Applied Geography 46 (2014) 147–157

Contents lists available at ScienceDirect

Applied Geography

journal homepage: www.elsevier.com/locate/apgeog

Patterns and causes of uncertainty in the American Community Survey

Seth E. Spielman [a,b,*], David Folch [b], Nicholas Nagle [c]

[a] Geography Department, University of Colorado, 110 Guggenheim Hall, Box 260 UCB, Boulder, CO 80309, USA
[b] Institute of Behavioral Science, University of Colorado, 110 Guggenheim Hall, Box 260 UCB, Boulder, CO 80309, USA
[c] Department of Geography, University of Tennessee, USA

Keywords:
American Community Survey
Census tract
Small area estimation
Margin of error
Uncertainty

A B S T R A C T

In 2010 the American Community Survey (ACS) replaced the long form of the United States decennial census. The ACS is now the principal source of high-resolution geographic information about the U.S. population. The margins of error on ACS census tract-level data are on average 75 percent larger than those of the corresponding 2000 long-form estimate. The practical implications of this increase is that data are sometimes so imprecise that they are difficult to use. This paper explains why the ACS tract and block group estimates have large margins of error. Statistical concepts are explained in plain English. ACS margins of error are attributed to specific methodological decisions made by the Census Bureau. These decisions are best seen as compromises that attempt to balance financial constraints against concerns about data quality, timeliness, and geographic precision. In addition, demographic and geographic patterns in ACS data quality are identified. These patterns are associated with demographic composition of census tracts. Understanding the fundamental causes of uncertainty in the survey suggests a number of geographic strategies for improving the usability and quality ACS.

Introduction

In 2010 the American Community Survey (ACS) replaced the long form of the United States decennial census as the principal source of high-resolution geographic information about the U.S. population. The ACS fundamentally changed the way data about American communities are collected and produced. The long form of the decennial census was a large-sample, low-frequency national survey; the ACS is a high-frequency survey, constantly measuring the American population using small monthly samples. One of the primary challenges for users of the ACS is that the margins of error are on average 75 percent larger than those of the corresponding 2000 long-form estimate (Alexander, 2002; Starsinic, 2005). The practical implications of this increase are that users often face data like those in Table 1, which shows the ACS median income estimates for African American households for a contiguous group of census tracts in Denver, Colorado. Income estimates range from around $21,000 to $60,000 (American Factfinder website accessed 7/15/2013). Without taking account of the margin of error, it would seem that Tract 41.06 had the highest income, however, when one accounts for the margin of error, the situation is much less clear – Tract 41.06 may be either the wealthiest or the poorest tract in the group.

Some degree of uncertainty is inherent in surveys like the ACS, however the amount of uncertainty in the ACS has far exceeded the United States Census Bureau's (USCB hereinafter) expectations. Initial expectations were that the amount of uncertainty (margin of error) in the ACS would be 33 percent greater than the decennial census long form (Navarro, 2012), This loss in precision was justified by the increase in timeliness of ACS estimates which are released annually compared to the once a decade long form. This tradeoff prompted Macdonald (2006) to call the ACS a "warm" (current) but "fuzzy" (imprecise) source of data. Unfortunately, those early expectations were too optimistic, the actual uncertainty in the ACS is much more than 33 percent greater than the census long form. While there are clear advantages to working with fresh data, the ACS margins of error are so large that for many variables at the census tract and block group scales the estimates fail to meet even the loosest standards of data quality.

The ACS is the primary national source of geographically and demographically detailed information about the American population. It is an essential resource for cartographers, geographers, or anyone interested in understanding neighborhood scale social and economic patterns. Savvy use of the ACS requires understanding the nature of the uncertainty in the ACS and its causes- the purpose

* Corresponding author. Geography Department, University of Colorado, 110 Guggenheim Hall, Box 260 UCB, Boulder, CO 80309, USA. Tel.: +1 303 492 4877.
E-mail address: seth.spielman@colorado.edu (S.E. Spielman).

0143-6228/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apgeog.2013.11.002

Table 1
ACS estimates of African-American median household income in a group of contiguous tracts in Denver County, Colorado.

Tract number	African-American median household income	Margin of error
Census Tract 41.01	$28,864	$8650
Census Tract 41.02	$21,021	$4458
Census Tract 41.03	$43,021	$14,612
Census Tract 41.04	$36,092	$3685
Census Tract 41.06	$60,592	$68,846

of this paper is to provide users with a readable account of the causes of uncertainty in the ACS and to suggest some potential solutions to the problem(s). This paper assumes readers have no prior background in survey statistics, however a basic familiarity with survey methods is necessary to understand uncertainty in the ACS, so this paper explains basic survey-statistical concepts in plain English (Section Uncertainty in surveys). Section The construction of the ACS, provides a high-level overview of the methods underlying the ACS and discusses how these methods contribute to uncertainty in ACS data. Section Exogenous sources of uncertainty in the ACS discusses how factors beyond the USCB's control contribute to uncertainty in ACS data. Finally, some potential solutions to the ACS's problems will be offered in Section Geographic strategies for improving the ACS. It is important to note that the causes of uncertainty in the ACS are not entirely clear, a full enumeration of the causes of uncertainty in the ACS would run many hundreds of pages. While not a complete picture of the causes and nature of uncertainty in the ACS this article provides a detailed framework and vocabulary for understanding uncertainty and suggests geographic strategies for dealing with it.

