UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

2015 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES #ACS15-RER-09

DSSD 2015 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS15-R-03

MEMORANDUM FOR ACS Research and Evaluation Advisory Group

From: Patrick J. Cantwell
Chief, Decennial Statistical Studies Division

Prepared by: Keith A. Albright ***signed 4/4/16***
American Community Survey Estimation Branch
Decennial Statistical Studies Division

Mark E. Asiala
Chief, American Community Survey Statistical Design
Decennial Statistical Studies Division

Subject: Investigating Methods to Support Subannual State-Level Estimates

Attached is the final American Community Survey Research and Evaluation report for "Investigating Methods to Support Subannual State-Level Estimates." This is a pilot project to develop methods by which American Community Survey (ACS) data could be used to support historical subannual estimates of health insurance coverage or other subannual estimates of interest. We develop a weighting methodology to independently weight monthly ACS samples and create monthly estimates of health insurance coverage as well as other monthly estimates in order to evaluate the methodology.

If you have any questions about this report, please contact Mark E. Asiala (301-763-3605) or Keith A. Albright (301-763-5999).

Attachments

cc:
ACS Research and Evaluation Work Group

Anthony G. Tersine	(DSSD)
Karen King	
Beth Clarke Tyszka	(ACSO)
Jerry J. Maples	(CSRM)
Tucker S. McElroy	
Eric B. Jensen	(POP)
Brian Ohara	(SEHSD)

American Community Survey Research and Evaluation Program

January 15, 2016

Investigating Methods to Support Subannual State-Level Estimates

FINAL REPORT

Keith A. Albright and
Mark E. Asiala
Decennial Statistical
Studies Division

United States Census Bureau

Table of Contents

1 Executive Summary 1

2 Introduction 1

3 Research Questions 2

4 Methodology 2

4.1 Data Used 2

4.2 Weighting Methodology 4

4.3 Estimates and Variance Estimation 13

5 Results… 15

5.1 Performance of the Weighting 16

5.2 Monthly Health Insurance Estimates 18

5.3 Other Monthly Estimates 21

6 Conclusions and Further Research 23

7 References 24

Attachment A. Monthly Health Insurance Estimates for the United States 26

Attachment B. Monthly Health Insurance Estimates for California 34

Attachment C. Monthly Health Insurance Estimates for Kentucky 40

Attachment D. Monthly Health Insurance Estimates for Massachusetts 46

Attachment E. Monthly Health Insurance Estimates for Mississippi 52

Attachment F. Monthly Health Insurance Estimates for North Dakota 58

Attachment G. Monthly Health Insurance Estimates for Texas 64

1 Executive Summary

Staff in the American Community Survey Estimation Branch (ACSEB), Decennial Statistical Studies Division (DSSD), took ACS data collected from April 2010 through June 2013 to create monthly estimates of health insurance coverage, along with other select characteristics. The ACS weighting methodology was modified to independently weight the monthly samples to produce estimates for the entire household population. The weighting methodology performed well by the standards that ACSEB uses to evaluate weighting in the yearly production data. Monthly estimates showed interesting and expected results at the national level. However, state-level estimates seem less useful, except for large states. This is due to variability in the data, which is particularly problematic in small states. The results of this study suggest that further research in this topic is worth pursuing based on the results of the weighting and the production of monthly estimates for large domains. We recommend, however, exploring subannual estimates for quarterly time periods or enhancing the monthly estimates using small-area methods as a means to mitigate the high variance in the small states.

2 Introduction

This research was intended to be a pilot project, using relatively simple methods, to develop weighting methods by which the ACS data could be used to support historical subannual estimates of health insurance coverage. The weighting methodology we developed could be applied independently to monthly samples. The intent was not to develop a method to produce subannual estimates on a flow basis, in conjunction with data collection, but rather to generate subannual estimates using the existing production microdata. These 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 ACS data pooled into yearly and multiyear samples.

The weights attached to the production ACS data are annualized. That is, they are designed to compute estimates that are actually averages over the periods of data collection that are defined by calendar years. Earlier attempts to use annualized weights for subannual estimates were not successful (King, 2009). King's efforts to produce monthly estimates simply took each month's tabulated data and multiplied the annualized weights by 12, using them to create monthly estimates of the total population and poverty rates; the resulting monthly estimates were erratic, even for demographic characteristics that should be stable throughout the year. It was concluded that annualized weights were not suitable for subannual estimates. However, it was not ruled out that subannual estimates could ever be produced, but a methodology for weighting monthly (or other time period) samples would be needed.

