Investigating Methods to Support Subannual Estimates in the American Community Survey

Keith A. Albright, Mark E. Asiala
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract[1]

The Census Bureau uses American Community Survey (ACS) data to produce, on an annual basis, one-year, three-year, and five-year estimates, for time periods defined by calendar years. The data for these estimates are collected in yearly samples that are equally distributed across monthly panels, but monthly response patterns are not equal. Numerous external and internal stakeholders have expressed interest in estimates that are based on shorter subannual time periods, such as monthly estimates. The variable monthly response patterns mean that annualized weights, from the pooled samples, cannot be used to produce subannual estimates. This paper presents results of a pilot project that attempts to optimize the existing weighting methodology to produce monthly estimates of health insurance coverage using data from the monthly ACS samples. If successful, these methods could be used to study trends in fast changing characteristics or possible seasonal patterns in the data in a way that is not currently possible with annualized weights.

Key Words: American Community Survey, weighting, variance estimation

1. Introduction

Each year, roughly 3.5 million addresses are selected for the ACS (U.S Census Bureau 2014, Ch. 4). The selected sample is randomly allocated among the 12 month of the year. For each month's sample, questionnaires are mailed and there is a three month window for data collection. Data collected in the first month is by mail[2] or internet response[3]. In the second month computer assisted telephone interviewing (CATI) is attempted to collect data from addresses that did not respond during the first month. In the third month, a sample of addresses that have still not responded is selected for computer assisted personal interview (CAPI) by Census Bureau field representatives. The data collected throughout the year is pooled together and put through a process of editing/imputation, disclosure avoidance procedures, and weighting. The weighted data is then used to produce estimates that cover one-year, three-year, and five-year time periods defined by calendar years. Stakeholders have expressed interest in estimates based on shorter time periods, such as monthly or quarterly estimates. King (2009) attempted to use annualized weights for subannual estimates and concluded that these weights were

[1]This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical and methodological issues are those of the authors and not necessarily those of the U.S. Census Bureau.
[2] Mail returns are accepted throughout the three month period.
[3] Internet data collection began in January 2013.

not suitable for subannual estimates. In this effort, the annualized weights from each month's interviewed sample were multiplied by 12, using them to create monthly estimates of the total population and poverty rates. The resulting monthly estimates were very erratic, even for demographic characteristics that should be stable throughout the year, with significant differences where there should not be any. It was not ruled out that subannual estimates could ever be produced, but a new methodology for weighting monthly (or other time period) samples would be needed. This research is the result of a pilot project, using relatively simple methods, to develop weighting methods by which the ACS data could be used to create subannual (monthly or some other time period) estimates. If successful, this data could then be used to study trends in fast changing characteristics like health insurance coverage. The methods developed could be used by researchers to study rapidly changing trends or possible seasonal patterns in the data in a way that is not currently possible with the production annualized weights. The weighting methodology was applied, independently, to each of the monthly tabulated ACS samples. We created monthly estimates of health insurance coverage, plus a few other characteristics, and some of these results are presented in this paper.

This paper is a summary of a much larger report in the ACS Research and Evaluation Program (Albright 2015), which is available online. That report has more detail on the methods and shows more monthly estimates that can be shown here.

2. Methods

2.1 Data Used

Throughout this paper, we refer to the concepts of sample month and tabulation month. The sample month for a selected address is simply the month whose sample the address was randomly allocated. The tabulation month for an address is the month in which data for that address was actually collected, whether a completed questionnaire or its status as a noninterview was determined. Because of the three-month data collection window, the selected and tabulated samples for any given month are not the same. The ACS estimates published each year are based on the year's tabulated samples, which include some cases that were actually selected for the prior year's sample.

ACS data from the April 2010 – June 2013 (39 months) tabulated samples was used to produce monthly estimates. These dates were chosen for two reasons. The first is that April 2010 is the earliest date for which independent monthly estimates of population and housing units, which are used in weighting the data, were available. June 2013 was chosen as the end date because we wanted monthly estimates that were not affected by the 2013 government shutdown (information about the samples from two months before and after these dates are also used in the early stages of the weighting methodology). The group quarters (GQ) sample was not included because the GQ sample is not uniform throughout the year. National-level monthly estimates that are created do not include Alaska because remote areas of Alaska are sampled only twice a year. Consequently, national estimates created in this project only include the 48 contiguous states, Hawaii, and the District of Columbia.

