2008 American Community Survey Research Memorandum Series
October 2008

Statistical Issues of Interpretation of the American Community Survey's One-, Three-, and Five-Year Period Estimates

Michael Beaghen
Decennial Statistical Studies Division
Lynn Weidman
Statistical Research Division

U S C E N S U S B U R E A U
Helping You Make Informed Decisions

LIST OF SECTIONS

1. INTRODUCTION 3

2. BACKGROUND 4
- **2.1 The Decennial Census Long Form 4**
- **2.2 The American Community Survey 4**

3. STATISTICAL ISSUES AND ISSUES OF INTERPRETATION FOR MULTIYEAR DATA 6

4. THE MULTIYEAR ESTIMATES STUDY 6

5. KEY CONCEPTS FOR DATA USERS 7
- **5.1 Multiyear Estimation: Key Concepts 7**
- **5.2 Period Estimates 8**
- **5.3 Relative Precision of the One-, Three-, and Five-Year Estimates for Totals . . . 9**
- **5.4 Precision of Estimates for Subpopulations 12**
- **5.5 Currency Versus Precision 13**
- **5.6 A Multiyear Estimate is not an Estimate of its Center Year 16**
- **5.7 Make Comparisons between Areas over the Same Time Periods 20**
- **5.8 Controlling Selected Estimates 20**
- **5.9 Additional Methodological Considerations 21**

6. INTERPRETING CHANGE OVER TIME WITH MULTIYEAR ESTIMATES 22
- **6.1 The Precision of Non-Overlapping Estimates over Time 23**
- **6.2 Comparing Estimates with Overlapping Time Periods 25**
- **6.3 Smoothing with Multiyear Estimates 27**
- **6.4 Patterns of Change Over Time Part I: Sampling Variation Versus Trend . . . 27**
- **6.5 Patterns of Change Over Time Part II: Smoothing Change over Time with a Series of Multiyear Estimates 37**

7. SUMMARY: WHEN AND WHEN NOT TO USE MULTIYEAR ESTIMATES 46

8. CURRENCY VERSUS PRECISION - A FUNDING ALLOCATION EXAMPLE 46
- **8.1 A State for Analysis 46**
- **8.2 Non-English Speakers in New State 47**
- **8.3 Question and Challenge 47**
- **8.4 A Hybrid Approach to Allocating the Funding 48**

9. CONCLUSION 50

10. REFERENCES 51

1. INTRODUCTION

The U.S. Census Bureau began implementation of the American Community Survey (ACS) at its full sample size in 2005. The ACS five-year estimates will be calculated for subject and geographic detail comparable to the U.S. Census Bureau's former decennial long form, but produced annually. While the long form represented a snapshot in time, April 1, 2000, the ACS interviews monthly samples (U.S. Census Bureau, 2006a) that will ultimately produce three types of period estimates: one based on one year of collected data, one based on three years, and one based on five years. The three-year and five-year estimates are referred to as multiyear estimates (MYE). In 2006 the U.S. Census Bureau published one-year estimates based on data collected in 2005, and in 2007 one-year estimates based on data collected in 2006. In 2008 the U.S. Census Bureau will publish its first three-year estimates based on data collected from 2005 through 2007 in addition to the 2007 one-year estimate. In 2010 the U.S. Census Bureau will publish its first five-year estimates based on data collected from 2005 through 2009, in addition to producing one-year and three-year estimates.

The U.S. Census Bureau expects that the five-year ACS estimates will be the default estimates to which users will first be directed. Thus a basic comprehension of the annual five-year MYEs is essential for all data users. But many users will also want to explore and understand the possibilities offered by the three-year and one-year estimates. While this paper aims at aiding all data users, it is more helpful to the latter.

The MYEs will provide new opportunities and challenges for those who have in the past used decennial long form data or the ACS one-year estimates. This document is part of a broader effort by the U.S. Census Bureau to prepare data users for the release of the ACS MYEs, an effort which will include a series of user handbooks targeted to specific data user communities such as state and local governments, and rural areas (U.S. Census Bureau, 2008a), and further yet-to-be determined materials posted on the U.S. Census Bureau ACS website. Its role will be to describe the statistical use and interpretation of the MYEs. In particular, it will focus on two new challenges and opportunities the ACS data present: the choice between one-, three-, and five-year estimates, and examining change over time with annually released ACS estimates. The document is not intended as a broad introduction to ACS data products, as documentation of the MYE estimation, or as a primer in statistical methods, as these topics are covered in other places.

This paper's purpose is both expository and exhortatory. We want users to understand certain key concepts, however, we also have recommendations as to how they should and should not use the data. Ultimately, one of the key points that we want to convey is that users have to think about their data needs and what is the best data available for them, for unlike the decennial sample data, they will have choices. We hope this paper informs them in their thinking. At the same time we recognize that as more multiyear estimates are released data users will develop best practices, and that users' knowledge and needs will grow perhaps in ways that the authors can't anticipate. Hence we expect to update the document over time in response to users' needs as they become apparent. It is intended to serve as a starting point in the development of data users practices, not an end point. The data users we are targeting include both statistically unsophisticated users and statistically sophisticated users who could benefit from an introduction to these new data products. While we strive to make the document helpful to less sophisticated users, we also strive to make the coverage of topics broad and in sufficient depth, which leads in some cases to material that may be too statistically advanced for the less sophisticated users. We hope overall to have achieved a good balance between accessibility and thoroughness.

The first several sections of the document are shorter and serve as further introduction and background. Section 2 starts with background on the ACS, Section 3 gives statements of the statistical issues that the paper focuses on, and Section 4 is a brief introduction to the MYE Study, which was the source of the examples. Since the two major new challenges for data users are the choice between one-, three-, and five-year estimates, and interpreting change in estimates over time, the bulk of the paper is dedicated to these topics in the next three sections. Section 5 highlights statistical issues relating to the choice between one-, three-, and five-year estimates. It discusses the relative precision of the period estimates, the trade-off between currency and reliability, the precision of estimates of subpopulations, and methodological topics of importance such as how controlling weights affects the precision of estimates and how changing geographic boundaries are handled. Section 6 discusses the interpretation of change over time using ACS estimates. Section 7, which is brief, summarizes the lessons of the previous two sections. Lastly, Section 8 illustrates with a more extensive example the challenges of using ACS data for decision making in the context of characteristics changing over time.

2. BACKGROUND

2.1 The Decennial Census Long Form

The decennial census long form questionnaire has been sent to a sample of households for every census since 1940. In 2000, approximately one in six households was in the long form sample. The long form contained all of the questions on the decennial census short form, as well as additional detailed questions relating to the social, economic, and housing characteristics of each individual and household (U.S. Census Bureau, 2002). Information derived from the long form is referred to as census sample data, and was tabulated for geographic entities as small as block group in the 1980, 1990, and 2000 census data products.

2.2 The American Community Survey

The ACS asks essentially the same questions as the Census 2000 long form. However, it offers different data products, and there are some differences in resulting estimates because of differences in reference periods and in how the data are collected (U.S. Census Bureau, 2006b). The ACS has been producing one-year estimates since 1997 for selected geographic areas. The ACS sample was increased to its full size starting in 2005, and starting with estimates for 2005 the ACS provides full sets of estimates annually for all states and for all communities of 65,000 persons or more. For less populous communities, such as rural areas, city neighborhoods, or very small population groups, the sample size is too small to make reliable estimates from one year of ACS sample. Geographic entities with populations of at least 20,000 receive three-year estimates (U.S. Census Bureau, 2006a). The ACS will produce its first three-year estimates with a fully implemented sample in 2008, then will produce three-year estimates every year thereafter.

With a sample of about three million addresses per year, it will take five years of accumulated sample data for the ACS to provide estimates for subject and geographic detail comparable to that of the U.S. Census Bureau's former decennial long form. Estimates based on five years of sample data will be published for all statistical, legal, and administrative entities, including census tracts, block groups, and small incorporated places such as cities and towns. The ACS will produce its first five-year estimates in 2010 and then will produce five-year estimates every year thereafter. For additional background on ACS data products and operational considerations see Torrieri (2007).

An important way that ACS estimates differ from decennial estimates is that they are period estimates. The ACS estimates the average of a characteristic over the year or period years, as opposed to the characteristic at a point in time. In a given year, one-year period estimates are based on data collected during the twelve months of the calendar year, three-year period estimates are based on data collected during the 36 months of the three most recent calendar years, and five-year period estimates are based on data collected during the 60 months of the five most recent calendar years. Table 1 shows the ACS data release strategy as it will unfold over the coming years along with the ACS data collection periods at full sample size and the population thresholds for the different period length estimates. There are further differences between ACS estimates and decennial estimates because of differences in residence rules. The ACS interviews people at their current residence, where current residence is defined by the two-month rule (see U.S. Census Bureau, 2003), whereas the decennial census interviews people at their usual residence.

Table 1: Year(s) of Data Collection for Producing ACS Estimates by Year of Data Release

Data Product	Population Threshold	2006 Release	2007 Release	2008 Release	2009 Release	2010 Release	2011 Release	2012 Release
One-year Estimates	Population 65,000+	2005	2006	2007	2008	2009	2010	2011
Three-year Estimates	Population 20,000+			2005-2007	2006-2008	2007-2009	2008-2010	2009-2011
Five-year Estimates	All Areas					2005-2009	2006-2010	2007-2011

As can be seen in Table 1, the ACS provides more current data than the long form in between decennial census years. It also has the advantages that it will unburden the decennial census of the long form operations, and that the continuous data collection will allow for the development of an experienced cadre of interviewers who can collect higher quality data. On the other hand, the ACS five-year estimates have a smaller sample than the Census 2000 long form. However, we believe the gain in the quality of the interviews partially offsets the loss in sample size.

The weighting for the ACS multiyear estimates will be similar to that which has been used to produce the ACS one-year estimates (Asiala and Tersine, 2007; U.S. Census Bureau, 2006a). Data for interviews completed in the months of the estimation period will be combined together and the weighting steps will be similar to those of the one-year weighting. The one-year ACS population and housing unit estimates are controlled by county or groups of counties to the U.S. Census Bureau's official estimates from the Population Estimates Program (U.S. Census Bureau, 2007a) for that year. In contrast, the ACS MYEs are controlled to the average of the U.S. Census Bureau's official estimates over the individual years of the multiyear estimation period.

3. STATISTICAL ISSUES AND ISSUES OF INTERPRETATION FOR MULTIYEAR DATA

The ACS has inspired much discussion inside and outside the U.S. Census Bureau on the interpretation and use of ACS data. Thinking on this topic from outside is diverse and includes work by the Committee on National Statistics (2007), the Transportation Research Board (2008), and Taeuber (2006, though this work makes reference to outdated MYE methodology). The crux of the current work is to draft guidelines to address the following issues and to develop examples and case studies to illustrate them.

1. How do we explain that MYEs are period estimates and what that means?
2. How can we describe in general terms for which kinds of applications we recommend one-year, three-year or five-year estimates? In particular, what key concepts can we convey to users to help guide them in this choice?
3. How do the variances of the one-year, three-year and five-year estimates compare across geographic areas of differing sizes and across subpopulations?
4. What is the trade-off between currency and sampling variability of ACS estimates?
5. How cautious must users be when comparing multiyear estimates to one-year estimates, such as the one-year estimate which is at the 'center' of the multiyear period?
6. How do we explain to data users how to use multiyear estimates to track change over time and what can we say about the sampling variability of ACS estimates in measuring change over time?
7. What cautions do we give users about comparing MYEs whose periods overlap?
8. When does a series of MYEs reveal changes that sampling variance obscures in a series of one-year estimates? That is, when are MYEs useful to smooth a series of estimates?

4. THE MULTIYEAR ESTIMATES STUDY

To demonstrate and discuss the various relationships among the estimates for the different length periods, we found examples from among the 34 counties in the Multiyear Estimates Study (MYE Study). These 34 counties are a subset of the 36 counties selected for the U.S. Census Bureau's test phase of the ACS from 1999 to 2001 (U.S. Census Bureau, 2004). Five of these 34 counties had a sampling rate of 3 percent and the rest had a sampling rate of 5 percent from 1999 to 2001. From 2002 to 2004 they were included in the ACS with a base sampling rate of 2.5%. In 2005 the 34 counties were included in the fully implemented ACS with a base sampling rate of 2.3%. Three-year estimates starting with the 1999-2001 estimate and five-year estimates starting with the 1999-2003 estimate for these 34 test counties were released in 2007 as part of the MYE Study. One-year estimates were also released starting with 2000 (1999 one-year estimates were not released because of operational limitations). For more details on the sample and estimation of the MYE Study see U.S. Census Bureau (2007b).

In preparation for the 2008 release of multiyear estimates and to help data users begin to understand the characteristics of multiyear estimates, the U.S. Census Bureau conducted the MYE Study. Data profiles that include demographic, social, economic, and housing characteristics were produced for a broad set of geographic areas in 34 counties. Data products are available for counties, place parts, minor civil divisions, school districts, Public Use Microdata Areas, ZipCode Tabulation Areas, American Indian areas, tracts, and block groups. An overview of the MYE Study can be found at U.S. Census Bureau (2007c), and details on the study, including a full inventory of the geographic areas and products available, can be found at U.S. Census Bureau (2007d).

The MYE Study data were released for research purposes only. It is important to recognize that the estimates included in these data profiles have not undergone the subject matter and technical review required for standard ACS data releases. For this reason these estimates are not official estimates. We strongly suggest that users consult with U.S. Census Bureau staff before drawing conclusions or taking actions based on these data.

5. KEY CONCEPTS FOR DATA USERS

In this section we discuss key concepts about MYEs that we want data users to comprehend. We defer to Section 6 the topic, “Interpreting Change Over Time with Multiyear Estimates,” which is extensive enough that we devote to it a section of its own.

5.1 Multiyear Estimation: Key Concepts

We have identified several key concepts that users should grasp when using MYEs and when deciding between the use of MYEs and one-year estimates.

- The ACS estimates are period estimates.
- There is a relationship between the precision of the one-, three-, and five-year estimates.
- Estimates of small subpopulations can have large standard errors even in large geographies.
- There is a trade-off between the currency and precision of estimates.
- A MYE is not an estimate of its center year.
- Comparisons should be made between areas using the same length of period estimates.
- ACS estimates are controlled to total population and housing unit counts from the Population Estimates Program.

More detailed discussion of these concepts follows.

5.2 Period Estimates

The first key concept users must grasp is that the ACS estimates are period estimates. Users should interpret the ACS estimates as an average over the collection period. They give equal weight to each month. Contrast this with the decennial census, which is interpreted to be a snapshot of April 1 of the census year. Thus the ACS one-year estimate for 2000 and the Census 2000 estimate do not measure the same time frame. Taeuber (2006) puts it well: "The one-year estimate is NOT an estimate of an implied beginning, midyear (such as July), or end-of-year characteristic".

Conceptually, a multiyear period estimate is an extension of a one-year estimate. Data are collected and combined for 36 or 60 months. They are controlled to an average of annual Population Estimates over the time period. To be clear, MYEs are not an average of one-year estimates, though the MYE will be approximately equal to the average of the one-year estimates included in its period. Further, they are not moving averages although they share some characteristics with moving averages.

The example in Table 2 illustrates the difference between an estimate of an average of a characteristic over a year and a point-in-time estimate of a characteristic. Because neither the census nor ACS produce monthly estimates this example has to be artificial. Imagine a town on the Gulf of Mexico whose population is dominated by vacationing retirees in the winter months but by year-round local residents in the summer months. Consider the characteristic 'in labor force (population 16 years and over)' by month. Retirees would be in the labor force at very low rates. The average percent over the year is 45, which is about what the ACS estimate would be (it wouldn't be exactly that because of sampling variation, missing data, and controlling). A census estimate for April 1 would be higher at 60 percent.

Table 2: Percent of Population 16 Years and Over in Labor Force - Winter Village, Artificial

Month	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sept	Oct.	Nov.	Dec.
In Labor Force	20	20	40	60	60	60	60	60	60	50	30	20

To illustrate the concept of a period estimate as it applies to multiyear estimates, consider the estimates in Table 3. It shows one-year estimates from 2001 to 2003 from the MYE Study. Compare these to the 2001-2003 estimate of 14.1 percent (with a margin of error of $\pm$ 0.3)[1]. Note that the three-year estimate does not correspond to the estimate of the middle year of 13.7, nor to that of the first or last year of the period. It is an average based on the 36 months of pooled data.

[1]In this document the margin of error is defined to be $\pm$ 1.645 times the SE. The bounds formed using this margin of error are equivalent to the bounds of a 90% confidence interval.

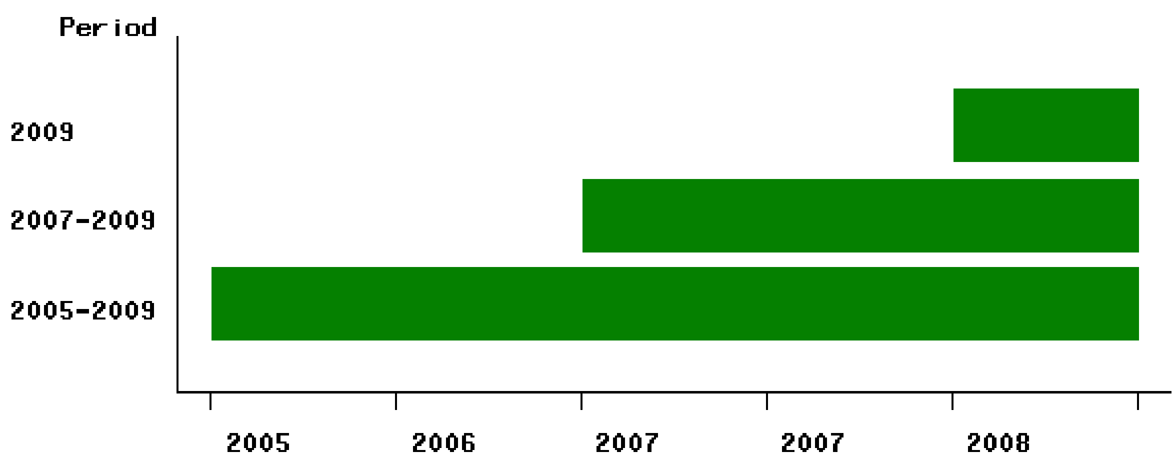
Table 3: Percent of Population 5 Years and Over who Speak Spanish at Home[2] with Margins of Error Based on 90% Confidence Intervals - Lake County, IL (There were 644,299 people 5 years and over in Lake County according to the 2005 PEP)

Period	2001	2002	2003	2001-2003
	13.6 (0.4)	13.7 (0.5)	15.1 (0.6)	14.1 (0.3)

5.3 Relative Precision of the One-, Three-, and Five-Year Estimates for Totals

Statistical measures of the precision of an estimate include its standard error (SE) and coefficient of variation (CV), which is the SE divided by the estimate. For uncontrolled estimates of totals the SE is a function of the sample size. A three-year MYE is based on three times as much sample data as a one-year estimate, and a five-year MYE on five times as much sample data. Figure 1 below shows the data collection periods for the ACS estimates that will be released in 2010, the first year that the five-year estimates will be released. The relationship between the sample sizes used for the estimates means that in many cases one can give an approximate relationship for the SE or CV of one-, three-, and five-year estimates of totals of persons, households, or housing units with certain characteristics. The SE's and the CV's of the three-year estimates are about one over the square root of three, or about 58%, of the one-year estimates; and the five-year estimates are about one over the square root of five, or about 45% of the one-year estimates.

