A DICUSSION OF THE QUALITY OF ESTIMATES FROM THE AMERICAN COMMUNITY SURVEY FOR SMALL POPULATION GROUPS

Charles H. Alexander
U.S. Census Bureau
July 17, 2002

This version is a draft designed for discussion and comment from Census Bureau Advisory Committees.

I. INTRODUCTION

This report discusses the quality and usefulness of estimates from the planned American Community Survey (ACS), for very small populations groups. The ACS is intended to replace the long form survey in the 2010 census. Since the long form is unique as a source of information about small population groups, the top objective of the ACS design has been to provide good information about the smallest groups.

The general premise of the ACS design is that by spreading the “long form” sample across the decade, it is possible to provide updated information for all sizes of population groups. In principle, this should be especially advantageous for small population groups, because there is currently very little information about how these populations change over time. Also, the ACS is expected to have more consistent quality because of the advantages of a continuous operation, which is especially important for small, “hard-to-enumerate groups.

However, concerns have been expressed about the quality of ACS estimates for very small population groups. These concerns are described in Section III, with responses in subsequent sections. We have described the ACS as replacing the long form “snapshot” with a “video”. Using this metaphor, the most widespread concerns are 1) that a “freeze frame” from the video is not as clear as a snapshot, and 2) that if the subject of the picture is small and fast-moving, the video may show a blur.

The response is, continuing the metaphor, that the freeze frame is almost as clear as the snapshot, and provides the advantage of being able to look at a freeze frame at any time. For fast-moving subjects, a video at least tells you that the subject is moving and in what direction, while a snapshot misses the action totally. Small population groups have the potential to change more dramatically than larger groups, so having a “video” is most valuable for small groups.

Sections I through V, along with Appendix 1, present the basic issues. Section VI and Appendix 2 provide some additional details. Appendix 3 points out some more complex statistical issues, which are not mentioned in the body of the report to keep the discussion simpler.

Comments on topics where the report may not have effectively explained the issues are welcome. There may be subsequent revisions of the report based on discussions with the Census Bureau Advisory Committees, or as new information is available from evaluations of the ACS and comparisons to Census 2000.

II. BACKGROUND ON THE OPERATIONS AND DESIGN OF THE AMERICAN COMMUNITY SURVEY

The ACS is part of a plan to re-engineer the 2010 census. Besides replacing the long form with the ACS, the plan includes modernizing the MAF/TIGER geographic systems, and early planning and research to design better and more accurate ways to count the population in 2010.

The ACS plan is to start in 2003, with an annual sample of 3 million addresses spread across the list of addresses in each census tract, covering all places (such as cities or towns), American Indian Reservations, Alaska Native villages, and Hawaiian Homelands. About 250,000 addresses will be contacted for the first time each month. No address will be in sample more than once in a 5-year period; most addresses can expect to go about forty years between ACS interviews.

Most addresses in the sample start out with a mail questionnaire in their first month, with a prenotice, a reminder card, and a targeted second mailing. In the second month, nonresponding addresses for which a telephone number is available are followed-up by a Computer-Assisted Telephone Interviewing (CATI) operation. In the third month a one-in-three sample of addresses which have still not responded are selected for follow-up by Field Representatives who use Computer-Assisted Personal Interviewing (CAPI). Mail responses with substantial amounts of missing data are designated for recontact by telephone in a "failed edit follow-up" operation. Units for which there is no usable mailing address skip the mail and CATI phases. A two-in-three sample of these units go straight to CAPI.

The combined effect is that the ACS will sample about 30 million addresses over a 10-year period, resulting in about 21.9 million interviewed housing units. This compares to about 19.2 million addresses in the Census 2000 long form sample.

Small governmental units will be sampled at a higher rate, depending on the population of the area, as was done for the last three census long form samples. In particular, the smallest governmental units will be sampled at a rate of 10 percent per year. Addresses in large census tracts sampled at a somewhat lower rate, except if they are, in a small governmental units.

The survey designers are considering a plan to oversample census tracts which have much lower-than-average mail response rates, by having a greater than one-in- three CAPI followup rate in those areas. To make up for this, the initial sampling rate would be reduced slightly in tracts with above-average mail response rates. Options for doing this are being discussed with stakeholders, for possible implementation during 2003.

A Puerto Rico Community Survey with similar design and sampling rates, is planned starting in 2003, pending congressional funding.

A crucial part of the ACS message is that the ACS provides the *characteristics* of the population, not *counts*. The census will continue to provide a complete count of the population every ten years. In the intercensal years, the official number of people will continue to come from the intercensal demographic estimates program, as part of the Federal/State Cooperative Population Estimates (FSCPE) program. However, information from the ACS will be used to improve these population estimates; research has begun on these improvements and how to implement them as a Program of Integrated Estimates."

To replace the long form estimates, the ACS will produce annually updated 5-year average estimates for geographic areas down to the block group level. For example, in 2008 there will be a set of tables, and public-use microdata files, covering the period 2003-2007. In 2009, the updated estimates will cover 2004-2008, and so forth. Each 5-year average may be thought of as replacing a hypothetical census long form in the middle year; for example, the 2003-2007 average would correspond to a "2005 long form estimate." The 2008-2012 average is the one most closely corresponding to the 2010 time period. These 5-year averages are the most important ACS data product for small population groups.

The ACS will also produce 3-year averages and 1-year average estimates. For large areas and population groups, meaning over 20,000 population for 3-year averages and 65,000 population for 1-year averages, these averages will be regularly available for the full range of tables.

For smaller areas and population groups, 3-year and 1-year averages will be available in a different format, possible a SAS file, "for research purposes". These research purposes include statistical analyses such as time series modeling or multiple regression analyses which pool information from a number of areas. These data will also be useful for "multi-year average interpretation", in other words for studying in more detail the changes that take place within a 5-year average. Note that below the 20,000 and 65,000 thresholds, the 3-year and 1-year averages for individual areas have high standard errors and are only useful for detecting large changes.

III. A SPECIAL CASE: "COUNTS" OF RACE/ETHNIC GROUPS INCLUDED ON THE SHORT FORM

For the race and ethnic groups included on the census short form, if the interest is in knowing the number of people in the group, then the ACS offers clear advantages. This includes not only the broad groups such as "Asian", "Hispanic", or "American Indian or Alaska Native", but detailed subgroups such as "Korean", Puerto Rican", or specific American Indian tribes.

The reason is that the ACS plan still includes a 100 percent enumeration every 10 years with a short form, so the counts of these groups will be collected in the census as always. In addition, the ACS will provide sample-based estimates of changes in the size of these population groups during the decade. This information will be incorporated into the intercensal population estimates program to improve the accuracy of the intercensal estimates.

The intercensal estimates program currently provides estimates only for the broad race groups or all Hispanic, not the detailed subgroups or tribes. Even for the broader groups, the intercensal estimates historically have not done well in reflecting changes in migration patterns below the state level. Information from the ACS, along with other improvements in the methodology for intercensal estimates, will improve the quality of sub-state estimates of the broad race/origin groups. In addition, the ACS 5-year averages will provide information about changes between the censuses in the detailed groups. As illustrated in subsequent sections, for small groups the ACS can only measure fairly dramatic changes, but dramatic changes are the ones that are most important to measure. Appendix 1 includes some examples of dramatic change between 1990 and 2000, suggesting that large percentage changes are fairly common for small groups.

For ancestry groups, the intercensal estimates program does not currently provide estimates. The ACS will update the estimates of the size and characteristics of ancestry groups.

IV. REASONS GIVEN FOR CONCERNS ABOUT ACS ESTIMATES FOR SMALL POPULATIONS

A. Each ACS 5-year average is based on a smaller effective sample size than the census long form. This means that the ACS estimates will have larger confidence intervals than long-form estimates. Since the long-form estimates already have large confidence intervals for small groups, this may make the data too noisy to be useful. (This is the “blurry freeze frame” in our metaphor.)

