Chapter 15.
Improving Data Quality by Reducing Nonsampling Error

15.1 OVERVIEW

As with all surveys, the quality of the American Community Survey (ACS) data reflects how well the data collection procedures address potential sources of nonsampling error, including coverage error, nonresponse and measurement errors, and errors that may arise during data capture and processing. Chapters 4 and 11 provide information regarding the steps the ACS takes to reduce sampling error while still managing costs.

There are four primary sources of nonsampling error (Groves, 1989):

- Coverage Error. The failure to give some units in the target population any chance of selection into the sample, or giving units more than one chance of selection.

- Nonresponse Error. The failure to collect data from all units in the sample.

- Measurement Error. The inaccuracy in responses recorded on survey instruments, arising from:

  - The effects of interviewers on the respondents' answers to survey questions.

  - Respondents' inability to answer questions, lack of requisite effort to obtain the correct answer, or other psychological or cognitive factors.

  - Faulty wording of survey questions.

  - Data collection mode effects.

- Processing Error. Errors introduced after the data are collected, including:

  - Data capture errors.

  - Errors arising during coding and classification of data.

  - Errors arising during editing and item imputation of data.

This chapter identifies the operations and procedures designed to reduce these sources of non-sampling error and thus improve the quality of the data. It also includes information about ACS Quality Measures, which provide data users an indication of the potential for nonsampling error. The ACS releases the survey estimates, as well as the Quality Measures, at the same time each year, so that users can consider data quality in conjunction with the survey estimates. The ACS Quality Measures are available on the American FactFinder (AFF) Web site <http://factfinder .census.gov/home/saff/main.html?lang=en> for ACS data beginning with 2007 (and all multiyear data). The Quality Measures for years 2000 to 2006 are located on the ACS Quality Measures Web site <http://www.census.gov/acs/www/UseData/sse/index.htm>.

15.2 COVERAGE ERROR

All surveys experience some degree of coverage error. It can take the form of under-coverage or over-coverage. Under-coverage occurs when units in the target population do not have a chance of selection into the sample; for example, addresses not listed on the frame, or people erroneously excluded from a household roster. Over-coverage occurs when units or people have multiple chances of selection; for example, addresses listed more than once on the frame, or people included on a household roster at two different sampled addresses. In general, coverage error can affect survey estimates if the characteristics of the individuals or units excluded or included in

error differ from the characteristics of those correctly listed in the frame. Over- and under-coverage sometimes can be adjusted as part of the poststratification process, that is, adjusting weights to independent population control totals. Chapter 11 provides more details regarding the ACS weighting process.

The ACS uses the Master Address File (MAF) as its sampling frame, and includes several procedures for reducing coverage error in the MAF. These procedures are described below. Chapter 3 provides further details.

- Twice a year, the U.S. Census Bureau receives the U.S. Postal Service (USPS) Delivery Sequence File (DSF) that includes the addresses including a house number and street name rather than a rural route or post-office box. This file is used to update the city-style addresses on the MAF.

- The ACS nonresponse follow-up operation provides ongoing address and geography updates.

- The MAF includes address updates from special census operations.

- The Community Address Updating System (CAUS) can provide address updates (as a counterpart to the DSF updates) that cover predominately rural areas where city-style addresses generally are not used for mail delivery. CAUS was put on hold in late 2006 and is expected to be back in 2010. CAUS was put on hold because of the address canvassing operation for the 2010 Census.

The ACS Quality Measures contain housing- and person-level coverage rates (as indicators of the potential for coverage error). The coverage rates are located on the AFF for ACS data for 2007 and beyond (including all multiyear data). Coverage rates for prior years (2000 to 2006) are available on the ACS Quality Measures Web site.