Independent monthly estimates of residential population and housing units, used as controls for post stratification adjustments in weighting, were provided by the Census

Bureau's Population Estimates Program (PEP)[4]. Housing unit estimates were provided at the county level. Population estimates were provided at the state level for the resident population, crossed by the following demographic variables: sex, single year of age, Hispanic origin, and six categories of race. We had to estimate the household population since the monthly PEP estimates are for the residential population and we excluded the GQ population. We assumed that the GQ population was stable throughout the year (we believe this is a reasonable assumption at the state level, but it clearly would not be for substate areas, especially in areas where there is a large GQ population). We then assigned, to each month, the GQ population from the corresponding yearly PEP estimates in order to estimate monthly HU population.

2.2 Weighting Methodology

In computing an estimate for a particular month, there is a question of which sample cases will be included in that estimate: those that were selected for the sample in that month or those that were tabulated in that month. In this research, monthly estimates are created from the monthly tabulated samples. The reason for this is that a sample HU/person's status for a characteristic is the status at the time of interview, not the time of sample selection. As noted in the introduction, the selected and tabulated samples for a given month are not the same because data for a month's selected sample is collected over a period of three months. For example, the data collected in March will consist of mail and internet responses for addresses in the March sample, CATI responses for addresses sampled in February, and CAPI responses for addresses sampled in January. So while the sample that was selected for a particular month is representative of the population, the sample actually interviewed (tabulated) in that month is not necessarily representative. But using the monthly tabulated samples to create estimates depends on the assumption that the data actually collected in a month is still representative of the overall population. Addressing this issue is one of two overall goals of any weighting methodology we develop. These goals are 1) make the month's interviewed sample representative of the population and 2) make monthly housing unit and population estimates agree with the independent controls.

We weighted each month's tabulated samples independently. The weighting process is a series of ratio adjustments. Some of these steps are modifications of current methods, while others are nearly identical to current methods, differing only in the level of geography and variables used to form ratio adjustment cells. Adjustment factors are calculated separately within each state. The steps of the weighting methodology are:

- Modify the initial base weight
- Variation in monthly response factor
- Noninterview adjustment factor
- Housing unit post-stratification adjustment factor
- Person post-stratification adjustment factor
- Final Housing Unit Weight

In describing these adjustments, we often briefly compare and contrast with the current ACS weighting methods. See U.S. Census Bureau (2014, Ch. 10) for more details about the current methods.

[4] The PEP program also produces annual estimates for counties, incorporated places, and minor civil divisions which are used as controls in weighting the yearly ACS sample.

2.1.1 Modify the initial base weight

The initial annualized base weights are multiplied by 12 since each month's sample represents 1/12 of the yearly sample. The base weight includes a subsampling adjustment to account for the selection of a sample of nonresponding housing units for computer assisted personal interview (CAPI). An additional adjustment is also applied to ensure that the total weight for cases tabulated in a month is the same as the total weight for cases selected for sample in the same month. We call the resulting weight *WSSFP*.

2.2.3 Variation in monthly response factor (VMS)

As noted earlier, data collected in any particular month consists of sample cases from that month and the previous two months. Furthermore, seasonal variations in monthly response patterns result in monthly weighted housing unit totals that vary across months. The *VMS* factor is designed to smooth out this monthly distribution. It is arguably the most important factor in ensuring that a given month's tabulated sample is representative of the overall population in the way that the month's selected sample is representative. This factor is also part of the current weighting methodology and is modified in this research.

The weighted total of housing units in a tabulation month *t,* after calculation of *WSSFP*, can be written as:

$$\sum WSSFP_t = \sum WSSFP_{t,t} + \sum WSSFP_{t-1,t} + \sum WSSFP_{t-2,t} \qquad (1)$$

where the subscripts on the right hand side indicate sample month and tabulation month, in that order.

It is desired that the weighted total of housing units tabulated in month *t* be the same as the weighted total for housing units that were sampled in month *t*. Additionally, it was noted by Bell (2013) that in monthly estimates in the form of (1), it's not clear what population is being estimated. Bell proposed an estimator that also makes use of information about the HUs that were sampled in month *t*, but tabulated in months *t+1* and *t+2*. Incorporation of this idea led to our modification of the VMS adjustment used in this research. To the second and third terms of (1), we apply *VMS* factors, denoted as *VMS1* and *VMS2,* calculated as:

$$VMS1 = \frac{\text{Total } WSSFP \text{ of cases sampled in month } t \text{ and tabulated in month } t+1}{\text{Total WSSFP of cases sampled in month } t\text{-}1 \text{ and tabulated in month } t} = \frac{\sum WSSFP_{t,t+1}}{\sum WSSFP_{t-1,t}} \qquad (2)$$

$$VMS2 = \frac{\text{Total } WSSFP \text{ of cases sampled in month } t \text{ and tabulated in month } t+2}{\text{Total } WSSFP \text{ of cases sampled in month } t\text{-}2 \text{ and tabulated in month } t} = \frac{\sum WSSFP_{t,t+2}}{\sum WSSFP_{t-2,t}} \qquad (3)$$

where the subscripts on the right hand side indicate sample month and tabulation month, in that order.