3 Research Questions

There were three main questions that we attempted to answer in this project. They were:

1) Can we develop a weighting methodology to independently weight the monthly samples to produce direct estimates?

2) Do the monthly estimates produced seem "reasonable" and stable?

3) Are variance estimates of monthly estimates low enough to make the estimates useful?

4 Methodology

This section summarizes the data and methodology used for this research. Section 4.1 describes the ACS sample data and independent estimates that were used. The independent estimates were provided by the Census Bureau's Population Division for use as post-stratification controls in weighting the sample data. Section 4.2 describes the weighting methodology, comparing and contrasting it with the production ACS methods.[1] Section 4.3 outlines the health insurance coverage and other estimates, including variances, which were produced to evaluate the estimation methodology.

4.1 Data Used

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[2]. Data for a month's selected sample is collected over a period of three months. For example, in the sample selected for January, data for internet[3] and most mail respondents[4] is collected in January, data for computer assisted telephone interview (CATI) respondents is collected in February, and data for computer assisted personal interview (CAPI) respondents is collected in March. Data tabulated in January includes CATI cases selected for December and CAPI cases selected for November. In this research, a month's estimates are based on the sample cases tabulated in that month. Our justification for this is that a sample housing unit (HU) or person's status for a characteristic is the status at the time of interview, not the time of sample selection. However, the appropriateness of the estimator depends on the assumption that the data actually collected in a month is still representative of the overall population.

---

[1] Throughout this report, the term 'production ACS' refers to, depending on the context, the weighting methodology as described in the ACS design and methodology report (U.S. Census Bureau, 2014), or the estimates produced by those methods.

[2] The tabulation sample for a month consists of the sample cases whose status, as an interview or noninterview, was determined in that month. This is different from the group of housing units that were selected for that month's panel and had questionnaires mailed to them because of the three-month data collection window in the ACS.

[3] Internet data collection began in January 2013.

[4] Mail returns are accepted throughout the three-month period.

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 the Population Division (POP) could provide us with monthly estimates of population and housing units, which are used in weighting the data. 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 is also used in the early stages of the weighting methodology).

The group quarters (GQ) sample was not included in this project from the outset. In the course of our research activities, we realized that more research would be needed to create monthly estimates for Alaska because of the nature of the remote Alaska samples. The GQ and remote Alaska samples are not uniform throughout the year (remote Alaska locations are only sampled in January and September). Consequently, national estimates created in this project only include the 48 contiguous states, Hawaii, and the District of Columbia.

We used the unswapped microdata from the production ACS for weighting and estimation. Swapping was not considered necessary since we are only producing estimates for the nation and states.[5] Additionally, the ACS swapping procedure affects the tabulation date of records. This is an important consideration because the relationship between the panel date and the tabulation date is important in the weighting process.

Vintage 2013[6] independent monthly estimates of population and housing units, used as controls for post-stratification adjustments in weighting, were provided by the Population Division's Population Estimates Program (PEP). These monthly estimates were derived using a combination of high-level monthly data and linear interpolation of the yearly estimates. Consequently, the monthly estimates steadily increase over the year rather than exhibit any monthly or seasonal cyclical trends. State-level population estimates were provided for the resident population, crossed by the following demographic variables: sex, single year of age, Hispanic origin, and race. The race categories are White Alone, Black Alone, American Indian/Alaska Native (AIAN) Alone, Asian Alone, Native Hawaiian/Other Pacific Islander (NHOPI) Alone, and multi-race. These are the same demographic variables for which the PEP provides yearly estimates that are used in the production ACS, with the exception of multi-race. The monthly PEP estimates only include a single estimate for the multi-race population, while the yearly PEP estimates provide estimates for all 31 combinations of the five major race groups. For this project, we combined the NHOPI and Asian categories since the NHOPI population is too small to be used for monthly controls.

Since the monthly PEP estimates are for the residential population, and we excluded the GQ population in this study, we had to estimate the household population. We assumed that the GQ population was stable throughout the year; the creation of the monthly estimates also uses this

[5] Prior to weighting the data that is used for publication, the microdata on a portion of sample units is swapped with the microdata of other sample units. This is one of several disclosure avoidance methods used in the ACS.

[6] The vintage of PEP estimates refers to the year they were created. PEP population and housing unit estimates are created each year and include revisions to PEP estimates from earlier years.

assumption[7]. We then assigned, to each month, the GQ population from the corresponding yearly PEP estimates in order to estimate monthly HU population.

