



THE BROOKINGS INSTITUTION

1775 Massachusetts Avenue, NW Washington, DC 20036-2103
Tel: 202-797-6000 Fax: 202-797-6004
www.brookings.edu

September 27, 2007

Director, U. S. Census Bureau
Room 8H001, Mail Stop 0100
Washington, DC 20233-0001

E-mail: douglas.w.hillmer@census.gov

RE: Request for comments regarding proposed American Community Survey Data Products

Dear Director Kincannon,

On behalf of the Metropolitan Policy Program (Metro) of the Brookings Institution, I am pleased to respond to the notice placed by the Department of Commerce in August 28, 2007 Federal Register asking for comments regarding plans for the type and release schedule of data products from the American Community Survey (ACS). We very much appreciate the opportunity to provide our comments and hope you find them of value.

Metro promotes innovative solutions to help the nation's metropolitan communities grow in more inclusive, competitive, and sustainable ways. From this perspective, we believe that the ACS is critically important to the health and well-being of metropolitan America and are quite excited about the prospect of obtaining annually updated multi-year estimates for small geographic areas.

The release of these estimates will represent the final stage in the transition from the decennial "long form." The promise, and premise, of the ACS has been to enable researchers and practitioners operating at any scope of geography, from the nation to the neighborhood, to easily obtain current, detailed, accurate data in a usable format. The multi-year products will help fulfill this promise by offering data on demographic, social, economic, and housing characteristics in census geographies below 65,000 in population, down to the block group. The products will be quite valuable to users in all sectors of public and private sector activity. While current ACS products meet many broad-based needs, the multi-year estimates products will increase the justification for and usefulness and cost-effectiveness of the ACS to federal, state, and local agencies, the Congress, and a wide range of data users.

We see the proposed release schedule as reasonable, meeting user needs for data access in the calendar year after data collection while giving the Census Bureau adequate time to review the data for quality and ensure product accuracy. We note that the Census Bureau plans to release its first five-year ACS estimates ACS in December 2010, the same month in which the release of counts from the 2010 Census are required by law to be released by law. We suggest informing

data users whether timing of the ACS multi-year data releases in 2010 will be affected by the Census Bureau's release requirements regarding the 2010 Census.

We applaud the overall intent of the proposed plan for ACS data products to meet data users' needs. In particular, we support plans to provide easy-to-use products for data users with limited statistical literacy, such as the Data Profiles, the Narrative Profiles, and Thematic Maps, as well as the array of detailed products for more experienced analysts.

We have attached detailed recommendations regarding the proposed data products. In summary:

- We suggest that the Census Bureau more clearly differentiate products for experienced analysts from those for casual data users, and products intended for those analyzing one or few geographic areas from those needing to download large amounts of data for geographic areas across a state, a multi-state region, or the nation.
- We ask that the data products provide confidence intervals. We are troubled that the Census Bureau continues to propose to show only the margin of error associated with each estimate rather than the confidence interval, despite many comments from analysts about the inefficiencies and heavy costs associated with this choice.
- The proposal leaves unclear some questions about the products in which the multi-year estimates will appear and about long-range plans for the presentation of data about the population living in group quarters. We suggest that multi-year estimates be provided wherever possible, and that distinct tables be provided for group quarters.
- As confidentiality of the data is so vital, and to ensure that data users will continue to be able to efficiently design their own custom tabulations from microdata, we recommend that the Census Bureau develop alternatives to the Public Use Microdata files.
- We propose additional ACS products, including ones related to geographical mobility to replace those that had been provided from the decennial censuses and an automated calculator of margins of error and statistical significance.
- We ask that the Census Bureau provide guidance to users about appropriately choosing among 1-, 3-, and 5-year estimates.

In conclusion, Metro strongly supports the continuing development of data products from the ACS and applauds the Census Bureau's efforts to release useful data products on a timely basis. We ask the Census Bureau to consider our recommendations for improvements as described in the attachment, hope you find our ideas of value, and thank you for the opportunity to provide them.