B. The 5-year averages are harder to interpret than a snapshot, especially when there are substantial changes in the population during the period of the estimate. (“This is the blurry video for fast-moving objects.”)

C. Because of the subsampling for nonresponse follow-up, populations with low mail return rates will tend to have a smaller number of ACS interviews than other groups. This will increase the standard errors for the groups with low mail response.

D. A small population group may be dispersed throughout the list of addresses. Given how small the ACS sample is each month, how can we be sure of getting a representative sample from the group? For example, some months there might be no one selected from the group.

The responses to the concerns will be summarized in Section V, with some additional details discussed in Section VI. The next Section discusses an important special case.

V. THE BASIC RESPONSES TO THE CONCERNS IN SECTION IV

This discussion will use the example of a population group that numbers 400 persons. This could represent either 1) the number of people in a particular population group in a particular area; or 2) the number of people in an area from a group who have a specific long-form characteristic, such as being employed in a particular industry teenage mothers enroll in school, or people who use a language other than English at home. This hypothetical example uses a relatively high, yet generally realistic, standard error for both the ACS and the long form. The relatives high standard error in the example would correspond to a characteristic which has the same value for all or most of the members of a household; Such characteristics tend to have higher-than-typical standard errors. As long form and ACS data become available for a wider range of characteristics, analyses like this one will be done using the actual standard errors for a variety of estimates, large and small.

A. Each *ACS 5-year average is based on a smaller effective sample size than the census long form. This means that the ACS estimates will have larger confidence intervals than long-form estimates. Since the long-form estimates already have large confidence intervals for small groups, this may make the data too noisy to be useful.* (This is the "blurry freeze frame" in our metaphor.) For a group of 400 people, the census long form would typically have a 90 percent confidence interval of roughly (280, 520).[1] An ACS 5-year average would have a larger slightly interval, on the order of (240,560). In other words a typical confidence interval for a hypothetical 2010 census long form estimate of 400 would be ± 120. A 2008-2012 ACS average estimate of 400 would have an interval of ± 160.

The basic premise of the ACS rolling sample is that this relatively moderate increase in the sampling error for one part of a decade is a good tradeoff to give the ability to update the 5-year average every year. If the size of the population changes, the 5-year average gives a more accurate picture of current conditions; for example, if the actual 5-year average population increases from 400 to 480, then the updated ACS 5-year average will be more accurate than the previous census.[2] As shown in Attachment 1, small population groups can easily change by much more than this, in which case the updated ACS estimate would give a much

---

[2] This would be the confidence interval will be centered on 400, if the estimate is 400. The actual estimate would not be exactly equal to the population value because of sampling error. See Appendix 3-A. The length of the interval depends on what characteristic is being measured. See Appendix 3-B.

more accurate reflection of current conditions, compound to continuing to use the previous census.

In addition, the ACS can provide information about when during the decade the changes took place. This can be important in trying to assess the reasons for the change and whether the change is likely to continue into the next decade. Because of the relatively small annual ACS sample size, this ability is limited to large changes, as discussed in Section VI.A and illustrated in Appendix 2.

The confidence intervals mostly reflect sampling error. This discussion has not taken into account potential improvements in non sampling error in the ACS, such as reduction in because the experienced interviewers and follow-up by telephone of mail forms with missing data. This is expected to compensate in part for the slightly large confidence intervals.

B. *The 5-year averages are harder to interpret than a snapshot, especially when there are substantial changes in the population during the period of the estimate. ("This is the blurry video for fast-moving objects.").* The bottom line of the argument in favor of the moving averages is as follows. If the population is not changing substantially, then a multi-year average is equivalent to a snapshot. If the population is changing substantially, then getting some information about the change is better than getting no information, as happens when data are collected only once in ten years.

A more detailed answer, depends on the specific situations. Section VI.B provides examples of ways a population might change over time, and how the ACS information would be used in these different situation. In every case, the series of 5-year averages are preferable to statistics for only having one year out of ten. In some situations, the fullest use of the information provided by the ACS would supplement the standard 5-year averages with information from the 3-year and 1-year average “research files.”

There are some cases in section VI.B where only a single 5-year average, with no updating, would not be as good as a decennial snapshot. In these cases, it is the regular updating that gives the ACS its advantage. Obviously, it would be ideal if we could collect the full long form sample, every year, but that is not an option because of the cost and public burden.

C. (*The 5-year averages are harder to interpret than a snapshot, especially when there are substantial changes in the population during the period of the estimate. (“This is the blurry video for fast-moving objects*.”) On average, about 60 percent of the population are represented by the ACS data collected by mail or CATI. For most of the remaining 40 percent of the population, the data from a one-in-three subsample

The ACS Survey is being designed to compensate for areas with lower-than-average mail response rates, so that the standard errors will be more similar for all small population groups. Studies have shown substantial variation in mail response rates by race and geography. Mail response rates in the testing phase have been in lower for tracts with high proportions of African American or Hispanic population. There is some evidence of substantially lower rates for tracts with high proportions of American Indian or Alaska Native population or Native Hawaiian and Other Pacific Islander population. There is also evidence that households with limited English proficiency, including non-Hispanic households, have a lower-than-average mail return rate. It is important to compensate for these variations, to bring the standard errors for all groups and areas in line with the overall expectations. We are developing plans to oversample geographic areas with lower mail response rates, and to provide more opportunities for people with limited English proficiency to respond to the ACS as another language, either by mail or telephone. These developments are important for meeting the standard error objectives for all population groups.

For many areas and population groups with lower-than-average mail response rates. The ACS is expected to have compensating advantages in nonsampling error Judging from 1990 census results, the completeness of long-form data collection in such “hard-to-enumerate” tends to be uneven. There is evidence from the ACS tests that the ACS has more consistently high completeness of data collection for the units in its sample, because of the smaller, more experienced interviewing staff and because of the opportunity to improve the collection over time in areas where these are collection problems. This is being studied further as part of the ACS evaluators.

It is important to keep the proportion of missing data low for all areas and population groups.

More complete data collection is important for two reasons. First, incomplete data increases the standard errors of the estimates; in other words, the confidence intervals are larger than they typically would be. Second, the incomplete data may result in a statistical “bias” in the estimates. The potential “bias” occurs because the survey systematically excludes some proportion of the population, namely the people whose data would be missed by the survey even if their address was selected for sample and follow-up.

A. (*How can a small sample represent a small, a small population group may be dispersed throughout the list of addresses*. *Given how small the ACS sample is each month, how can we be sure of getting a representative sample from the group? For example, some months there might be no one selected from the group widely dispersed population group*?) Although sampling statisticians rely on the "laws of probability" to select survey samples that are "representative" and have a certain "margin of error," the intuition behind these "laws" seems less plausible in some cases than in others. It is easy for people to visualize how a systematic sample, for example taking every sixth address, gives good representation for a population group that is clustered in a particular geographic area. However, it is hard to visualize how a representative sample can be "guaranteed" if the population group is scattered randomly without any particular pattern. This is hardest to visualize when the population group is small and the sampling rate is relatively small.

These intuitive concerns about the "guarantee statisticians" of representativeness relate to a legitimate issue. The laws of probability make "guarantees" only within a certain "margin of error," or "confidence interval." When the sample and population group are both small, the margin of error can be large, as a percentage of the actual estimate. The "laws of probability" do not have any remarkable property of giving very precise estimates from small samples. The only thing that is remarkable about a "scientifically" drawn sample is that the laws of probability allow statisticians to be quite specific about how large the margin of error due to sampling is likely to be.[2] Section VI.C has more discussion about how it is possible to know the margin of error for a 5-year average, even with a very small sample each month. The explanation is in brief “the law of averages”: Each month’s estimates may be very unpredictable, but averaged over 60 months, the results are reasonably stable.