4.2 Weighting Methodology

The weighting process has two main goals. The first goal is to ensure that the tabulated sample for a month is representative of the sample that was selected for that month. As noted earlier, the selected and tabulated samples for any given month are not the same due to the three-month data collection window. So while the selected sample for a given month is representative of the population[8], the tabulated sample for the same month is not. The second goal is for estimates of housing units and population totals, for certain demographic groups, to equal the PEP estimates.

The methodology we developed for weighting monthly samples is, as in the production ACS, a series of ratio adjustments. Some of these steps are nearly identical to what is done in the production ACS, differing only in the level of geography and variables used to form ratio adjustment cells. Others are new or substantially modify a step used in the production methodology. The steps of the weighting methodology are

1. Calculation of the initial base weight and CAPI subsampling weight
2. Calculate and apply the CAPI correction factor
3. Calculate and apply the variation in monthly response factor
4. Calculate and apply the noninterview adjustment factor
5. Calculate and apply the housing unit post-stratification adjustment factor
6. Calculate and apply the person post-stratification adjustment factor
7. Calculate the final housing unit weight

These weighting steps were performed independently on the tabulated samples for each of the months April 2010 through June 2013. All ratio adjustments were calculated separately within each state and the District of Columbia. For simplicity, throughout this section, we describe the adjustments and give formulas without reference to state, with the understanding that these adjustments are calculated within each state.

4.2.1 Initial Base Weight (*BW*) and CAPI Subsampling Weight (*WSSF*)

For this study, the initial annualized base weights from the production ACS, *BW*, were multiplied by 12 since each month's sample represents 1/12 of the yearly sample. Then the CAPI subsampling factor, *SSF*, was applied in the same manner that is currently done in the production ACS. The weight after the CAPI subsampling factor is called *WSSF*.

[7] This assumption is made by necessity, although it's known that it's probably not realistic.

[8] The ACS sample is selected on a yearly basis. The sample selected for a year is randomly allocated among the 12 months, with each month having approximately 1/12 of the total yearly sample.

4.2.2 <u>CAPI Correction Factor (*SSFP)*</u>

This factor was devised for use in this research and is not used in the production ACS. In the production ACS, the total *WSSF* weight in each monthly sample has historically been, on average, about one percent lower than the total base weight due to a slight bias in the weights introduced by late self-response returns. We corrected for this by applying another adjustment to CAPI cases that we call the CAPI correction factor, or *SSFP*. For a given tabulation month *t*, *SSFP* is calculated as

$$SSFP_t = \frac{\sum_{i=1}^{n_t} I*BW_{t,i}}{\sum_{i=1}^{n_t} I*WSSF_{t,i}} \qquad (1)$$

where n_t is the number of HUs tabulated in month *t*, $BW_{t,i}$ and $WSSF_{t,i}$ are weights of the i^{th} HU in tabulation month *t* ; I_i is an indicator variable based on HU *i* where

$$I_i = \begin{cases} 1 \text{ if the interview status for HU } i \text{ is resolved in the CAPI mode} \\ 0 \text{ otherwise} \end{cases}$$

Note that I_i = 1 for all cases that are resolved through CAPI including interviews, non-interviews, and those that were not selected for follow-up but had no late self-response return. *SSFP* is 1 for mail and CATI cases. Application of this factor yields the adjusted weight *WSSFP = WSSF*SSFP*. The total of *WSSFP* in each tabulation month is then equal to the total of *BW* for the same tabulation month.

4.2.3 <u>Variation in Monthly Sample (*VMS)*</u>

The Variation in Monthly Sample Factor, or *VMS,* is used in the production ACS to smooth the distribution of weighted housing unit totals by tabulation month. This factor is important because, for the production ACS, the goal is to produce an annual estimate which is an average across the months within the calendar year. Seasonal variations in monthly response patterns, especially in the self-response modes, result in monthly weighted HU totals that can vary widely from month to month. The *VMS* adjustment corrects for this by adjusting the weights so that the weighted total of HUs tabulated in a given month after the adjustment equals the total *WSSFP* weight of HUs that were sampled in that month. In the production ACS, all sample HUs tabulated in the same month, in a given geographic area, receive the same *VMS* adjustment. For this research, we developed an alternative *VMS* that is applied at a more detailed level than in the production ACS, while still achieving the equality of weighted totals described above.