Sincerely,



Andrew Reamer, Fellow
Metropolitan Policy Program

Attachment -- Recommendations for ACS Products

Deliverables that differentiate among the experience level of data users

- *Products for casual data users vs. “tool kits” for experienced analysts* – Paul Overberg of USA TODAY compares the current and proposed ACS products to what parents have the night before a winter holiday– a box full of toy parts with complex directions to stick Tab A into slot Z. Everything is there for making the toy if you know how to use the tools and have all night to put it all together. However, our observation is that casual data users will not try to find the associated documentation – they want the “finished toy.” Moreover, the proposed ACS products could do much more to help experienced data users. Consequently:
 - We suggest that the Census Bureau distinguish between products intended for casual data users and those that are for analysts with knowledge of basic statistical concepts and appropriate uses and limitations of data. We perceive that current ACS products appropriate for experienced data analysts often confuse those who seek data only occasionally.
 - Specifically, we recommend that Narrative Profiles, Data Profiles, and Thematic Maps be identified as products most appropriate for data users with little statistical knowledge.
 - We recommend that the Census Bureau list the ACS products in order of complexity on the American FactFinder and explicitly guide inexperienced data users to the simpler products.
 - With regard to problem noted by Paul Overberg regarding difficulties in using more complex products, we provide specific examples below and suggest improvements for helping experienced analysts better understand the potential and limitations of this new data set, particularly how its content and methodology differ from the decennial census long form in ways that affect their analyses. Essentially, we would like to see the more complex products be seen as a “tool kit” of both data and easily accessible documentation.
- *Hyperlinks on the American FactFinder (AFF) for documentation* – As an example of the “tool kit” approach, we ask that the AFF include a hyperlink to the ACS home page for a complete list of technical documentation.

Narrative Profiles, Data Profiles, and Thematic Maps – As noted, we view these three products as most appropriate for data users with little statistical knowledge.

- *Product order in AFF* – We recommend that AFF list these three products first, beginning with “Narrative Profiles.”
- *Display of the MOE* – It seems suitable to show only the Margin of Error (MOE) in the Data Profiles, as this appropriately warns inexperienced users that they are using estimates, not counts, but does not burden them with information they are unlikely to use

further. By providing a hyperlink that explains MOE, the Census Bureau is wisely taking the opportunity to educate and train new data users.

- *Graphic display of confidence intervals in Narrative Profiles* – We recommend that the charts in the Narrative Profiles show the 90-percent confidence intervals, especially if there are comparisons across time or geographic areas. The income and poverty reports show graphics in the style of high-low-close stock charts, a consumer-focused approach we think works very well.
- *Consistency of variable names across tables, use of plain English, access to definitions* – Paul Overberg of USA TODAY provides several examples of these issues:
 - The subject table about “teenagers” is for 15-to-19-year olds but, by anyone’s definition of that term, should be for 13-to-19-year olds.
 - Plain English for “journey to work” is “commuters.”
 - In subject table S1602 for “linguistic isolation,” there is no hyperlink that defines the term. The S1602 is for households. Without guidance, data users are confused by the closely related, but measure for persons shown in ranking table (R1603) that provides data for the "Percent of People 5 Years and Over Who Speak English Less Than "Very Well."
- *Training for inexperienced data users* – We view hyperlinks in the products are de facto training tools for new data users open to increasing their understanding of how to use data and gaining an understanding of basic statistical concepts. Therefore, we suggest that Census staff who write materials for the hyperlinks in these products make the terms and concepts accessible for the less experienced and use plain English.
- *Thematic Maps* – The Thematic Maps product meets the needs of both experienced and inexperienced data users by offering the option of displaying the statistical uncertainty of the estimates, which facilitates comparisons. We urge continuation of that function.

Ranking Tables and Geographic Comparison Tables

- *Sortable GCTs for all geographic areas* – Geographic ranking tables are useful products for both experienced and inexperienced data users. However, at present, ranking tables are available only for states. On the other hand, geographic comparison tables (GCTs) are available for a number of substate geographies, for the nation as a whole and each state, but cannot be organized in any way but alpha-numeric order.

It would be helpful to Members of Congress, for example, to quickly see how their Districts compare with others. At present, however, comparison is only possible by eyeballing the poverty rates from 535 rows or downloading the GCT data, computing the standard errors, running statistical tests for all comparisons, and knowing how to use text formulas in a spread sheet to reformat from page/reading format to a rectangular format.

Another anomalous difference between the ranking tables and GCTs is that charts, with confidence intervals, are available only for the former and thematic maps are available only for the latter.

The current set of comparison products is confusing to users and limits their ability to quickly make comparisons. Therefore, we recommend that the two products—the rankings tables and the GCTs—be integrated into one product, sortable GCTs that offer the ability to sort by alpha-numeric or by rank and provide functionality to create a chart, with confidence intervals, and thematic maps.

- *Multi-year estimates in the Sortable GCTs* – The examples provided indicate the Census Bureau intends to allow users of ranking/comparison tables to click on previous years. We recommend that this function include multi-year estimates as well. It is valuable to be able to see the statistical significance of these comparisons.
- *Naming Ranking Tables* -- If the current distinction between ranking tables and GCTs is retained, we suggest that the former should more plainly be called what they are, “State Ranking Tables.”