Whether a survey’s sample size is adequate depends on whether the confidence intervals for the survey estimates are small enough to allow data users to learn what they need to learn from the data. A common way to think about the adequacy of confidence intervals is to consider how large a difference in survey estimates it would take to be “statistically significant”. With census long form estimates for two groups of about 400, each having a confidence interval of ± 120 estimates would be statistically significant unless the two estimates were as different as about 315 for one group versus 485 for the others.

With the larger ACS confidence interval of ±160 for a 5-year average, the difference between averages of 315 and 485 would not quite be statistically significant. It would take a difference of 297 versus 513 to be significant. This indicates the price paid because the proposed ACS has a smaller sample size in a 5-year period than the long form has in the census year. To reiterate the points made in Section V.A., for this price, data users get the ability to look at changes during the decade, and reductions in nonsampling error will offset a portion of this price.

[2] This contrasts with non-sampling errors such as nonresponse, undercoverage, or misunderstanding of questions, where it is hard to quantify how large the resulting error in the estimates is likely to be.

The ultimate question for users of data for small groups is whether the long form's somewhat slightly greater precision, for comparing groups at census time, has practical importance that is worth giving up the opportunity to learn about substantial changes in the size or characteristics of the small group over time. The premise of the ACS design is that the ability to learn about substantial changes over time is very important for small groups, and worth a moderate loss of precision at any given point in time.

As an example, consider the potential use of estimates of the number of children age 0-5 who speak a language other than English at home, to prepare school systems for the need to provide appropriate educational opportunities for these children in coming years. The series of ACS 5-year averages can monitor trends in the number of children in this group, Also discussed in the next section, the 1-year and 3-year averages can detect sudden large changes. By contrast, a single decennial estimate or a single 5-year average, whether 400±120 or 400±160, does not have the precision and timeliness to be much help in planning.[1]

VI. MORE DETAILED EXAMPLES AND DISCUSSION

(*More about standard errors and change estimates*). The figures in Appendix 2 illustrate the ACS standard errors as they affect the measurement of small populations which change over time. Figure 1 shows ACS 5-year averages for a population which changes from 400 in the year 2010 to 1400 in the year 2020. The graphs start with the year 2012, when the population value is ___.

[1] Note that the 2010 long form data would be available in late 2012, in time for planning the 2013-2014 school year. The 2007-2011 ACS average, available in mid-2012, would not by itself be any more timely for planning the 2013-2014 school year. It is the series of 5-year languages, and the ability of shorter averages to more quickly detect large changes, that make the ACS more useful.

In each figure, the diamond symbols indicate the assumed population values for each year. The solid lines indicate the upper and lower bounds for a probable estimate from the ACS averaged for the past 5 years and updated each year. For each year, the graph shows the most recent information that would be available in that year. For example, the bounds for the year 2018 show the range that has a 90 percent probability of containing the 2013-2017 average estimate for the ACS sample, given that the population values are the diamonds.

The increasing spread between the upper and lower bounds in figure 1 occurs because the number of people with the characteristics is increasing. Larger estimates tend to have larger standard errors, although the standard error grows smaller as a percentage of the estimate.[1]

Figure 2 represents the corresponding probable ranges for the long form sample.

Although the 5-year averages in Figure 1 tend to lag slightly behind the population values, and there is substantial sampling error is the long form in Figure 2X, the 5-year moving averages obviously tend to be closer to the current population value in most years than the long form in Figure 2. The ACS 5-year averages reflect the direction of the actual trend, unlike the long form.

[1] For example, the standard error for the 2012 estimate is____, which is___ percent of the middle of the range of values, which is___. For the 2022 estimate, the standard is ___, which is ___ percent of the mid range value which is ____.

Figures 3 and 4 show the same example, using ACS 3-year and 1-year averages. The 3-year averages are a reasonable alternative to the 5-year averages for many uses of the data, where The smaller time lag would compensate for the higher sampling error. The single-year ACS has a much larger range of probable error, and is not as useful unless there is a very large change.

Figures 5 through 8 show the same information as Figures 1 through 4, except that the true values in the example have a sudden jump from 400 to 1,400 at the end of 2014. The 5-year ACS averages in Figure 5 pick up the changes with a few years, much sooner than the decennial long form in Figure 6. The changes is fully reflected it in the 2015-2019 average. However, the 5-year averages give the impression that there is a steady increase starting in 2015, rather than the sudden jump. This is better than the actual information than that form, the long form, but not the best picture of the change.

This is an example where a more detailed analysis using 3-year and 1-year averages is needed to get a full picture. Having decided from looking at the 5-year averages that there is an important change in the population during the decade, an analyst can learn more by looking at the 3-year and 1-year averages. In this extreme example, comparing each 1-year average to the previous year would give a good indication of the timing of the change. After learning from the 1-year numbers that there might be an unusual jump in 2015, the 3-year averages give a better idea of the size of the jump without overly “smoothing” the change as the 5-year averages do. Having considered all three ACS charts, the change (up or down), the analyst would know the direction of the magnitude of the increase, and that it took place over a few years in the middle of the decade. The analyst might still be uncertain whether the change took place all in one year or over several years. None of this information would be available from two measurements taken ten years apart (Figure 6).

The practical implications for policy decision are obvious. The ACS allows informed decisions to be made in response to changing conditions. The decennial census documents historical changes after they have occurred over a decade.

A. *Multi-year averages for changing populations.* If the population does not change meaningfully over a 5-year period, then there is no issue about interpreting the 5-year average. For different patterns of change over time, illustrated below, the average may relate in different ways to the single year estimates. With the continuously collected ACS data, it is possible to get considerable information about the magnitude and direction of change over time. Because of the sampling error, however, it will not be possible to be sure of picking up a slight trend, or whether a strong trend is steady or somewhat irregular.

The examples below address the question of how useful it would be to know only the information available from, compared with knowing one individual value out of ten. To keep the examples simple, the tables below do not include the margins of error, as did the graphs Figure 1-8. Appendix 2 provides a discussion of some important statistical points for those who want a detailed technical discussion.

In all the examples, averages which cannot be calculated from the data for the years shown in the tables, are left blank to make the example easier to follow. These rules would be available from the ACS, once it has been fully in place.

In most of the examples, the census year is the fifth year shown in the table, so data before and after the census are shown. In some examples, to illustrate what would have happened if the pattern of change had occurred one year earlier compared to the census, there is an additional row of numbers showing what would be measured by a census in the sixth year.

1. **Steady Trend.** If there is a steady increase or decrease in the size of the group being measured over a 5-year period, then the 5-year average corresponds to the value in the middle year of the average. For example, the second row shows that the average for years 1 through 5, which is available in year 6, is 440. This is equal to the actual size of the group in year 3 as, shown in the first row.

EXAMPLE 1: A STEADY TREND

Year(y)	1	2	3	4	5	6	7	8	9	10	11	12

Actual Size of group in year y	400	420	**440**	460	480	500	520	540	560	580	600	620
Average of previous 5 years	___	___	___	___	___	**440**	460	480	500	520	540	560
Previous census (year 5 census)	___	___	___	___	___	___	480	480	480	480	480	480

If the steady increase continues, the series of averages will give an accurate description of the trend, albeit with a three-years behind. The decennial snapshot, in the third row, misses the trend and becomes steadily more out of date, and fails to measure the trend. If the trend of annual data is somewhat irregular, the moving average will tend to smooth out the irregularities, making the trend look more steady than it actually is. A smooth increase in the moving averages means that the actual population is generally trending upwards, but not necessarily as steadily as the averages suggest.

2. **Sudden jump or drop.** The 5-year averages will show an increase when there is

   sudden jump, but they will smooth it out, masking the suddenness of the change. This is illustrated in Figures 5 through 8, and in the second row of following example. Two possible census years are shown in the example, a “year 5” census illustrating a jump which occurs right after the census, and a “year 6” Census illustrating a jump which occurs right before the census.

