

**OPEN ACCESS**

via Creative Commons 3.0

**RESEARCH**

## **Handling Data Quality Information of Survey Data in GIS: A Case of Using the American Community Survey Data**

David W. Wong<sup>a</sup> and Min Sun<sup>a</sup>  
George Mason University

### **ABSTRACT**

Many survey datasets have relatively high levels of error. The American Community Survey (ACS) data disseminated by the U.S. Census Bureau offer critical socio-demographic data of the U.S. and include a margin of error to indicate the reliability of an estimate for each census unit. Such feature-based reliability data should be incorporated when using ACS data in GIS such that users are informed of the data quality. This paper provides a brief review of the ACS data. Challenges to the use of ACS data quality information in GIS are discussed. The paper briefly reviews a set of mapping tools packaged as an ArcGIS extension to tackle some challenges in incorporating the data quality information in ACS data. These mapping functions can also be used to handle other survey spatial datasets with reliability information. The extension provides a foundation to build additional GIS tools to handle and analyze survey data.

**KEYWORDS:** Feature-based data quality information, decennial census, american community survey, margin of error, ArcGIS extension

### **INTRODUCTION**

As the U.S. Census Bureau is disseminating the data collected for the 2010 Census, census data users should know that no detailed socioeconomic data will be available from the 2010 Census datasets. Detailed socioeconomic and certain demographic data used to be gathered through the long-form questionnaires in previous censuses now have to be obtained from the American Community Survey (ACS), a continuous survey program that provides more timely snapshots of the American society. Such a change in data gathering strategy means that many socioeconomic and demographic analyses have to rely on ACS data in the future.

However, ACS data are gathered by surveys and therefore derived estimates have errors and may be unreliable. Previous long-form data were more reliable in general, but not without errors. ACS data should be examined carefully to determine if the errors in estimates are too large to be usable before they are adopted for a study. This should be a general practice for using any survey data inside and outside of GIS.

---

<sup>a</sup>George Mason University

**Corresponding Author:** David Wong, Geography and GeoInformation Science, George Mason University, Fairfax, VA 22030  
E-mail: [dwong2@gmu.edu](mailto:dwong2@gmu.edu)

To document the quality of ACS estimates, the U.S. Census Bureau provides a margin of error (MOE) accompanying an estimate for each variable in each census unit. This type of data quality information is in contrast to the data quality information for the entire geospatial database using summary measures, in accordance to the metadata standard for geospatial databases promulgated by the U.S. Federal Geographic Data Committee (1998). The MOEs in ACS data are in essence feature-based or per-feature quality information for attribute accuracy. This concept of feature-based data quality measures has been proposed sometime ago (Wong and Wu 1996), and its importance is supported by the fact that data quality varies spatially. Thus, feature-based quality measures have been used to document and depict the spatial variability of data quality (Zandbergen 2008).

Prototype systems have been proposed to handle per-feature quality information (e.g., Gan and Shi 2002, Qiu and Hunter 2002, Devillers et al. 2005). However, these systems focus mainly on managing and storing positional accuracy information. On the other hand, many cartographers and Geographic Information scientists have investigated different approaches and methods to visualize attribute uncertainty of geospatial datasets (e.g., Buttenfield 1993, MacEachren et al. 1998) and the literature is quite rich along this line of research (readers may refer to Sun and Wong 2010 for a concise review on this topic). Still, how per-feature attribute accuracy data can be incorporated into GIS databases and should be taken into account in spatial analysis and mapping have not been thoroughly addressed.

Although the U.S. Census Bureau began disseminating certain versions of ACS data in early 2000s, the use of ACS data in GIS and geographical research has been limited, partly due to some particular geographical characteristics of the data, and partly due to the difficulties in incorporating the accompanied data quality information in the analysis. Having per-feature attribute accuracy information for each attribute is an unusual data structure for GIS, a

minor technical issue that can be resolved. The major impediment lies on how the data reliability information may affect or be incorporated into spatial analysis, even as simple as mapping. Detailed socioeconomic and demographic analyses in the future have to rely on ACS data and the impediments to using ACS data have to be tackled. In principle, methods and frameworks that can successfully tackle some of these impediments to using ACS data can also be used to deal with other survey-type geospatial datasets with feature-based accuracy information.

In this paper, we will provide an overview of ACS data in the light of the 2010 Census data in the next section. In the third section, we will discuss difficulties in using ACS data in GIS and geographical research, with a special emphasis on data quality issues. In section four, we will present our effort in developing an ArcGIS extension to tackle some of the challenges to use ACS data in GIS. Finally, we briefly discuss the potential and limitations of the extension in dealing with ACS and other survey data in the conclusion section.

## OVERVIEW OF THE ACS DATA

The overview of ACS data provided here is not intended to be comprehensive, but to highlight certain characteristics of ACS data that may discourage the wider adoption of the data in GIS and geographical research. Although the conceptualization and testing of the ACS program started way before 2000, the first full implementation of the survey was in 2005 and produced the 2006 12-month data. ACS gathers detailed socioeconomic and demographic data from a sample of U.S. households every month. The survey questions are very similar to those in the decennial census long-form questionnaires used in the 2000 and previous censuses, but some selected questions ask for more detailed data. Part of the reason to introduce ACS is to provide more timely data instead of relying on the decennial long-form data which were available once every ten years.

While ACS provides more timely data by surveying the population more frequently, such higher data gathering frequency is traded off by the relatively small sample size for each period as compared to the long-form samples in previous censuses (MacDonald 2006).<sup>1</sup> Data gathered from the long form, which was received by approximately 1 in 6 households, were used to compile census data sets such as Summary Files (SFs) 3 and 4 to support many socioeconomic analyses. While these long-form data are not error-free, their quality and precision were usually not a major concern. Thus, numerous studies, including many spatial analytical studies, involving mapping or not, used these data without much concern of their reliability.