Subject Tables

- *Usefulness of Subject Tables* – We believe that the subject tables have the potential to be very valuable for a range of users. We view the current topic list as good but limited. While the summary tabulations are those that have been provided traditionally, we think they could be improved with, for example, age categories, that are frequently required by policymakers and funders. As the subject tables are geared towards policy issues, we recommend that the ACS Office meet with federal and state agencies and the Congress to better determine the topics and content of the summary tabulations that would best meet their needs.
- *Multi-year estimates in Subject Tables* – The example shows that the Census Bureau will continue to allow data users to click on previous years for subject tables. We recommend that this will include multi-year estimates as well.

Detailed Tables

- *Revised set of Detailed Tables based on subject expert guidance* – We find that the Census Bureau has not given data users adequate time to review the detailed tables. Consequently, we recommend that the current set be considered as preliminary rather than as final and that the Census Bureau convene expert data users in each subject area for thorough review and recommendation, leading to a final set of detailed tables in 2010.
- *“C” Tables* – The ACS 5-year estimates are the replacement for the decennial long-form estimates; much of the detail in the tables will be suppressed for the 1- and 3-year estimates for smaller areas, resulting in relatively little published data for these areas.

While the detailed tabulations are needed for many purposes at higher geographic levels, we recommend increasing the number of the “C” or collapsed tables so that basic tabulations are available for all areas. The added advantage of the C tables is that the MOEs are computed, and are more accurate as they are based on individual data rather than grouped data.

- *Additional tabulations* – We propose that certain commonly-used tabulations be added to the proposed tables. For example, we strongly recommend the addition of a basic age table with totals and collapsed age groups for the detailed race and Hispanic origin groups. While these data can be gleaned after much clicking from the subject tables, the percentage distributions across groups will not add up because the percentages are based on group data; for example, the estimated percentage of Non-Hispanics 18 years and older shown in the subject table for Total Non-Hispanics is quite different from that derived from summarizing that percentage from the data shown in each table for the separate groups (White Non-Hispanic 18+; Black Non-Hispanic 18+; and so forth). Data users also want additional tables on all children in the household, not just "own children."
- *Display of margins of error vs. confidence intervals* – We are disappointed to see that the Census Bureau continues to propose showing only a column for the MOE of the estimate, despite many comments from users that they prefer to see the 90-percent confidence interval. Experienced analysts use the confidence interval both for a quick review of what likely can be legitimately compared once formally tested and for formal statistical tests for published statements of comparisons. Users’ need to compute the standard error and the confidence intervals for every estimate they want to use, which could easily be provided by the Census Bureau, results in the waste of hours by analysts and substantial needless added costs. The Census Bureau provided the 90-percent confidence intervals in data products prior to 2005, then changed to MOE without discussion with data users. Many users have indicated to the Census Bureau that this change has been harmful. We strongly recommend that the Census Bureau reinstitute the publication of the 90-percent confidence interval.
- *Data products about the population living in group quarters and in households* – There is no statement in the proposed plan about data products on the population living in group quarters. We find that users desire distinct tables on group quarters. They also desire the capacity to analyze data for households only. Consequently, we recommend that the Census Bureau provide data products, for total population, “household only,” “group quarters” each year. Regarding group quarters data products, we further recommend that the Census Bureau hold discussions with data users about the appropriate tradeoffs between the detail of the tabulations and the detail of the geography for which such data are shown.
- *Add hyperlinks on the tables to relevant documentation* – The detailed tables are used most appropriately by experienced analysts. While it is helpful that each table provides links to relevant subject-matter definitions and information on statistical methodology, even experienced data users have experienced difficulty finding other relevant documentation. As part of facilitating the toolkit concept, we recommend that the Census

Bureau hold discussions with data users about additional links to technical products that are generally useful, such as the ACS questionnaires, and to informational links for specific tables, such as errata.

- *On-line calculator for confidence intervals and statistical testing.* The calculation of uncertainty and the testing of comparisons is an essential step for analyses of survey data. We strongly recommend that the Census Bureau provide an on-line calculator for confidence intervals and statistical testing. Such a calculator would be a very useful addition to a tool kit for ACS data users who often must add and subtract categories provided in the detailed reports. While we understand that an automatic calculator could produce only an approximation of the standard error, as it must use the grouped data, we see that data users have no other choice; whether the approximation is produced by hours of labor on an excel sheet or by on-line calculator, the end result of the calculation is the same.