EXAMPLE 2: SUDDEN JUMP

Year (y)	1	2	3	4	5	6	7	8	9	10	11	12
Actual Size of group in year y	400	400	400	400	400	600	600	600	600	600	600	600
Average of previous 5 years.	—	—	—	—	—	400	440	480	520	560	600	600
Previous census (year 5 census)	—	—	—	—	—	—	400	400	400	400	400	400
Previous census (year 6 census)	—	—	—	—	—	—	—	600	600	600	600	600

To detect the fact that the change is much more sudden than the 5-year averages indicate, it is necessary to look at the 3-year and 1-year averages, as part of studying and interpreting the changes. (As in Figures 7 and 8.)

The uses of decennial census data, in a situation of sudden change like this, depend very much on the year that the large change takes place. If , as with the September 11 attacks in 2001, the census comes shortly before an event of dramatic change, the census will provide a valuable profile of the area, but it will be 10 years before the impact of the event will be measured. This is illustrated in the third row of the table. If the census occurs after the change, as illustrated in the fourth row, it will instead provide a useful “after” profile. If the census occurs during a period of dramatic change, for example right after a natural disaster, the census be disrupted by the event and the data may have limited value.

The ACS 5-year averages would give a baseline profile for the small group before the dramatic change year 6, and eventually would give a 5-year average profile after the change (year 12). In that sense, it combines the information of a census before the change and a census after the change. However, as the example indicates, the picture given by the averages that cross the change year (years 7 through 11) requires careful interpretation.

The entire series of ACS moving averages have a clear advantage over a decennial snapshot in such situations. They provide before-and-after profiles of the area. The 1-year and 3-year averages will give a useful earlier measure of the change if it is large.

3. **Irregular, seemingly patternless, change.** This seems at first like the most difficult situation for which to interpret an average, but it is actually the most natural situation to use an average. Averages would often be used in such situations, even if there were a census every year and no concern about sampling error, because the average over a period of time provides a more stable description of the area.

For this example, consider the populations in the group of interest in tow areas, each with considerable variation from year to year:

EXAMPLE 3: IRREGULAR CHANGE POPULATION IN FIRST AREA

Year (y)	1	2	3	4	5	6	7	8	9	10	11	12
Size of group in year y	159	263	226	367	117	253	79	298	234	64	159	162
Average of previous 5 years	—	—	—	—	—	226	245	208	223	196	186	167
Previous census (year 5 census)	—	—	—	—	—	—	117	117	117	117	117	117
Previous census (year 6 census)	—	—	—	—	—	—	—	253	253	253	253	253

EXAMPLE 3: IRREGULAR CHANGE POPULATION IN SECOND AREA

Year (y)	1	2	3	4	5	6	7	8	9	10	11	12
Size of group in year y	491	355	317	513	458	270	534	394	468	373	539	347
Average of previous 5 years	—	—	—	—	—	427	383	418	434	425	408	462
Previous census (year 5 census)	—	—	—	—	—	—	458	458	458	458	458	458
Previous census (year 6 census)	—	—	—	—	—	—	—	270	270	270	270	270

In this example, the second area generally has a higher population for the group of interest. As the averages in the second row show, the first area's population averages somewhere around 200 and the second area somewhere around 400. The 5-year averages in row 2 show this general relationship of the areas more clearly than the "unsmoothed" single year numbers in row 1.

A decennial snapshot can give a very misleading picture in such situations. In this example, if the snapshot is in year 5, it would give an usually large impression of the difference between the areas (117 compared to 458 from row 3). If the census were in year 6, it would show the areas as being almost the same (253 compared to 270 from row 4).

4. **A single-year spike**. This is an extreme case where a decennial snapshot could be advantageous under very restricted circumstances, but in general the ACS information would be preferable, although far from perfect. In this example, members of a population move into an area for one year.

EXAMPLE 4

Year (y)	1	2	3	4	5	6	7	8	9	10	11	12
Size of group in year y	0	0	0	0	0	400	0	0	0	0	0	0
Average of previous 5 years	—	—	—	—	—	0	80	80	80	80	80	0
Previous census (year 5 census)	—	—	—	—	—	—	0	0	0	0	0	0
Previous census (year 6 census)	—	—	—	—	—	—	—	—	400	400	400	400

If the group happened to be in the area at census time, then there would be a full one-in-six sample to provide a description of the group with an error only on the order of ±120 as in line 4 of the example. However, this may not be a desirable outcome, unless there is independent information that the group was in the area only for one year. It would not necessarily be desirable to use the census data to describe the area in future years. If the group happened to be in the area outside of the census year, then the census would give no indication of the characteristics of the group, and no indication that it ever was in the area, as in line 3.

In this example, the ACS 5-year averages in line 2 would indicate that the group was in the area but would give a "blurred" picture of exactly when the group was there. The 5-year average would be something like 80 ± 72.[4] The 1-year estimates would give the most useful description of the situation. The ACS 1-year estimate would be based on a one-in-forty sample and would give a confidence interval something like 400 ± 360. The margin of error is the same percentage of the estimate for both the 1-year and 5-year averages: 72 divided by 80 is the same as 360 divided by 400. This is because the 5-year average estimate in this example depends entirely on the data from the one year when the group was in the area. After looking at the 1-year series, the analyst would

[4] These intervals illustrate the margin of error and general magnitude of the estimate. An actual sample would probably give an estimate different than 80. See the discussion in Appendix 3-1.

recognize the blurring effect of the 5-year averages, which gives a l-year estimate spread over 5 years.

The ACS provides some useful information. Even though the 5-year averages give a confusing picture of the timing, the 1-year data would indicate that the group was in the area. and when the survey estimates of the characteristics of such a small one-year group will have a large confidence interval and will be of little use, unless the members of the group have very similar characteristics, such as being employed in a particular occupation.

In 9 years out of 10, a decennial snapshot would totally miss the group. If the census happens to be in the year the group is in the area, then it gives much better estimate of the group's characteristics for that single year than the ACS. However, in future years, continuing to use this snapshot would give an erroneous picture of the areas. So the Census would be advantageous only if the group is in the area during census year and if there is independent information that this was a short-term event. Otherwise the ACS provides more useful information.

C. *(Details on Proposal for oversampling areas with low mail response.)* Specific proposals are given in the paper Tersine, et al (2002) listed in the references.

There is increasing evidence of the completeness of data collected by the ACS field staff in "hard-to-enumerate" areas. This was seen in a study of the ACS data in the Bronx NY ACS test sites (Salvo and Lobo, 2002), as well as U.S. Census Bureau (2002). Studies for additional test sites will be available next year.

D. (*How does probability sampling work for small samples*?) The discussion starts with a simplified analogy to illustrate the basic principles. Then it presents a more exact theoretical model, and then describes how confidence intervals are calculated in practice.

E. *A simplified analogy*. To give an example with round numbers suppose the group of interest had 240 people instead of 400. Each month the ACS selects a sample of about one in 480 people in the population, or about one in 40 for the year. So the expectation is that the sample will have about one person from the group in every two months of sample, or an expected "one-half person each month," calculated as 240/480.

Looking at 60 months of sample, this is somewhat analogous to tossing 60 coins, each with a fifty-fifty chance of heads, giving an expected number of heads equal to one-half on each toss of the coin. The total number of people from the group who fall into sample in the 60 months would correspond to the total number of heads on the 60 coin tosses.

To get from the number of people in the group who fall into sample, to the estimated number of people in the group who are in the population, it is important to remember that the survey data are weighted. Since the ACS samples about one person in 40 every year, each sample person in any given year is counted as 40 people in the population for that year's estimate. To get the 5-year average, the 5 annual estimates are added together and divided by 5. This means that the 5-year average would be 8 times the number of people in sample in the 5-year period (counted as 40 in the annual estimate, and then divided by 5 to get the 5-year average).

This means that if exactly 30 of the 60 "coin tosses" were "heads" then the 5-year average estimate would be 8x30=240, the "expected" result. If a large number of different samples were selected, this would be the average result from all the sample estimates combined.