Data from ACS are collected from much smaller samples. In 2010, approximately 250,000 housing addresses in selected U.S. counties received the ACS questionnaires every month. Over the year, a total of about 3 million addresses, which was about 2.3 percent of all the addresses in 2005, were being sampled. Such a sample size is relatively large for national surveys in general, but is still quite small as compared to the sample sizes receiving the long forms in the past. In addition, the follow-up procedure for non-responding households in ACS is different from that used in collecting census long-forms. Therefore, estimates provided by ACS are subject to sampling errors much larger than those for data gathered from the census long forms. In general, areas with larger samples produce more reliable estimates with smaller margins of error (MOEs), and areas with smaller samples likely produce estimates with higher levels of variability reflected by the relatively large MOEs. The reliability levels of ACS estimates could vary tremendously across census units at the same census geography level and across different levels of census geography.

The MOEs provided in ACS are at 90% confidence level (CL), meaning that 90% of the chance that the true value of the estimate is found within the range of the estimate  $\pm$  MOE. MOEs at other confidence levels can be computed based upon the standard error, which can be derived algebraically from the MOE at 90% CL. Although

the Bureau does not officially regard the MOEs as the metadata for data quality, these MOEs are essentially per-feature data quality measures and therefore can be regarded as spatial metadata from a spatial database perspective (Wong and Wu 1996). Apparently, estimates with smaller MOEs are more reliable and desirable.

Several data products are generated from the ACS data. Different data products have to meet certain disclosure rules in order to protect the confidentiality of respondents. The 1-year estimates (12 months of collected data) are only available for areas with at least 65,000 people and the 3-year estimates (36 months) are available for areas with at least 20,000 people. ACS estimates are not available for smaller areal units not meeting these population thresholds. The 5-year estimates (60 months) are available for areas as small as census block groups, but many estimates for such small areas tend to be quite unreliable with relatively large MOEs. The Census Bureau also generates other ACS products, such as the Public Use Microdata Sample (PUMS) product, and the ACS tabulations by zip code tabulation areas (ZCTAs).

Some reports and studies have offered some guidelines for using ACS data. The publication by Citro and Kalton (2007) summarizes major findings about the use of ACS data from a study conducted by the *Panel on the Functionality and Usability of Data from the American Community Survey*, a panel convened under the National Statistics of the National Academies. In addition to provide many recommendations to the U.S. Census Bureau in terms of improving certain technical aspects of the ACS program, this publication also offers guidelines on how ACS data should be used for different groups of potential ACS users, including government officials and academic researchers. The U.S. Census Bureau also published a set of handbooks for different groups of stakeholders, including the general public users, politicians, federal and local government officials, businessmen, and researchers.<sup>2</sup> These handbooks cover certain general issues about the use of ACS data and highlight particular important issues relevant to the specific groups. A limited number of articles

have appeared in the literature about the use of ACS data in several social science disciplines (e.g., Gates 2006, Hough and Swanson 2006, Thomson et al. 2009). However, very few publications so far specifically address the use of ACS data in geography, spatial science and GIS (e.g., MacDonald 2006, Sun and Wong 2010).

## CHALLENGES TO THE USE OF ACS DATA IN GIS

Decennial census data have been used frequently in mapping exercises and projects (e.g., Allen and Turner 1997, Brewer and Suchan 2001, Peters and MacDonald 2004, Turner and Allen 1991, Xiao and Hawley 2010). Bringing these data into GIS for mapping and analysis does not involve overwhelming manipulation processes. Using ACS data in GIS and geographical research involves some difficult conceptual issues. For example, many ACS data products are averages over several years, but census boundaries also changed over the period. Then users should caution of using ACS data for very local scale analyses, which may be affected by inconsistent zonal boundaries. On the other hand, due to the particular data structure of ACS, the limited geographical coverage for certain data products, and the survey nature of the data, using ACS in GIS-based research has to deal with technical challenges associated with these characteristics of ACS data, particularly in interpreting the data with reliability information.

### Data Structure Issues

Bringing decennial census data into GIS requires several major steps, but overall, the process is quite straight-forward. Boundary data for census areal units in GIS-compatible formats have to be identified. These boundary data, which are derived from TIGER/Line files, are often provided by the U.S. Census Bureau.<sup>3</sup> Then, the appropriate census data table has to be modified into a format that it can be merged or joined with the corresponding boundary file. To map ACS data, the process is slightly more complicated. ACS spans between census years, and boundaries

of census units change occasionally. While census boundary files for different years are available, using ACS dataset for a specific time frame needs to identify the corresponding version of boundary data to be merged with the ACS tables.<sup>4</sup> The additional complication of using ACS data is to include the associated MOE values, which reflect the reliability of the estimates. Each column of estimates in an ACS table is followed by another column of MOEs, and both columns should be present in the GIS attribute table.

Related to this unusual data structure in ACS is the format of the estimate labels. For each ACS variable and the associated MOE, a field name and a field description are provided. The field names are usually codes without explicit meanings. They are useful for data manipulation and processing, but not understandable to most users. The field descriptions provide meanings of the estimates, but they are usually quite lengthy, possibly exceeding the length limit of field names in a GIS attribute table. Thus, the original field names need to be changed to much shorter ones.

Different from the decennial census data tables that fill with standard numeric inputs, ACS tables include several special symbols and characters to indicate various issues and conditions related to sampling and estimation. To accommodate all these special symbols and different value types in a GIS attribute table is not a trivial task. Table 1 shows selected census tract records for Virginia from the 2005-2009 (5-year) ACS variable of median age of total population. The median age of several tracts were reported as “-666666666”, indicating that estimates cannot be computed. The accompanied MOEs are represented by “\*\*\*”, one of many special characters used in the ACS data. Developing procedures to handle these special characteristics and values in GIS is quite tedious.

### Geographical Coverage

Partly due to the minimum population thresholds that meet the disclosure rules and partly due to the survey sampling framework, the 5-year data products include estimates for all counties, but the 1-year and 3-year data products do not have

complete coverage of the entire U.S. even at the county level. Missing ACS estimates for specific geographical units of interest is definitely undesirable. But even if estimates for concerned units are available, the limited geographical coverage of certain ACS data products creates holes of missing values on maps, posing significant difficulties in performing spatial analysis.