These two factors are applied selectively to each sample housing unit, depending on the lag between the month that the case was selected for sample and the tabulation month. For sample cases tabulated in month *t*, but sampled in month *t*–1, the weight after *VMS* is *WVMS = WSSFP*VMS1*. For sample cases tabulated in month *t*, but sampled in month *t*–2, the weight after *VMS* is *WVMS = WSSFP*VMS2*. Then the estimate of total housing units using the *WVMS* weight is:

$$\sum WVMS_t = \sum WSSFP_{t,t} + VMS1 \sum WSSFP_{t-1,t} + VMS2 \sum WSSFP_{t-2,t}$$

$$= \sum WSSFP_{t,t} + \sum WSSFP_{t,t+1} + \sum WSSFP_{t,t+2} \qquad (4)$$

where the subscripts on the right hand side indicate sample month and tabulation month, in that order.

So after application of the *VMS* factors, the weighted total of cases tabulated in month *t* is the same as the *WSSFP* weighted total of cases sampled in month *t*. The *VMS* factor currently used in the production ACS yields the same overall result. However, in the production ACS, a single *VMS* factor is applied to all cases within the same tabulation date and geographic area. In our modification, the *WVMS* and *WSSFP* totals are equalized among cases with the same amount of lag time between their sample and tabulation months.

2.2.4 Noninterview adjustment factor (NIF)

The noninterview adjustment factor adjusts weights of interviewed housing units to account for valid housing units for which no interview is completed.[5] The production ACS uses census tract, building type (single vs multi-unit), and tabulation month to form adjustment cells because these variables have been shown to be related to housing unit response in other surveys (Weidman, Alexander, Diffendal, & Love, 1995). The production ACS actually uses two successive noninterview adjustments to account for these three variables. In this research, tabulation month is already being taken into account since we are weighting monthly samples independently. But within each tabulation month, there are three sample months which we can use in forming adjustment cells. Using census tract isn't practical with monthly samples because few tracts would have enough sample cases each month to support NIF adjustment cells, even if no other variables were used. We decided that within each tabulation month, we would use state, sample month, and building type to form adjustment cells.

If a state/sample month/building type cell does not meet the criteria to stand on its own, then we collapse across building type. However, we did not collapse any further than that (across sample month). For a cell to stand, it must have: 1) at least ten sample cases or 2) at least one sample case, with no noninterviews. When collapsing across building type, the resulting cell is defined only by tabulation month and sample month. The size of the ACS sample guarantees that this cell will have enough sample cases so it won't be necessary to collapse further. Then for each cell, the *NIF1* factor is calculated as the ratio

[5]Vacant units deleted units are not included in the NIF adjustment and it is assumed that these units are correctly identified in the field. Deleted or out-of-scope HUs consist of: (1) those that have been demolished, condemned, or are uninhabitable, (2) addresses that do not exist, and (3) addresses that identify commercial establishments, units being used permanently for storage, or group quarters (U.S. Census Bureau, 2014).

of the total *WVMS* weight of interviews and noninterviews to the total *WVMS* weight of interviews. Note that for cells with no noninterviews, *NIF1* = 1. For noninterviews, *NIF1* = 0. The *NIF1* factor is not applied to vacant or deleted housing units, so *NIF1* = 1 for those cases. Then for interviewed housing units, the weight after application of the *NIF1* factor is *WNIF1 = WVMS*NIF1*. After applying *NIF1*, the total weight of sample cases in each cell is the same as the total weight using *WVMS*. Since *NIF1* cells were defined by sample month, mode of data collection was implicitly taken into account when calculating *NIF1*, something that is not done in the production ACS. After the *NIF* adjustment, the production ACS applies a mode bias factor to account for the fact that the characteristics of CAPI cases are different from other cases (Weidman et al., 1995) and that most noninterviews occur among the CAPI sample (U.S. Census Bureau, 2014). The mode bias factor is not necessary in this weighting methodology since mode of data collection was taken into account.

2.2.5 Housing unit post-stratification factor (HPF)

The housing unit post-stratification adjustment factor is a simple ratio adjustment that equalizes the total weight of all housing units to the independent monthly estimates of housing units. This allowed us to apply *HPF* at a substate level, which helps account for differential substate coverage. We used the 2,130 ACS 1-year weighting areas (a single county or group of small counties) as the geographic level for which *HPF* was computed. The *HPF* used in the production ACS is calculated the same way, differing only in the level of geography that it's applied to. The production ACS calculates the adjustment for sub county areas (incorporated places and minor civil divisions). For each month, a weighting area's *HPF* is calculated as the ratio of the PEP estimate of housing units to the total *WNIF1* weight of all interviewed cases. The weight after application of the *HPF* is given by *WHPF = WNIF1 * HPF*.