Selected Population Profiles

- *Race and Hispanic origin displays* – While the current selected population profiles product is of use to data users, its value is limited because it provides the static display of “Total population” and only one selected population group. We think that the product would be improved if data users could select multiple groups for display in the table. For example, it is not unusual to want a table for “Total population of Hispanic origin” compared with “Total population not of Hispanic origin” or a table that compares the “Total population not of Hispanic origin” with the major race groups not of Hispanic origin (e.g., White NH, Black NH, Asian NH, Other race groups not of Hispanic origin).
- *Age group detail for the total population* – While age groups are embedded in the detailed tables, they are not sufficient for many purposes. We believe that data users want a product with more detailed age groups and less subject matter detail than the detailed tables; the table outlines now used for the current selected population profiles would meet that need if they provided more detailed age groups as well as the population “Under 18 years” and “18 years and older.”
- *90-percent confidence intervals* -- Because the selected population profiles are used so frequently for comparisons with other groups, it is inconvenient, time consuming, costly, and inefficient to show only the MOE for these tables. We strongly recommend showing the lower and upper bounds of the 90-percent confidence interval for this product.

Comparison Profiles

- *All data points* – We appreciate that the Census Bureau proposes to re-institute area profiles that show the data over multiple years; this will allow users to more easily identify trends and patterns. We strongly encourage the Census Bureau to provide access to 1-, 3-, and 5-year averages. Some statisticians use all the data points available in regressions and it is a convenience to have all the data for an area available in one place.

Public Use Microdata Samples (PUMS)

- *Alternatives that meet the needs for microdata and protect confidentiality.* The PUMS data are useful products for the most sophisticated analysts who want to explore relationships among variables and differences among population groups that they cannot do with summary tabulations. Data users want continued access to microrecords. Over the years, however, as the Census Bureau has understandably had to impose increasingly onerous confidentiality edits, the file has become less useful. We urge the Census Bureau to consider long-range solutions to this problem by investing in research in and exploring alternatives to providing useful microdata while protecting confidentiality.

We note that the Census Bureau's LEHD program has developed a continuing research program to improve synthetic estimates. We suggest there be more interaction between the two programs to the benefit of both.

We also recommend that the Census Bureau continue to pursue development of tools helpful to users interested in microdata, such as DataFerrett, the Advanced Query System, and the Microdata Access System.

- *Combined population and housing PUMS* – We suggest that the Census Bureau explore the possibility of a combined population+housing PUMS for each state and the nation. While large file size may not make this practical in the near term, as bandwidth continues to grow and computing and storage costs decrease, this may be less of an issue for users over time.

ACS Summary File (ACS-SF)

- *Plans for an ACS SF* – The Census Bureau previously said that it plans an ACS-SF. It is unclear whether that continues to be a planned deliverable and, if so, how often it will be released. Is the 2006 ACS SF the final product or a prototype? We recommend an annual release of the ACS-SF.

Additional Products and Functionality

- *Geographic options in American FactFinder* – We suggest that when downloading detailed data tables from FactFinder, there should be an option to download for metropolitan areas only (and, likewise, micros only), rather than having to do all CBSAs together.
- *On-line calculator for MOEs, confidence intervals, and tests of significance* – Data users are constantly combining and subtracting categories and calculating percentage distributions and ratios – these require recalculation of the MOE for the new estimate. They must also do testing of the statistical testing of comparisons. This is an enormous barrier to all but the most statistically-savvy data users, and even for the last group, it is an enormous time waster to set up spreadsheets and download data to do the calculations. We recommend that the Census Bureau provide an on-line calculator for MOEs,

confidence intervals, and tests of significance. We understand that an automated calculator would be based on generalized variances and grouped data, but that is all data users have anyway.

- *Display of MOEs* - MOEs are shown in the format “+/- xxx” where the “+/-“ prefix makes it difficult to read with standard statistical package and fortran format statements. A better alternative would be to show the value as “xxxx” with a note in the table heading and documentation to show this represents a + or – value from the point estimate.
- *Need for Geographical Migration products* - The 2010 census will not have a migration question to provide researchers the traditional “county to county” or “state to state” migration files and associated products (see <http://www.census.gov/population/www/cen2000/migration.html>). Therefore, we suggest that the Census Bureau plan for substitute public use migration products based on the ACS migration question (“residence one year ago”). These products might include flow matrices with a limited number of attributes (broad age groups, gender, race and Hispanic origin) or in- and out-migration flow profiles for levels of geography that pass disclosure review. The most useful files would involve the 3-year or 5-year aggregated data for maximum geographic disclosure. We recommend that the Census Bureau convene a panel of state, federal, and academic researchers familiar with the migration uses of the previous decennial census products to suggest possible ACS geographical migration products.