With 60 coin tosses, there might not be exactly 30 with heads up. However, the range of results that are likely to occur is quite predictable. There is a 90 percent probability that out of 60 coin tosses, the number of heads will be at least 23 and more than 37.

Likewise, the exact results from 60 months of ACS sample may not be the "expected" number, but the likely range of result for each statistic can be predicted. Multiplying by 8, this give an estimated number of people in the range 184 to 296. There is a these gives a 10 percent chance that the estimate will be outside this range, but when that happens, it is rarely far outside the range. In tossing 60 coins, there is a 99 percent probability that the weighted estimate will be in the range 160 to 320, and a 99.9 probability that it will be in the range 136 to 344.

This predictability of the range of results that will be obtained from a large number of random outcomes is the basis for the science of sample survey. It is the reason that sample units are selected using random numbers rather than someone's judgment of what would be a good sample. The same principle is used in applications as varied as actuarial life tables, random clinical trials, and predicting payouts at casinos. Anyone who would like a more vivid illustration of how random sampling works is encouraged to actually perform the experiment of shaking 60 coins in a cup, pouring them out onto a table, and counting the number of "heads", that come up. Repeat this 5 or 10 times to see how the results vary, but stay within a predictable range.

A more realistic model. The number of people in a small population group who are part of the ACS sample is, of course, not produced by tossing a coin to get either one person or zero people in sample each month. If the population has 400 people in the group of interest, they actually are divided up into households, some with one person from the group, some with two and some with many more. Some of these households will not respond by mail or telephone, so even if the household is selected there is a two-thirds chance that it will be dropped from sample and a one-third chance that it will be selected for follow-up, and weighted to count in the estimates as three households with the same characteristics as the sample household.

A more appropriate model would be what probability theorists sometimes call an "urn model". Imagine filling an urn with as many beads as there are housing units in the population. Most of the beads are labeled "zero," but some are labeled with one, two, or some other number, with the numbers totaling to 400. Some of the beads are labeled “N” for “nonresponse by mail or telephone.” A sample of 1 in of 480 beads is selected each month. For each one of the sample beads which is labeled “N”, a random number is chosen to give a one-third chance of keeping the bead, and multiplying its number by three, and a two-thirds chance of discarding the bead.

Mathematicians have analyzed experiments like this theoretically, and confirmed their analysis with small-scale experiments with actual beads and boxes, and larger-scale experiments with computer-generated random numbers. Such experiments show the same sort of predictability of results as coin tosses. The exact numerical results giving the likely range of values are more complicated to work out than for coin tosses, but for reasonably large samples, the same mathematical approximations can be used to work out the likely range of values for the estimates from a sample of 480 “beads” from any such “urn”.

The same rules of arithmetic that apply to beads with numbers on them also apply to households with a number of people in each one.assigning an identification label to each address, and using a random process to select which labels are in sample, the numerical results of the sample of households from a list of addresses are very nearly the same as if beads were being drawn at random from urns.

Modern methods for estimating standard errors. Up until the 1960’s, survey confidence intervals were estimated by learning enough about the population to deduce the standard error theoretically from mathematical models such as the urn models described above. For each characteristics, it was necessary to make assumptions about the pattern of “clustering” of the characteristic within households. Highly clustered characteristics such as ancestry, where all members of a household often have the same characteristic, tend to have higher standard errors than less clustered characteristics such as disability.

Since the 1970’s the more usual procedure has been to use “ replication methods”. These methods work by splitting the sample into pieces, so that each piece is a microcosm or “ replicate” of the full sample. Then the variation among the pieces is measured, and the standard error is mathematically deduced from the measured variation among the pieces.

Mathematical theory, and tests with simulated populations, show that these methods give numerical results similar to the more traditional “urn model” approaches for deriving the standard errors. The replication methods can more readily reflect the actual clustering patterns found in the actual population, because they are calculated from the actual sample data, not a theoretical model. The ACS and the census long form both use replication methods to calculate standard errors, so the standard errors reflect the actual clustering found for various population groups.

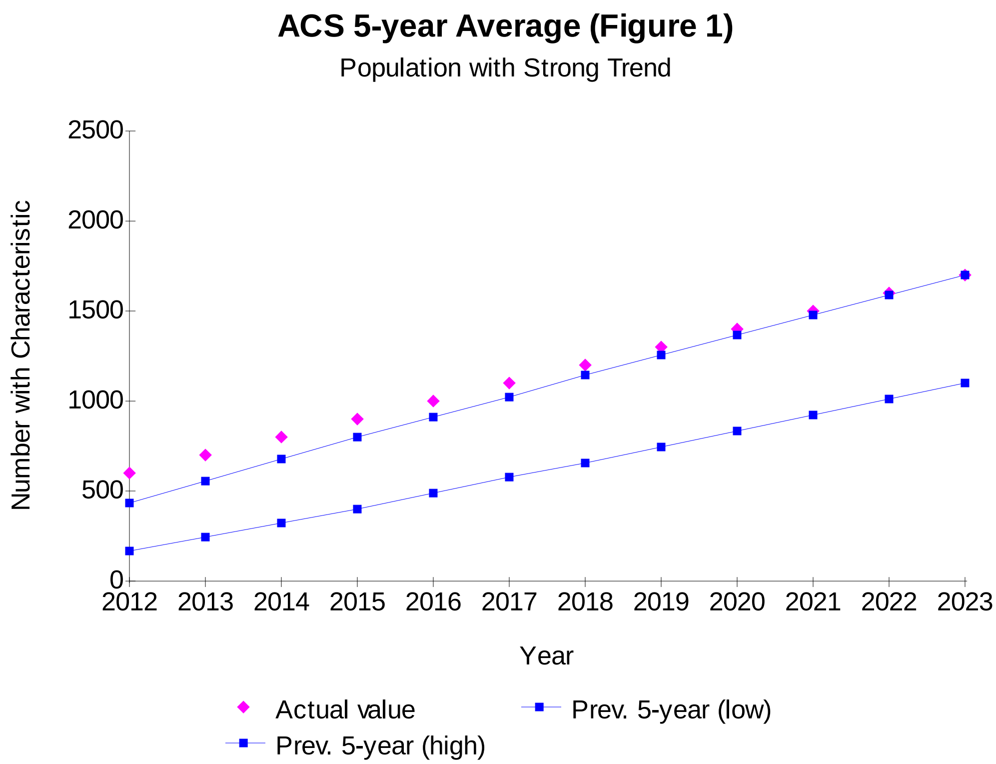
In summary, probability sampling works because of the predictability of random events in the aggregate. No single month of the ACS sample is very predictable, but 60 months is enough of an aggregate for the results to be predictable within a calculatable margin of error. The calculation of error is based on statistical methods that have been developed to deal with random events that are more complicated, but basically similar, to tossing coins and counting the number of times heads comes up.