2.2.6 Person post-stratification factor (PPSF)

The person post-stratification adjustment factor assigns weights to persons using an iterative two-dimensional raking-ratio estimation procedure (iterative proportional fitting). This procedure is what the ACS used before 2009, when subcounty population controls, a third dimension, were introduced into the raking process. The first dimension is called family equalization and enforces internal consistency between estimates that, logically, should be equal. No independent estimates are used as controls for the marginal totals in this dimension. The marginal totals are determined using the *WHPF* weights for occupied housing units. The combined estimates of spouses and unmarried partners equals the combined estimates of married-couple and unmarried-partner households and the estimate of householders equals to estimate of occupied housing units. The second dimension makes population estimates equal to the independent controls within cells defined by age, sex, and race/Hispanic origin. Cells for *PPSF* are defined by a full cross of the categories in the two dimensions. The person weight is then calculated as *WPPSF = WHPF*PPSF*.

2.2.7 Final housing unit factor (HHF)

The final housing unit weight is assigned in this step, which is identical to what is done in the production ACS. In each occupied housing unit, a householder factor (*HHF*) is assigned. *HHF* is simply the *PPSF* for that housing unit's householder. The householder is defined as the reference person. *HHF* is intended to account for householder characteristics and gives an indication of under coverage for households whose householders have the same demographic characteristics (U.S. Census Bureau, 2014

Ch.10). *HHF* is set to 1.0 for vacant housing units. The weight after application of *HHF* is given by *WHHF = WHPF*HHF*. This weight is then rounded by a controlled rounding procedure that results in the final housing unit weight.

2.5 Variance Estimation

Variances of estimates were computed the same way they are for published estimates in the production ACS, by using replicate weights that were created using the Successive Differences Replication (SDR) method (Wolter, 1984; Fay & Train, 1995; Judkins, 1990). The SDR method has been used to for variance estimation in the ACS since it began. It is useful for systematic samples where the sort order is important, like the geographic sort of the ACS sample. Eighty replicate base weights are assigned to each sample housing unit. Details of how these replicates are created and used are given in the variance estimation chapter of the ACS Design and Methodology Report (U.S. Census Bureau, 2014). The weighting process is rerun for each set of replicate weights to produce 80 final replicate weights for each sample housing unit and person, but collapsing patterns are not determined again for each replicate. For each estimate, 80 replicate estimates are computed using the replicate weights. Then the variance of an estimate Y, with replicate weights Y_r, $r=1,\ldots,80$, is given by $\mathrm{Var}(Y) = 0.05 * \sum_{r=1}^{80}(Y_r - Y)^2$

3. Results

3.1 Performance of the Weighting

The modified weighting methodology performed well overall. The sizes of the adjustment factors, from *VMS* through *PPSF* compared favorably to the corresponding factors in the production ACS. Distributions of these factors, over all tabulation months and states, are shown in Table 1 along with the distributions of corresponding factors from the 2012 ACS. The 1st through 99th percentiles of the factors were generally comparable to what is observed in the production ACS. But the production ACS typically has more extreme values. The exception to this is with the housing unit post-stratification adjustment factor (*HPF*). This factor had a larger range of values than it does in the production ACS. In the 2012 1-year ACS for example, the HPF values ranged from 0.83 to 1.36, with 98% of values falling between 0.93 and 1.14. As we stated earlier, Alaska is being excluded from the monthly estimates shown in this report. More research will be needed to devise procedures for weighting monthly remote Alaska samples.

Table 1. Distribution of Weighting Factors (over all states and months)											
	Percentile										
Factor	Min	1st	5th	10th	25th	50th	75th	90th	95th	99th	Max
VMS1	0.48	0.64	0.73	0.78	0.86	0.96	1.18	1.31	1.39	1.51	1.84
VMS2	0.84	0.89	0.93	0.95	0.98	1.01	1.03	1.06	1.08	1.16	1.27
VMS in 2012 ACS	0.48	0.74	0.82	0.86	0.93	1.01	1.10	1.20	1.28	1.50	2.89
NIF (non-unit values)	1.00	1.00	1.00	1.00	1.00	1.01	1.06	1.11	1.15	1.24	1.76
NIF in 2012 ACS	1.00	1.01	1.02	1.03	1.06	1.10	1.17	1.26	1.34	1.55	4.96
HPF	0.39	0.73	0.82	0.86	0.94	1.02	1.12	1.25	1.37	1.64	3.95
HPF in 2012 ACS	0.83	0.93	0.96	0.98	1.00	1.02	1.04	1.07	1.09	1.14	1.36
PPSF	0.16	0.49	0.65	0.74	0.90	1.04	1.26	1.54	1.77	2.36	7.24
PPSF in 2012 ACS	0.00	0.47	0.68	0.78	0.91	1.04	1.25	1.56	1.81	2.55	46.82
Source: 2010-2013 American Community Survey experimental data 2012 American Community Survey											

