Presented to ACS Data Users Group
Discussant: Dave Hubble, Westat
Date: September 1, 2016
Overview

• Review of Comparison Results

• Contributing Factors to Differences (i.e., limitations associated with aggregated 1-year estimates)

• New Supplemental Files
  — 2014 1-year Estimates for 20,000+ Populations
  — 2010-2014 Variance Replicate Estimates

• Micro-data Analysis System (MAS) Development

• Other potential approaches: Model-based Estimates
Review of Results: Tabulations

• For geographies of 65,000+, aggregated 3-year estimates and MOEs do a fairly reasonable job of representing corresponding 3-year period estimates and MOEs.

• This is generally as expected as the aggregation of 1-year estimates into multi-year estimates was the original plan for the ACS prior to development of period estimate approach.

• Processes are fairly straightforward, but possibly laborious.
Review of Results: Tabulations (2)

• Recall that one of the limitations of these “custom” MOEs is that the covariance between the 1-year estimates is not accounted for in the approximation formulas.

• Interestingly, the observed differences in the MOEs were both positive and negative.

• One implication is that the missing covariance did not result in the measures of reliability exhibiting large-scale systematic under- or over-estimation.

• However, it also means they are not consistently conservative.
Review of Results: PUMS

- The general consistency seen in tabular aggregation is equally true for estimates generated from concatenated 1-year PUMS files relative to 3- and 5-year period PUMS files.

- PUMS files concatenation is also fairly straightforward and not quite as laborious.

- Suggest that relative differences be included when making comparisons.

- Suggest looking at measures of reliability, as well.
But Why Are There Differences?

• Don’t the exact same set of interviews (and non-interviews) contribute to both sets of estimates.

• For period estimates, all processed together (at the same time)

• For aggregate estimates, processed separately (at different times) and then joined

• Of the sources of differences, can anything be done to mitigate?
  — Answer: No, Yes, and Maybe
What is it about Period Estimates that Contributes to Observed Differences?

• “Pooling” of data records allows for more refined non-response (NR) adjustment cells and corresponding NR factors.

• Current-Year “Vintage” applied along a few dimensions:
  — Population Controls - previous years adjusted to current year vintage (most sensitive when period crosses census year)
  — Geography - current definitions applied to all years in period
  — Variable Definitions - standardized across years
  — Inflation - previous years income and housing values adjusted to current year dollars (Compass Handbook - Appendix 5 provides guidance on use of “All Items CPI-U-RS Annual Averages” to adjust for this difference)
Other Factors beyond “Vintage”

• For multi-year period estimation only, a “model–assisted” weighting step to control to sub-county populations estimates is employed. 1-year estimates do not currently employ this step.

• For multi-year period variance estimation, a finite population correction factor (FPC) is applied to appropriately reduce estimates of variance to reflect the proportion of addresses in sample. Variances for 1-year estimates do not currently employ this adjustment.

• With this expanded usage of 1-year estimates, possibly these estimation differences should be revisited?
So, a Few Things to Note

• A more comprehensive assessment would come from aggregated 3-year estimates from geographic areas with 20,000-65,000 in population
  — Smaller sample sizes will contribute to greater deviations between 3-year period and aggregated estimates.
  — Differences from some of the contributing factors listed in the previous slides will be more sensitive within 20,000-65,000 population geographies relative to 65,000+ population geographies.
  — Especially, from the weighting and variance estimation differences related to sub-county estimates and high sampling rate geographies.
So, with that in mind…

• The first 1-year (2014) supplemental estimates have been released.
  — For geographies with 20,000+ population
    — Includes 23 geography levels: nation, state, county, place, metropolitan areas, congressional & school districts
    — Includes 58 high-level detailed tables

• A review of adequacy should be conducted
Great Start, But…

• Prior to waiting two more years
  — Generate & release corresponding 2012 and 2013 1-year supplemental estimates to avoid a two year gap in the 3-year series for 20,000-65,000 population geographies.

  — In fact, releasing 2011 1-year supplemental estimates would support repeating PRB’s 3-year comparison efforts (period vs. aggregated), but now for the more directly relevant 20,000-65,000 population geographies.

  — Greater variation in estimates and MOEs likely to be observed in these smaller population geographies.
In Addition,…

• Spotted release of variance replicate estimates tables for 2010-2014 5-year estimates.

• Provided to allow users to calculate MOEs for their own aggregated estimates across geographies or categories within a table more accurately than from approximation methods in ACS documentation and employed in PRB report.

• So, knowing the intended use is to aggregate, is there merit in supplying this kind of table package for the supplemental 1-year estimates?
Looking Ahead…

• Possibly, all these supplemental estimates and ad hoc aggregation efforts could be made moot by other approaches.

• In particular, the Micro-data Analysis System (MAS) comes to mind

  — Online remote access system to allow users to request custom tables from underlying micro-data

  — Possibly, expand to allow for aggregation over time in addition to geography and variable categories
Micro-data Analysis System (MAS)

• However, the current MAS development is very much challenged by desire to maximize utility while simultaneously maintaining confidentiality standards.

• Mostly driven by complexities associated with likely need to apply greater degree of data perturbation.

• But I wonder if restricting use to “aggregation over time” with fixed tables (i.e., fixed geographies & categories) would overcome confidentiality concerns and expedite development/release, at least for this capability of aggregating over time?
Another Consideration…

• Model-Based Estimates


• Used 1-year and 5-year period estimates to model 3-year estimates for every county in the US
5-, 3-, and 1-Year Period Estimates of Median Household Income (2013)

Each period estimate has a relatively large measure of uncertainty.
(Bayesian) Model-based based estimates (g) and (h) use the 1-year period estimates and the 5-year period estimates from the previous slide, but do not use the 3-year period estimates.
Going further...

• Might be interesting to compare modeled results for counties with 20,000-65,000 populations with corresponding 3-year period published estimates

• Also wonder about how the modeling would be “improved” with inclusion of the new supplemental 1-year estimates for these 20,000-65,000 population counties being made available
Some Resources: Census Bureau

• ACS General Compass Handbook

• Design & Methodology Report - 2014
  — https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

• Accuracy of the Data – PUMS 2010-2014
  — http://www2.census.gov/programs-surveys/acs/tech_docs/pums/accuracy/2010_2014AccuracyPUMS.pdf
Some Resources: Census Bureau (2)

• Table and Geography Changes

• Supplemental Materials
Additional Resources…

  — https://fcsm.sites.usa.gov/files/2016/03/J3_Schar_2015FCSM.pdf

Thanks for the opportunity.

DavidHubble@Westat.com