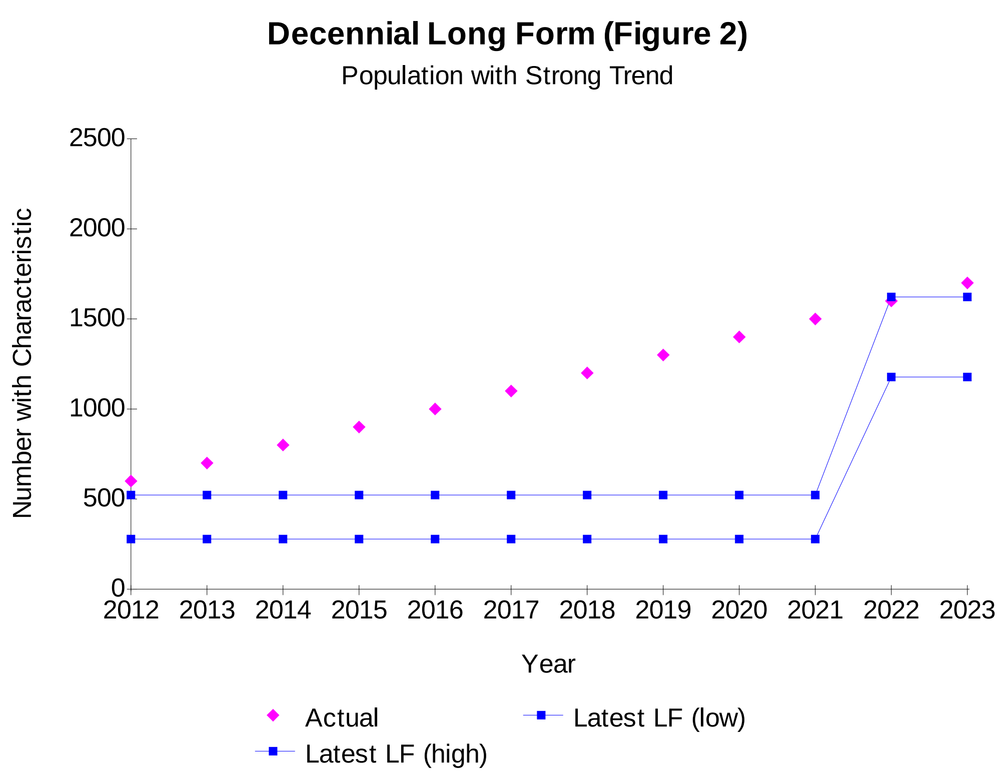
Appendix 1

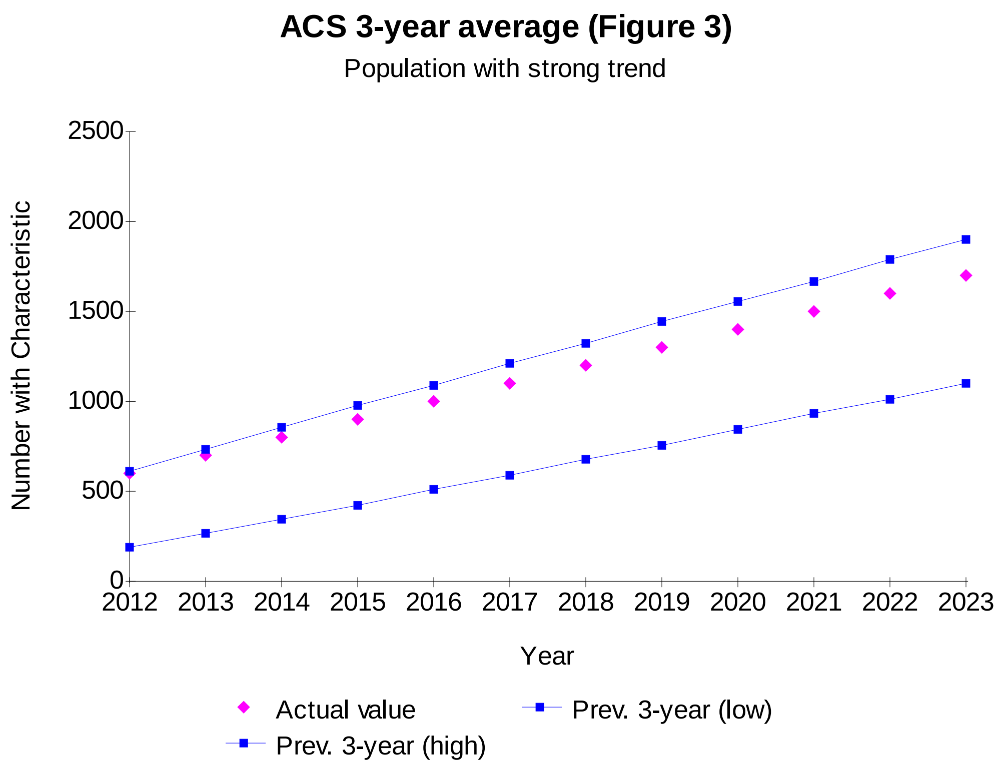
Examples of Large Growth For Small Population Groups in the ACS Comparison Counties

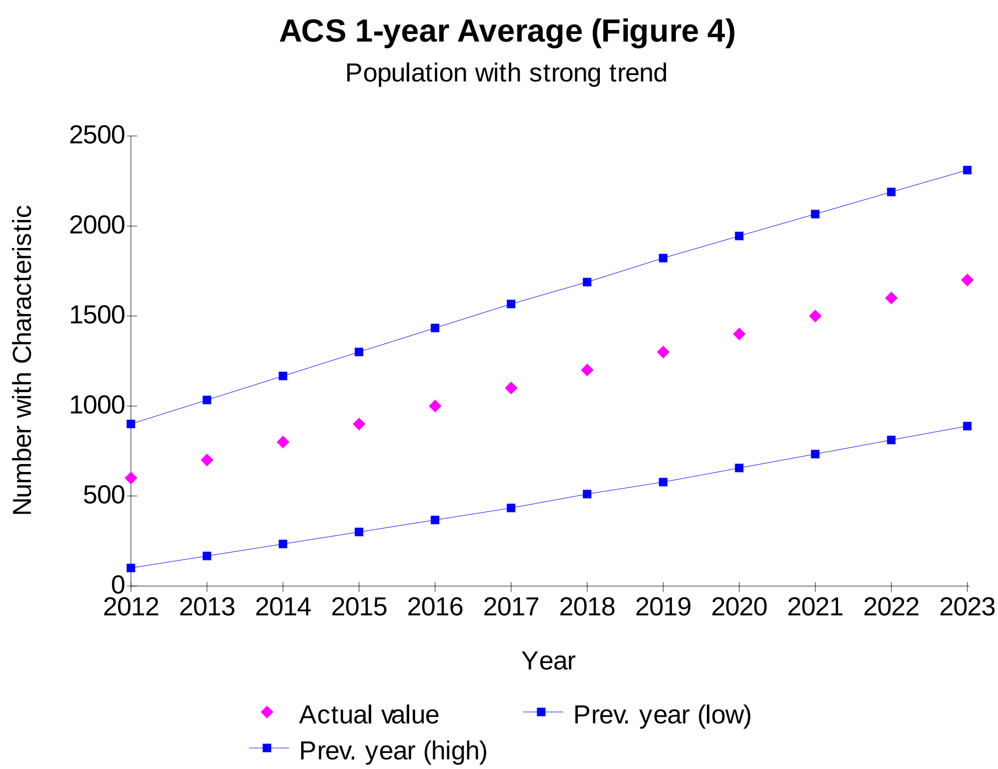
Population Group	County	1990 Estimate	2000 Estimate	Sources (1990/2000)
Asian Indian	Pima, AZ	1,041	2,105	STF-1/SF-1
Chinese	Ft Bend, TX	4,072	10,500	STF-1/SF-1
Korean	Lake, IL	1,923	4,089	STF-1/SF-1
Vietnamese	Douglas, NE	529	1,122	STF-1/SF-1
Black or African American	Schuylkill, PA	842	3,147	STF-1/SF-1
American Indian or Alaska Native	Bronx NY	6,069	11,383	STF-1/SF-1
American Indian or Alaska Native	Lake, IL	1,198	1,801	STF-1/SF-1
Native Hawaiian and Other Pacific Islander	Bronx NY	541	1,383	STF-1/SF-1
Other Micronesian	Multnomah, OR	181	505	STF-1/SF-1
Dominican	Broward, FL	3,489	8,869	STF-3/ACS
Salvadoran	Douglas, NE	52	414	STF-3/ACS
Arab	Broward, FL	5,174	9,461	STF-3/ACS
Ukranian	Multnomah, OR	1,524	5,469	STF-3/ACS