3.2 Estimates of Health Insurance Coverage

Monthly estimates of health insurance coverage that we computed include estimates of the number of uninsured persons and persons with insurance. Estimates included the overall uninsured as well as for uninsured rates for groups defined by categories of age, race/Hispanic origin, and poverty index[6]. Rates, for those with insurance, of coverage by public and private insurance plans were also included. Table 2 contains these rates for the United States (we present these in tabular form in order to maximize the number of subgroups shown). These estimates were also created for states which are presented in Albright (2015). While these are not official estimates and it's not our intent to draw conclusions about the rates of health insurance coverage, some features of these rates stand out that we feel merit some discussion. The overall rate of uninsured persons shows evidence of a downward trend with a few jumps or drops in the rates. In 2011, the uninsured rate jumps June 2011, dropping again in July to the same level as in May. This jump appeared driven by a jump in the uninsured rate among persons who were part of the April sample and interviewed in June. There are also notable drops in the uninsured rate in January 2011 and January 2012. It is not known at this time if these are real changes or due to anomalies in the data that could be addressed in the weighting process. The rates for private insurance and public insurance coverage also have points of interest. A spike in private insurance coverage occurs in January 2013 with a corresponding drop in public insurance coverage. At this time, the cause of this phenomenon isn't known. It's possible that it's related to the introduction of Internet response in January 2013, with the window for self-response being longer in that month than in other months. Persons with private coverage may have a higher propensity for response than those with public coverage. It could also be related to the open enrollment of private insurance plans under the Affordable Care Act. The opposite situation occurs in March 2013, a drop in private coverage along with a spike in public coverage.

[6] The poverty index is the percent that a persons income is above the poverty threshold.

Table 2. Estimates of Rates of Uninsured and Private/Public Coverage in the US										
		Age		Race/Hispanic Origin			Poverty Index		Insured	
Month	Overall	0-18	19-64	White	Black	Hispanic	0-138	138-399	Private	Public
Apr-10	15.8	8.8	21.9	13.3	19.7	32.8	30.9	20.9	79.7	24.3
May-10	15.8	8.3	22.1	13.3	19.5	33.2	31.0	20.6	79.4	24.3
Jun-10	15.7	8.8	21.8	13.1	19.5	32.9	31.8	20.1	79.9	23.8
Jul-10	15.5	8.4	21.5	12.8	19.6	32.4	31.4	19.8	80.1	23.8
Aug-10	15.7	8.3	22.0	13.2	19.8	32.4	31.2	20.3	79.7	24.1
Sep-10	15.7	8.5	21.8	13.1	19.9	32.6	31.2	20.1	79.8	24.0
Oct-10	15.8	8.6	21.9	13.5	19.5	32.9	31.2	20.1	79.6	24.2
Nov-10	15.3	8.2	21.4	13.0	19.4	31.5	30.8	19.6	79.5	24.2
Dec-10	15.5	8.1	21.6	13.1	19.1	32.1	30.6	20.0	79.4	24.4
Jan-11	15.0	8.0	21.0	12.6	18.9	31.1	30.1	19.8	79.8	24.0
Feb-11	15.1	8.0	21.1	12.7	18.2	31.8	29.7	19.9	79.5	24.3
Mar-11	15.1	7.8	21.1	12.6	19.5	31.6	30.3	19.6	79.3	24.6
Apr-11	15.1	7.7	21.2	12.7	18.9	31.4	29.9	19.8	79.0	24.8
May-11	15.1	7.8	21.2	12.9	18.6	30.7	29.5	20.0	79.1	24.7
Jun-11	15.5	8.3	21.6	12.9	18.7	32.6	31.1	20.2	79.1	24.7
Jul-11	15.1	8.0	21.1	12.7	19.1	30.9	30.6	19.4	79.4	24.5
Aug-11	15.0	7.6	21.1	12.6	18.6	31.0	30.2	19.2	79.2	24.7
Sep-11	15.4	8.2	21.5	13.0	19.1	32.0	30.7	19.6	79.0	25.0
Oct-11	15.1	7.8	21.2	12.8	19.1	30.8	29.7	19.4	78.9	24.9
Nov-11	15.2	7.9	21.4	12.9	19.0	31.5	30.3	19.8	79.1	24.9
Dec-11	15.2	7.8	21.3	13.0	18.5	31.1	30.3	19.4	79.0	24.9
Jan-12	14.8	7.7	20.8	12.4	18.5	30.7	29.7	19.2	79.2	24.6
Feb-12	14.9	7.7	21.0	12.6	18.6	31.0	29.4	19.9	79.0	24.9
Mar-12	15.0	7.7	21.1	12.7	18.7	30.9	30.2	19.4	78.7	25.1
Apr-12	15.2	7.6	21.4	13.0	18.9	30.7	29.4	19.9	78.7	25.1
May-12	14.9	7.5	21.0	12.7	18.4	30.3	28.9	19.4	78.3	25.5
Jun-12	14.9	7.5	21.0	12.5	18.1	31.0	29.7	19.4	78.9	25.0
Jul-12	14.7	7.7	20.7	12.2	18.0	31.2	29.5	19.1	78.9	24.9
Aug-12	15.0	7.7	21.2	12.7	19.1	30.8	29.5	19.7	78.6	25.2
Sep-12	14.9	7.5	21.1	12.5	18.8	30.8	29.2	19.6	78.4	25.3
Oct-12	14.8	7.3	21.0	12.6	18.2	30.5	28.6	19.6	78.6	25.3
Nov-12	14.8	7.5	20.9	12.6	18.3	30.6	29.4	19.3	78.8	25.1
Dec-12	14.6	7.6	20.6	12.2	18.7	30.2	28.7	19.1	78.8	25.0
Jan-13	14.5	7.6	20.5	12.2	18.5	30.0	29.4	19.4	79.5	24.4
Feb-13	14.7	7.8	20.6	12.4	18.7	30.2	29.4	19.1	79.0	24.9
Mar-13	14.8	7.5	20.9	12.7	18.4	30.2	28.6	19.5	78.1	25.8
Apr-13	15.0	7.6	21.2	12.7	19.4	30.4	28.9	19.9	78.5	25.4
May-13	14.6	7.4	20.7	12.4	18.6	30.1	28.2	19.4	78.4	25.4
Jun-13	14.8	7.6	20.9	12.5	18.3	30.6	28.9	19.2	78.2	25.6
Source: 2010-2013American Community Survey experimental data										

