COMPARISON OF ACS AND ASEC DATA ON AGE AND SEX

Karen Humes
Julie Meyer
Denise Smith

Population Division
U.S. Census Bureau

July 16, 2007

This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on the statistical and methodological issues in this report are those of the authors and not necessarily those of the U.S. Census Bureau.

INTRODUCTION

This report compares national age and sex distributions based on data from the 2004 American Community Survey (ACS) with those based on data from the 2004 Annual Social and Economic Supplement (ASEC) to the Current Population Survey. In this analysis, we compare the 2004 ACS and the 2004 ASEC distributions, look for differences that are both statistically and substantively different and, for those found, offer possible explanations.

It is important to note that the age and sex data from the ACS and the ASEC are controlled to intercensal population estimates of sex, age, race, and Hispanic origin. The intercensal population estimates are used as a method to calibrate both surveys, thereby creating sex, age, race, and Hispanic origin distributions that are reflective of a demographic framework of population accounting. Under this framework, the intercensal population estimates are based on the previous decennial census enumeration and the incorporation of data representing components of population change – births, deaths, and net migration. While this paper contains a description of differences between the two surveys in data collection methodology, item nonresponse, and data editing and imputation, the fact that the age and sex data are controlled to intercensal population estimates for the universe utilized by each survey, exerts the strongest influence on the resulting distributions. Thus, the overall result is that both the ACS and the ASEC have similar age and sex distributions, although there are a few statistically significant differences.

METHODOLOGY

The table included in this report compares the most commonly tabulated data on age and sex from the ACS and the ASEC. Comparisons consist primarily of percentage-point differences between the two distributions. The table displays the ACS and the ASEC estimates, the margin of error from which the 90-percent confidence intervals of the estimates can be derived, and the difference in the estimates. In the case of frequency distributions, the difference is calculated as the percent difference between the two estimates. In the case of relative frequency distributions, the difference is calculated as the percentage-point difference between the two estimates. An asterisk (*) denotes statistically significant differences. The remainder of this section examines the differences in methodology between the ACS and the ASEC.

Sample Frame

The 2004 ACS surveyed a national sample of housing units, both occupied and vacant. Data were collected in a total of 1,240 counties out of the 3,141 counties in the United States. The sample is designed to provide estimates of housing and socio-economic characteristics for the nation, all states, most areas with a population of 250,000 or more, and selected areas of 65,000 or more.

The 2004 ASEC surveyed a national sample of housing units and noninstitutional group quarters in 754 primary sampling units. The sample is designed primarily to produce estimates of socio-economic characteristics, including labor force characteristics of the

civilian noninstitutionalized population 16 years of age and older for the nation and all states[1].

One difference between the two survey universes is that the 2004 ASEC, unlike the 2004 ACS, includes individuals living in noninstitutional group quarters (e.g. college and university dormitories, workers' dormitories, group homes, etc.).

Sample Size and Mode of Data Collection

The 2004 ACS interviewed a total of 534,383 households. Data were collected continuously throughout the year using a combination of mail-out/mail-back questionnaires, Computer-Assisted Telephone Interviewing (CATI), and Computer-Assisted Personal Interviewing (CAPI). Each month a unique national sample of addresses receives an ACS questionnaire. Addresses from this group that do not respond are telephoned during the second month of collection if a phone number for the address is available, and personal visits are conducted during the third and last month of data collection for a subsample of the remaining nonresponding units of this group. The 2004 ACS achieved an overall survey response rate, calculated as the initially weighted estimate of interviews divided by the initially weighted estimate of cases eligible to be interviewed, of 93.1 percent[2].

[1] For a detailed explanation of the ASEC sampling frame, see the following Internet site: <www.census.gov/prod/2002pubs/tp63rv.pdf>.