As noted earlier, data collected in any given month comes from sample units belonging to three different monthly samples: the sample selected for the same month and the samples that were selected for the two previous months. Consider an estimate of some characteristic *Y* in a given tabulation month *t,* which is calculated by summing the weights of sample units that were tabulated in month *t*. This estimate can be expressed as the sum of three component estimates

$$\hat{Y}_{.t} = \hat{Y}_{tt} + \hat{Y}_{(t-1)t} + \hat{Y}_{(t-2)t} \qquad (2)$$

where $\hat{Y}_{st}$ is the weighted total of sample units sampled in the sth month and tabulated in month t and $\hat{Y}_{.t}$ is the weighted total of all sample units tabulated in month t and

s$\in [t-2, t]$, t = 1,…,12 since for any given tabulation month, the sample unit can only come from one of three sample months. Note a negative value of s corresponds to a month in the previous calendar year where s = –1 corresponds to December of the previous year and *s* = –2 corresponds to November.

One difficulty regarding estimates from monthly estimates using the annualized weights in the form of (2) is that it isn't entirely clear what population is being estimated (Bell, 2013). As we noted in the beginning of this section, the sample tabulated in a given month is not representative of the population we are attempting to estimate. Ideally, equation (2) would be based on the sample month *s* rather than the tabulation month *t* since the sample from a given sample month is representative of the nation. Bell proposed a type of post-stratification estimator that leverages the representativeness of the sample month but makes use of the characteristic data collected in the tabulation month. His proposed estimator was expressed in terms of component estimates separated by mode of data collection (mail, CATI, and CAPI). We substituted sample month in place of mode of data collection because mail, CATI, and CAPI cases from a given sample month do not separate neatly into different tabulation months. Late mail returns can come in during the following two months. Some CATI-like cases from telephone questionnaire assistance are tabulated in the month they were sampled.

Bell's proposed estimator takes the weighted total of all cases that were *sampled* in a given month and modifies its component estimates. This total, for a given sample month *s*, can be written as

$$T_{s.} = T_{ss} + T_{s(s+1)} + T_{s(s+2)} \tag{3}$$

where T_{st} is the weighted total for all sample units sampled in month *s* and tabulated in month *t* and $T_{s.}$ is the weighted total for the sample month regardless of the tabulation month.

If we were to make an estimate for *Y* in month *t* we would multiply each component of $T_{s.}$ by the weighted proportion of characteristic *Y* that is measured for that component.

$$\hat{Y}_{.t} = T_{tt}P_{tt} + T_{t(t+1)}P_{t(t+1)} + T_{t(t+2)}P_{t(t+2)} \tag{4}$$

where $P_{st} = \frac{Y_{st}}{T_{st}}$ is the weighted proportion of all sample units that have the characteristic *Y* in sample month *s* and tabulation month *t*.

However, using the weighted proportions for the characteristic *Y* from multiple tabulation months would create a composite estimate across months whose mixture rates would be dependent on the variation in self-response rates from month to month. Thus Bell proposed substituting the proportions for sample month *s* in equation (4) with the corresponding proportions that are observed in tabulation month *t* under the assumption that the bias introduced by this substitution would be small.

With this substitution in the proportions, Bell's alternative estimator to (2) is given below.

$$\hat{Y}'_{.t} = T_{tt}P_{tt} + T_{t(t+1)}P_{(t-1)t} + T_{t(t+2)}P_{(t-2)t} \qquad (5)$$

For this final estimator, all of the component totals T_{st} come from the sample month s equal to the tabulation month t, and all of the component proportions come from the tabulation month t. It is in this manner that the estimator can be thought of as post-stratified estimator where the components are the strata.

This post-stratification is made more clear by rearranging the fractions in (5)

$$\hat{Y}'_{.t} = \frac{T_{tt}}{T_{tt}}\, Y_{tt} + \frac{T_{t(t+1)}}{T_{(t-1)t}}\, Y_{(t-1)t} + \frac{T_{t(t+2)}}{T_{(t-2)t}} Y_{(t-2)t} \qquad (6)$$

Note that the Y_{st} terms in (6) are the same as those in (2). The fractions in (6) can be thought of as adjustment factors to the terms in (2).

These adjustment factors are then used to define our revised *VMS* adjustment. We define three different adjustments based on the relation between sample and tabulation month. For a given sample month s and tabulation month t, *VMS* is applied as follows

$$VMS_{st} = \begin{cases} 1; & \text{if } s = t \\ \dfrac{\sum_{i=1}^{n_{t(t+1)}} WSSFP_{t(t+1)i}}{\sum_{i=1}^{n_{(t-1)t}} WSSFP_{(t-1)ti}}; & \text{if } s = t-1 \\ \dfrac{\sum_{i=1}^{n_{t(t+2)}} WSSFP_{t(t+2)i}}{\sum_{i=1}^{n_{(t-2)t}} WSSFP_{(t-2)ti}}; & \text{if } s = t-2 \end{cases} \qquad (7)$$

Where n_{st} = number of sample HUs in sample month s and tabulation month t and $WSSFP_{sti}$ is the i^{th} sample HU in sample month s and tabulation month t.