Figure 1: Data Collection Periods for One-, Three-, and Five-Year Estimates



For an illustration of how SEs decrease as the period lengthens consider the 2000-2004, 2002-2004, and 2004 estimates of relationship in Black Hawk County, IA (from the MYE Study) which are seen in Table 4. The SEs of the three-year estimates range from about 54% to 69% of

[2] For the topic definition of "population 5 years and over who speak Spanish at home" and others referred to in this document, see the U.S. Census Bureau website, "How to Use the Data: Subject Definitions", at **http://www.census.gov/acs/www/UseData/Def.htm**.

the one-year estimates. For example, the SE for the 2002-2004 three-year estimate for Householder, 414, is 54.3% of the 2004 one-year estimate, 762. The SEs of the 2000-2004 five-year estimates range from about 34% to 56% of the 2004 one-year estimates. The observed three-year and five-year SE's in this example differ from the predicted 45% and 58% of the one-year SE's because of sampling variation and the effects of controlling, among other reasons.

Table 4: Relationship to Householder - Comparisons of Standard Errors - Black Hawk County, IA

Relationship to Householder	2000-2004 Total	2000-2004 SE	As a Percent of 2004 SE	2002-2004 SE	As a Percent of 2004 SE	2004 SE
Householder	51,457	263	34.5	414	54.3	762
Spouse	25,305	320	45.5	405	57.6	703
Child	31,751	239	36.4	409	62.3	657
Other relatives	3,827	282	56.1	349	69.4	503

While these relationships hold approximately for estimates of totals, they don't hold up as well for estimates of ratios or averages because they involve estimates in both the numerator and the denominator.

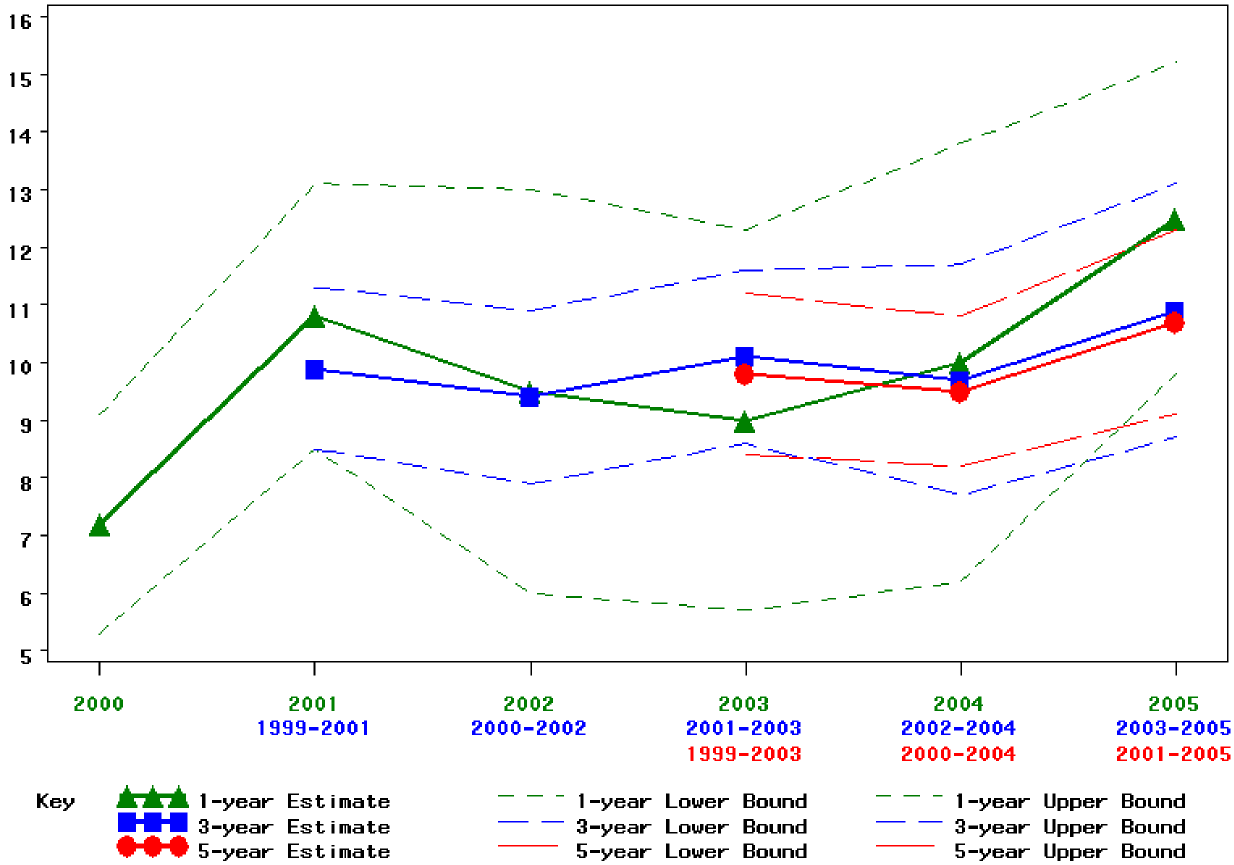
In Table 5 one sees another example of how SE's shrink with the MYEs, but in this case the estimates are of percentages. Note that the percentage of families below poverty level is the ratio of two estimates: the number families below poverty level and the families. Thus this is a situation where we don't expect the relationship between the standard errors to hold up as well. One sees in the columns titled 'As a Percent of 2004 SE' that the relationships between the size of the SEs of the 2000-2004 and 2002-2004 estimates to the SEs of the 2004 estimates are not far off from the approximation given above. However, the proportions do tend to be lower than the 45% and 58%. For example, the SE of the 2000-2004 estimate of the percent married couple families with related children under 18 years below poverty level is 1.2%, or 30.8% of 3.9%, the SE of the 2004 estimate.

Table 5: Percent of Families Below Poverty Level by Family Type for Sevier County, TN

	Census 2000	2000-2004 ACS	2000-2004 ACS	2000-2004 ACS	2002-2004 ACS	2002-2004 ACS	2002-2004 ACS	2004 ACS	2004 ACS
	Total Family Types	Percent Below Poverty Level	SE	As a Percent of 2004 SE	Percent Below Poverty Level	SE	As a Percent of 2004 SE	Percent Below Poverty Level	SE
All families	20,836	9.5	0.8	34.8	9.7	1.2	52.2	10.0	2.3
With related children under 18 years	9,494	15.3	1.5	33.3	16.5	2.4	53.3	17.8	4.5
Married-couple families	16,884	5.8	0.7	35.0	5.4	0.9	45.0	7.9	2.0
With related children under 18 years	6,968	7.7	1.2	30.8	7.3	1.7	43.6	12.1	3.9
Female householder, no husband present	2,883	27.2	3.0	41.7	26.7	4.8	66.7	19.0	7.2
With related children under 18 years	1,876	40.2	4.9	37.7	40.4	6.8	52.3	38.3	13.0

One can also compare the one-year, three-year, and five-year estimates from 2000 through 2005 graphically in Figure 2, which plots estimates of percent below poverty level over time for all families. The estimation period is on the horizontal axis and the three estimates are aligned by the last year of the estimation period. That is, the 2005 estimate, the 2003-2005 estimate, and the 2001-2005 estimate are placed in the same horizontal position. We compare the estimates based on the last year of the estimation period because these estimates are released at the same time, 2006, and they represent the choice of estimates data users will be presented with. The lines connecting the estimates are drawn to help visualize the change over time. The dashed lines are upper and lower 90% confidence intervals or equivalently, margins of error. In this example the one-year estimates have, as expected, larger confidence intervals than the multiyear estimates, a reflection of the one-year estimates' greater sampling variation. The decision to use the one-, three-, or five-year estimates is discussed in later sections, notably in Section 5.5 which talks about the trade-off between precision and currency.

Figure 2: Percent[3] of Families Below Poverty Level - Sevier County, TN
(There were 20,836 families in Sevier County according to Census 2000)



5.4 Precision of Estimates for Subpopulations

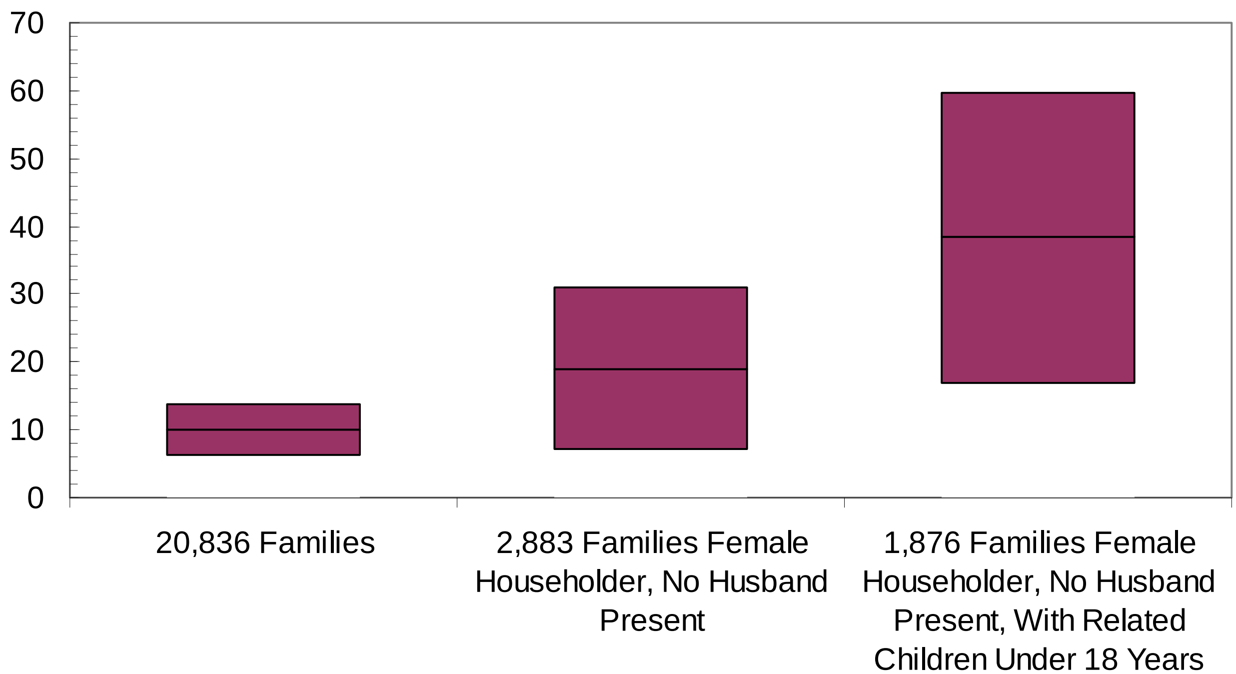
ACS data users need to be cautioned when working with estimates of subpopulations. The standard error of an estimate depends on the sample size upon which it is based, among other factors. An estimate of a characteristic can have an acceptably small standard error when it applies to the full population of a published geography. However, the estimate of that characteristic may have an unacceptably large standard error when it applies to a subpopulation of the same geography because it is the size of the subpopulation which determines how large the sample is for the estimate. For example, consider Sevier County, Tennessee, which had an estimated population of 71,170 according to the Census 2000. This total is larger than the ACS's 65,000 cutoff for publishing one-year estimates for geographic areas. However, some

[3]Note that Figure 2 and most of the other figures in this document do not start at zero percent on the vertical axis. Instead, the graph displays only the range of the observed data. Restricting the range allows the graph to display to best advantage the confidence intervals and the year-to-year changes. However, the reader should be aware that graphs which truncate the lower ranges can be misleading as they do not allow one to compare the variations of the estimates with the overall magnitude of the characteristic of interest.

subpopulations will be much smaller than 65,000. In Table 5 above one sees that there are 20,836 families based on the Census 2000; but the number of families with a female householder, no husband present, with related children less than 18 years, is only 1,876. Not surprisingly, the SE for the 2004 one-year estimate of the percent below poverty level for this subpopulation is large, 13.0%. For such small subpopulations users obtain much more precision using the three-year or five-year estimate. In this example the five-year estimate has a SE of 4.9%, and the three-year estimate has a SE of 6.8%.

One can view the relative size of the standard errors graphically in terms of 90% confidence intervals in Figure 3. (Recall that a confidence interval is a function of the standard error). In short, *MYEs are typically preferable to one-year estimates for examining estimates based on small subpopulations*.

Figure 3: 90% Confidence Intervals for Percent Below Poverty Level by Family Type for 2000-2004 ACS Estimates[4] - Sevier County, TN



5.5 Currency Versus Precision

This section addresses the question of which ACS estimate the data user should use when they have a choice between the one-, three-, and five-year estimates, or between the three- and five-year estimates. For estimates of the smaller geographies the data user will have no choice

---

[4]Note that Figure 3 is not drawn to the same scale as most of the figures in this document in order to accommodate a larger range of observed estimates. Caution should be exercised when comparing graphs which are not drawn in the same scale.

but to use five-year estimates. But for estimates from geographies with populations of 20,000 or larger they will have a choice. In this section we present a key principle to guide the data user in this situation.

The concept that the data user needs to understand is that the key trade-off between one-year estimates and MYEs is currency versus precision: *MYEs yield smaller CV's but use less current data.* In general, data users will want to use a shorter period estimate as it uses data more relevant to what is happening currently. However, if that estimate is not precise enough to answer their questions, currency must be traded for the additional precision of a multiyear estimate. Table 6 illustrates how the most current five-year MYE can differ from the most current one-year estimate. It shows the percentage of people 5 years and over who speak Spanish at home for the test counties Broward, FL, and Lake, IL. The differences between the 2004 and 2000-2004 estimates are not small at 1.4% and 1.5%, and they are clearly larger than their margins of error (MOE) and statistically significant.

Table 6: Percent of Population 5 Years and Over who Speak Spanish at Home

County	2004	MOE of 2004	2000-2004	MOE of 2000-2004	Difference	MOE of Difference
Broward Co., FL	19.9	0.4	18.5	0.2	1.4	0.4
Lake Co., IL	15.9	0.4	14.4	0.2	1.5	0.4

In Table 7 one sees how the lack of currency is most important when there is a strong linear trend over time. In the first row, one sees percent Spanish spoken at home as measured by one-year estimates increasing from 13.1% in 2000 to 16.8% in 2005. In the second row one sees how the four three-year estimates lag in revealing the increase. Not unexpectedly, the five-year estimates lag even further. To illustrate what is meant when it is said the MYEs lag, consider the three estimates whose last estimation year is 2005; the one-year estimate is at 16.8%, while the three-year estimate lags about a year behind at 15.9% (about the same as the 2004 one-year estimates), and the five-year about two years behind at 15.1% (about the same as the 2003

Table 7: Percent of Population 5 Years and Over who Speak Spanish at Home with Margins of Error Based on 90% Confidence Intervals[5] - Lake County, IL (There were 644,299 people 5 years and over in Lake County according to the 2005 PEP)

One-year Estimate	2000	2001	2002	2003	2004	2005
	13.1 (0.5)	13.6 (0.4)	13.7 (0.5)	15.1 (0.6)	15.9 (0.4)	16.8 (0.5)
Three-Year Estimate		1999-2001	2000-2002	2001-2003	2002-2004	2003-2005
		13.2 (0.2)	13.4 (0.3)	14.1 (0.3)	14.9 (0.3)	15.9 (0.3)
Five-Year Estimate				1999-2003	2000-2004	2001-2005
				13.7 (0.2)	14.4 (0.2)	15.1 (0.3)

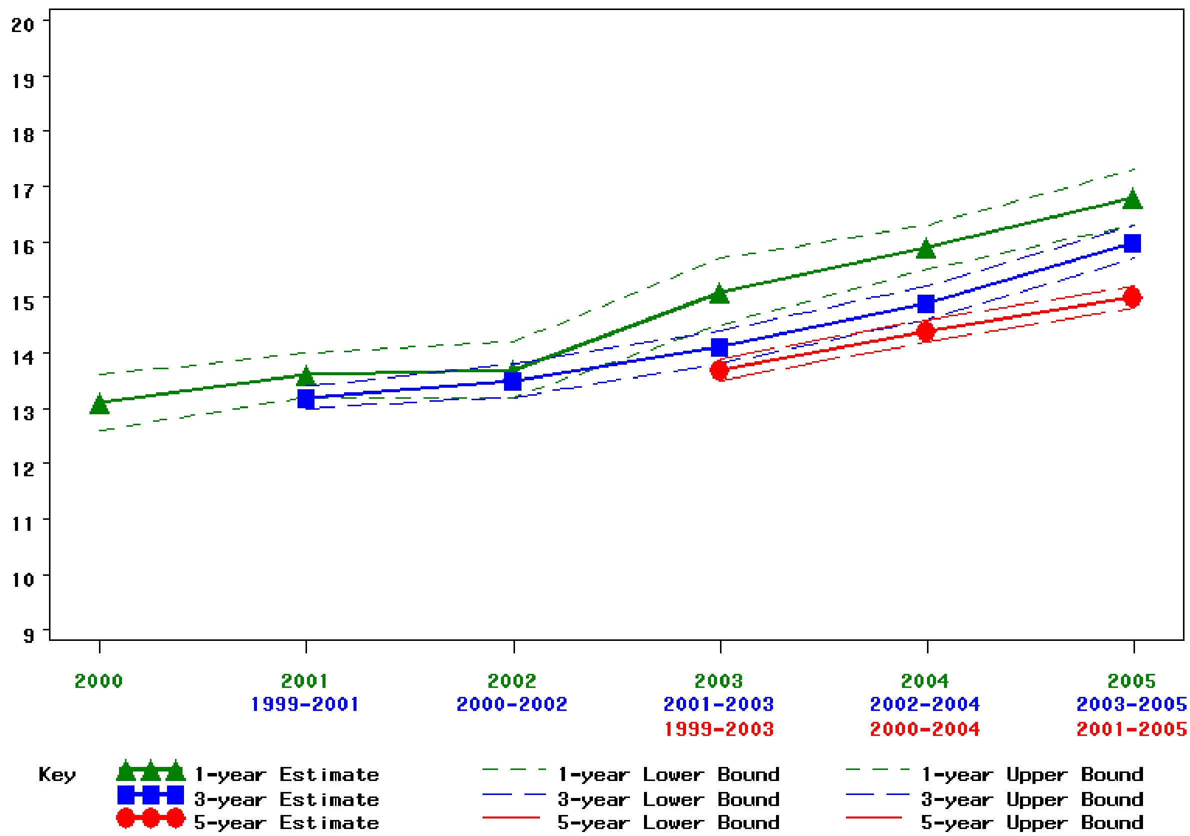
[5]The margins of error are the values in parentheses.

one-year estimate). In Section 6 we discuss in more detail looking at change over time with MYEs. (Note that the differences between the one-year, three-year, and five-year estimates released in the same year are statistically significant, with the one exception of the comparison between the 2002 estimate and the 2000-2002 estimate).