[2] As a result of a reduction in funding in 2004, ACS dropped the telephone and personal visit followup operations for the January 2004 panel, thus only allowing mail respondents to contribute to the overall response for that panel. Dropping the nonresponse followup operations for that single panel month reduced the annual response rate by about four percentage points. If we exclude the January panel from the calculation, the annual response rate rises to 97.3%

The 2004 ASEC contained interviews from about 78,000 housing units and noninstitutional group quarters. The number of noninstitutional group quarters included in the sample was very small (59). The ASEC interviews were collected over a three-month period in February, March, and April 2004 as a supplement to the basic monthly Current Population Survey conducted during those months, with most of the data collected in March. All of the ASEC data is collected via Computer-Assisted Telephone and Personal Interviews (CATI/CAPI), with interviews conducted during one week each month. The response rate for the 2004 ASEC was 91.8 percent.

Both the ACS and the ASEC employ experienced, permanent interviewers for CATI and CAPI data collection.

Residence Rules

The ACS and the ASEC employ different residence rules to determine which individuals in a household are eligible for interview. The ACS uses the concept of current residence while the ASEC uses a version of usual residence. This difference may contribute to variation in the universes for which the social characteristics are described.

The ACS collects interviews from everyone in the housing unit on the day of interview who is living or staying there for more than two months, regardless of whether or not they maintain a usual residence elsewhere. If a person who usually lives in the housing unit is away for more than two months at the time of the survey contact, he or she is not considered to be a current resident of that unit. This rule recognizes that people can have

more than one place where they live or stay over the course of a year, and these people may affect the estimates of the characteristics of the population for some areas.

The ASEC interviews everyone staying in the housing unit or noninstitutional group quarters at the time of the interview who considers the housing unit or noninstitutional group quarters as their usual residence or who has no usual residence elsewhere. The ASEC also includes temporarily absent individuals who consider the housing unit or noninstitutional group quarters to be their usual residence. The universe for this analysis includes people in the civilian noninstitutionalized population.

The different residence rules create one notable difference between the universes of these two surveys. Because the 2004 ACS excluded group quarters from the sample frame and interviewed individuals at their current residence, college students living in residence halls and dormitories were not included in the ACS universe. In contrast, the ASEC interviewers are instructed to include as household members any college students who are temporarily absent from the household, including those who are currently residing in college residence halls and dormitories. The resulting ASEC sample universe was expected to include more college-age students than the ACS sample universe, because the ASEC includes those who live most of the year away from their parent's home in a college or university residence hall or dormitory. This difference could cause estimates for certain segments of the age distribution to vary between the two surveys.

Question Wording and Reference Periods

There are slight differences in presentation and wording of the sex, date of birth, and age questions between the ACS and the ASEC. Additionally, reference periods for these topics in the ACS and the ASEC will be discussed.

Sex.

The 2004 ACS collected data on sex on the mail out/mail back questionnaire by asking the following question of all respondents:

What is this person's sex?

☐ *Male*

☐ *Female*

The 2004 ACS CATI/CAPI interviewers were instructed to record the respondent's and household members' sex by inference. This means that ACS interviewers, if possible, recorded sex using their judgment, which could be based on how a respondent appeared visually, how a respondent's voice sounded over the telephone, or by a respondent's answers to other questionnaire items such as name and/or relationship. The interviewers received the instruction below and asked the following question only if they felt that it was necessary:

If not obvious, ask

(Is <Name>/Are you) male or female?

○ *1. Male*

○ *2. Female*

Similar to the ACS CATI/CAPI interviewers, the ASEC interviewers received the instruction below when collecting the respondent's and household members' sex:

Ask only if necessary

What is (name's/your) sex?

<1> Male

<2> Female

ASEC interviewers, if possible, also recorded sex using their judgment, which could be based on how a respondent appeared visually, how a respondent's voice sounded over the telephone, or by a respondent's answers to other questionnaire items such as name and/or relationship.

Since sex is a basic demographic item, and is not complicated for respondents to report, slight differences in the wording of the sex question between the ACS and the ASEC most likely cause little difference in the estimates from the two surveys. For both surveys, CATI and CAPI interviewers ask the sex question of the respondent only when they deem it necessary, which means that the sex data are collected in an inconsistent manner and most respondents do not have the opportunity to self-report a response.

The ACS and the ASEC sex data have different reference periods. After data collection and during weighting, the ACS data are controlled to intercensal estimates of county population by age, sex, race, and Hispanic origin as of July 1. The ASEC data are also controlled to intercensal estimates, but use estimates of state population by age, sex, race,

and Hispanic origin as of March 1. This is because the ASEC data are collected primarily in March, but also in February and April.