Note that the numerator is always based on a component of the sample month $s = t$ and the denominator is always based on a component of the tabulation month t. Applying the VMS adjustment defined in (7) to the *WSSFP* weights yields the updated HU weight *WVMS*. The total weight for all HUs in tabulation month t after applying the *VMS* factor will equal the total *WSSFP* weight for all HUs in sample month $s=t$. This is the same result as the single *VMS* adjustment factor per tabulation month used for the production ACS. However, the *VMS* adjustment developed here also preserves the total *WSSFP* weight for each of the three values of lag between sample and tabulation month. For example, the total of *WVMS* for HUs tabulated in month t and sampled in month t -2 equals the total of *WSSFP* for HUs sampled in month t but not tabulated until month t + 2. Both of these groups of cases have a lag of two months between sample and interview month. This is an important consideration because there can be significant differences in the characteristics of HUs with differing lags between sample and tabulation month. Applying the *VMS* adjustment in this fashion helps ensure that the tabulated sample for a month is representative of the population in the same way that the sample selected for that month is representative.

4.2.4 Noninterview Adjustment Factor (NIF)

The noninterview adjustment factor adjusts weights of interviewed HUs to account for valid HUs for which no interview is completed. 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 HU response in other surveys (Weidman, Alexander, Diffendal, & Love, 1995). The production ACS 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. All adjustment cells must meet one of the following two conditions (which are identical to what is used in the ACS production)

- At least ten sample cases
- At least one sample case, with no noninterviews

Using census tract is not practical with monthly samples because very few tracts would have enough sample cases to meet the conditions, 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, we collapse across building type. However, we did not collapse any further than that (across sample month).

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 a given sample month *s*, tabulation month *t*, and building type *b*, *NIF1* is computed as follows

$$NIF1_{stb} = \frac{\sum_{i \in \text{Interviews}} WVMS_{stbi} + \sum_{i \in \text{Noninterviews}} WVMS_{stbi}}{\sum_{i \in \text{Interviews}} WVMS_{stbi}} \qquad (8)$$

where $WVMS_{stbi}$ is the *WVMS* weight of the i^{th} sample HU in sample month *s*, tabulation month *t*, and building type *b*.

Note that for cells with no noninterviews, *NIF1* = 1.0. Deleted and vacant HUs are not included in the computation of the *NIF1* adjustment[9]. The weights for these HUs remain unchanged during this stage of the weighting process since it is assumed that all vacant and deleted units are properly identified in the field and therefore are not eligible for the noninterview adjustment (U.S. Census Bureau, 2014). The updated HU weight, *WNIF1*, is then defined as

[9] 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).

Interview Status	*WNIF1*
Occupied and temporarily occupied HU	*WVMS*NIF1*
Vacant and deleted HU	*WVMS*
Noninterview	*WNIF1* = 0

After applying *NIF1*, the total weight of sample HUs in each *NIF1* adjustment cell is the same as the total weight obtained using *WVMS*. Consequently, the totals of the *WNIF1* and *WVMS* weights in a tabulation month are the same.

Since *NIF1* cells were defined by sample month, mode of data collection was implicitly taken into account in the calculation of *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, Alexander, Diffendal, & Love, 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 already taken into account.

4.2.5 Second Noninterview Adjustment Factor

We considered using a second *NIF* adjustment after *NIF1* for this research. This second adjustment was developed specifically for this research and does not correspond to any weighting adjustment in the production ACS. In this adjustment, census tracts were grouped together into clusters that had similar levels of health insurance coverage. Tract-level insurance coverage rates were determined using the 2013 ACS 5-year data. *NIF* adjustment cells were then formed by state and tract cluster. Use of this adjustment would lead to two sets of final weights. One weight would incorporate this second *NIF* adjustment and only be used for estimates of health insurance coverage. The other weight would only incorporate the first *NIF* adjustment and be used for other estimates. We compared health insurance coverage estimates based on both weights and found virtually no difference between the estimates. Thus, we decided not to use this factor and it will not be considered further in this report.