The data from Table 7 are displayed in Figure 4, where one sees the lag clearly for both the three- and five-year estimates. As in Figure 2, the three estimates are aligned by the last year of the estimation period so as to compare those estimates that are released at the same time. For these data the one-year estimate is preferred for its greater currency.

We contrast the Lake County example (Table 7 and Figure 4) with the example of percent below poverty level of families in Sevier County, TN (Figures 2 and 3). In Figure 2 one sees more sampling variability than in Figure 4. If there is change over time in the percent below poverty level it is masked by the variability and cannot be detected. For the data in Figure 2 one should use the five-year data for its better precision.

Figure 4: Percent of Population 5 Years and Over who Speak Spanish at Home - Lake County, IL
(There were 644,299 people 5 years and over in Lake County according to the 2005 PEP)



5.6 A Multiyear Estimate is not an Estimate of its Center Year

Usually, data users are interested in the current state of affairs and are focused on the most recent data available. Most of the discussion in these guidelines takes this perspective - it is for this reason that when graphically comparing one-, three-, and five-year estimates we align them by the last year of their estimation period. However, in some situations one-year estimates could be unavailable or of insufficient reliability, or the interest may be in historical data, in particular, in a time series. In such circumstances data users may want to use the MYE as an estimate of the middle year of the MYE period. This way of viewing MYEs merits a fair amount of discussion. This section emphasizes the point that an MYE is not an estimate for any single year and also introduces the discussion of looking at a series of MYEs over time. We take up this latter topic again in Section 6.

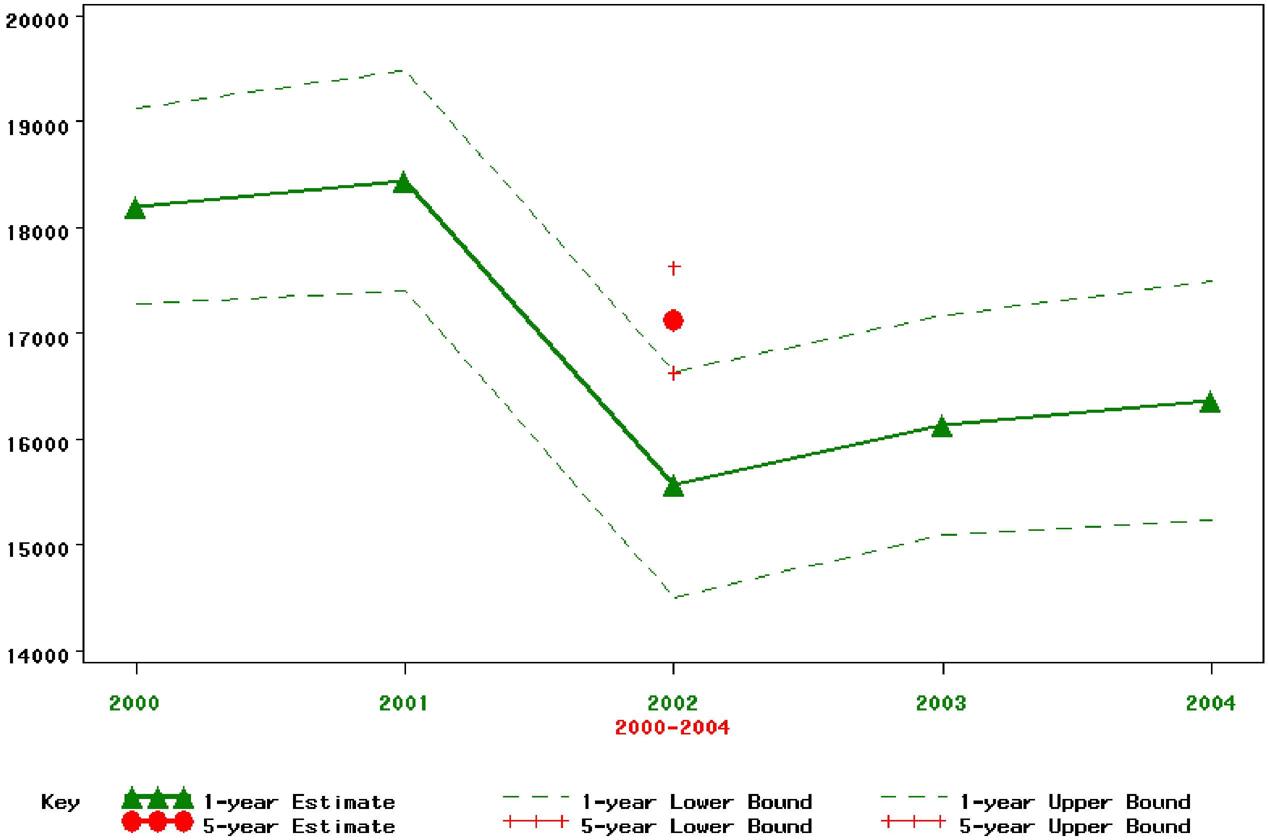
We start the discussion with a cautionary example. In Table 8 we compare the one-year estimate for 2002 versus the 2000-2004 MYE for the percent of population 5 years and over who speak Spanish at home in Tulare County, CA. One sees that the percent for 2002 is 2.8% higher than for 2000-2004, no small difference. A researcher who used the MYE as an estimate of the middle year could come to a misleading conclusion.

Table 8: Percent of Population 5 Years and Over who Speak Spanish at Home - Tulare County, CA
(There were 360,347 people 5 years and over in Tulare County according to the 2004 PEP)

2002 Percent	2002 MOE	2000-2004 Percent	2000-2004 MOE	Difference Percent	Difference MOE
43.3	1.3	40.5	0.6	2.8	1.5

In fact, it is the pattern of change of the estimates across years that determines if the MYE is close to the estimate for the middle year of the period. *If the change over time is roughly linear, the MYE may be a reasonable estimate of the middle year of the period. If it is not linear, especially around the year of interest, it can be a poor estimate of the middle year.*

Figure 5: Civilian Veterans: One-Year Estimates and Five-Year Estimate - Schuylkill County, PA[6]
(The population of Schuylkill County was 146,209 according to the 2005 PEP)



The next two examples will illustrate these points. The first of them again shows the danger of thinking of an MYE as the middle year of its period. In Figure 5, 'Civilian Veterans - One-Year Estimates and Five-Year Estimate - Schuylkill, PA ', one sees one-year estimates from 2000 through 2004 along with the five-year estimate for 2000-2004. There is a sharp drop in the estimate from 2001 to 2002. Not surprisingly, the five-year MYE, 17,127, based on the 2000 to 2004 data, is significantly different from the one-year estimate for 2002, 15,570. One can see from the confidence intervals of these estimates in Figure 5 that the differences are statistically significant.

[6]Note that Figure 5 is not drawn to the same scale as most of the figures in this document as it displays counts and not percentages. Caution should be exercised when comparing graphs which are not drawn in the same scale.

Figure 6: Civilian Veterans – One-year Estimates and Multiyear Estimates - Schuylkill County, PA[7]
(The population of Schuylkill County was 146,209 according to the 2005 PEP)

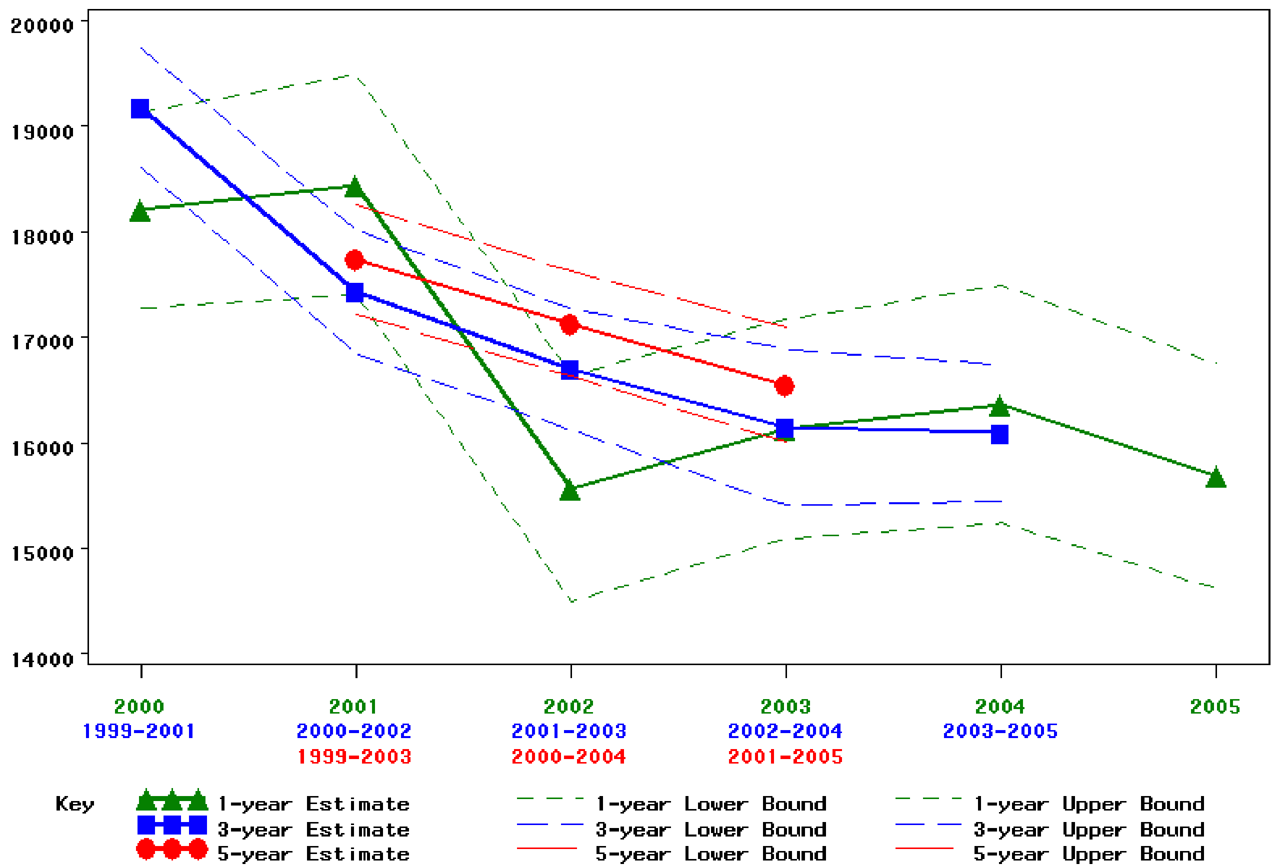
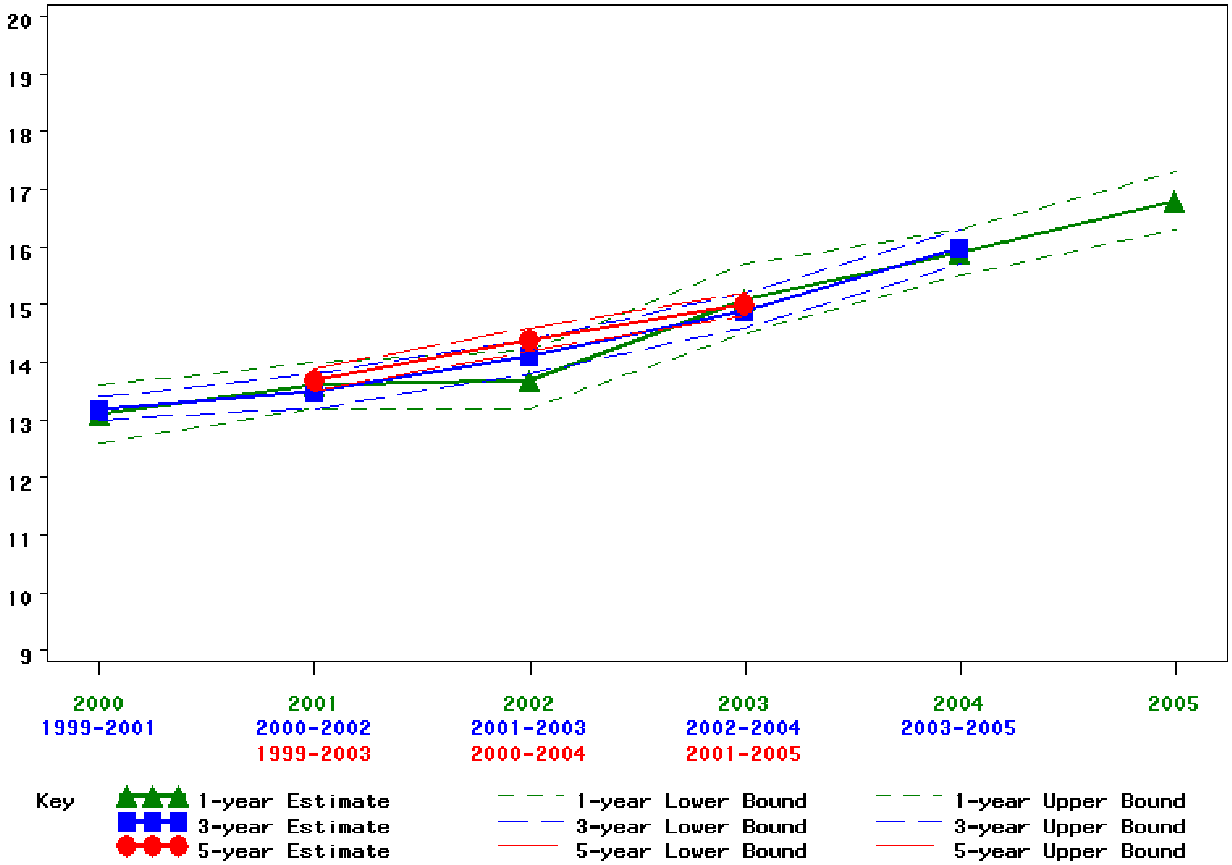


Figure 6 includes the MYEs from 2000 through 2005. Notice that the one-, three-, and five-year estimates are aligned based on the middle year of the estimation period (unlike Figures 2 and 4). One sees that the MYEs, both the three-year and the five-year, smooth over the sharp drop seen in 2002. Generally, they don't track the one-year estimates.

[7]Note that Figure 6 is not drawn to the same scale as most of the figures in this document as it displays counts and not percentages. Caution should be exercised when comparing graphs which are not drawn in the same scale.

Figure 7: Percent of Population 5 Years and Over who Speak Spanish at Home - Lake County, IL
(There were 644,299 people 5 years and over in Lake County according to the 2005 PEP)



In contrast, in Figure 7, 'Percent of Population 5 Years and Over who Speak Spanish at Home - Lake County, IL ', one sees an upwards linear trend in the one-year data. As in Figure 6 we align the estimates based on the middle year of the estimation period. However, here we see the MYEs are generally close to the one-year estimates. This is an example where the MYEs are a good estimate of their middle year. In short, while an MYE is not an estimate of any single year of its data collection period, using MYE as an estimate of the middle year can be reasonable if the interest is in a historical time series and if the trend over time is roughly linear. See Section 6 for a related discussion about analyzing change over time with MYEs.

5.7 Make Comparisons between Areas over the Same Time Periods

When comparing estimates across geographies or subpopulations, users should compare the same period length for each estimate. For example, say one wishes to compare the percent of population 5 years and over who speak Spanish at home) between a suburban area, Lake County IL, and a rural area, Upson County GA. (In a more realistic example one might want to pick neighboring suburban and rural counties, but the MYE Study has only 34 counties to choose from). The most recent Lake County published estimates are for the periods 2001-2005, 2003-2005, and 2005, whereas Upson County has only 2001-2005 estimates available; see Table 9. The correct comparison is Lake County's five-year estimate with Upson's five-year estimate: 15.1% versus 1.4%. To compare Lake County's one-year estimate for 2005, 16.8%, with Upson County's five-year estimate, 1.4%, would confound differences due to geography with differences due to time frame.

Table 9: Percent of Population 5 Years and Over who Speak Spanish at Home - Lake County, IL and Upson County, GA

	Lake County, IL	Upson County, GA
One-Year Estimate, 2005	16.8 (0.5)[8]	unavailable
Three-Year Estimate, 2003-2005	15.9 (0.3)	unavailable
Five-Year Estimate, 2001-2005	15.1 (0.3)	1.4 (0.4)

5.8 Controlling Selected Estimates

The ACS constrains total housing unit (HU) and person weights to conform with annual estimates from the Census Bureau's Population Estimates Program by demographic groups defined by sex, age, race and Hispanic origin and by housing unit totals. Controlling estimates in this manner reduces bias due to undercoverage of HUs and people within HUs relative to the Population Estimates, and reduces the variability of the ACS HU and person estimates. The level of geography at which the controlling is done is the weighting area, which is usually a county but can be two or more counties when a smaller county is grouped with others (though in the MYE Study all controlling was done at the level of the individual county). Note that for the characteristics used as controls we would never recommend obtaining estimates from the ACS. Users should obtain basic demographic and HU estimates from the Population Estimates Program whenever possible.