Date of Birth and Age

ACS

In 2004, the ACS mail out/mail back questionnaire collected date of birth and age by asking the following question of all respondents:

What is this person's age and what is this person's date of birth?

Print numbers in boxes.

Age (in years)

Month *Day* *Year of birth*

For the 2004 ACS CATI/CAPI interviews, date of birth and age were collected in the following manner:

What is (<Name>'s/your) date of birth?

- ○ *Enter birth month.*
- ○ *Enter birth day.*
- ○ *Enter birth year.*

Since the ACS CATI/CAPI utilizes an automated instrument, date of birth is collected first. When date of birth is provided, the automated instrument calculates the age as of the interview month and confirms the age with the respondent, by asking:

That makes (<Name>/you) (<Calculated age>) this month/less than one year old. Is this correct?

As long as a valid month of birth and year of birth, or at a minimum a valid year of birth, were provided, an age could be calculated, which was then confirmed with the respondent. When only a valid year of birth was provided, respondents were given two choices when asked to confirm a calculated age. They were asked whether one or none of the following were correct: a calculated age assuming the person had not had a birthday in the current year or a calculated age assuming the person already had a birthday in the current year.

If the respondent indicated that the calculated age was incorrect, the interviewer rechecked the date of birth with the respondent to ensure that it had been entered correctly. When the respondent indicated that the calculated age was still incorrect or a full or partial date of birth was unknown or refused, the ACS CATI/CAPI interviewers asked respondents to guess:

What is your best estimate of (<Name>'s/your) age?

If respondents still cannot provide an age, the ACS CATI/CAPI interviewers try to identify an age range:

(Is <Name>/Are you):

- ○ *1. Less than 3 years old*
- ○ *2. 3 or 4 years old*
- ○ *3. 5 to 14 years old*
- ○ 4. *15 years old and older*

The ACS uses multiple methods to collect an age for each member of the household because age is critical for determining the correct paths of questions within the ACS automated data collection instrument. Values of 0 to 115 were accepted for age in the 2004 ACS.

ASEC

Similarly, for the 2004 ASEC CATI/CAPI, date of birth and age were collected by asking the following question of all respondents:

What is (name's/your) date of birth?

Enter month ===>

Enter day ===>

Enter year ===>

Just as with the 2004 ACS, when date of birth is provided, the ASEC automated instrument calculated age and the interviewer confirmed it with the respondent in the following manner:

that would make (name/you) ((age)/approximately (age)/less than 1/over 98) (year/years) old. Is that correct?

<1> Yes, age is correct.

<2> No, age is not correct.

Similar to the 2004 ACS, the 2004 ASEC instrument can calculate an age, as long as a valid month of birth and year of birth were provided. If the respondent indicated that the calculated age was incorrect, the interviewer rechecked the date of birth with the respondent to ensure that it had been entered correctly. When the respondent indicated

that the calculated age was still incorrect or a full date of birth, or the minimum partial date of birth information, was unknown or refused, ASEC interviewers asked respondents to guess:

Even though you don't know (name's/your) exact birthdate, what is your best guess as to how old (you/he/she) (were/was) on (your/his/her) last birthday?

<099> 99 years or older

<0098> 0 to 98 years old

The 2004 ASEC also uses an additional question to collect age information by asking respondents who still cannot provide an age the following:

ASK IF NECESSARY:

(Are/Is) (you/he/she) under 15?

<1> Yes

<2> No

For the 2004 ASEC, just as for the 2004 ACS, it was critical to collect an age, because it determined the paths of questions interviewers ask of respondents. Single years of age were accepted in the range of 0 to 98. Any age reported above 98 was captured in the category "99 years and older" in the 2004 ASEC.

The ACS and the ASEC age data have different reference periods. Similar to sex data, after data collection and during weighting, the ACS data are controlled to intercensal estimates of county population by age, sex, race, and Hispanic origin as of July 1. The ASEC data are also controlled to intercensal estimates, but use estimates of state

population by age, sex, race, and Hispanic origin as of March 1. This is because the ASEC data are collected primarily in March, but also in February and April.

