 Decennial Census, 2008-2012 ACS 5-year)
- 2011 (PEP v2012 year 2011, 2008-2012 ACS 5-year)
- 2012 (PEP v2012 year 2012, 2008-2012 ACS 5-year)