4.2.6 Housing Unit Poststratification Factor (HPF)

The HU post-stratification adjustment factor is a simple ratio adjustment that equalizes the total weight of all HUs to the monthly PEP estimates, which were provided at the county level. 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 a given tabulation month *t,* the *HPF* for a weighting area applied as follows

$$HPF_t = \frac{H_t}{\sum_{i=1}^{n_t} WNIF1_i} \qquad (9)$$

where H_t is the PEP estimate of housing units in the weighting area for month t and $WNIF1_i$ is the *WNIF1* weight of the i^{th} sample HU in tabulation month t.

4.2.7 Person Post-stratification Factor (PPSF)

The person post-stratification adjustment factor is used to assign weights to sample persons. This factor is also used in the production ACS. For this research, the *PPSF* is computed using methodology very similar to what is currently used in the production ACS. It is computed using an iterative two-dimensional raking-ratio estimation procedure (iterative proportional fitting). It is designed so that

- First dimension -- 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 HUs.
- Second dimension -- Estimates for defined demographic groups equal those derived from PEP estimates.

These two dimensions are also used in the production ACS, with only minor differences in this research. The production ACS has a third dimension, using population controls for incorporated places and minor civil divisions,[10] which we do not do in this research. The first dimension is also known as family equalization and equalizes estimates in the first bullet point, while the second dimension equalizes the estimates in the second bullet point. The marginal totals (controls) for the first dimension don't come from an independent source, but are determined using the *WHPF* weights for occupied housing units. This is to enforce internal consistency for estimates that should, logically, be equal. The marginal totals for demographic groups are derived from the PEP estimates.

This section provides an overview of the raking process, including how sample cases are assigned to raking cells and how race/age/sex groups are collapsed to form demographic cells. For more detailed information about this process, see Chapter 11 of the ACS design and methodology report (U.S. Census Bureau, 2014).

There are four family equalization cells, whose marginal totals are determined by *WHPF* weights and the total population from the PEP estimates. These categories, along with their marginal totals are shown in Table 1 below.

[10] This third dimension was introduced into the ACS weighting methodology in 2009.

Table 1. Family Equalization Categories

Family Equalization Cell	Marginal Total
Householder in a married couple or unmarried partner relationship	Total *WHPF* weight of HUs containing such a householder
Spouse or unmarried partner of householder	Equals marginal total of first cell
Other householders	Difference between total *WHPF* weight of all occupied HUs and sum of marginal totals in first two cells
All other persons	Difference between PEP total population estimate and sum of marginal totals in the other three cells

The demographic groups are formed by crossing race, Hispanic origin, sex, and age. First, race and Hispanic origin combined into six categories we refer to as *weighting race group (WRG).* The *WRGs* are

1. White Non-Hispanic
2. Black Non-Hispanic
3. AIAN Non-Hispanic
4. Asian/NHOPI Non-Hispanic
5. Multi-race Non-Hispanic
6. Hispanic

The production ACS also uses six race groups, with minor differences.[11] Within each *WRG,* persons are placed in age/sex groups (*ASG*) formed by crossing sex by the following 13 age categories: 0-4, 5-14, 15-17, 18-19, 20-24, 25-29, 30-34, 35-44, 45-49, 50-54, 55-64, 65-74, and 75+. These are the same age/sex groups that are used in the production ACS.

This yields 156 potential *WRG/ASG* combinations. The detailed PEP estimates were grouped into these cells to create marginal totals to use for raking in the demographic dimension. Naturally, many combinations do not have enough sample cases (or have a zero population estimate) to stand alone and need to be collapsed with other cells. For a cell to stand, it must satisfy both of the following

- Have at least 10 sample persons
- The ratio total *WHPF* weight to the PEP estimate is between 1/3.5 and 3.5

These two requirements are also used in the production ACS. We first test the sample size and ratio for each weighting race group without regard to age/sex group. If the *WRG* fails, it is

[11] In the production ACS, Asian and NHOPI are separate categories. It also does not use a multi-race category because the yearly PEP population estimates include estimates for detailed multi-race combinations that allow multi-race persons to be assigned to a single race group for weighting purposes.

collapsed with another race. The resulting race groups after any required collapsing are referred to as *collapsed weighting race groups (CWRG)*. Then with each *CWRG*, the age/sex groups cells are tested and collapsed as necessary. The resulting age/sex cells are referred to as *collapsed age/sex groups (CASG)*.