ITEM NONRESPONSE

Item nonresponse occurs when an individual does not provide complete and usable information for a data item. Item allocation rates are often used as a measure of the level of item nonresponse. These rates are computed as the ratio of the number of eligible people for whom a value was allocated during the editing process for a specific item to the number of people eligible to have responded to that item.

Sex

Sex is one of the best-reported demographic items from the two surveys and their allocation rates were similar. For the 2004 ACS, the sex allocation rate was 0.2 percent. For the ASEC, the sex allocation rate was 0.1 percent.

Age

The age allocation rates were very different for the two surveys. For the 2004 ACS, the age allocation rate was 0.9 percent. For the ASEC, the age allocation rate was 2.5 percent. The reason for the dissimilarity between the allocation rates for age is unclear, however there were a few differences between the two surveys that could contribute to the higher ASEC age allocation rate.

The ACS allocated age if an age value was left blank, out of range, or failed a set of consistency checks. For example, consistency checks ensured that the age of a reference

person was 15 or older. Additionally, the ACS consistency checks ensured that the difference in ages for people in the household was logical by applying a host of generational checks utilizing relationship responses (i.e. ensuring that the age difference between parents and children are logical, etc.).

There were similarities, as well as a difference, when comparing ASEC and ACS age allocation rules. Similar to the ACS, the ASEC used consistency checks to ensure that the reference person and their spouse and/or parent had ages that were 15 or older. If these checks failed, an age was allocated. The ASEC differs from the ACS in that the ASEC also allocated an age longitudinally when age was unknown or refused, yet age data for that person existed from a previous ASEC interview. Thus, the previously collected age was adjusted to reflect what would be the person's age as of the current interview.

There are other aspects of the ACS and the ASEC that may contribute to the difference in the age allocation rates. The ASEC, which is a voluntary CAPI/CATI survey, had to allocate age if it could not be collected (age or age estimate) during the current interview. In contrast, the ACS is a mandatory survey that makes an additional effort to collect data missing from the mail-out/mail-back data collection, which is the ACS Failed Edit Follow-up operation. In this operation, when processed mail-out/mail-back questionnaires are determined to have either coverage failure or content failure, an attempt is made to contact that household to obtain missing information or to resolve inconsistencies. Coverage failure occurs when large households cannot provide detailed information on the questionnaire that is designed for a 5-person household or the reported

number of people in the household differs from the number of people for whom responses are reported. Content failure occurs when responses to two critical demographic items (such as age) or other required items have not been provided. Results from the follow-up for this mandatory survey could improve ACS' ability to collect age data.

DATA EDITING AND IMPUTATION PROCEDURES

The ACS and the ASEC edit and imputation rules are designed to ensure that the final edited data are as consistent and complete as possible. These rules are used to identify and account for missing, incomplete, and contradictory responses. In each case where a problem is detected consistent, pre-established edit rules govern its resolution.

The ACS and the ASEC employ two principal imputation methods, relational imputation and hot deck allocation. Relational imputation assigns values for blank or inconsistent responses on the basis of other characteristics on the person's record or within the household. Hot deck allocation supplies responses for missing or inconsistent data items from the responses of similar housing units or people who did respond to the survey.

In terms of editing and allocating sex and age, both surveys use hot deck allocation. The ACS makes extensive use of relational imputation, while the ASEC does so minimally.

Sex

For both surveys, when the edit and allocation process begins and a response to the sex item is missing, initial efforts use other information from the person's record to assign a

sex. The ACS attempts to assign sex based upon the person's first name. For example, if the person with no response to the sex item is named "Kimberly" there is a strong likelihood that this person is female, thus that sex would be assigned. Also, the ACS utilizes the response to the fertility item for those 15 years and older to assign a missing sex. So, for example, when fertility information is provided, "female" would be the assigned sex when it is missing.

The ASEC relies upon two methods that utilize a person's additional information to provide a sex value when it has not been reported, as well as conducts limited consistency checks for sex when it has been reported for spouses. First, if the household was previously in the sample and a sex for that person was provided during the previous interview, that sex is allocated longitudinally for the current interview. A second method to assign a sex when it has not been provided utilizes military status. For example, if sex is not available from a previous interview and the person without a response to the sex item is not a civilian, there is a strong likelihood that person is male, thus that sex would be assigned. When the same sex has been reported for the householder's spouse, the spouse's sex value is blanked and the opposite sex is assigned.