Assessment of the size of variances was done using the coefficient of variation (CV). The CV of an estimate Y is given by *CV(Y) = 100*SE(Y)/Y*, where *SE(Y)* is the standard error. Table 3 shows median CVs for the monthly estimates of health insurance coverage, both for national and state level estimates (all states except Alaska are included). The states were grouped into four categories based on population size and the median CVs for each category are shown. Generally, we consider CVs less than 10 as good, between 10 and 30 as moderate, and greater than 30 as large. The CVs for the national level estimates are generally very low, but moderate for the smallest population groups. The median CVs for state level estimates fare much worse. In the smallest states, the median CV is good for only a few estimates. For the smallest groups, the median CVs are large even in large states. The large variances for many state level estimates indicate that monthly estimates for health insurance coverage won't be adequate. But it may be worth considering subannual estimates based on larger periods, such as quarters. Since sample sizes for quarterly samples are about three times that of monthly samples, CVs would be reduced by roughly 0.58 ($1/\sqrt{3}$).

Table 3. Median CVs for Health Insurance Estimates

	Nation	**States by Population Category**			
Uninsured		< 2 million	2-5 million	5-10 million	> 10 million
Total Uninsured	0.7	11.7	5.9	4.7	3.0
Age 0-18	2.0	34.6	18.5	14.3	10.6
Age 19-64	0.6	11.4	5.6	4.4	2.9
White Non-Hispanic	1.0	14.0	7.9	6.6	4.5
Black Non-Hispanic	1.7	56.5	17.9	11.6	7.5
Hispanic	1.2	39.7	13.0	10.4	5.2
Poverty index 0-138	1.2	20.7	9.7	8.4	5.4
Poverty Index 139-399	1.1	17.8	10.1	7.7	4.8
Poverty Index 400+	2.2	31.1	19.4	14.5	8.6
Private Insurance					
Age 0-64	0.3	3.4	2.2	1.5	1.1
Age 0-18	0.5	6.9	4.4	2.9	2.1
Age 19-64	0.2	3.1	2.0	1.3	1.0
Public Insurance					
Age 0-64	0.6	9.8	5.3	4.3	2.7
Age 0-18	0.8	12.6	6.4	5.5	3.3
Age 19-64	0.7	11.7	6.8	5.0	3.3

Source: 2010-2013American Community Survey experimental data

Another assessment of the variances, related to performance of the weighting, is to compare the CVs to what would be expected if the production ACS only had 1/12 of its sample size. It's expected that the CV would increase by a factor of $\sqrt{12}$. The 1-year ACS data for 2010-2013 was used to calculate the CV for the percent of the residential population without health insurance. These yearly CVs were multiplied by $\sqrt{12}$ and compared to the mean of the monthly CVs using the relative difference. The relative difference between two numbers, a and b, is given by $(a-b)/((a+b)/2)$. These relative differences were computed for the CVs of the percent uninsured population. Table 4 shows the relative differences for the nation (excluding Alaska). At the national level, the CVs are actually smaller than would be expected. Table 5 gives a yearly summary of the relative differences among the states including the results of a sign test.