Coverage rates for the total resident population are calculated by sex at the national, state, and Puerto Rico geographies, and at the national level only for Hispanics and non-Hispanics crossed by the five major race categories: White, Black, American Indian and Alaska Native, Asian, and Native Hawaiian and Other Pacific Islander. The total resident population includes persons in both housing units (HUs) and group quarters (GQ). In addition, these measures include a coverage rate specific to the GQ population at the national level. Coverage rates for HUs are calculated at the national and state level, with the exception of Puerto Rico because independent HU estimates are not available.

The coverage rate is the ratio of the ACS population or housing estimate of an area or group to the independent estimate for that area or group, multiplied by 100. The Census Bureau uses independent data on housing, births, deaths, immigration, and other categories to produce official estimates of the population and HUs each year. The base for these independent estimates is the decennial census counts. The numerator in the coverage rates is weighted to reflect the probability of selection into the sample, subsampling for personal visit follow-up, and is adjusted for unit nonresponse. The weight used for this purpose does not include poststratification adjustments (weighting adjustments that make the weighted totals match the independent estimates), since the control totals serve as the basis for comparison for the coverage rates. The ACS corrects for potential over- or under-coverage by controlling to these official estimates on specific demographic characteristics and at specific levels of geography.

As the coverage rate for a particular subgroup drops below 100 percent (indicating under-coverage), the weights of its members are adjusted upward in the final weighting procedure to reach the independent estimate. If the rate is greater than 100 percent (indicating over-coverage), the weights of its members are adjusted downward to match the independent estimates.

15.3 NONRESPONSE ERROR

There are two forms of nonresponse error: unit nonresponse and item nonresponse. Unit nonresponse results from the failure to obtain the minimum required data from an HU in the sample. Item nonresponse occurs when respondents do not report individual data items, or provide data considered invalid or inconsistent with other answers.

Surveys strive to increase both unit and item response to reduce the potential for bias introduced into survey estimates. Bias results from systematic differences between the nonrespondents and the respondents. Without data on the nonrespondents, surveys cannot easily measure differences between the two groups. The ACS reduces the potential for bias by reducing the amount of unit and item nonresponse through procedures and processes listed below.

- Response to the ACS is mandated by law, and information about the mandatory requirement to respond is provided in most materials and reinforced in any communication with respondents in all stages of data collection.

- The ACS survey operations include two stages of nonresponse follow-up: a computer-assisted telephone interview (CATI) follow-up for mail nonrespondents, and a computer-assisted personal interview (CAPI) follow-up for a sample of remaining nonrespondents and unmailable addresses cases.

- The mail operation implements a strategy suggested in research studies for obtaining a high mail response rate (Dillman, 1978): a prenotice letter, a message on the envelope of the questionnaire mailing package stating that the response is "required by law," a postcard reminder, and a second mailing for nonrespondents to the initial mailing.

- The mailing package includes a frequently asked questions (FAQ) motivational brochure explaining the survey, its importance, and its mandatory nature.

- The questionnaire design reflects accepted principles of respondent friendliness and navigation, making it easier for respondents to understand which items apply to them, as well as providing cues for a valid response at an item level (such as showing the format for reporting dates, or using a prefilled '0' to indicate reporting dollar amounts rounded to the nearest whole number). Similarly, the CATI and CAPI instruments direct interviewers to ask the appropriate questions.

- The questionnaire provides a toll-free telephone number for respondents who have questions about the ACS in general or who need help in completing the questionnaire.

- The ACS includes a telephone failed-edit follow-up (FEFU) interview with mail respondents who either failed to respond to specific critical questions, or who indicated a household size of six or more people. (The mail form allows data for only five people, so the FEFU operation collects data for any additional persons.)

- The ACS uses permanent professional interviewers trained in refusal conversion methods for CATI and CAPI.

- Survey operations include providing support in other languages: a Spanish paper questionnaire is available on request, and there is a Spanish CATI/CAPI instrument. Also, there are CATI and CAPI interviewers who speak Spanish or other languages as needed.