If there is enough sample in a weighting area, the ACS exactly controls estimates by 156 combinations of sex, age, race, and Hispanic origin. However, for most weighting areas the 156 combinations are collapsed into a smaller number of combinations which are controlled exactly. These exactly controlled estimates are assumed to have no standard error at the weighting area level because population estimates are assumed to have no sampling error. However, published ACS sex, age, race and Hispanic origin estimates typically have modest standard errors because the published groupings are not the same as the groupings that are exactly controlled.

[8] The values in parenthesis are the margins of error.

Since the MYEs are an estimate for the multiyear period, their controls are the average of the annual independent population estimates over that period. Thus the controls for MYEs are not those for a particular year; e.g., the 2005-2007 controls are the average of the 2005, 2006, and 2007 independent population estimates.

Estimates strongly correlated with controlled variables also have lower SEs and CVs at the level to which they are controlled. Consider the estimates for occupied housing units, which are highly correlated with housing unit totals, estimates to which the ACS controls. The estimated number of housing units in the Bronx in 2005 is 502,211; this estimate has a SE of zero since it comes from the Population Estimates Program. In Table 10 one sees the estimate for the number of occupied HUs, 468,210, has only a small SE of 1,774. The estimate for the number of renter-occupied HUs, 369,974, has a SE of 2,950, which, while still modest, is larger than that of occupied housing units because its correlation with the total number of housing units is not as strong.

Table 10: 2005 Estimates of Occupied Housing Units - Bronx County, NY

	Total	SE	Percent	SE
Occupied HUs	468,210	1,774	not applicable	
Owner-occupied	98,236	2,403	21	0.5
Renter-occupied	369,974	2,950	79	0.5

In short, because the ACS controls its estimates of totals of housing units and totals of persons with certain demographic characteristics, it is not valuable for estimating those very totals - data users should be referred to the official population estimates. On the other hand, characteristics highly correlated with controlled demographic characteristics, such as housing unit tenure, will have small SEs.

5.9 Additional Methodological Considerations

In addition to controlling, there are two other methodological considerations which users should be aware of. First, for tabulation areas that can change from year-to-year, the legal boundaries used to determine the housing units and persons residing there are those reported in the Boundary and Annexation Survey (U.S. Census Bureau, 2008b) as of January 1 of the last year included in the estimation period. Second, to account for price changes, inflation factors are used to adjust dollar amounts reported for all years in the estimation period to the average of the last estimation year. For more information on inflation adjustments see U.S. Census Bureau (2006a).

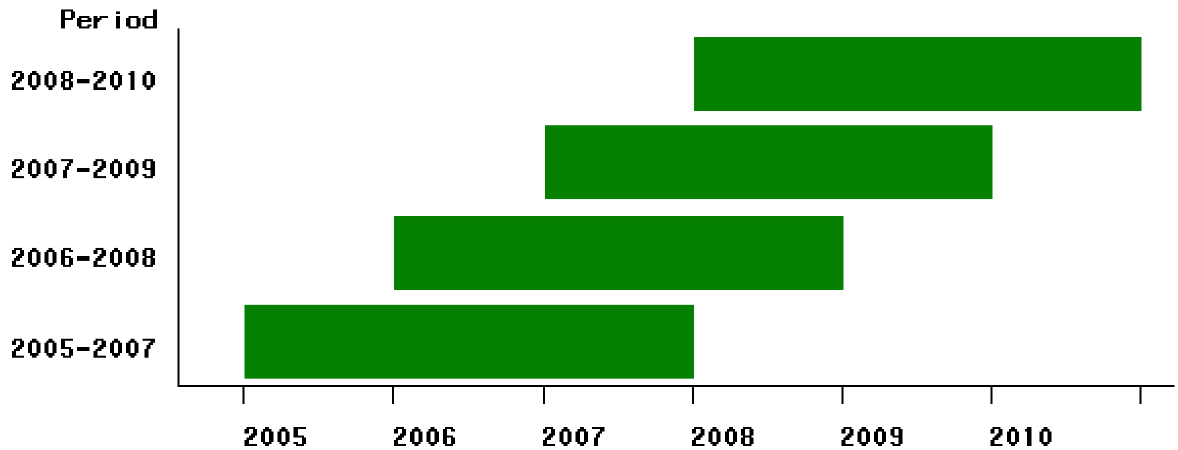
The implementation of these methods is illustrated in the following two examples. Consider first the three-year estimates published in 2008 for the years 2005, 2006, and 2007. The boundaries used are those reported in the Boundary and Annexation Survey as of January 1 of 2007. So that the dollar amounts throughout the three years are treated as if they were reported in the same year, inflation factors are used to adjust dollar amounts reported for the prior two years to amounts for the average of 2007. Next consider the five-year estimates published in 2010 for the years 2005, 2006, 2007, 2008, and 2009. The boundaries used are those reported in the Boundary and Annexation Survey as of January l of 2009. So that the dollar amounts throughout the five years are treated as if they were reported in the same year, inflation factors are used to adjust dollar amounts reported for the prior four years to amounts for the average of 2009.

6. INTERPRETING CHANGE OVER TIME WITH MULTIYEAR ESTIMATES

The ACS will release one-year and multiyear estimates annually, in contrast with the decennial census which produced only one set of estimates every ten years. These annual estimates will present data users with opportunities to analyze estimates over time that they didn't have with census estimates. Interpreting change over time with MYEs is a topic broad enough that we devote a section to it. Recall that we have already touched on this topic in previous sections. In the discussion of currency versus precision in Section 5.5 we pointed out the lack of currency of the MYEs. In Section 5.6 we pointed out that when examining a historical time series of MYEs under certain circumstances it may be reasonable to think of each MYE as an estimate of its middle year. In the following sections we discuss these points and others in more depth.

When considering change in ACS estimates over time, an important distinction is whether or not the time periods being compared overlap. Analyzing change by examining estimates of non-overlapping time periods is both mathematically simpler and more straightforward to interpret, and thus recommended to users. However, there may be circumstances when overlapping time periods are worth examining, such as when there is an interest in smoothing a series of estimates. We start the discussion in Sections 6.1 and 6.2 by explaining the importance of this distinction. Then in Sections 6.3, 6.4, and 6.5 we put more emphasis on interpreting patterns of change over time and using series of MYEs over time to smooth the effects of sampling variation.

Figure 8: Overlapping versus Non-Overlapping Time Periods



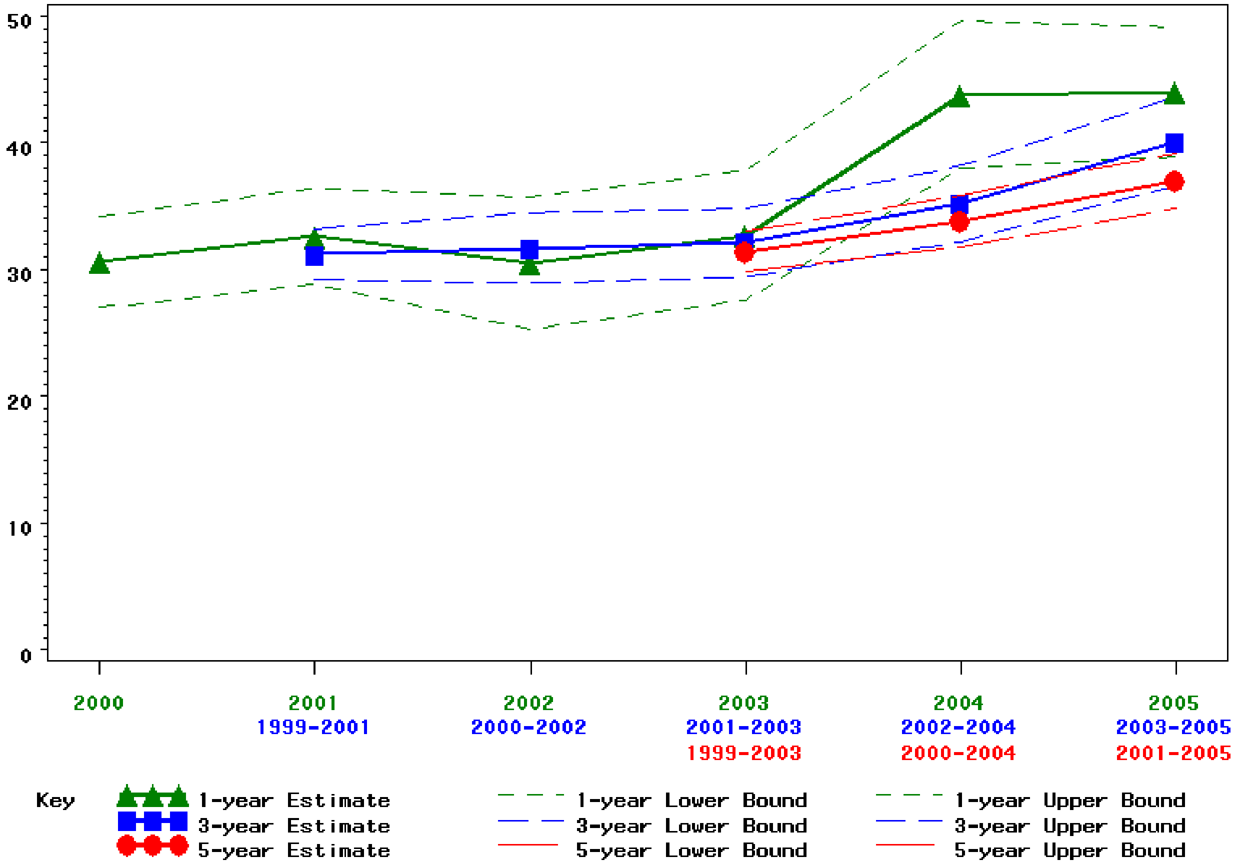
6.1 The Precision of Non-Overlapping Estimates over Time

By non-overlapping estimates we mean that the estimates do not have data collection years in common. Figure 8 displays the data collection period for a series of three-year MYEs from 2005 to 2010. It is easy to see in Figure 8 that the 2005-2007 MYE does not overlap with the 2008-2010 MYE. Since estimates for non-overlapping years are nearly independent, a very good approximation to the variance for a difference between non-overlapping estimates is just the sum of the variances of the two estimates. This is an approximation because the way ACS selects its sample results in a small correlation between estimates of non-overlapping years. In terms of SEs, the approximation to the SE of the difference is the square root of the sum of the squared SEs (variances). That is, if X_1 and X_2 are two non-overlapping MYEs, or any two one-year estimates, then a close approximation to the standard error of the difference between X_1 and X_2 is given as follows:

$$SE(X_1 - X_2) = \sqrt{\left[SE(X_1)\right]^2 + \left[SE(X_2)\right]^2}$$

For example, compare the non-overlapping three-year estimates for percent below poverty level among female-headed households, no husband present, with related children younger than 18, in Multnomah County, OR, for 1999-2001 and 2002-2004: 31.2% with a SE of 1.2% versus 35.2% with a SE of 1.8%. The difference is 4.0% and the standard error of the difference obtained by the formula above is 2.2% (thus the difference is statistically significant). These estimates are displayed in Figure 9.

Figure 9: Percent Below Poverty Level for Families with Female Householder with Related Children under 18 Years[9] - Multnomah County, OR
(There were 77,302 families with a female householder, related children under 18 years in Multnomah County according to Census 2000)



When changes over time can be analyzed with sufficient precision with a series of one-year estimates, then a user may have no reason to consider MYEs. *However, if the differences in a series of one-year estimates are large but not statistically significant, one should consider comparing differences between non-overlapping MYEs which have smaller SE's.* In Figure 9 the year-to-year variations one sees are generally not significant, the one exception being the change in percent below poverty level between 2003 and 2004. For a less variable view of change over time one can consider non-overlapping three-year or five-year estimates, such as 2000-2002 versus 2003-2005.

[9]Note that Figure 9 is not drawn to the same scale as most of the figures in this document in order to accommodate a larger range of observed estimates. Caution should be exercised when comparing graphs which are not drawn in the same scale.

6.2 Comparing Estimates with Overlapping Time Periods

By overlapping estimates we mean that the data collection periods of the two estimates share years in common and consequently also share the housing unit and person data collected in those years. For example, one sees in Figure 8 that the 2005-2007 MYE shares two years of data, 2006 and 2007, in common with the 2006-2008 MYE. The estimates of differences between overlapping MYEs suffer from difficulty in interpretability. Users must be advised to be cautious when examining overlapping MYEs, as *the difference between overlapping MYEs is driven by the difference between the non-overlapping years.*

To illustrate this point we can approximate the differences between overlapping MYEs by assuming that a MYE is equal to the average of the one-year estimates in its period (Committee on National Statistics, 2007). For example, consider two overlapping five-year estimates, M_1, and M_2, that are averages of years one through five, Y_1, Y_2, Y_3, Y_4, Y_5, and years two through six, Y_2, Y_3, Y_4, Y_5, Y_6, respectively. Then one has the following approximations, $M_1 = (Y_1 + Y_2 + Y_3 + Y_4 + Y_5)/5$, and $M_2 = (Y_2 + Y_3 + Y_4 + Y_5 + Y_6)/5$, which lead to $M_2 - M_1 = (Y_6 - Y_1)/5$. Note that approximations to the standard errors of the differences between overlapping MYEs are discussed in the Multiyear Estimate Study's Accuracy Statement (U.S. Census Bureau, 2007b) and by the Committee on National Statistics (2007).

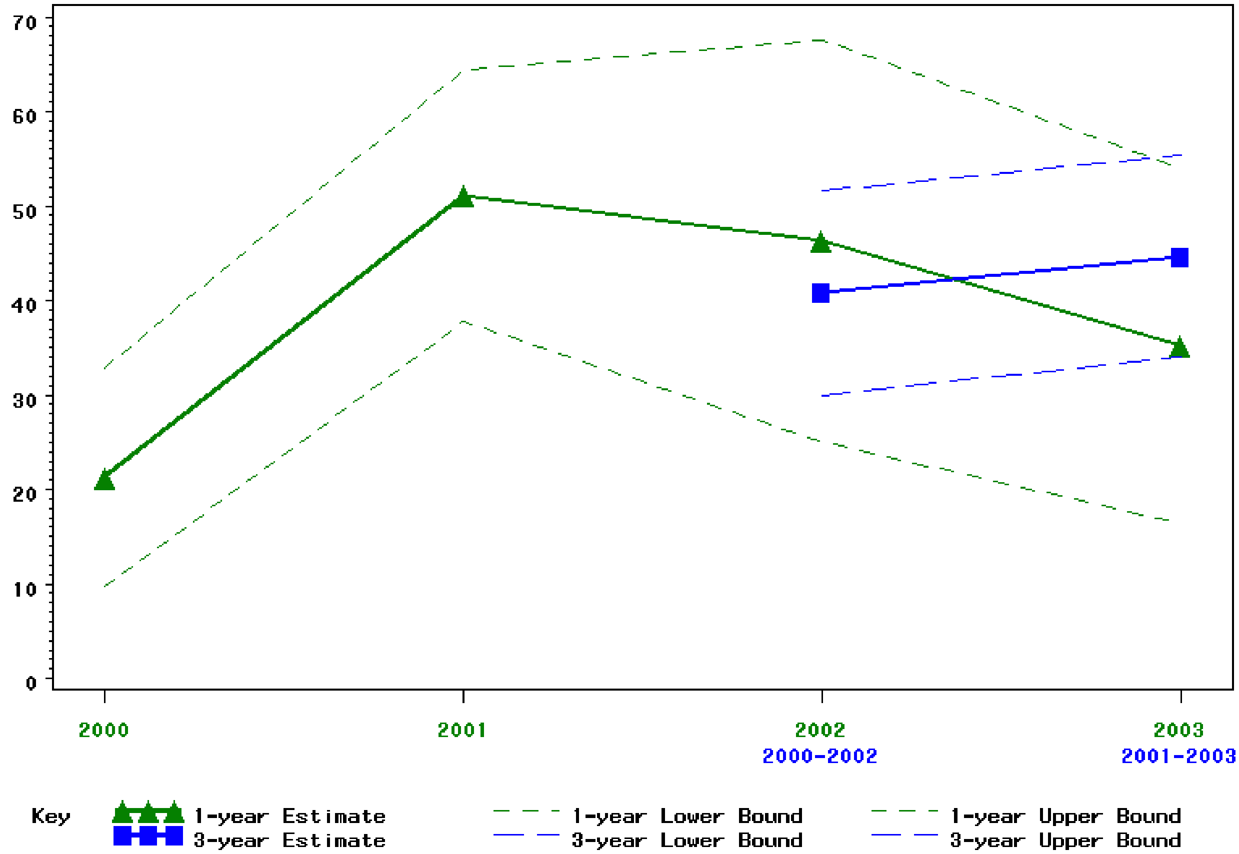
It would be easy for a naive data user to come to incorrect conclusions when directly comparing overlapping estimates. They might interpret the difference between consecutive MYEs as the difference between the most recent years in the estimation periods, or the middle years of the estimation periods. For example, they might see the difference between the 2000-2004 and 2001-2005 estimates as indicative of the difference between 2004 and 2005. The example with artificial data in Table 11 illustrates how such comparisons could be misleading.

Table 11: Percent Below Poverty Level, Artificial County

One-Year Estimate in Percent	2005 11.0	2006 17.0	2007 17.0	2008 14.0
Three-Year Estimate in Percent			2005-2007 15.1	2006-2008 16.1

Consider the two three-year MYEs from 2005-2007 and 2006-2008. The estimate for 2005-2007 is 15.1% and for 2006-2008 is 16.1%. If one viewed the difference as a reflection of the change in the most recent years, 2007 to 2008, one would conclude incorrectly that the percent below poverty level had increased by one percentage point in that one-year period, though it can be seen that the percent below poverty level declined by three percentage points from 2007 to 2008. Similarly, if one viewed the difference between the 2005-2007 and 2006-2008 estimates as the difference between the middle years, 2006 and 2007, one would conclude they had increased by 1% when they remained the same.