Two maps show the median household income using two sets of ACS data products: 1-year estimates for 2009 and 3-year estimates for 2006-

2008 (Figure 1). For the 2009 1-year ACS estimates, 2428 counties out of a total of 3233 counties in the U.S. have no data. For the 2006-2008 3-year estimates, 1346 counties have no data. One may contemplate combining estimates from multiple years for a more completed geographical coverage. However, estimates collected in different single-years are temporally mismatched and comparing estimates from overlapping years could be a challenging task (U.S. Census Bureau, 2008). Therefore, the geographical completeness of ACS data creates a pragmatic challenge to use the data in geographical research and GIS.

*Table 1. Selected census tract records from 2005-2009 ACS estimates and MOEs of median age in Virginia*

GEO_NAME	Bo1002_1_EST	Bo1002_1_MOE
<i>Geography</i>	<i>Median age (Estimate)</i>	<i>Median age (Margin of Error)</i>
Census Tract 4702, Fairfax County, Virginia	-666666666	**
Census Tract 0, York County, Virginia	-666666666	**
Census Tract 0, Norfolk city, Virginia	-666666666	**
Census Tract 63, Norfolk city, Virginia	-666666666	**
Census Tract 0, Virginia Beach city, Virginia	-666666666	**
Census Tract 105.02, Amherst County, Virginia	28.5	43.2
Census Tract 601, Richmond city, Virginia	39.5	25.3
Census Tract 8501, Prince George County, Virginia	30.6	24.4
Census Tract 39, Norfolk city, Virginia	50.9	19.7

## Data Reliability Issues

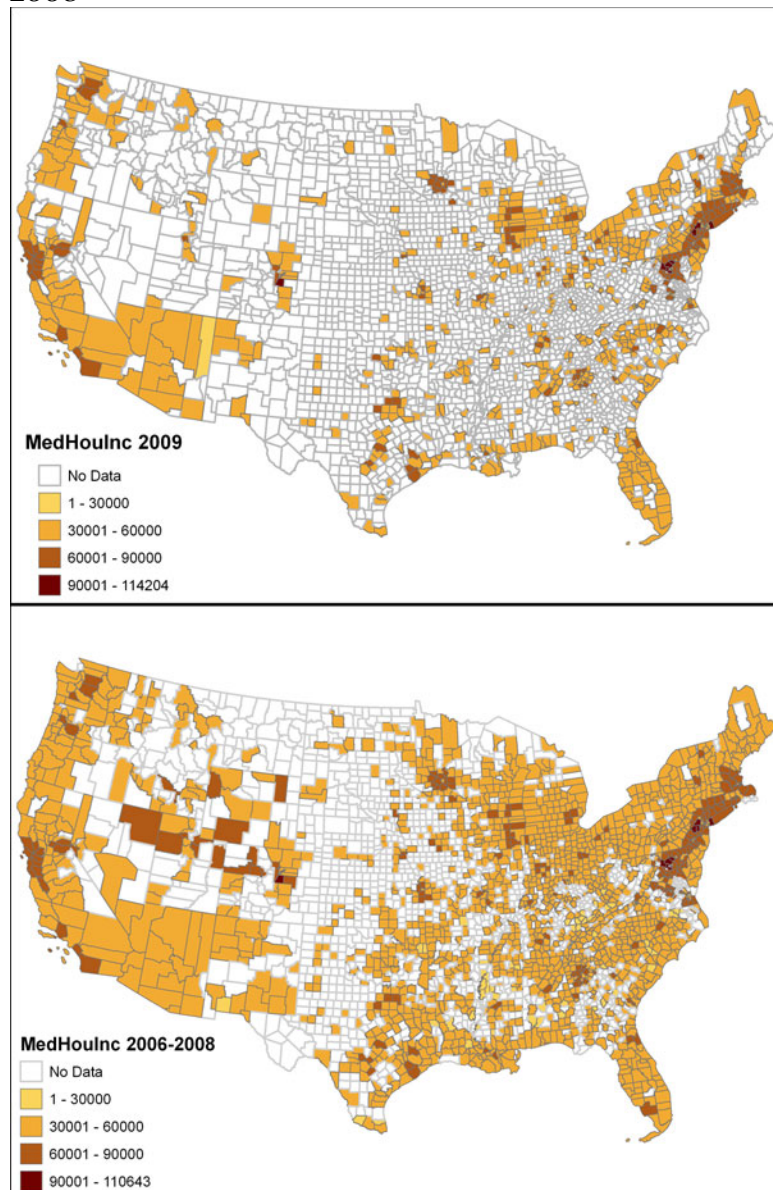
The 5-year ACS data cover the entire U.S. and are available down to the census block-group level. However, ACS data for detailed geographical levels have relatively high levels of unreliability. In their report, the first guideline that Citro and Kalton (2007) offer to ACS users is “always examine margins of error before drawing conclusions from a set of estimates” (2007: 3). The essence of this recommendation is that ACS estimates that are different numerically may not be truly different. Using the accompanied MOEs, estimates can be compared statistically to determine if their differences are significant or not.

The accuracy of ACS estimates also varies by

geographical levels in the census geography hierarchy. In general, larger tabulation areas include larger samples. Therefore, estimates from larger areas should be more reliable. ACS data at the county level are quite reliable overall. But when tabulation units become smaller, the reliability of an estimate likely becomes worse, as is reflected by the larger associated MOE. The 5-year ACS products include data for block groups, the smallest tabulation areas supported by the ACS program. However, the reliability of block-group level data is a concern. Even at the census tract level, the reliability of estimates may not be at a level acceptable to many users. Users may have to trade-off between using spatially more aggregated data with higher reliability and data with higher geographical details, but less reliable.

Several selected tracts in Virginia have alarming levels of MOE for the median age variable (Table 1). For tract 105.02 in Amherst County, the MOE for the median age estimate was 43.2, which is larger than the estimate of 28.5. If one creates a 90% confidence bound of the estimate, the lower bound of this estimate will be negative - a value that is difficult to interpret. Two other tracts in Richmond City and Prince George County have smaller MOEs, but they are still relatively large as compared to their corresponding estimates.