Unit Nonresponse

The Census Bureau presents survey response and nonresponse rates as part of the ACS Quality Measures. The survey response rate is the ratio of the units interviewed after data collection to the estimate of all units that were eligible to be interviewed. Data users can find survey response and nonresponse rates on the AFF for ACS data for 2007 and beyond (including multiyear estimates). The same rates for data years 2000 to 2006 are available on the ACS Quality Measures Web site. The ACS Quality Measures provide separate rates for HUs and GQ persons. For the HU response rate, the numerator includes all cases that were interviewed after mail, telephone, and personal visit follow-up. For the GQ person response rate, the numerator includes all interviewed persons after the personal visit. For both rates, the numerator includes completed interviews as well as partial interviews with adequate information for processing.

To accurately measure unit response, the ACS estimates the universe of cases eligible to be interviewed and the survey noninterviews. The estimate of the total number of eligible units becomes the denominator of the unit response rate.

The ACS Quality Measures also include the percentage of cases that did not respond to the survey by the reason for nonresponse. These reasons include refusal, unable to locate the sample unit, no one home during the data collection period, temporarily absent during the interview period, language problem, insufficient data (not enough data collected to consider it a response), and other (such as "sample address not accessible"; "death in the family"; or cases not followed up due to budget constraints, which last occurred in the winter of 2004). For the GQ rates, there are two additional reasons for noninterview: whole GQ refusal, and whole GQ other (such as unable to locate the GQ).

Item Nonresponse

The ACS Quality Measures provide information about item nonresponse. When respondents do not report individual data items, or provide data considered invalid or inconsistent with other answers, the Census Bureau imputes the necessary data. The imputation methods use either rules to determine acceptable answers (referred to as "assignment") or answers from similar people or HUs ("allocation"). Assignment involves logical imputation, in which a response to one question implies the value for a missing response to another question. For example, first name often can be used to assign a value to sex. Allocation involves using statistical procedures to impute for missing values. The ACS Quality Measures include summary allocation rates as a measure of the extent to which item nonresponse required imputation. Starting with the 2007 ACS data (including ACS multiyear data), the Quality Measures include only two item allocation rates: overall HU characteristic imputation rate and overall person characteristic imputation rate. These rates are available on the AFF at the national and state level. However, the ACS releases imputation tables on AFF that allow users to compute allocation rates for all published variables and all published geographies. Allocation rates for all published variables from 2000 to 2006 are available on the ACS Quality Measures Web site at the national and state level.

15.4 MEASUREMENT ERROR

All surveys encounter some form of measurement error, which is defined as the difference between the recorded answer and the true answer. Measurement error may occur in any mode of data collection and can be caused by vague or ambiguous questions easily misinterpreted by respondents; questions that respondents cannot answer or questions where respondents deliberately falsify answers for social desirability reasons (see Tourangeau and Yan (2007) for information on social desirability); or interviewer characteristics or actions such as the tone used in reading questions, the paraphrasing of questions, or leading respondents to certain answers.

The ACS minimizes measurement error in several ways, some of which also help to reduce nonresponse.

- As mandated in the Census Bureau Standard "Pretesting Questionnaires and Related Materials for Surveys and Censuses (Version 1.2)" <http://www.census.gov/quality/S11-0_v1.2 _Pretesting.pdf>, ACS pretests new or modified survey questions in all three modes before introducing them into the ACS.

- The ACS uses a questionnaire design that reflects accepted principles of respondent friendliness and navigation.

- The ACS mail questionnaire package includes a questionnaire instruction booklet that provides additional information on how to interpret and respond to specific questions.

- Respondents may call the toll-free telephone questionnaire assistance (TQA) line and speak with trained interviewers for answers to general ACS questions or questions regarding specific items.

- Differences among the mail, CATI, and CAPI questionnaires are reduced through questionnaire and instrument design methods that reflect the strengths and limitations of each mode of collection (for example, less complicated skip patterns on the mail questionnaire, breaking up questions with long or complicated response categories into separate questions for telephone administration, and including respondent flash cards for personal visit interviews).