Figure 10: Overlapping Periods - Percent Below Poverty Level of Families with Female Householder, No Husband Present, with Related Children Under 18 Years[10] – Sevier County, TN
(There were 1,876 families with a female householder, related children under 18 years in Sevier County according to Census 2000)



This same point is illustrated with real data in Figure 10. One sees the 2001-2003 MYE is higher than the 2000-2002 MYE (the difference is not statistically significant); however, the one-year estimates from 2002 to 2003 decrease (again not statistically significant).

[10]Note that Figure 10 is not drawn to the same scale as most of the figures in this document in order to accommodate a larger range of observed estimates. Caution should be exercised when comparing graphs which are not drawn in the same scale.

6.3 Smoothing with Multiyear Estimates

Because of the complexities of interpretation as discussed above, we discourage direct comparisons between estimates for overlapping time periods. However, plots of a series of overlapping periods can be useful to smooth out short-term fluctuations that can obscure longer-term trends or cycles. This is because a series of MYEs mathematically resembles a series of simple moving averages. Figure 9 presents a good example of the smoothing effects of MYEs. In this graph one sees that the year-to-year change in the one-year estimates is much smaller than the sampling error, so they are generally not significantly different; the only year-to-year change that is statistically significant is the jump from 2003 to 2004. The smoothing is apparent in the three-year estimates where the increase from 2003 to 2004 is spread over several years. In Figure 6 one sees another good example of the smoothing effect of MYEs. These examples typify the effects of the smoothing of the MYEs: sampling variation is suppressed but real changes may be drawn out over time or seen with lag. We will discuss smoothing again and provide more examples in Section 6.6 in the context of examining patterns of change over time.

6.4 Patterns of Change Over Time Part I: Sampling Variation Versus Trend

The purpose of this section is to illustrate some common data patterns over time. While the patterns a data user will encounter are of course particular to the characteristics of interest, there are several features that will commonly arise and are worth pointing out. Most importantly, the data user should understand the nature of sampling variation, sometimes referred to as noise, when examining patterns over time. The patterns we will illustrate and discuss are the following.

- Sampling variation with no clear trend
- Increasing trend with sampling variation
- Decreasing trend with sampling variation
- Increasing trend with milder sampling variation
- Decreasing trend with milder sampling variation

Figures 11 and 12 illustrate patterns of change over time dominated by sampling variation, i.e., they look ‘noisy’. In Figure 11 one sees ‘Percent Occupied Housing Units Renter Occupied - Schuylkill County, PA’. This characteristic is one that we would expect to be stable over time, perhaps indicating a gradual increase or decrease. Hence the bumps up and down over time can be attributed to sampling variation. Even the increase between 2004 and 2005, which is statistically significant, may be due to sampling variation. (Recall that a margin of error based on a 90% confidence interval implies that if there were no real difference, about 10% of the time one would still see a difference that looked statistically significant). Figure 12, ‘Percent of Population 5 Years and Over Who Speak Spanish at Home - Tulare County, CA’ shows a similar pattern of sampling variation. No trend is discernable in either set of data, though it is possible that weak trends are being masked by the sampling variation.

The best indicator of the strength of sampling variation is the width of the confidence bounds (all the examples use 90% confidence bounds): narrow confidence bounds indicate less sampling variation, and wider confidence bounds indicate more sampling variation. This relationship is illustrated by comparing Figures 13 and 14. In Figure 13 one sees a one-year jump from the 2002 to the 2003 estimates of percent of population 25 years and over with a bachelor's degree or higher in Black Hawk County, IA, though the estimates are accompanied by a fair amount of sampling variation. In Figure 14, percent of population 5 years and over who speak Spanish at home, Lake County, IL, one sees a clear, upward trend. The confidence bounds in Figure 14 are narrower than in Figure 13 because of the larger population of Lake County, IL. In Figure 14 the sampling variation does not obscure the trend because of the narrower confidence bounds and the magnitude of the change over time.

The patterns in Figures 13, 15 and 16 all indicate what seems to be increasing or decreasing trends masked with noise. In Figure 15, 'Percent Employed of Population 16 Years and Over - Hampden County, MA' one sees a downward trend with sampling variation and in Figure 16, 'Percent of Population 25 Years and Over with Bachelor's Degree or Higher - San Francisco, CA', one sees an upward trend with sampling variation. If the data user believes that the underlying trend should be gradual then they may prefer to examine the series of three-year or five-year MYEs, which are effectively smoothed. See the following section, Section 6.6, for more discussion on this point.

Figure 17, 'Percent Employed of Population 16 Years and Over - San Francisco, CA', displays a clear downward trend followed by a leveling off. This pattern corresponds to what is commonly known to have happened in the American economy when the 'dot-com' industry faltered, which had a strong adverse impact on the San Francisco economy. There is no apparent benefit from smoothing as the sampling variation is small and the trend is clear and strong.

In Figure 18, 'Civilian Veterans - Schuylkill County, PA', the graph is dominated by a sharp drop in civilian veterans from 2001 to 2002. This precipitous drop might reflect real events or it could also be that there is a more gradual downwards trend that looks sudden because of the effects of sampling variation. One could speculate that World War II veterans are moving out or dying. However, it would take a demographer with local knowledge of Schuylkill County to interpret this data definitively. This example serves as a reminder of the importance of subject matter knowledge when analyzing data.

Figure 11: Percent of Occupied Housing Units Renter Occupied - Schuylkill County, PA (There were 60,984 occupied housing units in Schuylkill County according to the 2005 ACS)

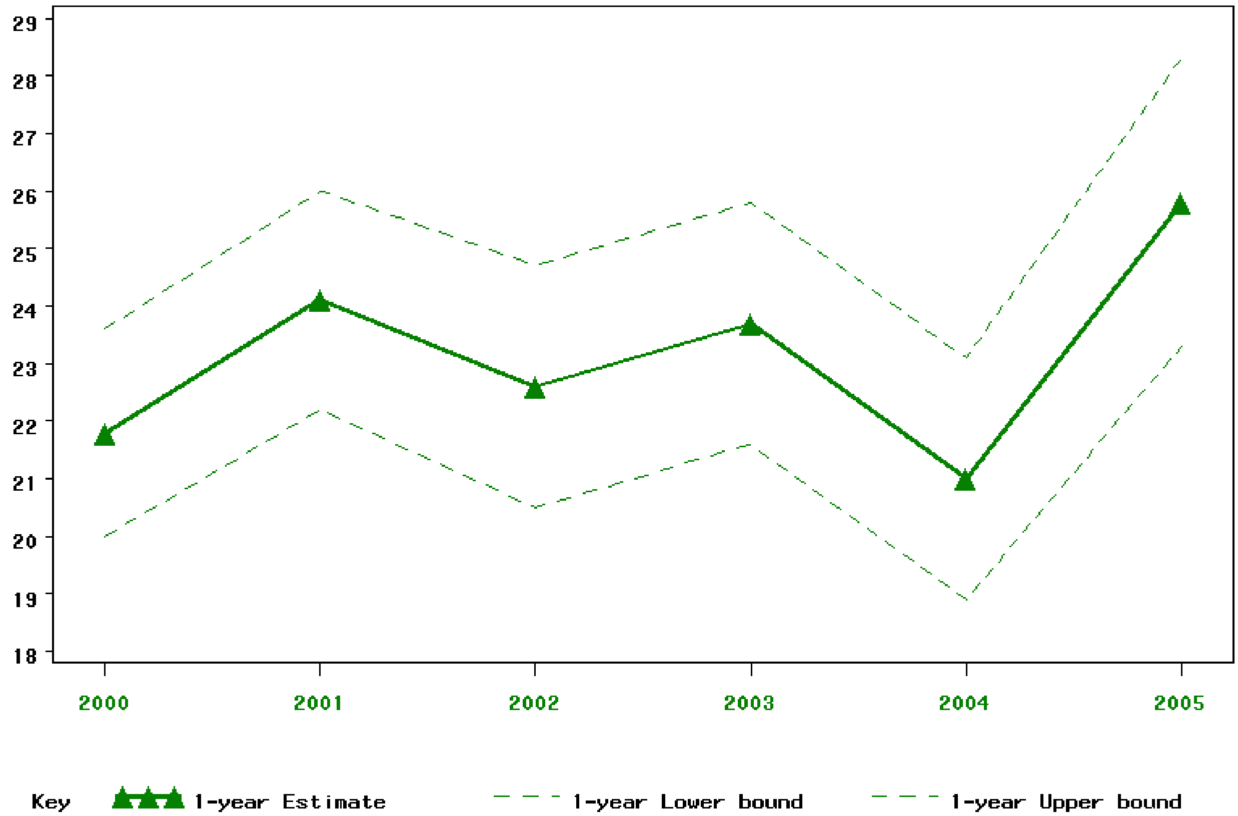


Figure 12: Percent of Population 5 Years and Over Who Speak Spanish at Home - Tulare County , CA
(There were 367,971 people 5 years and over in Tulare County according to the 2005 PEP)

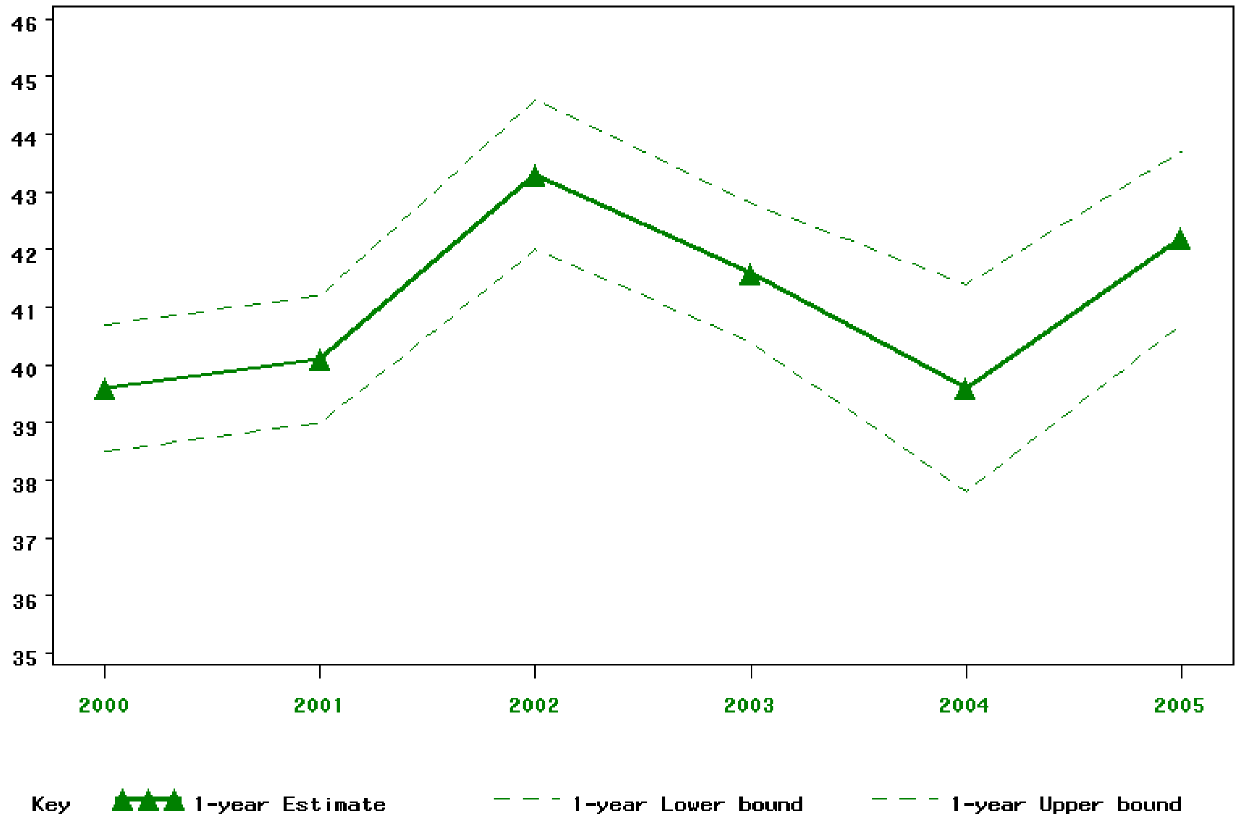


Figure 13: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - Black Hawk County, IA
(There were 78,994 people 25 years and over in Black Hawk County according to the 2005 PEP)

Figure 14: Percent of Population 5 Years and Over Who Speak Spanish at Home - Lake County, IL
(There were 644,299 people 5 years and over in Lake County according to the 2005 PEP)

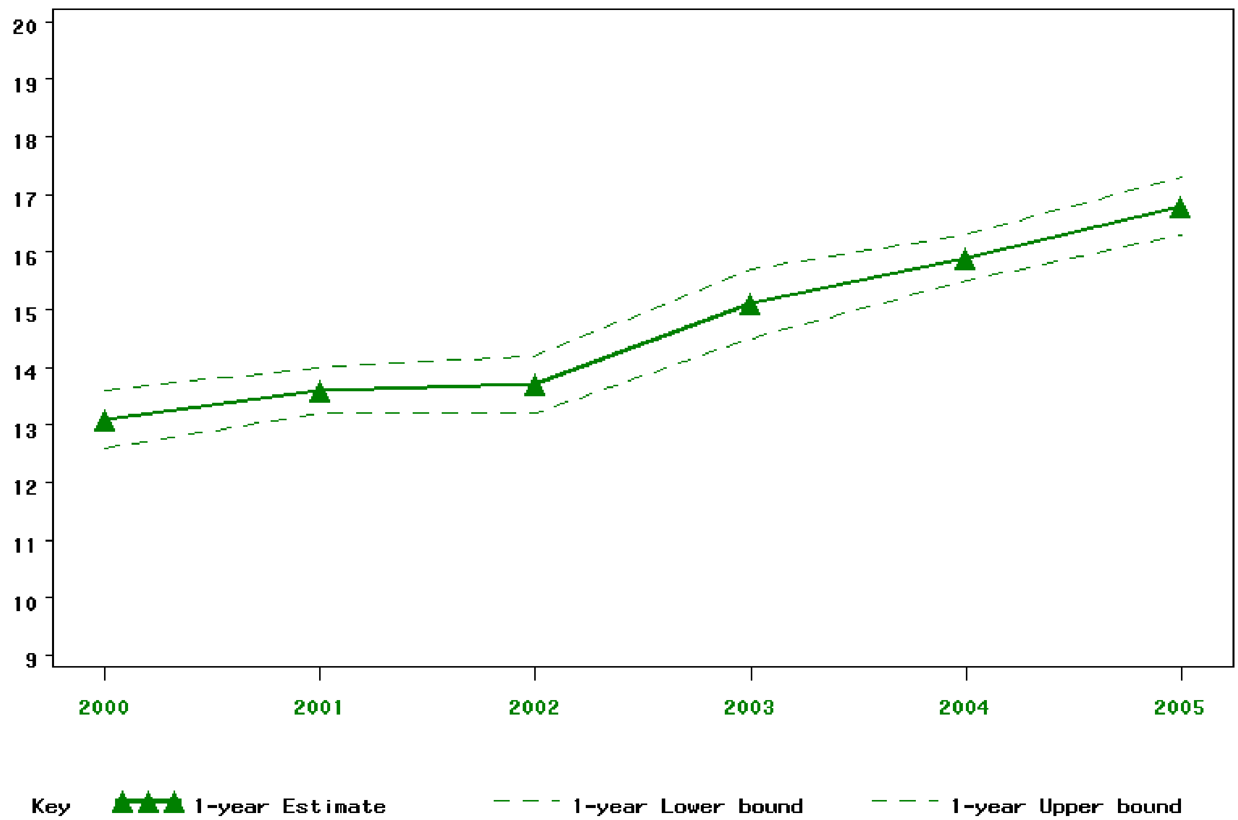


Figure 15: Percent Employed of Population 16 Years and Over - Hampden County, MA (There were 360,326 people 16 years and over in Hampden County according to the 2005 PEP)

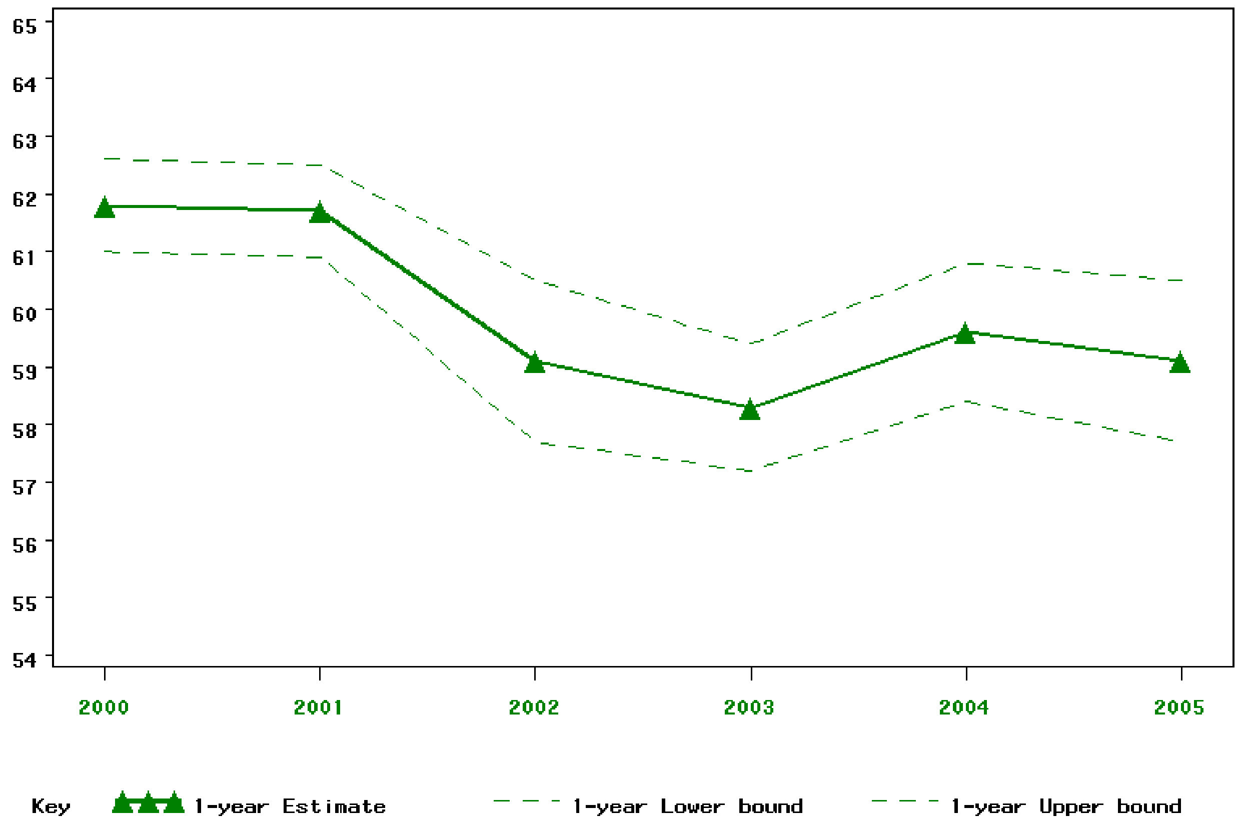


Figure 16: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - San Francisco, CA
(There were 588,864 people 25 years and over in San Francisco according to the 2005 PEP)