*Figure 1. Median household income (MedHouInc) estimates from ACS data: Upper: 2009; Lower: 2006-2008*



Given the quality of these estimates, the usability of these data is something in question. For many studies conducted at the community level, data for smaller tabulation areas are required. While most of the ACS estimates may be of reasonable quality, alarming levels of unreliability for selected estimates may create challenging situations. More resourceful and technical users may develop their own aggregation schemes along the attribute and/or geographical dimension to derive ACS

estimates with reasonable levels of reliability. A similar approach is to employ spatial estimation methods, leveraging the existing spatial structure in the data. A less technical approach, which was adopted in developing the ArcGIS extension described below, is to highlight unreliable estimates to warn the audience.

Another implication of the guideline provided by Citro and Kalton (2007) on examining the margin of error is related to the ways that maps are used. A major objective of creating a map is to explore and determine if the concerned data depict a spatial pattern. A spatial pattern emerges when differences in values exhibit a systematic spatial arrangement. The spatial arrangement of the differences may reflect, in general terms, the presence of a spatial trend, a clustering pattern or a dispersed pattern. However, when any such pattern is recognized visually, we assume that the differences between values are real, and thus the resultant spatial pattern is real. But if the values are subject to sampling errors such that differences between values may not be real, then the revealed spatial pattern may not exist. The bottom line is that when two ACS estimates are different numerically, we cannot immediately conclude that the two estimates are statistically different without considering their quality information or MOEs.



Two pairs of counties in the state of New Jersey are selected for the illustration (Table 2). The data are per capita income estimates with their MOEs at the county level from the 2008 1-year ACS data. We compared the per capita income estimates between Union and Sussex counties, and between Morris and Hunterdon counties. Between Union and Sussex, their difference in income was \$1,611. While their 90% confidence bounds overlap, the z-value of their difference is 1.5711, smaller than the critical value of 1.645, and therefore failing to conclude that the two estimates are significantly different at the 90% confidence level. When comparing Morris with Hunterdon counties, the difference was \$4,005, much larger than the difference for the previous pair of counties. As the MOE of Hunterdon County is relatively large, the resultant confidence bounds of the two counties do overlap. However, the z-value of the difference in estimates is 1.9904, larger than the critical values of 1.645 at the 90% confidence level and 1.96 at the 95% confidence level. Therefore, the two counties have significantly different per capita income estimates even their confidence bounds overlap.

Examples in Table 2 highlight the need to incorporate MOEs or related data quality information when estimates are compared either explicitly or implicitly in mapping analysis. These examples not only demonstrate that numerical differences between estimates alone cannot be used to decide that the estimates are different, but also show that even comparing confidence bounds of estimates is not sufficient to render a correct conclusion.

Therefore, a major challenge to the use of ACS data in GIS is to provide information to users and map readers about the reliability of estimates so that they can avoid making erroneous conclusions when estimates are compared implicitly or explicitly.

## AN ArcGIS EXTENSION SUPPORTING THE USE OF ACS DATA

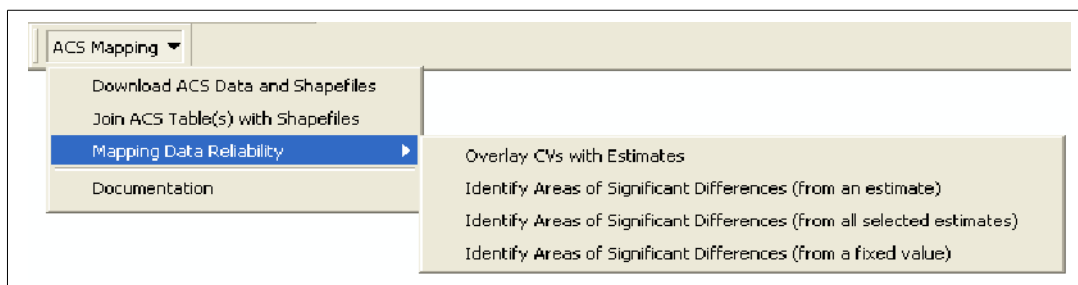
To address some of the challenges to the use of ACS data in geography and GIS discussed above, we have developed an extension for ArcGIS. This extension, the ACS Mapping Extension, provides three types of functions: instructions to obtain ACS data and boundary shapefile data, a tool to merge ACS tables with corresponding shapefiles, and a set of mapping functions that incorporates estimate reliability information (Figure 2). The extension provides step-by-step instruction to download various types of ACS tables from the American FactFinder (AFF), a web-based data discovery and dissemination system developed by the Census Bureau for users to access census data products. Instruction is also provided to obtain the associated boundary files in shapefiles format from the U.S. Census Bureau website.

The rest of this article will focus on the last two types of functions in the extension.

*Table 2. Per capita income levels and margins of error (MOEs) for two selected pairs of New Jersey counties, 2008 ACS*

<i>County</i>	<i>Per Capita Income</i>	<i>MOE</i>	<i>Lower Bound</i>	<i>Upper Bound</i>	<i>County Difference</i>	<i>z-value</i>
Union vs. Sussex	33,379	1,095	32,284	34,474	1,611	1.5711
Morris vs. Hunterdon	47,075	1,390	45,685	48,465	4,005	1.9904
	51,080	3,004	48,076	54,084		

Figure 2. Interface and menu items of the ACS Mapping Extension



After downloading ACS tables and the corresponding shapefiles, users can join them together. The set of ACS tables and shapefiles should have a shared unique identity field that can be used to link ACS tables with the attribute table in the shapefiles. Using this shared table item, ACS tables can be merged with the boundary shapefiles to form a new set of shapefiles. The ACS Mapping Extension includes a tool to merge one or multiple ACS tables with the corresponding shapefiles. The merged data are saved as a new set of shapefiles. In a form, users input parameters for the merging process (Figure 3).