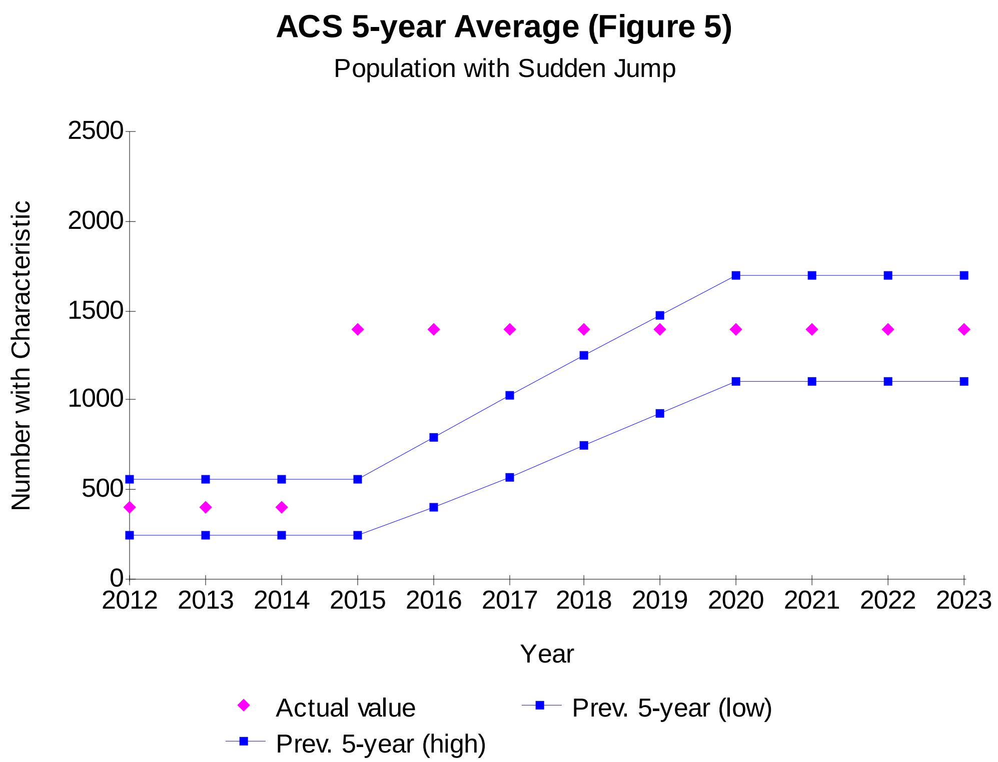
NOTES: The 2000 estimates for race or Hispanic origin are for those marking one race or one origin. The ancestry estimates are first ancestry. Census counts have been used when available. For Native Hawaiian and Other Pacific Islander count for in 1990, the detailed race tables were used. For the ACS estimates, the lower bound of the confidence interval for the 2000 data is shown. This is the most conservative estimate, and the actual growth is likely to have been larger than that shown in the tables.

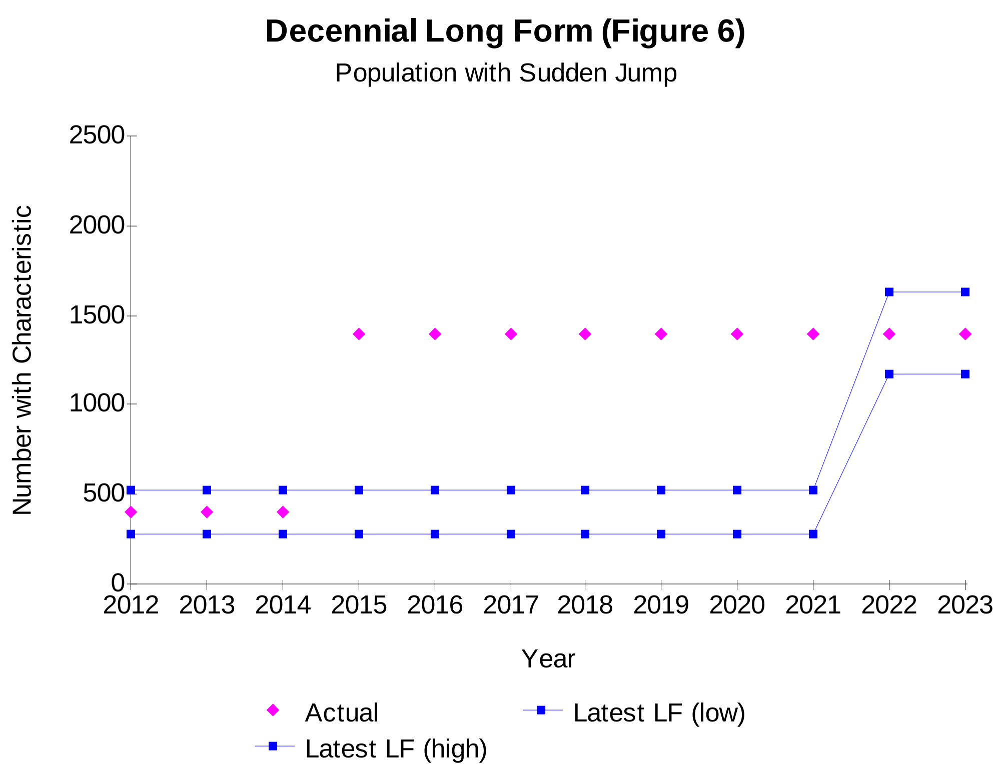

ACS 5-year Average (Figure 1)
Population with Strong Trend
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. 5-year (low)
Prev. 5-year (high)

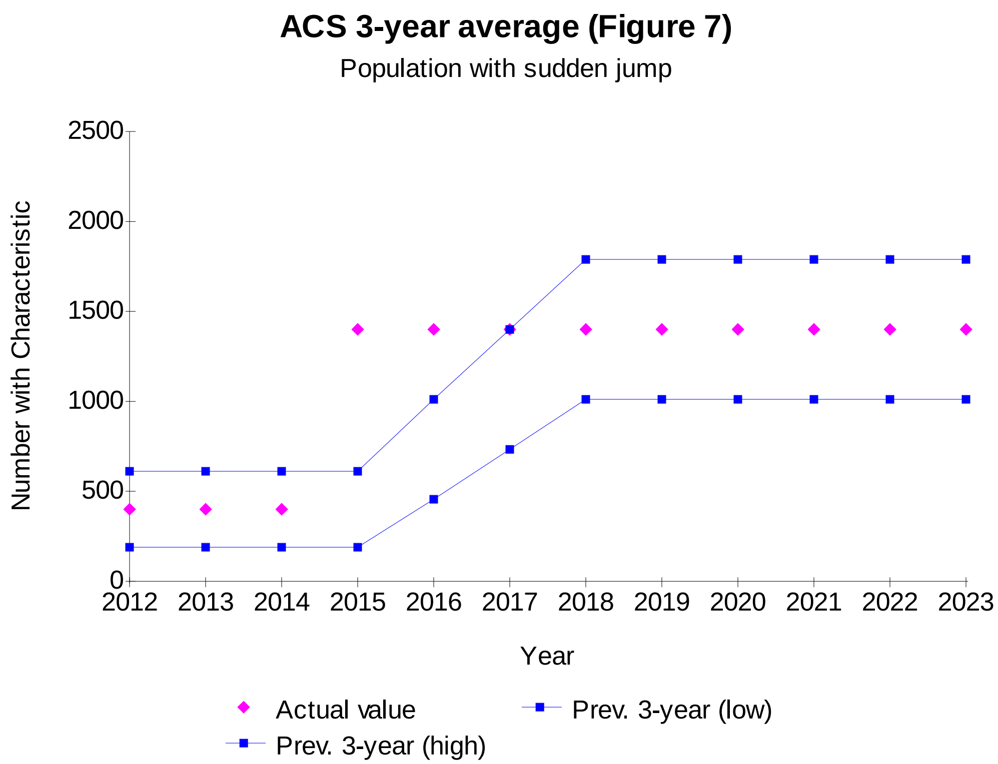

Decennial Long Form (Figure 2)
Population with Strong Trend
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual
Latest LF (low)
Latest LF (high)