Uncertainty in surveys

Like the decennial long form before it, the ACS is a sample survey. Unlike complete enumerations,[1] sample surveys do not perfectly measure the characteristics of the population—two samples from the same population will yield different estimates. This sample-to-sample variability creates some uncertainty about a population's true characteristics, therefore survey-based estimates are usually accompanied by a margin of error. While it was not commonly acknowledged, even the decennial census long form data came with instructions for estimating margins of error. In the ACS, the margin of error for a given variable expresses a range of values around the estimate within which the true value is expected to lie. The margin of error reflects the variability that could be expected if the survey were repeated with a different random sample of the same population. This variability is referred to as *sampling error* and is measured as *standard error* (SE). Calculating standard errors for a complex survey like the ACS is not a trivial task, the USCB uses a simulation procedure called Successive Differences Replication to produce variance estimates (Fay & Train, 1995; Judkins, 1990; Wolter, 1984). The margins of error reported by the USCB with the ACS estimates are simply 1.645 times the standard error.

Sampling error has two main causes. The first is the sample size – the larger the sample the smaller the standard error, intuitively more data about a population leads to less uncertainty about its true characteristics. The second main cause of sampling error is heterogeneity in the population being measured (Rao, 2003). Consider two jars of U.S. coins, one contains U.S. pennies and the other contains a variety of coins from all over the world. If one randomly selected 5 coins from each jar, and used the average of these 5 to estimate the average value of the coins in each jar, then there would be more uncertainty about the average value in the jar that contained a diverse mixture of coins. If one took repeated random samples of 5 coins from each jar the result would always be the same for the jar of pennies but it would vary substantially in the diverse jar, this variation would create uncertainty about the true average value.[2] While the ACS is much more complicated than pulling coins from a jar, this analogy helps to understand the standard error of ACS estimates. Census Tracts and block groups are like jars of coins. If a tract is like the jar of pennies, than the estimates will be more precise, whereas if a tract is like the jar of diverse coins, then the estimate will be less precise.

While the simple example is illustrative of important concepts it overlooks the central challenge in conducting surveys; many people who will be included in a sample will choose not to respond to the survey. While a group's odds of being included in the ACS sample are proportional to its population size, different groups of people have different probabilities of responding. Only 65% of the people contacted by the ACS actually complete the survey (in 2011, 2.13 million responses were collected from 3.27 million samples). Some groups are more likely to respond than others, this means that a response collected from a hard to count group is worth more than a response from an easy to count group. These differential response rates are controlled by *weighting* each response. In the ACS each completed survey is assigned a single weight through a complex procedure involving dozens of steps. The important point, as far as this paper is concerned, is that these weights are *estimated* and uncertainty about the appropriate weight to give each response is an important source of uncertainty in the published data.

The concepts of sampling error and weighting are central to understanding uncertainty in the ACS but they are not the entire story. Some of the factors affecting uncertainty in the ACS are the result of decisions and tradeoffs made by the USCB, whereas other factors affecting uncertainty are the result of circumstances beyond the control of the USCB. Section The construction of the ACS provides a high-level overview of the methods used to construct the ACS and discusses how these methodological choices affect uncertainty and Section Exogenous sources of uncertainty in the ACS discusses how factors beyond the control of the USCB further shape the quality of the survey.

The construction of the ACS

As the primary source for high-resolution, socio-spatial information about the U.S., the ACS has a profound impact on research and practice across the social sciences. It is extremely difficult to produce timely, detailed data for small geographic areas in a country as large and diverse as the United States. The data-quality problems in the ACS are the direct result of decades of innovation in national surveys. Four innovations in particular are combined in the ACS; the use of sampling, the provision of small-area estimates, the release of annual estimates, and the use of weighting to adjust the

[1] A complete enumeration of large and diverse population may not be possible. The decennial census while a complete enumeration in principle, has always under/over counted particular populations (Freedman, Pisani, & Purves, 2007).

[2] The USCB generally is not actually estimating the "average" value, they are estimating the "total" value of coins in the jar. Repeatedly grabbing five coins and computing the average will over many samples get you a very precise estimate of the average value, but it will give you no information on the total value. To get the total value, you need a good estimate of the average AND a good estimate of the total number of coins in the jar. The loss of cotemporaneous population controls caused by decoupling the ACS from the Decennial enumeration means that the census does not have information about the number of coins in the jar. This is discussed in Section The construction of the ACS.