This function in the extension has several features, addressing some of the difficulties in using ACS data in GIS. First multiple ACS tables can be included in the merging process. Users can download multiple ACS tables of the same type (detailed tables, data profiles tables or subject tables) for the same geography. All these tables can be joined together with the corresponding shapefiles. Second, different types of ACS tables have slightly different field formats. The tool that joins data tables and shapefiles together can handle all three popular table formats. Third, the tool allows users to replace the original code-based field names with more understandable customized field names in the resultant shapefiles. The tool can also handle all special values and characters used in the ACS data to indicate various sampling issues and conditions about the estimates. Users do not have to deal with them explicitly. The data structure of a column of estimates followed by a column of MOEs is also accommodated. Both types of columns will be merged with the

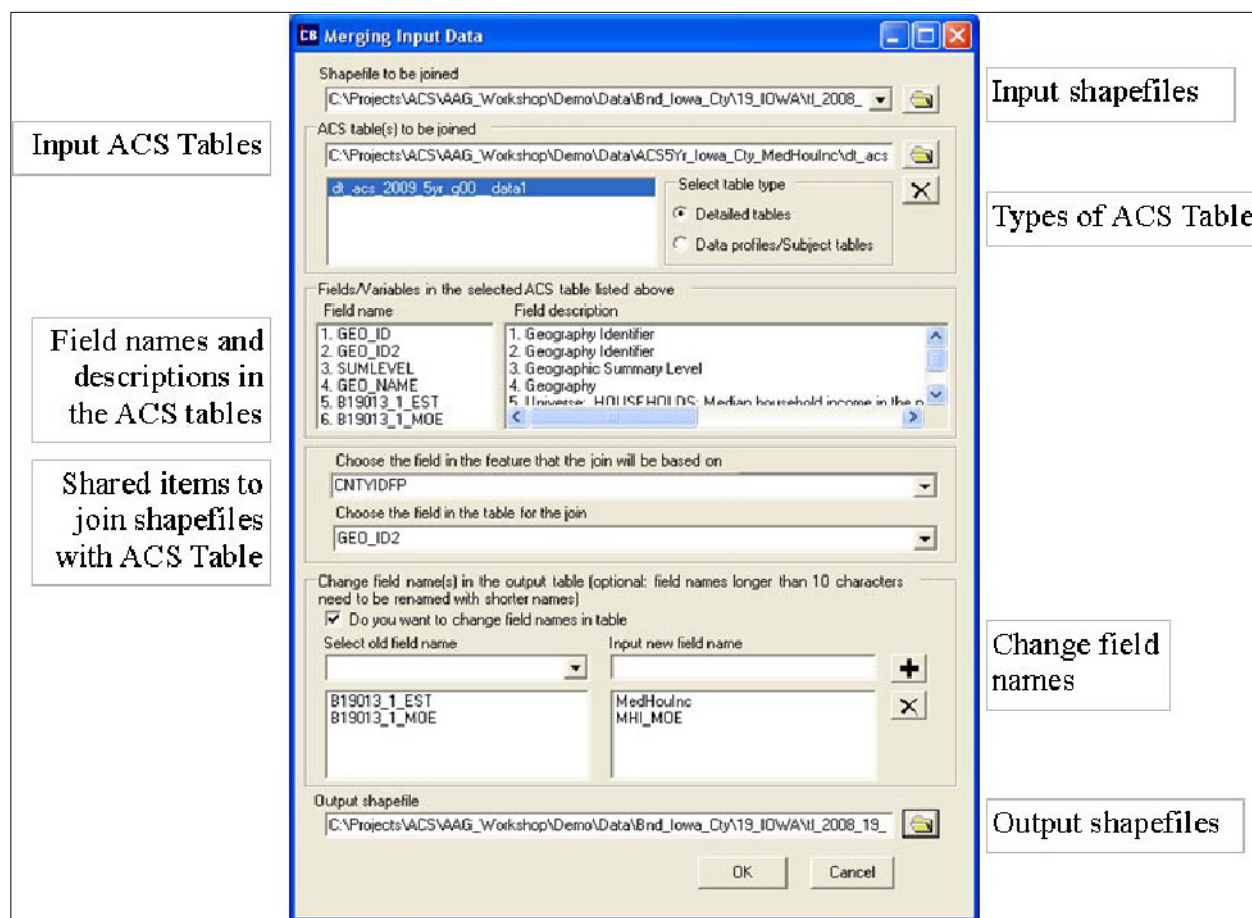
corresponding shapefiles. All these features streamline the process of bringing ACS data into GIS. This tool can also be used to merge any shapefiles with other census datasets, including the popular Summary Files from the decennial censuses, and other generic attribute tables to be used in GIS.

The final set of functions focuses on mapping ACS data by incorporating data reliability information. The main objective is to assist users when comparisons are made between specific estimates. Several methods of incorporating data reliability information in mapping ACS data have been proposed (Sun and Wong 2010), but some of them require graphical capabilities not easily supported by most GIS. The first mapping function in the extension (Overlay CVs with Estimates in Figure 2) allows users to overlay the coefficients of variation (CVs) on the top of estimates using texture patterns. In this case, CVs are used instead of MOEs because MOEs are sensitive to the scale of estimates. Estimates with higher values (such as income in New York City) tend to have larger MOEs, which are derived from standard errors. On the other hand, CVs are standard errors scaled by estimates, and therefore are not sensitive to the absolute magnitude of estimates. CVs are computed by the mapping tool, added to the attribute table, and used to reflect the reliability of estimates.

A bivariate legend approach is adopted to combine estimates with CVs, as this approach has been proven to be effective to visualize data quality information (MacEachren et al., 1998, 2005; Brewer and Pickle 2002; Pickle 1996). Such



Figure 3. Interface of the tool in the ACS Mapping Extension to merge ACS data tables with corresponding shapefiles