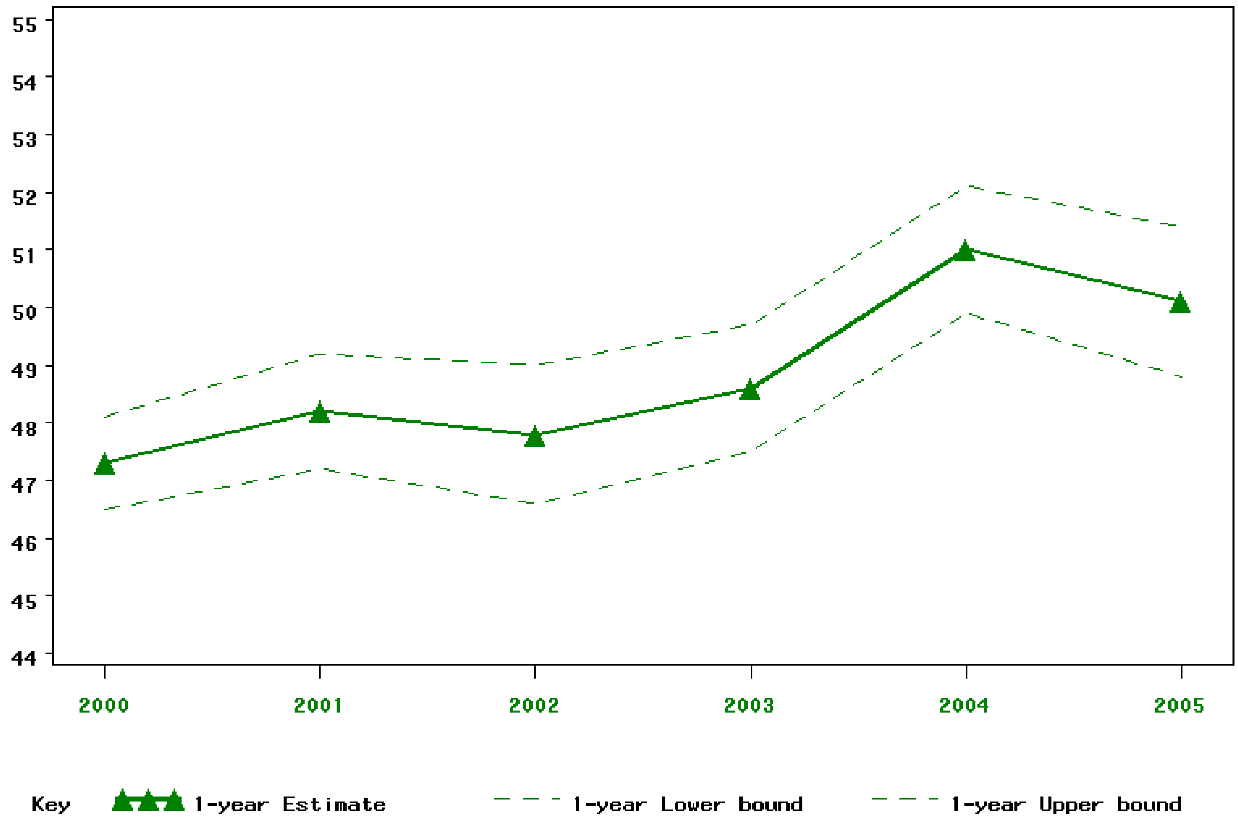


Figure 17: Percent Employed of Population 16 Years and Over - San Francisco, CA (There were 656,769 people 16 years and over in San Francisco according to the 2005 PEP)

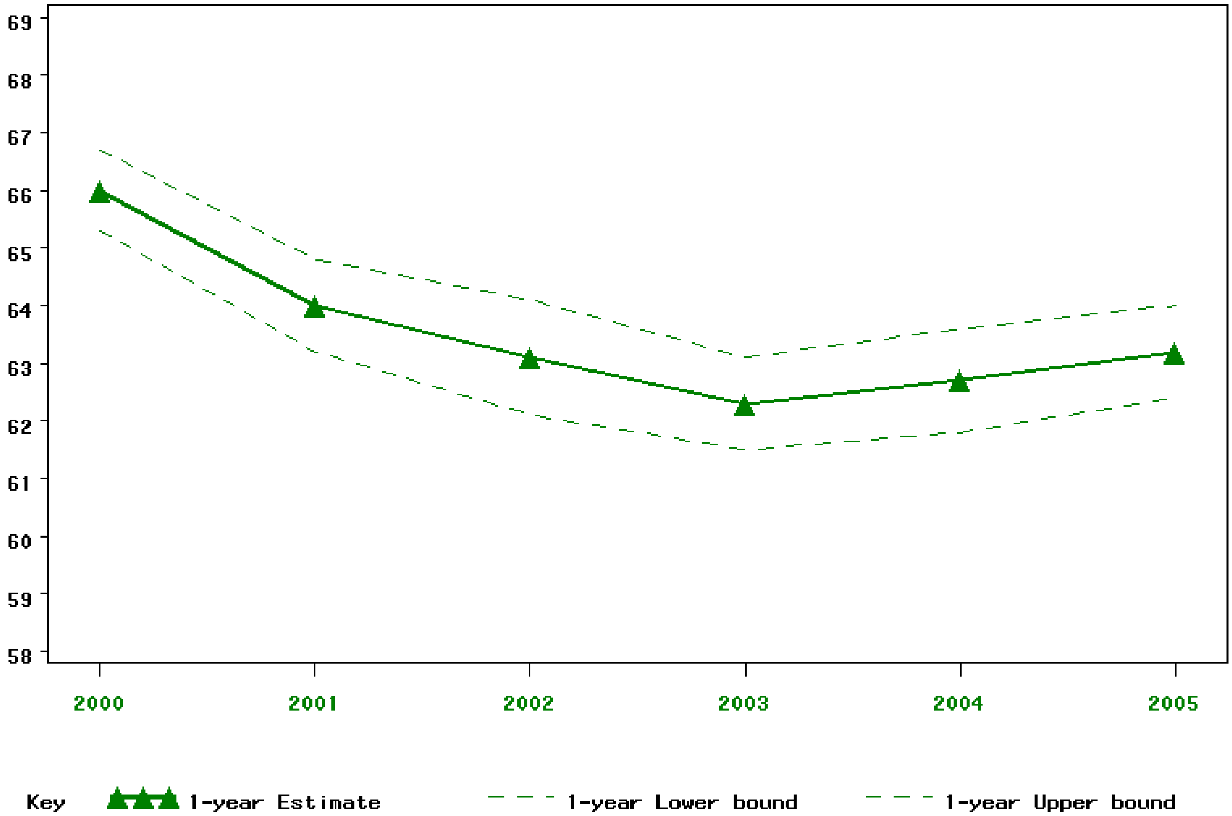
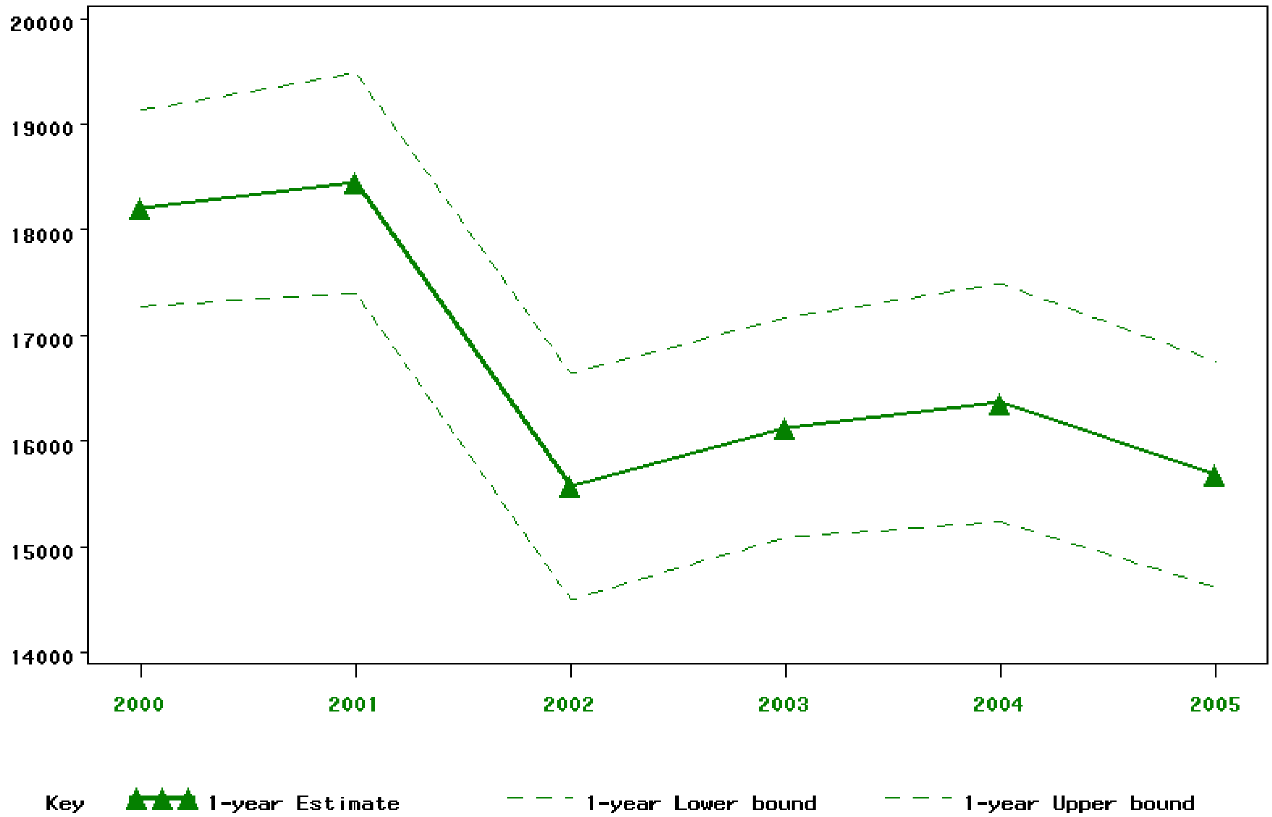


Figure 18: Civilian Veterans[11] - Schuylkill County, PA
(The population of Schuylkill County was 146,209 according to the 2005 PEP)



[11]Note that Figure 18 is not drawn to the same scale as most of the figures in this document as it displays counts and not percentages. Caution should be exercised when comparing graphs which are not drawn in the same scale.

6.5 Patterns of Change Over Time Part II: Smoothing Change over Time with a Series of Multiyear Estimates

The main point of this section is to illustrate how a series of MYEs can have the effect of smoothing a series of one-year estimates over time. The MYEs smooth estimates for two reasons.

1. Most importantly, as a result of having more years of data, MYEs have smaller sampling variation, that is, smaller standard errors and narrower confidence bounds. This point is true for both overlapping and nonoverlapping series of MYEs.
2. Also, overlapping MYEs share data in common, thus the MYEs change less from year to year than one-year estimates. This point of course doesn't apply to nonoverlapping MYEs.

We present the three-year and five-year estimates as well as the one-year estimates for four of the examples shown in Section 6.5. Notice that we display these comparisons in two ways, as it will help data users to better think about the properties of a series of MYEs over time.

1. Compare by the last year of the estimation period. This display is appropriate if one thinks in terms of the choices that a data user will have in any given year.
2. Compare one-, three-, and five-year estimates by the middle year of the estimation period. This display is appropriate if one is interested in a time series and thinks the trends are roughly linear.

Observations on the patterns that a series of MYEs follow.

- In Figures 19 and 20, 'Percent of Housing Units Renter Occupied - Schuylkill County, PA', the three-year estimate shows a relatively flat pattern. This is not surprising, as the bouncing seen in the series of one-year estimates is likely sampling variation and the underlying trend is likely flat or close to flat.
- In Figures 21 and 22, 'Percent of Population 25 Years and Over with a Bachelor's Degree or Higher - Black Hawk County, IA', the upward trend is clearly smoothed. The smoothed picture might be a better portrayal of what is happening, but it would require local knowledge to make that determination. For example, if an analyst with local knowledge had reason to expect a gradual change in this characteristic, then he or she would prefer the smoothed trend as a better portrayal of what is happening over time.
- In Figures 23 and 24, 'Percent Employed of Population 16 Years and Over - Hampden County, MA', one sees a smoothed downward trend in the three-year estimates. As in the previous example, the smoothed picture might be a better portrayal, though local knowledge is always important when interpreting trend data.
- In Figures 25 and 26, 'Percent of Population 25 Years and Over with a Bachelor's Degree or Higher - San Francisco, CA', one sees a smoothing of the upward trend. The comments from the previous two examples apply here.

- In all of these examples in this section the three-year data are sufficient to show a substantial smoothing. For data with more sampling variation five-year estimates may be required for more adequate smoothing.
- The multiyear estimates of Civilian Veterans in Schuylkill County, PA (see Figure 6) serve as a counterexample to Figures 21-26. Here the smoothed series of MYEs may be masking a sharp drop from 2001 to 2002.

Figure 19: Percent of Occupied Housing Units Renter Occupied - Comparisons Based on Last Year of Estimation Period - Schuylkill County, PA
(There were 60,984 occupied housing units in Schuylkill County according to the 2005 ACS)

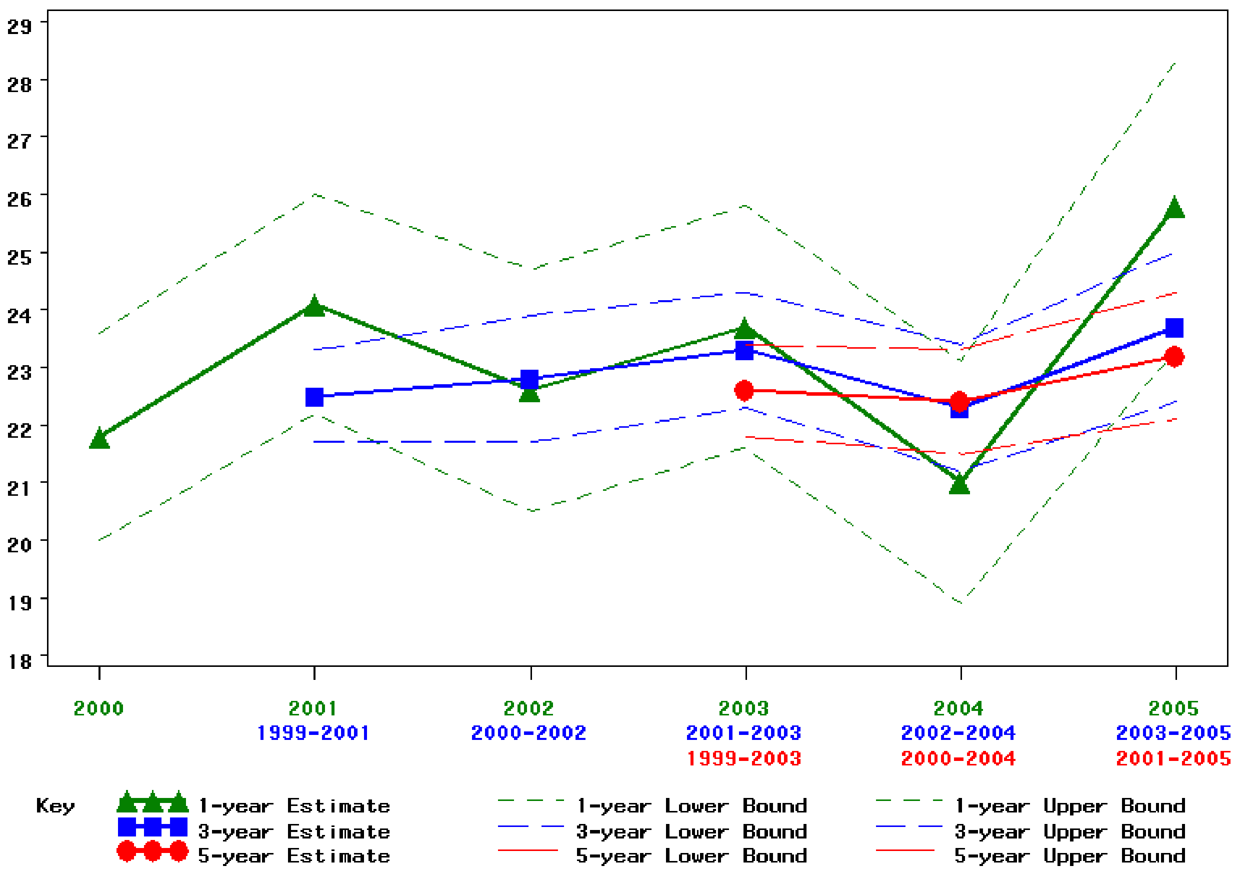


Figure 20: Percent of Housing Units Renter Occupied - Comparisons Based on Middle Year of Estimation Period - Schuylkill County, PA
(There were 60,984 occupied housing units in Schuylkill County according to the 2005 ACS)

