

COMMENTS ON PROPOSAL FOR ACS MULTI-YEAR DATA PRODUCTS

As usual, my comments are from the perspective of a private sector provider of demographic data, where we have been long time heavy duty users of the nationwide small area long form data, and now look to the ACS to provide similar capabilities.

I have found the proposal somewhat difficult to react to. The summary of the types of products to be provided and the basic schedule for providing them is about as expected. However, it is not clear to me what the Summary File 4 product is. Is it the ACS Summary File product proposed earlier, but which was described as being patterned after SF3? If so, why is it called Summary File 4? If not, where does the SF3-style product fit in this proposal?

And why would the summary file be provided only once every five years? I understand the view that change in 5-year average data should be viewed only after five years of additional data collection, but with new 5-year averages becoming available every year, why not provide SFs annually? If I read this correctly, the plan is still to provide new five year average data (detailed tables) every year. I work with people who will be using the detailed tables of 5-Year data every year. If the summary files are released only once every 5 years, we will probably ignore them, and acquire the tables we need through the Download Center. Maybe this would work just as well – I'm not sure.

Also, are there plans for summary files for the 1-Year and 3-Year data? If not, why not? The alternative to summary files is to acquire tables through the Download Center, so the smooth functioning of that option may be critical. Currently, I find the download function to be a little cumbersome, but still very usable – so long as I only need a few tables.

Whether acquired through summary files or the Download Center, it is the content of the detailed tables that is most important to my applications. Products such as Narrative Profiles, Subject Tables, Ranking Tables, and Comparison Profiles may be valuable for many users, but we will make little use of them where I work.

Some users are concerned that the *Federal Register* notice provides insufficient time for a thorough review of the detailed tables, but it is not clear the notice seeks such a review. There is also concern that by suppressing some data (even for collapsed tables), the data use rules will limit the usefulness of the 1-year and 3-year data, and if so, a thorough review of detailed table content might help minimize that problem. I can just imagine trying to get the diverse community of users to agree on revisions to the detailed tables, and again, my read is that the notice does not call for that type of input. However, the flexibility to adapt to changing needs has been described as an ACS advantage, so I hope the Census Bureau will maintain a dialogue with users on detailed table content even after products are “finalized.”

Moving further beyond the scope of the *Federal Register* notice, a few other things are on my mind as we prepare for multi-year ACS data. The use of Census estimates as ACS controls is reassuring in some ways (putting the data on somewhat familiar turf, even if it is a bit of an illusion), but it also raises interesting questions. There is the issue of controlling data based on ACS residence rules to estimates based on decennial residence rules, and concern that some uncontrolled ACS data on characteristics might be more accurate than estimates-based controls. These are important questions, but I find myself just as focused on questions of timing. Let's take the first set of 5-year data as an example (forgetting that one of those years lacks GQ data because that issue will pass). Some of what I have heard suggests (or maybe I'm misreading) that estimates for this period would be weighted or controlled to county estimates for July 2009 – the most recent available at that time. However, with the data reflecting the 2005-2009 period, it would seem to make more sense to weight either with July 2007 (the middle year) or better yet, the average of the estimates for 2005, 2006, 2007, 2008, and 2009. I believe the materials with the multi-year “study” indicate that the average of these estimates would be used. That approach seems to make sense.

If county controls are based on averages, I follow how each new set of 5-year average data (each year) would be controlled to the new 5-year average of the relevant county estimates. It's the transition to a new census base that could be difficult. For example, when 5-year data span data collection from 2008-2012, what county numbers would be used as controls for the 5-year average data to be released in 2013? It takes a while before the flow of county estimates resumes following a new census, so would we have 2011 and 2012 estimates as soon as we need them. And would 2008 and 2009 estimates, revised for consistency with 2010 counts, be ready in time? For many of us, it is important to maintain the annual flow of 5-year average data, but I can see that that could be challenging during the transition period to a new census. Of course, another potential issue with this transition could be the transition to the new tracts and block groups defined for the 2010 census. A 2008-2012 average for a new tract would require specifying the pre-2010 data for the new tract, and I don't know how big an obstacle that presents.

Some users have expressed concern that the number of tables and the amount of category detail to be provided in the 5-year average data will be dramatically reduced relative to what we have been getting from the long form. If this is the case, it would severely undermine ACS support among many data users. Admittedly, I have only looked at some of the detailed tables, but so far, I have not seen anything that suggests a dramatic reduction in content. I wonder if the content provided in the multi-year “study” might give the impression of such a reduction. My read is that the “study” was provided as a way to give users examples of multi-year data to try on for size, but without indicating the scope of multi-year products. The bottom line is that many users will expect the multi-year data to be provided for tables and category detail similar to what was provided from the long form.

I will conclude by taking the opportunity express concern with the sample size issue – that the ACS 5-year sample is much smaller than the long form sample (especially in terms of actual responses), and does not increase as total population and households increase. My sense is that many knowledgeable users are becoming nervous about the quality of the 5-year data for small areas. Withholding such data would abandon the objective of long form replacement – the premise on which a lot of ACS support has been built. Without such data, the ACS is a substantial downgrade for many users. But if the small area data are conspicuously problematic, users may regard them as an inadequate replacement for the long form. Either way, I am concerned that we risk a significant erosion of user support for the ACS when multi-year average data are released (especially the 5-year data). Support appears strong now, but the one-year data are the easy part. The true test is when the ACS tries to look like SF3. Only then will we see how it is received by many users. As difficult as it might be to get, a significant boost in sample size likely will be required (as suggested by the National Academies report). In the absence of such an increase, there might be a temptation to redistribute ACS sample to fortify the data for small government units. Of course, doing so would diminish the data for small statistical geographies that are used by large numbers of users.

Thank you for the opportunity to comment. I am already making careful use of the 1-year ACS data, and look forward to a continuing dialogue as we prepare for multi-year data.