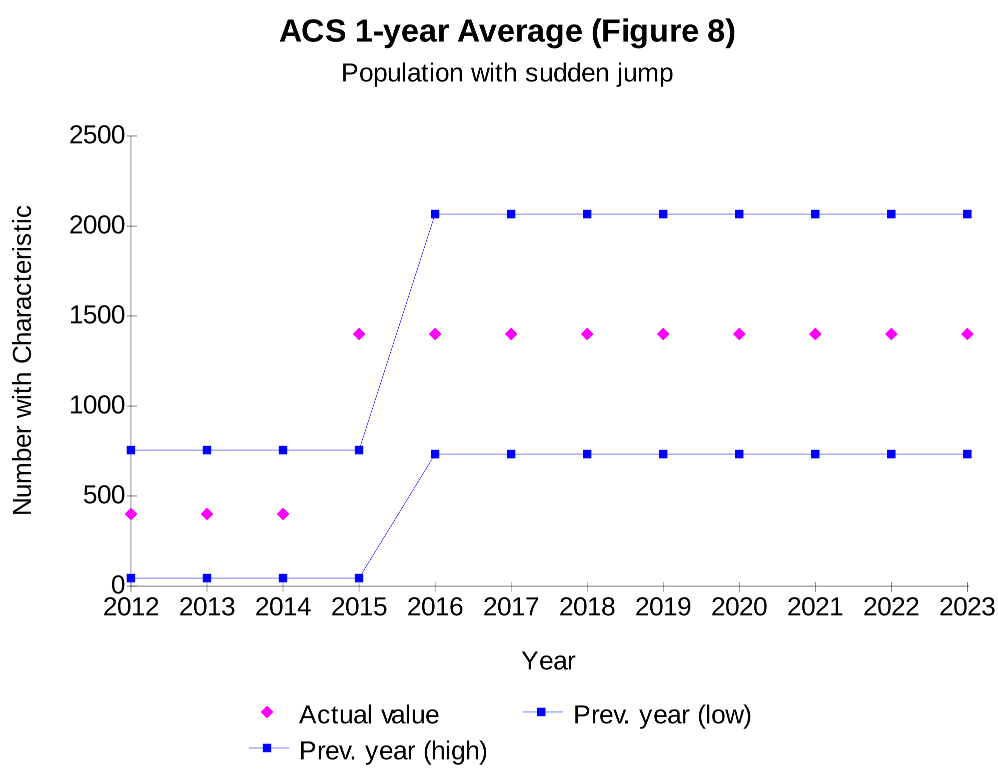

ACS 3-year average (Figure 3)
Population with strong trend
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. 3-year (low)
Prev. 3-year (high)


ACS 1-year Average (Figure 4)
Population with strong trend
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. year (low)
Prev. year (high)


ACS 5-year Average (Figure 5)
Population with Sudden Jump
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. 5-year (low)
Prev. 5-year (high)


Decennial Long Form (Figure 6)
Population with Sudden Jump
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual
Latest LF (low)
Latest LF (high)


ACS 3-year average (Figure 7)
Population with sudden jump
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. 3-year (low)
Prev. 3-year (high)


ACS 1-year Average (Figure 8)
Population with sudden jump
Number with Characteristic
2500
2000
1500
1000
500
0
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Year
Actual value
Prev. year (low)
Prev. year (high)

A. Population values and estimates. The discussions in Sections V and VI gloss over an important distinction made in theoretical statistics. In Figures 1-8 the graphs illustrate the range of likely estimated values, *assuming that the population has 400 people in the group of interest*. In Section V, the discussion considered what margin of error would be associated with a survey estimate, *assuming that the estimated number of people was 400*. Since the goal of the discussion was only to give a general idea of how ACS and long form standard error compare, and not a statistics lesson, the subtle distinction was not emphasized.

The distinction will be illustrated with an example. Suppose that the population of the group is a constant 400 people for the years 2008-2012, so that the actual 5-year average population is 400 and also the 2010 count would be 400. The estimate from a long form sample could (with 90 percent probability) range from 280 to 520. The corresponding ACS range is 240 to 560. This range is what the graphs in Appendix 2 illustrate.

The actual confidence interval would be centered around the estimated value, not around the population value 400. If the long form gave an estimate at the upper end of its range, the confidence interval would be 520±137, or (383, 657). If the ACS estimate were at the upper end of its range, it would be 560±189, or (371,749). Alternatively, the ACS or the long form could give estimates at the low end of the range. Exactly where the estimated value falls depend on which sample of addresses happens to be selected. The margin of error is larger or smaller depending on the size of the estimated value, as well as the sample size.

The example in the text simplistically assumed that the estimated value was 400. The midpoint of the range of estimates is 400, so in that sense 400 is a “typical” value for an estimate, and 400±120 or 400±160 is used to illustrate a typical confidence interval. However, the actual situation is more complicated as indicated in the previous paragraph.

B. **Total Error.** The total survey error is usually partitioned into “variance” and “bias”. The variance mainly refers to the sampling error, which is expressed by the margin of error or the confidence interval.“Bias” is defined as the difference between the “true” value and the average of all possible samples.

In the example in the second paragraph of Section V. A, the long form estimate is biased because it was based on a sample selected from a population where the value was 400, so 400 was the average of all possible samples, but the “true” value has now grown to 480. In this example, the ACS has no bias, as far as estimating the updated values of 480, but it has a larger standard error.

The usual measure of total error is the Mean Squared Error (MSE) given by the formula

$$\text{MSE} = (\text{Standard error})^2 + (\text{Bias})^2$$

At some point, as the population grows beyond the initial value of 400, and the bias in the outdated estimate grows, the term $(\text{Bias})^2$ outweighs the term $(\text{Standard Error})^2$ so the Mean Squared Error of the long form estimate is larger than that of the ACS. In this example, 480 is just beyond the break-even point.

Of course, if the “truth” is defined to be the most recent population value rather than the previous 5-year average, then the ACS 5-year average also has a bias because of population growth during the 5 years. Defining the “truth” this way increases the MSE of the ACS, but it tends to increase the MSE of the outdated long form even more, because the bias term is squared. As an example, of the last point, consider the steady trend in Example 1. By ear 10, 5 years after the census the most recent 5-year average would be 520. Compound to this, the census has a “bias” of –40 = 480-52, or a squared bias of 1600. If the “truth” assumed to be 520, the 5-year average has zero bias. If the “truth is taken to be the actual value in year 10, which is 600, than the 5-year average has a bias of –80=520-600 or a squared bias of 6,400. The census has a bias of –120=480-600, giving a squared bias of 14,400. The squared bias for the census therefore exceeds the squared bias of the 5-year average by 8,000 rather than 1,600.

C. "Margin of Error" and "Confidence Interval". The term "90 percent confidence interval" is precisely defined in the statistical literature as an interval from sample data, calculated in such a way that there is a 90 percent probability that the interval will contain the population value. A longer interval, for which the probability of containing population value is 95 percent, would be called a "95 percent confidence interval." The interval can be expressed[1] either the form 400±160 or (240,560).

The term "margin of error" has several meanings. It is sometimes used as a general term, synonymous with "the plus or minus amount in any confidence interval" and sometimes with a specific restricted meaning used mainly in political polling. The term is also used as a general reference to uncertainty in the estimate, with no specific numerical measure in mind.

In political polling, "the" margin of error for a poll is the margin of error for a 95 percent confidence interval, assuming that each candidate has 50 percent of the vote. The pollsters give one margin of error for the entire poll, rather than a different margin of error for each estimate. For a typical poll, a candidate polling at 50 percent might have a confidence interval of 50±3, while one polling at 10 percent would have an interval of 10±2. In this case, the poll would be said to have a margin of error "3 points, plus or minus", because that is the largest margin of error[1] In the more general use of the term, the first estimate would be said to have a margin of error of ±3 and the second estimate a margin of"±2.

---

[1] It turns out that estimates greater than 50 percent have a smaller margin of error that the 50 percent estimates. For example as estimate of 90 percent would have a confidence interval of 90±2.