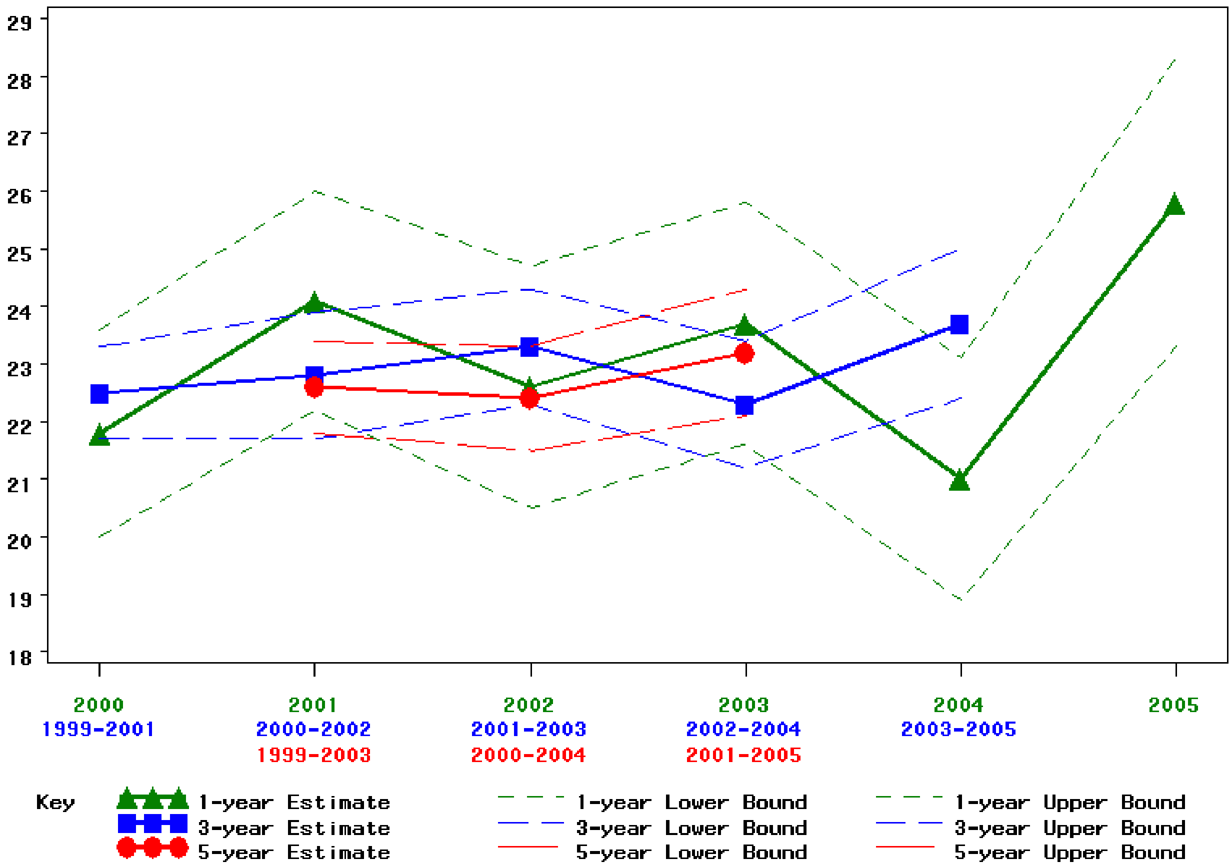


Figure 21: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - Comparisons Based on Last Year of Estimation Period - Black Hawk County, IA (There were 78,994 people 25 years and over in Black Hawk County according to the 2005 PEP)

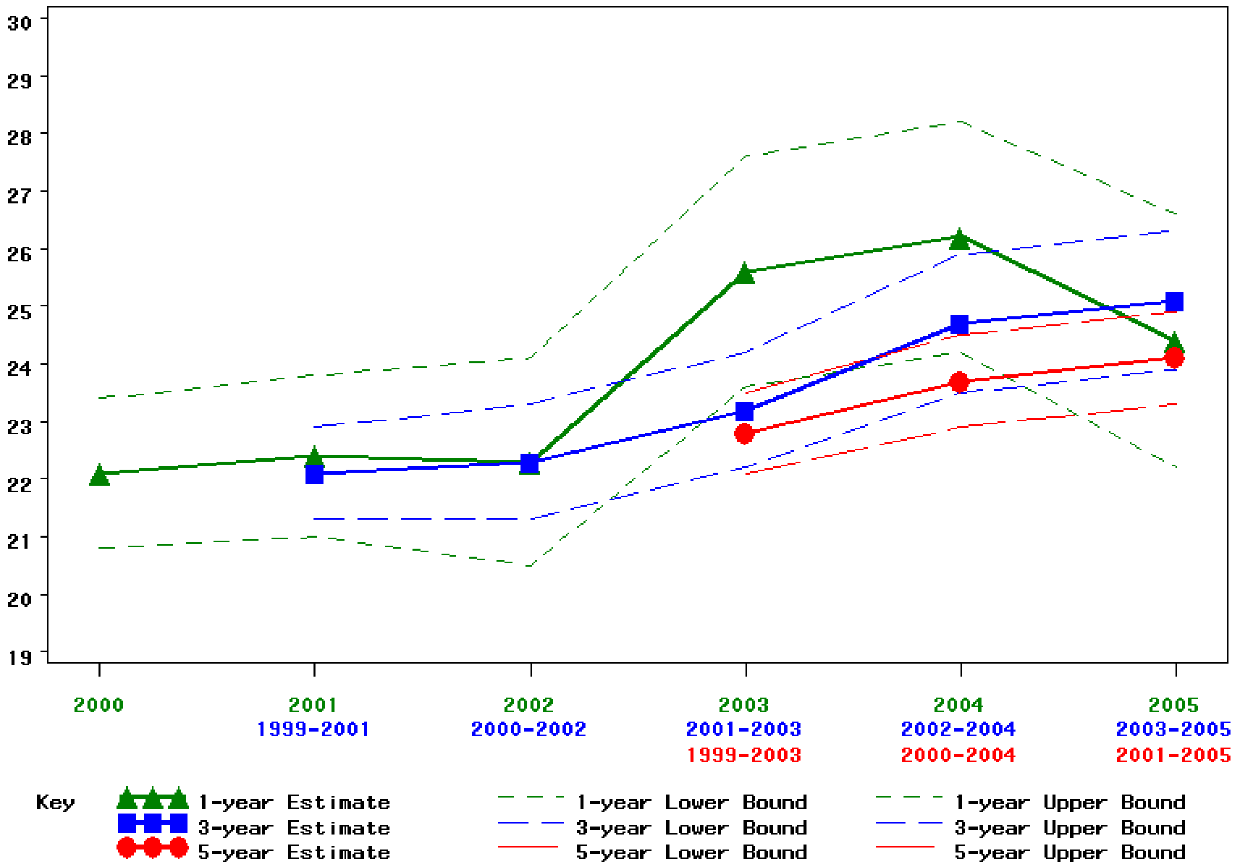


Figure 22: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - Comparisons Based on Middle Year of Estimation Period - Black Hawk County, IA (There were 78,994 people 25 years and over in Black Hawk County according to the 2005 PEP)

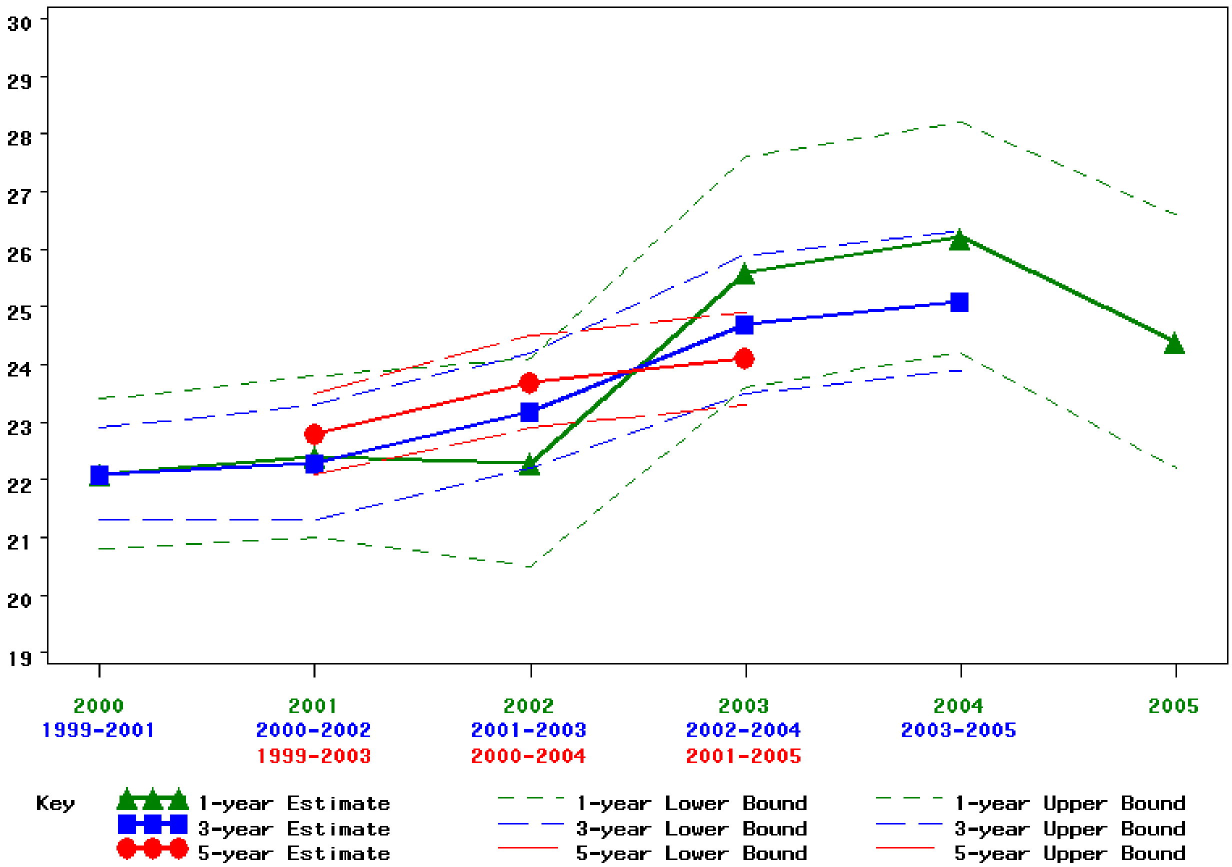


Figure 23: Percent Employed of Population 16 Years and Over - Comparisons Based on Last Year of Estimation Period - Hampden County, MA
(There were 360,326 people 16 years and over in Hampden County according to the 2005 PEP)

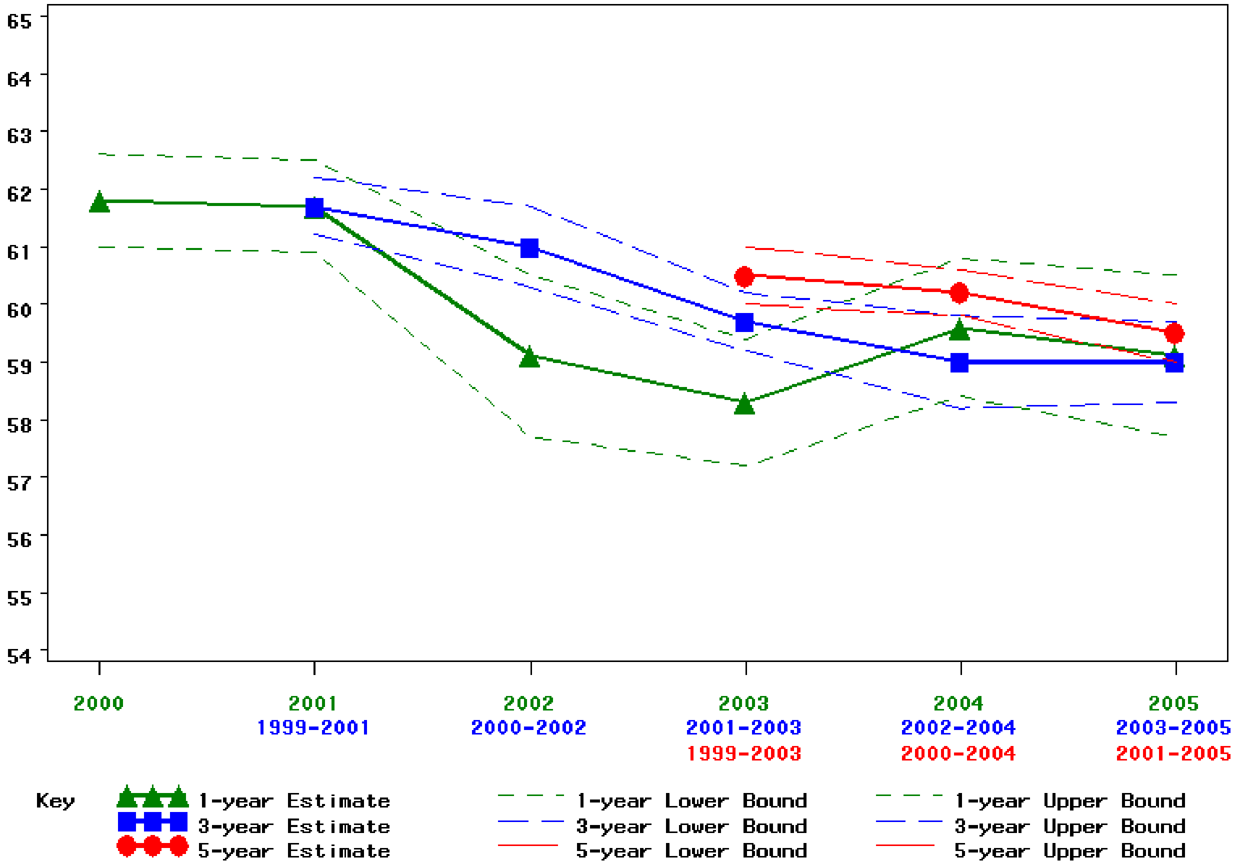


Figure 24: Percent Employed of Population 16 Years and Over - Comparisons Based on Middle Year of Estimation Period - Hampden County, MA
(There were 360,326 people 16 years and over in Hampden County according to the 2005 PEP)

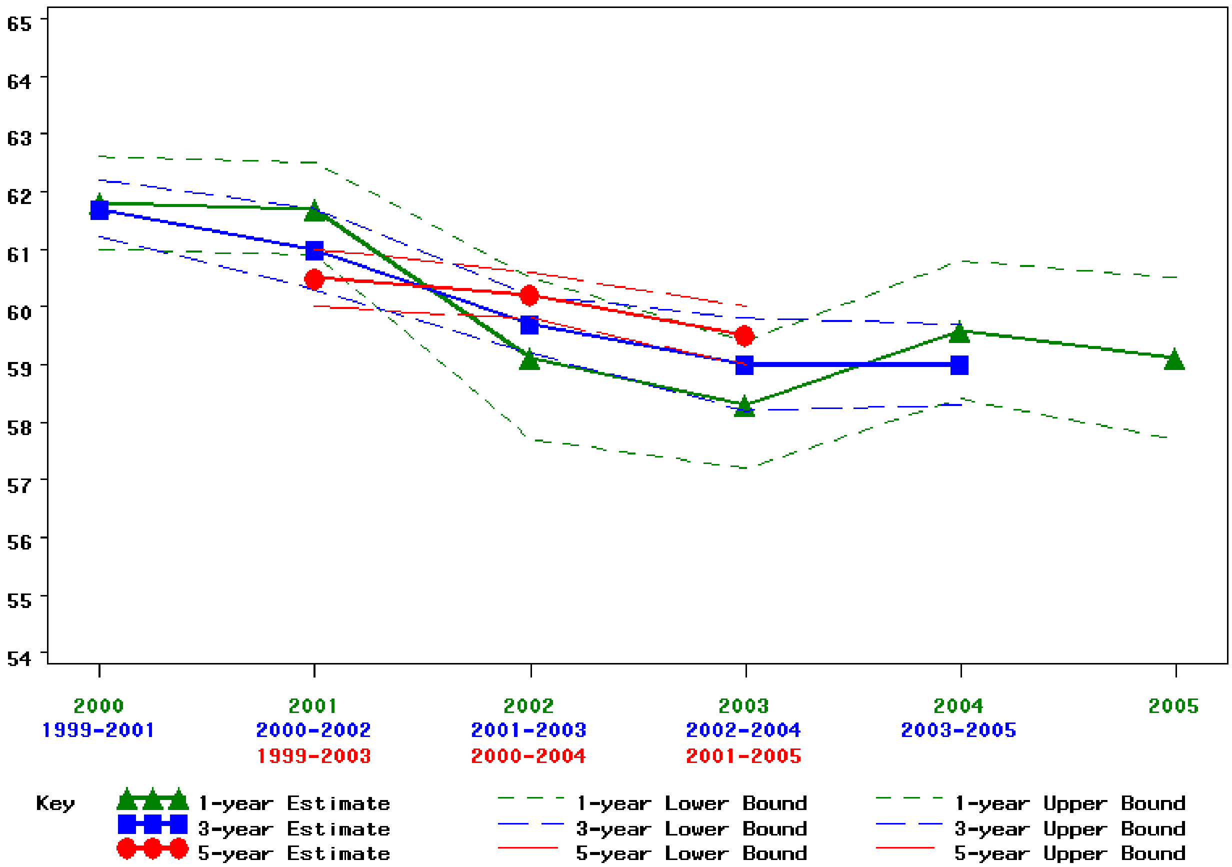


Figure 25: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - Comparison Based on Last Year of Estimation Period - San Francisco, CA (There were 588,864 people 25 years and over in San Francisco according to the 2005 PEP)