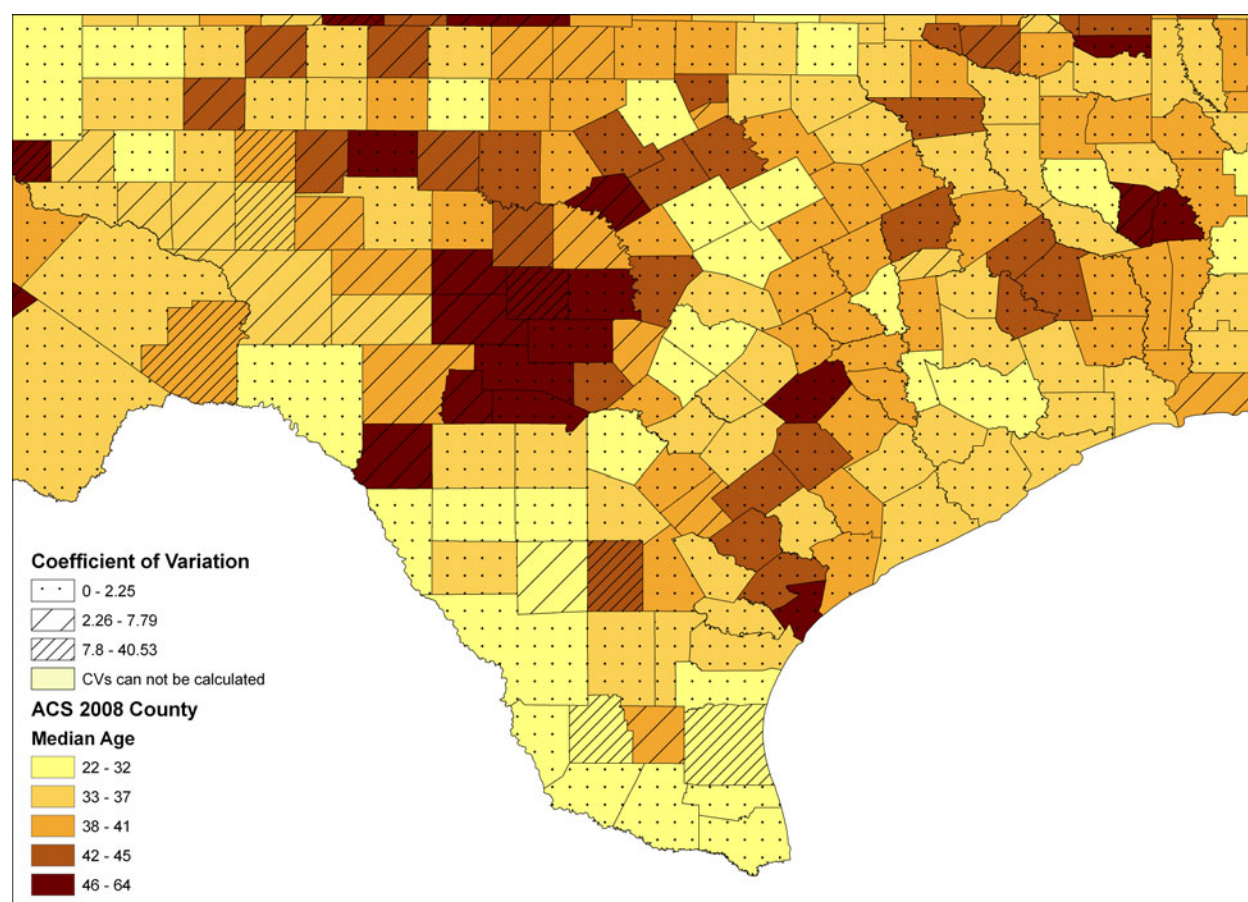


bivariate legends allow map readers to recognize the reliability levels of estimates. High CV values indicate unreliable estimates and vice versa. Figure 4 shows part of the county-level map of the entire U.S., using 2008 ACS county-level estimates of median age with the CV overlay. The map is reasonably effective in indicating different levels of reliability of the median age estimates across counties. Such approach to integrate attribute quality information in mapping can also be used for other datasets with feature-based quality information.

The other mapping functions in the ACS Mapping Extension are variations of comparing estimates, indicating if the differences between estimates are statistically significant or not. The second mapping function (Identify Areas of Significant Differences (from an estimate) in Figure 2) allows

users to determine if the estimate of a chosen areal unit is different from other estimates on the map. Users have to first select an areal unit within the study region as the referenced unit. The mapping tool will compare the estimate of the selected unit with estimates in all other units. Given a chosen level of significance, the tool will identify units with an estimate significantly smaller or larger than the estimate of the referenced unit. The extension includes some default patterns for the texture overlay. Users can alter these pattern selections. Such mapping results are useful for local comparison of estimates and can assist users to discern if estimates are really different or not, even they are in different classes of the thematic map. This function is similar to the Thematic Maps function on the ACS page of the legacy AFF, but it provides the flexibility of choosing different

Figure 4. ArcGIS overlays the coefficients of variation (CVs) on ACS estimates using the ACS Mapping Extension function