All sample persons are then assigned the *WHPF* weight of their housing unit and placed in family equalization cells and race/age/sex cells. Then the *WHPF* weight is used to calculate initial totals for all combinations of family equalization and race/age/sex cells. With these initial totals created, the iterative raking process is then carried out. An iteration of the raking consists of two successive ratio adjustments to marginal totals. To illustrate, consider the sample raking-ratio matrix in Table 2. For this simple example, we use only four categories of race/age/sex. In the table entries, T_{ij} is the total for persons in family equalization cell i and race/age/sex cell j at the beginning of an iteration. F_i and D_j are the marginal totals for family equalization category i and race/age/sex category j, respectively. For the first iteration, the T_{ij} are computed using the *WHPF* weights for each sample person.

Table 2. Example of Raking-Ratio Matrix

Family Equalization Cell	Race/Age/Sex Cell				Family Equalization Marginals
	1	2	3	4	
1	T_{11}	T_{12}	T_{13}	T_{14}	F_1
2	T_{21}	T_{22}	T_{23}	T_{24}	F_2
3	T_{31}	T_{32}	T_{33}	T_{34}	F_3
4	T_{41}	T_{42}	T_{43}	T_{44}	F_4
Race/Age/Sex Marginals	D_1	D_2	D_3	D_4	PEP Estimate of Total Population

The first adjustment in an iteration ratio adjusts the totals in each family equalization category to the family equalization marginal totals to get revised totals:

$$T'_{ij} = \left(\frac{F_i}{\sum_j T_{ij}}\right) T_{ij} \qquad (10)$$

The second adjustment in the iteration takes the revised totals in (10) and ratio adjusts them to the race/age/sex marginal totals to get new revised totals:

$$T''_{ij} = \left(\frac{D_i}{\sum_i T'_{ij}}\right) T'_{ij} \qquad (11)$$

The revised totals in (11) are used to start the next iteration. The iterated adjustment process continues until a convergence criterion is met. After the last iteration, *PPSF* adjustment factors are computed for each combination of family equalization and race/age/sex categories. For a given combination of family equalization category i and race/age/sex category j in tabulation month t, *PPSF* is calculated as:

$$PPSF_{ijt} = \frac{T^*_{ijt}}{\sum_{k=1}^{n_{ijt}} WHPF_{ijk}} \qquad (12)$$

where n_{ijt} is the number of sample persons in family equalization category *i*, race/age/sex cell *j*, and tabulation month *t*. $WHPF_{ijk}$ is the *WHFP* weight of the k^{th} sample person in family equalization category *i*, race/age/sex cell *j*, and tabulation month *t*.

and

T^*_{ij} is the total for family equalization category *i*, race/age/sex cell *j*, and tabulation month *t* after the last iteration of the raking process.

Then for a person in family equalization category *i*, race/age/sex category *j*, and tabulation month *t*, the person weight is calculated as $WPPSF = WHPF_t * PPSF_{ijt}$. These weights are rounded to yield the final person weight which is used to calculate estimates.

4.2.8 Final Housing Unit Factor (HHF)

This step assigns the final housing unit weight and is identical to what is done in the production ACS. A householder factor (*HHF*) is assigned for each housing unit. *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). *HHF* is set equal to the *PPSF* adjustment of the reference person (householder) in occupied units. It is set to 1.0 for vacant housing units. The housing unit weight after HHF is then calculated as *WHHF = WHPF*HHF*. These weights are then rounded to yield the final housing unit weight.

4.3 Estimates and Variance Estimation

This section describes the monthly estimates that were calculated to evaluate the methods described in the preceding sections. We focus on estimates of health insurance coverage, which was one of the largest reasons that this research was conducted, but we also look at a few other characteristics to help assess the validity of the methods.

4.3.1 Estimates of Health Insurance Coverage

Monthly estimates of health insurance coverage include estimates of the rate of uninsured persons and rates of public/private coverage among persons with insurance. Estimates of uninsured rates include the overall rate as well as uninsured rates for groups defined by categories of age, race/Hispanic origin, and poverty index. The microdata variable *HICOV* (binary indicator for coverage) was used to determine a person’s coverage status. Coverage by private and public health insurance plans was determined using the microdata variables *PRIVCOV* and *PUBCOV* (binary indicators for private and public coverage, respectively). The microdata variables *RCGP* (race group) and *HSGP* (Hispanic origin group) were used to classify persons into race and Hispanic origin categories. The microdata variable *POVPI* was the source each person’s poverty index (ratio of income to the poverty threshold). Table 3 shows the categories of persons for who estimates of insurance coverage were computed.