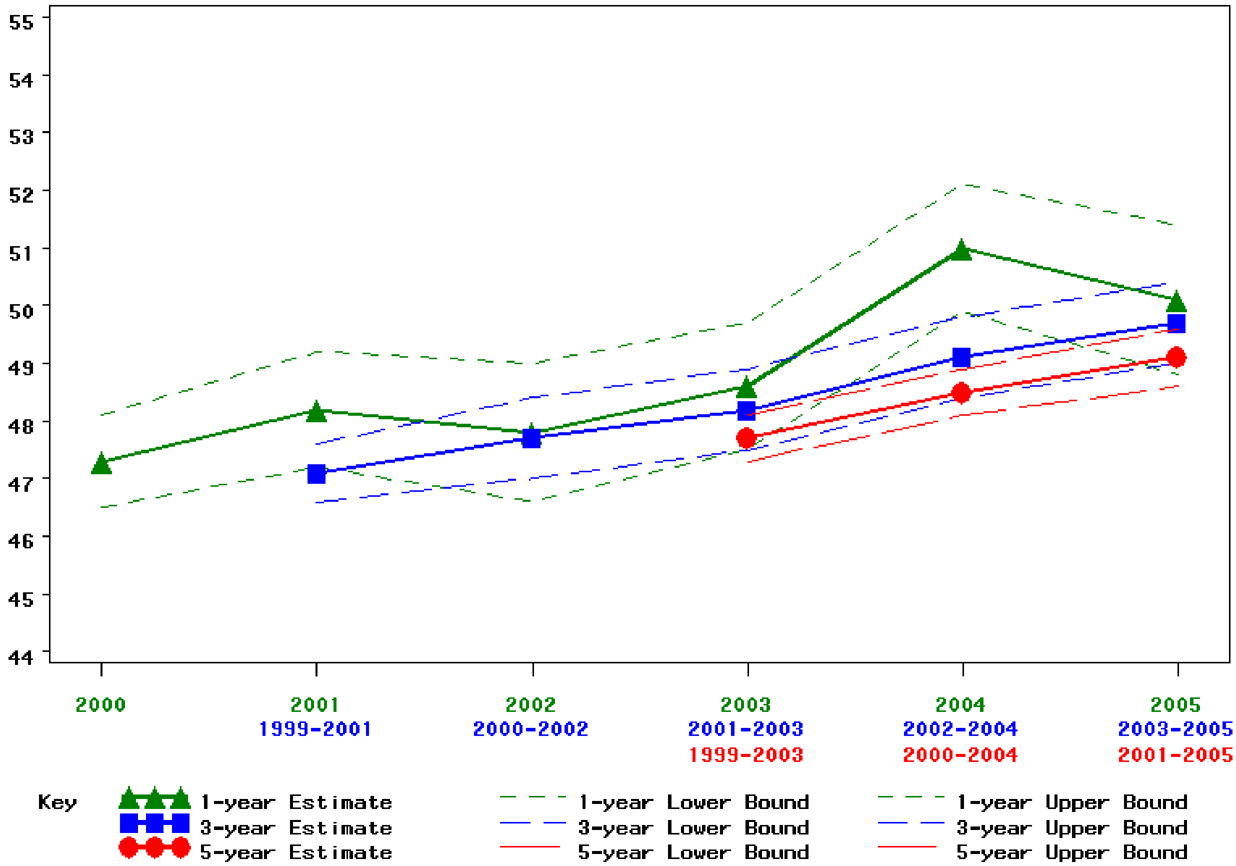
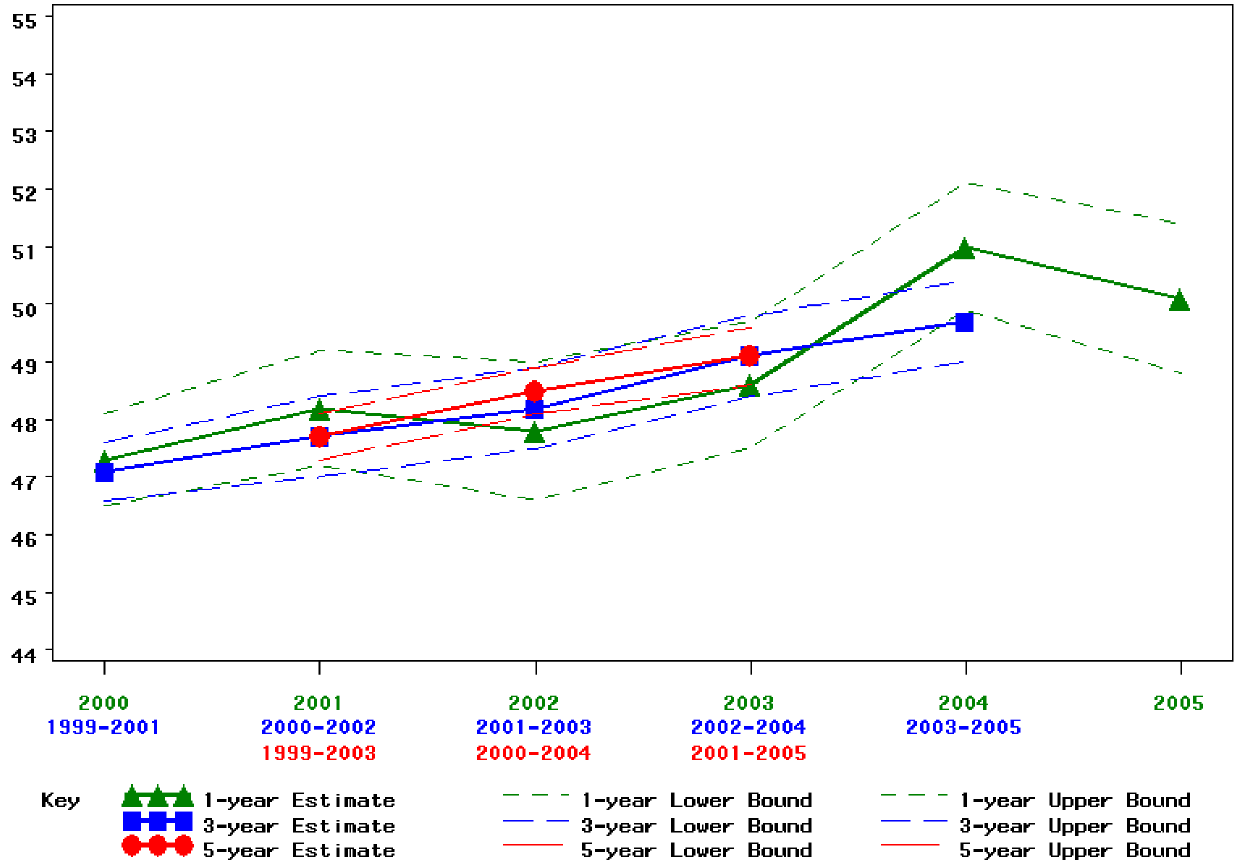


Figure 26: Percent of Population 25 Years and Over with a Bachelors Degree or Higher - Comparisons Based on Middle Year of Estimation Period - San Francisco, CA (There were 588,864 people 25 years and over in San Francisco according to the 2005 PEP)



7. SUMMARY: WHEN AND WHEN NOT TO USE MULTIYEAR ESTIMATES

At this point we present a summary of the previous sections on when to use multiyear estimates and when to use one-year estimates.

When to Use Multiyear Estimates

- For tracts and other smaller geographies with a population less than 65,000 because one-year estimates are not available.
- To obtain estimates with lower standard errors.
- For smaller subpopulations.
- For more precise comparisons of change over time with non-overlapping comparisons.
- For smoothing data over time.

When to Use One-Year Estimates

- For larger geographies and populations.
- For variables with small SEs because they are highly correlated with variables that are controlled.

8. CURRENCY VERSUS PRECISION - A FUNDING ALLOCATION EXAMPLE

This section presents an example of how ACS estimates could be used to distribute funding among counties. Its purpose is to demonstrate how different decisions to use one-, three-, or five-year estimates can affect results in the context of characteristics that are changing over time.

8.1 A State for Analysis

The example combines the 34 counties in the MYE Study to create a 'New State' of 9,813,462 residents according to the ACS 2000-2004 MYE, which was calculated from the average of the Official Population Estimates from 2000 through 2004 (U.S. Census Bureau, 2006a). Let's suppose that New State would like to strengthen English as a Second Language (ESL) programs. Table 12 shows the breakdown of these 34 counties into three groups defined by the shortest period estimates that are released, which is determined by population size. Table 12 also shows the number and proportion of limited English speakers as determined by the 2000-2004 ACS estimates, where 'limited English speakers' are defined as persons 5 years and over who speak a language other than English at home and speak English 'less than very well'. In Table 12 note the breakdown of the 34 counties; there are

- 19 counties with a population of 65,000 or more and thus having one-, three-, and five-year ACS estimates, with 96.6% of the limited English speakers in 2000-2004;
- 11 counties with a population of 20,000 to 64,999 and thus having three- and five-year ACS estimates, with 3.0% of the limited English speakers in 2000-2004; and
- 4 counties with a population less than 20,000 population and thus having five-year ACS estimates only, with 0.4% of the limited English speakers.

Table 12: Limited English Speakers by Size of County (2000-2004 ACS) - New State

Shortest Period/County Size	Number of Counties	Number of Limited English Speakers	Proportion of Limited English Speakers
Five-Year/ Under 20,000	4	4,213	0.4%
Three-Year/ 20,000-64,999	11	34,952	3.0%
One-Year/ 65,000 or more	19	1,197,032	96.6%
Total	34	1,236,197	100.0%

8.2 Non-English Speakers in New State

In the simulated example New State passes legislation providing $10,000,000 in funding to help limited English speakers by subsidizing English as a Second Language (ESL) classes. The legislation stipulates New State should allocate the $10,000,000 by county proportional to the number of limited English speakers, as determined by ACS data. Table 13 shows the number of limited English speakers in those New State counties for which the ACS produces one-year estimates. It is these data, which show an increase in limited English speakers, that motivated the state to pass the legislation subsidizing ESL classes.

Table 13: Number of Limited English Speakers in the 19 New State Counties with One-Year ACS Estimates

	Limited English Speakers
2000	1,027,959
2001	1,059,002
2002	1,111,733
2003	1,153,066
2004	1,197,032

8.3 Question and Challenge

The question facing New State is, “Which ACS estimates should be used to allocate funds among counties: one-year, three-year, or five-year estimates?” New State would like to use the most current data, but the ACS doesn’t provide one-year data for all counties. New State sees two obvious approaches:

- use the most current estimate available for each county;
- use five-year data for each county.

A drawback of the first approach is that the allocations to those counties not among the largest 19 are lower due to the mix of older data and the resulting lag effect (assuming the number of persons of limited English proficiency is increasing in the same way as in the largest 19). One sees this in Table 14 – if we used the most current estimates for each county, then the allocation for the smallest counties would be $34,080, instead of $36,883 when only using five-year estimates, a difference of almost 8%.

Table 14: Distribution Between Three Groups of Counties

Shortest Period/County Size	Most Current Estimates Available for Each County	Five-Year Estimates (2001-2005)	Percent Difference
Five-Year/ Under 20,000	$34,080	$36,883	7.60
Three-Year/ 20,000-64,999	$282,738	$297,543	4.98
One-Year/ 65,000+	$9,683,182	$9,665,574	-0.18

The drawback of the second approach is that it uses older data to allocate funds among the larger counties, which have the bulk of the limited English speakers. We will see this point clearly later in Tables 15 and 16.

8.4 A Hybrid Approach to Allocating the Funding

There are several alternative approaches that may be more satisfactory than the obvious ones outlined in Section 8.3. For example, consider the following hybrid method of allocation.

- Form three groups of counties based on the most recent data available, i.e., one-year counties, three-year counties and five-year counties.
- Divide up the funds among these three groups according to five-year data.
- Within each group allocate funds based on the most recent shortest period estimates available.

This approach would be fair to smaller counties with only five-year data, avoiding the losses pointed out in Table 14. However, it also uses the most recent data available for larger counties to distribute funds among themselves. To see the difference consider Tables 15 and 16. In Table 15 one sees that using the most recent one-year data would allocate about 7 percent more to Lake County, IL, and 48 percent more to Sevier County than using the five-year data. What one sees in Table 16 is analogous to Table 15, except here one sees the effects of using one-year data versus three-year data on the middle-sized counties. They are less dramatic in most counties than those seen in the larger counties.

Table 15: Allocation of Funds among 19 Large Counties Based on Two Schemes

County	By Most Recent Single-Year (Hybrid)	By Five-Year MYE	Difference	Percent Difference
Pima, AZ	$668,643	$641,704	$26,939	4.03%
Jefferson, AR	$1,946	$4,351	($2,405)	-123.59%
San Francisco, CA	$1,413,122	$1,527,932	($114,810)	-8.12%
Tulare, CA	$610,724	$633,159	($22,436)	-3.67%
Broward, FL	$1,763,165	$1,706,211	$56,954	3.23%
Lake, IL	$594,760	$554,202	$40,559	6.82%
Black Hawk, IA	$30,038	$31,158	($1,120)	-3.73%
Calvert, MD	$5,765	$6,470	($704)	-12.22%
Hampden, MA	$311,188	$343,207	($32,019)	-10.29%
Madison, MS	$9,455	$8,107	$1,349	14.26%
Flathead, MO	$6,670	$4,220	$2,450	36.73%
Douglas, NE	$198,603	$189,616	$8,987	4.53%
Bronx, NY	$2,607,810	$2,532,690	$75,121	2.88%
Rockland, NY	$300,239	$297,674	$2,564	0.85%
Franklin, OH	$390,957	$376,947	$14,010	3.58%
Multnomah, OR	$460,479	$497,979	($37,500)	-8.14%
Schuylkill, PA	$9,205	$12,685	($3,480)	-37.81%
Sevier, TN	$14,236	$7,459	$6,777	47.60%
Yakima, WA	$268,570	$289,804	($21,234)	-7.91%
Total	$9,665,574	$9,665,574	$0	

Table 16: Allocation of Funds among 11 Medium-Sized Counties Based on Two Schemes

County	By Most Recent Three-Year MYE (Hybrid)	By Five-Year MYE	Difference	Percent Difference
Upson. GA	$2,060	$1,830	$230	11.18%
Miami, OH	$1,822	$1,987	($166)	-9.09%
De Soto Parish, LA	$1,251	$1,156	$96	7.65%
Washington, MO	$1,090	$306	$783	71.88%
Lake, MT	$2,120	$3,012	($892)	-42.08%
Otero, NM	$48,064	$46,723	$1,340	2.79%
Starr, TX	$236,199	$235,945	$253	0.11%
Petersburg City, VA	$868	$1,418	($550)	-63.33%
Ohio, WA	$2,613	$3,038	($424)	-16.24%
Oneida, WI	$758	$875	($118)	-15.55%
Vilas, WI	$698	$1,252	($554)	-79.34%
Total	$297,543	$297,543	$0	

There are other reasonable approaches that would be worth considering. For example, instead of allocating among the three groups by five-year data, one could aggregate by one-year data. To do this one would have to request custom tabulations from the Census Bureau as one-year data is not published for counties under 65,000.

9. CONCLUSION

This paper is intended to accompany the release of the first ACS MYE data products in 2008. It highlights key statistical aspects of the multiyear estimates that we feel users should grasp, focusing on what is novel to users of decennial sample data and ACS one-year data, namely, choosing between one-, three-, and five-year estimates, and interpreting estimates over time. It benefits from being illustrated with examples using real ACS data from the MYE Study. The success of the document's approach will become apparent to the U.S. Census Bureau by the reaction of data users to it, and in particular by the questions they ask about it or they feel are unanswered by it. Further, we expect that as users become more familiar with the ACS data their practices and needs will grow, perhaps in ways that the authors cannot anticipate. We expect more statistical developments will suggest themselves as the needs of data users approach the limits of current practices and methodologies. When their needs are clearer to us we expect to update the paper. It is intended to be a living document that will be relevant to both current and future data users, and to be a starting point for thinking and discussion about the use of the MYEs.

10. REFERENCES

Asiala, M., and Tersine, A. (2007). "Methodology for the Production of American Community Survey Multiyear Estimates". *Proceedings of the Survey Research Methods Section.* American Statistical Association.

Committee on National Statistics (2007). "Chapter 6: Weighting and Interpreting ACS Multiyear Estimates." *Using the American Community Survey: Benefits and Challenges.* National Academies Press.

Taeuber, C. M. (2006). "Chapter 9: Multiyear estimates." *American Community Survey for Community Planning.* Trafford Publishing.

Torrieri, N. (2007), "America is Changing, and So is the Census: The American Community Survey," *The American Statistician, 61, 16–21.*

Transportation Research Board (2008). "A Guidebook for Using American Community Survey Data for Transportation Planning." *National Cooperative Highway Research Program Report #588.* http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_588.pdf

U.S. Census Bureau (2002). "Census 2000 Basics." http://www.census.gov/mso/www/c2000basics/00Basics.pdf

U.S. Census Bureau (2003). "American Community Survey Operations Plan" http://www.census.gov/acs/www/Downloads/OpsPlanfinal.pdf

U.S. Census Bureau (2004). "Meeting 21st Century Demographic Data Needs - Implementing the American Community Survey: Report #8: Comparison of the American Community Survey Three-Year Averages and the Census Sample for a Sample of County and Tracts." http://www.census.gov/acs/www/AdvMeth/acs_census/creports/Report08.pdf

U.S. Census Bureau (2006a). "Design and Methodology: American Community Survey." *Technical Paper 67.* Issued May 2006; http://www.census.gov/acs/www/Downloads/tp67.pdf

U.S. Census Bureau (2006b). "2006 Data Users Handbook The American Community Survey" http://www.census.gov/acs/www/Downloads/Handbook2006.pdf

U.S. Census Bureau (2007a). Census Bureau Website on Population Estimates. http://www.census.gov/popest/estimates.php

U.S. Census Bureau (2007b). Multiyear Estimates Study, "Accuracy of the Data". http://www.census.gov/acs/www/Downloads/Accuracy_of_the_Data_MYE.pdf

U.S. Census Bureau (2007c). "About the Data (Methodology): The Multiyear Estimates Study: Overview". http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/overview.html

U.S. Census Bureau (2007d). "About the Data (Methodology): The Multiyear Estimates Study: About the Multiyear Estimates Study". http://www.census.gov/acs/www/AdvMeth/Multi_Year_Estimates/study_details.html

U.S. Census Bureau (2008a). "A Compass for Understanding and Using American Community Survey Data." http://www.census.gov/acs/www/UseData/Compass/compass_series.html

U.S. Census Bureau (2008b). "2008 Boundary and Annexation Survey". http://www.census.gov/geo/www/bas/bashome.html