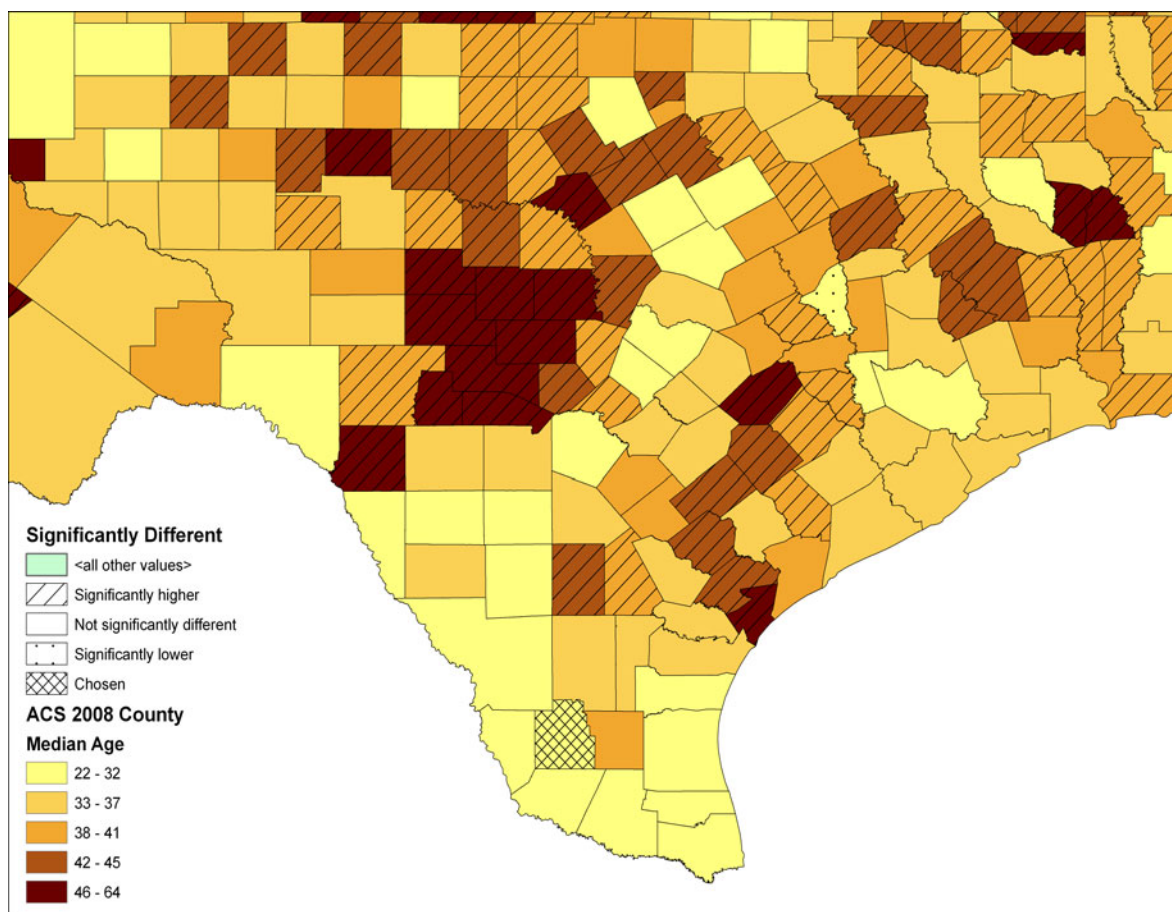
levels of significance in determining differences.

Using the same data as in Figure 4, a map was generated from the second mapping function (Figure 5). Jim Hogg County, a county near the southern tip of Texas, was selected as the referenced unit. This county, according to the 2008 ACS data, had a median age estimate of 32 and an MOE of 7, falling into the lowest category of the map classes (22-32). Counties with median age estimates significantly lower and higher than that in Jim Hogg County at the 90% confidence level were identified (Figure 5). Brazos County in east central Texas was assigned to the same map class as Jim Hogg County, but its median age estimate, 23, was significantly lower. On the other hand, many counties in the higher two classes (33-37 and 38-41), including the one east of Jim Hogg County, have median

age estimates not significantly different from that in the referenced county. Such a map shows that counties in different classes may not have significantly different ACS estimates.

In some analyses, area of interest may include more than one areal unit, such as several adjacent census blocks representing a neighborhood. One may be interested to find out if the concerned neighborhood has a socioeconomic status different from other neighborhoods. Then multiple units need to be selected as the references to be compared with units in the rest of study region, and the third mapping function supports such comparison (Identify Areas of Significant Differences (from all estimates) in Figure 2). The estimate of an areal unit in the study region may be significantly lower, significantly higher, or significantly different from *all* selected units. To

Figure 5. ArcGIS identifies areas with attribute values significantly lower, higher and not different from the attribute value of the selected Jim Hogg County in Texas using the ACS Mapping Extension function



illustrate this type of comparison, two neighboring Texas counties, McMullen and La Salle, were selected as a region to be compared with all other counties (Figure 6). The two counties have median age estimates of 42 and 32, respectively. Some counties have median age estimates significantly higher than both of these selected counties, and some have median age estimates significantly lower. However, some counties have median age estimates between these two county estimates, but are significantly different from both of them (Figure 6).

Finally, estimates in the study region may be compared with a fixed value, such as a national or regional average, or a specific

threshold or criterion value (Identify Areas of Significant Differences (from a fixed value) in Figure 2). Users can supply the fixed value on a form similar to those input forms for other mapping functions. Areas will be identified to have estimates significantly higher and lower or not different from the user-provided value at the chosen confidence level.

## SUMMARY AND CONCLUSION

In this article, the concepts of feature-based data quality information and spatial metadata were reiterated, and the concepts were linked to the reliability information of ACS estimates provided by the U.S. Census Bureau. A set of



challenges to the use of ACS data in geographical research and GIS was elucidated. The article also provides an overview of functions found in the ACS Mapping Extension for ArcGIS. Functions in the extension are intended to tackle some of the challenges to the use of ACS data in GIS. The main objective of the mapping functions in the extension is to assist users to discern if estimates are statistically different. Being able to decide if differences between estimates are statistically significant or not is still far from determining if the spatial pattern exhibited by the estimates is attributable to sampling error or not.

Current version of the extension includes only a limited number of functions to tackle a subset of challenges of using ACS in GIS, particularly on incorporating reliability information when mapping ACS data. Some challenges, such as the reliability issues for small areal units and the incompleteness of geographical coverage, limit the use of ACS data in GIS. The proposed functions may be used to evaluate the usability of ACS data, particularly those estimates for smaller census areal units with larger sampling errors. One may argue that GIS and spatial analytical techniques may be used to improve the quality of ACS estimates or to fill in the “holes” in the map. However, such procedures have to be evaluated stringently, particularly by statistical experts in the U. S. Census Bureau, before they are adopted by the public. On the other hand, the developed extension can streamline the process of merging attribute tables other than ACS tables with shapefiles, facilitating the process of creating geospatial data in GIS. The distinctive value of the extension is its capabilities to facilitate the handling and mapping of feature-based attribute quality information. In order words, the extension can be used for non-ACS datasets with attribute reliability information for each observation, including many survey datasets, such as the Current Population Survey,<sup>5</sup> in which the reliability of each estimate is included.

More functions for spatial data handling and

mapping could be added to the extension, including some ACS-specific mapping functions proposed by Sun and Wong (2010). The extension by no mean removes all impediments of using ACS data in GIS and geospatial research. However, it is the beginning to help GIS users, researchers and practitioners live with errors in geospatial data, a situation that Openshaw envisioned more than two decades ago (Openshaw 1989). Users of the extension or any spatial analytical tool should recognize that these tools cannot turn low quality geospatial data into reliable geographic information. Understanding the statistical nature and limitations of geospatial data is critical to ensure the correct use of these tools. Interested readers may download the ACS Mapping extension from <http://gesg.gmu.edu/census>.