The sign test results show no evidence for the monthly CVs being systematically higher than what would be expected due to having a smaller sample.

Table 4. Comparing Mean Monthly CVs to CV from 1-year ACS Data (National Uninsured Rate)

Year	1-year ACS CV * $\sqrt{12}$	Mean Monthly CV	Relative Difference
2010	1.01	0.70	-0.36
2011	1.04	0.70	-0.39
2012	0.91	0.67	-0.31
2013	0.95	0.64	-0.38

Source: 2010-2013 American Community Survey experimental data
2010-2013 American Community Survey 1-year data

Table 5. Comparing Mean Monthly CVs to CV from 1-year ACS Data (State-level Uninsured Rates)

Year	Minimum Relative Difference	Maximum Relative Difference	Mean Relative Difference	P-Value of Sign Test for Mean = 0
2010	-0.1551	0.2241	-0.0024	0.6718
2011	-0.1455	0.2550	0.0004	1.0000
2012	-0.2076	02611	0.0094	0.8877
2013	-0.1678	0.1444	-0.0486	0.0009

Source: 2010-2013 American Community Survey experimental data
2010-2013 American Community Survey 1-year data

3.3 Other Estimates

We also created monthly estimates for a few other characteristics not related to health insurance, in order to help assess the quality and face validity of monthly estimates. Figure 1 shows national monthly estimates of the total number of African-Americans along with ACS 2010-2013 1-year estimates. The pattern shows gradual increase through time, as expected. Comparisons of the monthly estimates and 1-year estimates from 2012 and 2013 are as one would expect. But for 2010 and 2011, most of the monthly estimate are lower than the 1-year estimates. A likely cause for this is the use of different vintage PEP estimates in the weighting. Vintage 2013 estimates were used for the monthly samples, while the yearly estimates were each weighted using the vintage estimates from the corresponding year.

Figure 1. Monthly Estimates of African Americans

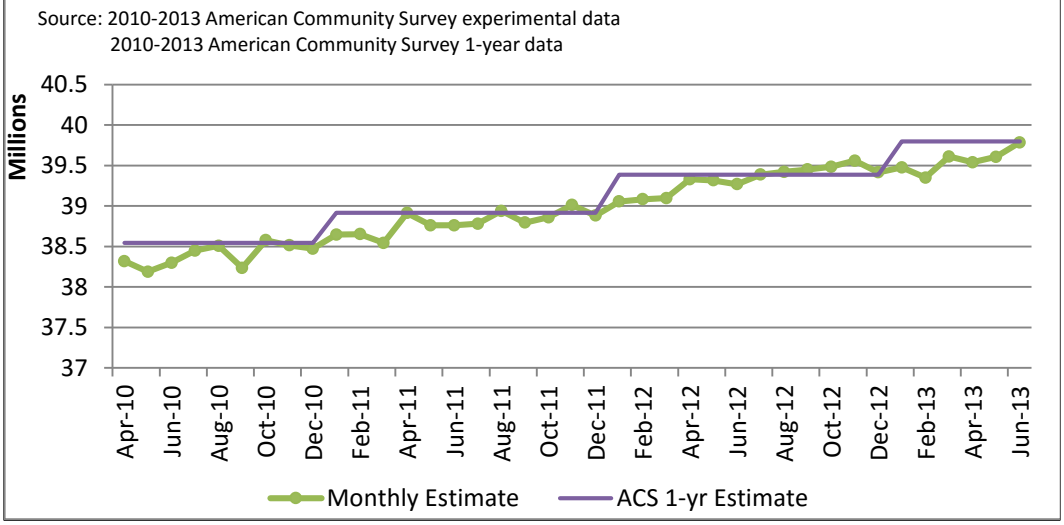
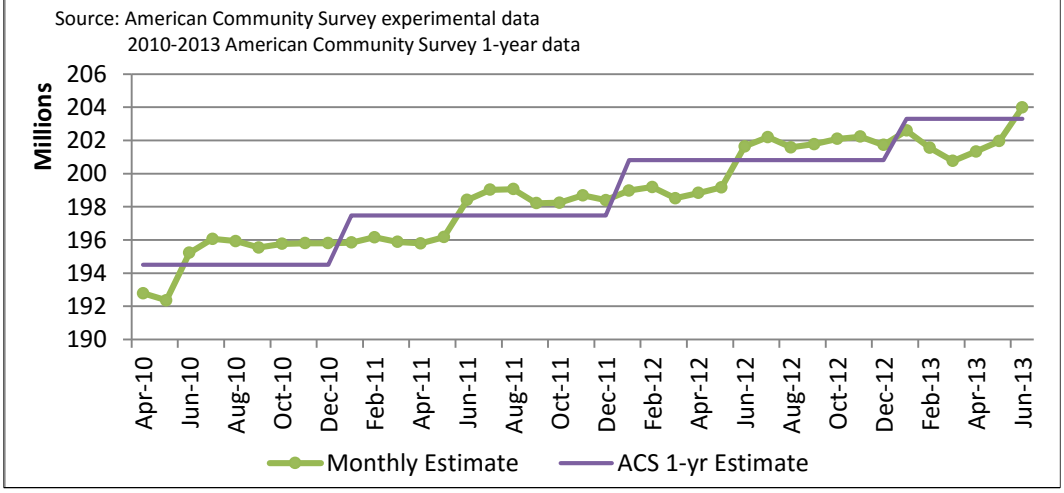