The ACS and ASEC may also utilize information provided for others in the same household to edit sex or to provide a value for sex when it has not been reported. For example, for both surveys, the reported sex of a person is used to assign the opposite sex to their spouse when their sex was not reported.

Finally, when efforts to utilize a person's additional information or the information of others in the household to assign a missing sex fails, a sex value is allocated from a hot deck. For the ACS, the allocation of sex from a hot deck is based upon whether or not age is reported. When age is reported, that information is used in conjunction with knowledge of stable demographic trends to allocate sex. Since early in life males outnumber females and, as aging occurs, that trend reverses so females outnumber males, information on age is used to allocate sex. The ASEC, however, allocates sex values from a hot deck on a rotational basis, regardless of age. Thus, ASEC alternates between allocating the values of male and female for persons that need a sex value retrieved from the hot deck.

Age

When the editing process for age begins, the ACS makes multiple efforts to check the consistency of the reported age with other information from the person's record, while the ASEC does not. The ACS generally begins this process by using the reported date of birth (if provided) to calculate the person's age as of the interview date. The calculated age is then compared with the reported age for consistency. If the two ages differ by no more than 3 years, they are considered to be consistent and the age calculated from date of birth is retained as the final age. If the two ages differ by more than 3 years, they are considered to be inconsistent. Inconsistencies can also occur between the reported age and responses to other selected items. For example, a person with an adult relationship value, such as grandparent, and a reported age of less than 15 years old would be inconsistent.

When inconsistencies between the calculated age and the reported age or between the reported age and responses to other selected items are discovered, additional information from the person's record is used to choose which age to retain. That person's responses to the educational attainment, school enrollment, labor force, marital status, and relationship items are used to determine the age that should be retained. For example, if a person's reported age is greater than or equal to age 15 and the computed age is less than age 15, and the person's relationship is not a reference person, spouse, parent, or unmarried partner, the computed age is accepted.

When inconsistencies still exist between a person's reported age and computed age after efforts to utilize additional information from that person's record have been exhausted, the ACS utilizes information from others in the same household to determine which age to retain. This involves comparing the person's reported age and his or her computed age with an acceptable reported age of someone else in the household – based upon relationship and general assumptions about age differences between those in a household. For example, if a person with inconsistent age information has the relationship of a child and the reference person has an acceptable reported age, comparisons are made. Specifically, if the reference person's age minus 25 is closer to the child's reported age, the child's reported age is retained. Conversely, if the reference person's age minus 25 is closer to the child's calculated age, the child's calculated age is retained. The ACS also conducts a series of additional consistency checks of a person's age by making comparisons based upon logical expectations and assumptions about members of

households. For example, when a reference person has an age that is between 0 and 14 years old, that age is blanked and allocated, since reference persons must be 15 years or older.

The ASEC, which does not utilize any of the relational imputation procedures described above for the ACS, makes one attempt to allocate a missing age using additional information for that person. If the household was previously in the sample, and an age for that person was provided during the previous interview, the age is allocated longitudinally by adjusting it to reflect the person's expected age as of the current interview date.

Finally, when efforts to utilize a person's additional information or the information of others in the household to allocate a missing age fail, an age is allocated from a hot deck. For the ACS, multiple hot decks exist to allocate a missing age. The use of a specific hot deck is primarily based upon the relationship and associated information of the person who needs an age allocated. The hot decks are stratified by key person and household information in order to improve the allocation of age from a pool of donors who have selected characteristics in common with the person whose age is missing. For example, various hot decks are designed to allocate an age to a husband based upon the wife's age and the presence of children; to allocate an age to a reference person based upon their household type, marital status, and the presence of children; and to allocate an age when relationship is unknown.