## Acknowledgements

The authors are grateful to the assistance provided by the two editors-in-chief of *Spatial Demography*. The authors would like to thank the U.S. Census Bureau for its support of the research leading to the development of this paper, particularly the guidance provided by Nancy K. Torrieri and Deborah H. Griffin of the Census Bureau, ACS Office. However, the authors are responsible for any error in this article.

## Notes

1. [http://www.census.gov/acs/www/Downloads/survey\\_methodology/acs\\_design\\_methodology\\_choose.pdf](http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology_choose.pdf).
2. [http://www.census.gov/acs/www/guidance\\_for\\_data\\_users/handbooks/](http://www.census.gov/acs/www/guidance_for_data_users/handbooks/)
3. [http://www.census.gov/geo/www/cob/bdy\\_files.html](http://www.census.gov/geo/www/cob/bdy_files.html)
4. <http://www.census.gov/geo/www/tiger/index.html>
5. <http://www.census.gov/cps/>

## References

- Allen, J. P., and E. Turner, E., 1997, *The Ethnic Quilt: Population Diversity in Southern California* (Northridge, CA: Center for Geographical Studies).
- Buttenfield, B.P., 1993, Representing Data Quality, *Cartographica*, 30(2), 1-7.
- Brewer, C. A., and L. Pickle, 2002, Evaluation of Methods for Classifying Epidemiological Data on Choropleth Maps in Series, *Annals of the Association of American Geographers*, 92(4), 662-81.
- Brewer, C. A., and T. A. Suchan, 2001, *Mapping Census 2000: The Geography of U.S. Diversity* (Redlands, CA: ESRI Press).
- Citro, C. F., and G. Kalton (Eds), 2007, *Using the American Community Survey: Benefits and Challenges* (Washington, DC, USA: The National Academies Press).
- Devillers, R., Y. Bédard, and R. Jeansoulin, 2005, Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use within GIS, *Photogrammetric Engineering and Remote Sensing*, 71, 205-15.
- Gan, E. and W. Shi, 2002, Error Metadata Management System. In Shi, W., P. F. Fisher, and M. F. Goodchild (Eds.), *Spatial Data Quality* (London: Taylor and Francis), 251-66.
- Gates, G. J., 2006, Same-Sex Couples and the Gay, Lesbian, Bisexual Population: New Estimates from the American Community Survey, *The Williams Institute on Sexual Orientation Law and Public Policy*, UCLA School of Law, April 25, 2011, <http://www.law.ucla.edu/williamsinstitute/publications/SameSexCouplesandGLBpopACS.pdf>.
- Hough, G.C., and D. W. Swanson, 2006, An Evaluation of the American Community Survey: Results from the Oregon Test Site, *Population Research and Policy Review*, 25, 257-73.
- MacDonald, H., 2006, The American Community Survey: Warmer (More Current), but Fuzzier (Less Precise) than the Decennial Census, *Journal of the American Planning Association*. *American Planning Association*, 72(4), 491-503.
- MacEachren, A.M., C. A. Brewer, and L. W. Pickle, 1998, Visualizing Georeferenced Data: Representing Reliability of Health Statistics, *Environment & Planning A*, 30, 1547-61.
- MacEachren, A.M., A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler, 2005, Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3), 139-60.
- Openshaw, S., 1989, Learning to Live With Errors in Spatial Databases. In: Goodchild, M. F. and S. Gopal (Eds.), *Accuracy of Spatial Databases*, (London: Taylor and Francis), 263-276.
- Peters, A. and H. MacDonald, 2004, *Unlocking the Census with GIS* (Redland, CA, USA: ESRI Press).
- Pickle, L.W., M. Mungiole, G. K. Jones, and A. A. White, 1996, *Atlas of United State Mortality*, (Hyattsville, MD, USA: National Center for Health Statistics).
- Qiu, J. and G. J. Hunter, 2002, A GIS with the Capacity for Managing Data Quality Information. In Shi, W., P. F. Fisher, and M. F. Goodchild (Eds.), *Spatial Data Quality* (London: Taylor and Francis), 230-250.
- Sun, M., and D. W. S. Wong, 2010, Incorporating Data Quality Information in Mapping the American Community Survey Data. *Cartography and Geographic Information Science*, 37 (4), 285-300.
- Thomson, E. F., B. Yu, A. Nuru-Jeter, J. M. Guralnik, and M. Minkler, 2009, Basic ADL Disability and Functional Limitation Rates Among Older Americans From 2000-2005: The End of The Decline? *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 64A (12), 1333-6.
- Turner, E., and J. P. Allen, 1991, *An Atlas of Population Patterns in Metropolitan Los Angeles County 1990*. (Northridge, CA, USA: Center for Geographical Studies).
- U.S. Census Bureau (USCB), 2008, *A Compass for Understanding and Using American Community Survey Data: What General Data Users Need to Know* (Washington, DC, USA: U.S. Government Printing Office). (<http://www.census.gov/acs/www/Downloads/handbo>

oks ACSGeneralHandbook.pdf).

U.S. Federal Geographic Data Committee (FGDC), 1998, Content Standard for Digital Geospatial Metadata (Washington, DC, USA: Federal Geographic Data Committee) FGDC-STD-001-1998.

Wong, D. W. S. and C. V. Wu, 1996, Spatial Metadata and GIS for Decision Support. Proceedings, 29th Hawaii International Conference on System Sciences (HICSS) Volume 3: Collaboration Systems and Technology.  
<http://doi.ieeecomputersociety.org/10.1109/HICSS.1996.493251>.

Xiao, N., and K. Hawley, 2010, Mapping the Census: An Introduction. *Cartography and Geographic Information Science*, 37(3), 167-168.

Zandbergen, P. A., 2008, Positional Accuracy of Spatial Data: Non-Normal Distributions and A Critique of the National Standard for Spatial Data Accuracy, *Transactions in GIS*, 12(1), 103-130.

### **Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.