- The CATI/CAPI instruments automate or direct skips, and show the interviewer only those questions appropriate for the person being interviewed.

- The CATI/CAPI instruments include functionality that facilitates valid responses. For example, the instruments check for values outside of the expected range to ensure that the reported answer reflects an appropriate response.

- Training for the permanent CATI and CAPI interviewing staff includes instruction on reading the questions as worded and answering respondent questions, and encompasses extensive role-playing opportunities. All interviewers receive a manual that explains each question in detail and provides detailed responses to questions often asked by respondents.

- Telephone interview supervisors and specially-trained staff monitor CATI interviews and provide feedback regarding verbatim reading of questions, recording of responses, interaction with respondents, and other issues.

- Field supervisors and specially-trained staff implement a quality reinterview program with CAPI respondents to minimize falsification of data.

- The CATI/CAPI instruments include a Spanish version, and bilingual interviewers provide language support in other languages.

Note that many of these methods are the same as those used to minimize nonresponse error. Methods that make it easier for the respondent to understand the questions also increase the chances that the individual will respond to the questionnaire.

15.5 PROCESSING ERROR

The final component of nonsampling error is processing error—error introduced in the postdata collection process of turning the responses into published data. For example, a processing error may occur in keying the data from the mail questionnaires. The miscoding of write-in responses, either clerically or by automated methods, is another example. The degree to which imputed data differ from the truth also reflects processing error—specifically imputation error. A number of practices are in place to control processing error (more details are discussed in Chapters 7 and 10). For example:

- Data capture of mail questionnaires includes a quality control procedure designed to ensure the accuracy of the final keyed data.

- Clerical coding includes a quality control procedure involving double-coding of a sample of the cases and adjudication by a third keyer.

- By design, automated coding systems rely on manual coding by clerical staff to address the most difficult or complicated responses.

- Procedures for selecting one interview or return from multiple returns for an address rely on a review of the quality of data derived from each response and the selection of the return with the most complete data.

- After completion of all three phases of data collection (mail, CATI, and CAPI), questionnaires with insufficient data do not continue in the survey processing, but instead receive a noninterview code and are accounted for in the weighting process.

- Edit and imputation rules reflect the combined efforts and knowledge of subject matter experts, as well as experts in processing, and include evaluation and subsequent improvements as the survey continues to progress.

- Subject matter and survey experts complete an extensive review of the data and tables, comparing results with previous years' data and other data sources.

15.6 REFERENCES

Biemer, P., and L. Lyberg. (2003) *Introduction to Survey Quality*, Hoboken, NJ: John Wiley and Sons.

Dillman, D. (1978) *Mail and Telephone Surveys: The Total Design Method*, New York: John Wiley and Sons.

Groves, R. M. (1989) *Survey Errors and Survey Costs*, New York: John Wiley and Sons.

Groves, R. M., M. P. Couper, F. J. Fowler, J. M. Lepkowski, E. Singer, and R. Tourangeau. (2004) *Survey Methodology*, Hoboken, NJ: John Wiley and Sons.

Tourangeau, R., and T. Yan. (2007) "Sensitive questions in surveys," *Psychological Bulletin*, 133(5): 859–883.

U.S. Census Bureau (2002) "Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: May 2002, Report 2: Demonstrating Survey Quality," Washington, DC.

U.S. Census Bureau (2004) "Meeting 21st Century Demographic Data Needs—Implementing the American Community Survey: Report 7: Comparing Quality Measures: The American Community Survey's Three-Year Averages and Census 2000's Long Form Sample Estimates," Washington, DC.

U.S. Census Bureau (2006) "Census Bureau Standard: Pretesting Questionnaires and Related Materials for Surveys and Censuses Version 1.2," Washington, DC.