Figure 2 shows national monthly estimates of the number of persons, age 16+, with a high school diploma or higher, along with the ACS 2010-2013 1-year estimates. The pattern shows clear jumps between May and June of each year, which are statistically significant at the 90% level, corresponding with high school graduation. The estimates before and after each yearly jump are stable from month to month. However, for state level data, significant differences between May and June were detected, for each year, in less than half of all states, at the 10% significance level: 26 states in 2010, 12 in 2011, 24 in 2012, and 21 in 2013. This is troublesome because the population being measured is large and indicates that state level monthly estimates for some characteristics may be too volatile to be reliable.

Figure 2. Monthly Estimates of High School Graduates, Age 16+, in the US



4. Conclusions and Recommendations

There are numerous operational and resource issues to overcome if we wish to produce subannual estimates on a flow basis instead of after all data for a year has been collected. Probably the most significant one is how imputation will be done with samples that are only 1/12 as large as the annual samples. As already noted, group quarters persons and Puerto Rico were initially excluded from this research because more research and consideration is needed to determine how they can be included in subannual estimates. More research will also be needed to determine how the remote Alaska samples can be treated so that subannual estimates for Alaska can be created. Variances for many state level estimates of health insurance coverage are, in our opinion, unacceptably large. The weighting methods used in this research can be applied to larger time periods as easily as they were applied to monthly samples. We should also consider using modelling, using small area methods, in conjunction with the direct estimates produced to create more reliable estimates.

Acknowledgements

The authors wish to acknowledge Census Bureau personnel who assisted in data analysis, reviewed our research, provided constructive criticism, and offered additional insight: William Bell, Eric Jensen, Richard Griffin, , Alina Kline, Jerry Maples, Tucker McElroy, Brian Ohara, Anthony Tersine, Beth Tyszka.

References

Albright, K (2015), Investigating Methods to Support Subannual State Level Estimates, American Community Survey Research and Evaluation Program. Available for download at http://www.census.gov/programs-surveys/acs/library.html

Bell, W (2013), Some thoughts on using ACS data to construct poverty estimates comparable to CPS estimates – DRAFT, unpublished notes. Washington, DC: U.S. Census Bureau.

Fay, R., & Train, G. (1995). Aspects of Survey and Model Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties. Joint Statistical Meetings: Proceedings of the Section on Government Statistics (pp. 154-159). Alexandria, VA: American Statistical Association.

Judkins, D. R. (1990). Fay's Method for Variance Estimation. Journal of Official Statistics, 6(3), 223-239.

King, K (2009), Issues Related to Adding Sub-annual Estimates to the Data Products Available from the American Community Survey, Presentation to the ACS Research Group Meeting at the U.S. Census Bureau on Oct. 1, 2009.

U.S. Census Bureau (2014), Design and Methodology: American Community Survey, U.S. Government Printing Office. Available for download at http://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html

Weidman, L., Alexander, C., Diffendal, G., & Love, S. (1995). Estimation Issues for the Continuous Measurement Survey. JSM Proceedings, Survey Research Methods Section (pp. 596-601). Alexandria, VA: American Statistical Association.

Wolter, K. M. (1984). An Investigation of Some Estimators of Variance for Systematic Sampling. Journal of the American Statistical Association, 79, 781-790.