The ASEC, in contrast, allocates age values from a hot deck on a rotational basis. The age range collected during the interview, which is a last effort to obtain some information about the person's age, stratifies this hot deck. For example, as persons in the age range of 25 to 49 need an actual age value allocated, the values retrieved from the hot deck would rotate through the ages of 25 to 49. Thus, a person in the age range 25 to 49 that needed an age value could have age 25 allocated, while the next person in the age range 25 to 49 that needed an age value could have age 26 allocated, and so on.

CONTROLS AND WEIGHTING

Some differences in the selection of controls and the calculation of weights between the two surveys may lead to differences in estimates. The ACS and the ASEC are both weighted to account for the probability of selection and housing unit nonresponse.

After the initial weighting, data from the ACS and the ASEC are both controlled to independent population estimates of sex, age, race, and Hispanic origin. For the ACS, data that have been initially weighted are divided into 13 age categories crossed by sex within selected race and Hispanic origin groups, creating a number of cells within a weighting matrix. These cells may be collapsed further in order to meet pre-established thresholds and they are then utilized in a raking process to ensure that the ACS data agree with aggregate intercensal population estimates. For the ASEC, data that have been initially weighted are divided into single-year age categories crossed by sex within selected race and Hispanic origin groups, creating a number of cells within a weighting

matrix. These cells are then aggregated to be used as population controls in the weighting process of the ASEC data.

Differences in the way the population controls are applied in each survey may lead to differences in the age and sex estimates. For example, data from the 2004 ACS are controlled, at the county level, to independent estimates of the household population and housing units in July 2004. The 2004 ASEC is controlled to independent national estimates of the civilian noninstitutionalized population in March 2004. In addition, the ACS presents the responses over a 12-month period, while the ASEC shows the responses of people for the February-April time period, although the population is controlled to March estimates. Because the ACS controls to both the total household population and the total number of housing units, the ACS files contain both person weights and housing-unit weights. The ASEC does not control to the total number of housing units and, thus, the ASEC files do not contain an independent housing-unit weight but instead use the weight of the householder as the weight of the housing unit.

RESULTS

Table 1 shows that data from both surveys in 2004 produced expected general demographic patterns for the distributions of age and sex at the national level. Across the two surveys, proportions for most age categories are not statistically different, but a few are. These general patterns are largely due to age and sex data from both surveys being controlled to independent population estimates.

Table 1 also indicates that while generally differences are not statistically significant between the 2004 ACS and the 2004 ASEC age relative frequency distributions, some notable differences are statistically significant at the national level. At one end of the relative frequency distribution for age, the 2004 ACS estimates of the percentages aged 15 to 19 (6.7 percent) and aged 20 to 24 (6.8 percent) were less than the 2004 ASEC estimates of the percentages aged 15 to 19 (7.0 percent) and aged 20 to 24 (7.1 percent)[3]. For the older ages, the 2004 ACS estimate of the percentage aged 60 to 64 (4.4 percent) was larger than the percentage estimated by the 2004 ASEC (4.2 percent). Conversely, the 2004 ACS estimate of the percentage aged 75 to 79 (2.5 percent) was less than the percentage estimated by the 2004 ASEC (2.6 percent).

These statistically significant differences are likely the result of the different universes utilized by each survey (the 2004 ASEC includes the civilian, noninstitutionalized group quarters population, while the 2004 ACS did not). Thus, the ASEC includes the population living in college residence halls and dormitories that tends to have high proportions of students in the age range of 18-to-24 years old. At the other end of the distribution, statistical differences exist, although very small, for the 60-to-64 years old and the 75-to- 79 years old categories. The reason for these differences is not clear.

SUMMARY

Data from the 2004 ACS on age and sex are generally consistent with age and sex data from the 2004 ASEC. This consistency is due largely to the fact that age and sex data

[3] For the 2004 ASEC, the percentages of the civilian, noninstitutionalized population that were aged 15 to 19 (7.0 percent) and aged 20 to 24 (7.1 percent) were not statistically different.

from both surveys are controlled to independent population estimates. However, there are some statistically significant differences in proportions of sex and specific age categories, namely estimates for those aged 15 to 19, aged 20 to 24, aged 60 to 64, and aged 75 to 79. While definitive explanations for all of these differences are lacking, some may be explained by the differences in the universe of the 2004 ACS (household population) and the 2004 ASEC (civilian, noninstitutionalized population).