These estimates of health insurance coverage were computed for the nation and states. Although we created estimates for all states, this report shows data for six selected states: California, Kentucky, Massachusetts, Mississippi, North Dakota, and Texas.

Table 3. Categories for Estimates of Health Insurance Estimates

Uninsured Rate
Overall
Age 0-18
Age 19-64
Age 0-64 by Race/Hispanic origin (White Non-Hispanic, Black Non-Hispanic, Hispanic)*
Poverty index 0-138
Poverty index 139-399
Poverty index 400+
With Insurance
Private Insurance, Age 0-64
Private Insurance, Age 0-18
Private Insurance, Age 19-64
Public Insurance, Age 0-64
Public Insurance, Age 0-18
Public Insurance, Age 19-64

*State-level estimates shown in this report only include White Non-Hispanic and all other minorities combined.

4.3.2 Other Estimates

Monthly estimates for a few other characteristics, not related to health insurance, were calculated to help assess the performance of the weighting methodology. These characteristics include ones that are expected to have slow steady increase over time, as well as ones that are expected to exhibit seasonal trends. These characteristic are:

- Total African American population
- Age 16+ with high school diploma or higher
- Vacant housing units

Monthly estimates for the African American population should be stable from month to month, with gradually increasing trends due to population growth. Monthly estimates of high school graduates should be stable from month to month, with a significant increase between May and June, due to new high school graduates. Seasonal trends are expected for estimates of vacant housing units in certain states such as Florida.

4.3.3 Variance Estimation

Variance 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 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. With the sample HUs in geographic sort order, 80 replicate base weights are assigned to each sample HU. Replicate factors used to create the replicate weights were re-assigned to the monthly tabulated samples rather than using the ones assigned in the production weighting. 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 HU and person. However, collapsing patterns for adjustment factors that may require collapsing are retained from the full sample weighting[12] and 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* is given by:

$$\mathrm{Var}(Y) = \frac{4}{80}\sum_{r=1}^{80}(Y_r - Y)^2 \quad (13)$$

where Y_r is the estimate computed using the r^th^ replicate weight.

The size of variance estimates will be evaluated using the coefficient of variation (CV). The CV is a measure of variance that is independent of the size and units of the estimate. It is the ratio of the standard deviation to the estimate, expressed as a percentage:

$$\mathrm{CV}(Y) = 100 * \frac{\sqrt{\mathrm{Var}(Y)}}{Y} \quad (14)$$

5 Results

There are three aspects to the results discussed in this section. Section 5.1 covers the performance of the weighting methodology, including comparisons to the production ACS. Section 5.2 covers monthly estimates of health insurance coverage. Charts showing national-level estimates are shown in this section, while charts showing state-level estimates are in the appendices. Section 5.2 also includes a summary of national and state-level variances of these estimates. Section 5.3 covers monthly national-level estimates for other characteristics where we expect to see seasonal trends or gradual change over time. Examining monthly estimates of these types of characteristics helps assess the validity of the weighting methodology.

[12] "Full sample" refers to the weights that are used to produce the point estimates.

5.1 *Performance of the Weighting*

5.1.1 Size and Distribution Weighting Adjustment Factors

The modified weighting methodology performed well overall. The sizes of the adjustment factors compared favorably to the corresponding factors in the production ACS. The percentile distributions of these factors, over all tabulation months and states, are shown in Table 4 along with distributions of the 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. The ranges of the factors are also smaller than those observed in the production ACS, with the exception of *HPF*. This factor had a larger range of values than it does in the production ACS. In the 2012 1-year ACS for example, *HPF* values ranged from 0.83 to 1.36, with 98 percent of values falling between 0.93 and 1.14.

Table 4. Distribution of Weighting Factors (over all states and months)

	Percentile										
Factor	**Min**	**1st**	**5th**	**10th**	**25th**	**50th**	**75th**	**90th**	**95th**	**99th**	**Max**
VMS1	0.480	0.638	0.732	0.783	0.863	0.961	1.175	1.305	1.389	1.513	1.839
VMS2	0.843	0.889	0.930	0.948	0.977	1.005	1.031	1.059	1.082	1.157	1.269
VMS in 2012 ACS	0.48	0.735	0.818	0.861	0.933	1.009	1.095	1.200	1.280	1.502	2.892
NIF1 (non-unit values)	1.000	1.001	1.002	1.002	1.004	1.013	1.063	1.113	1.145	1.244	1.758
NIF1 in 2012 ACS	1.001	1.008	1.022	1.030	1.061	1.104	1.168	1.261	1.342	1.551	4.955
HPF*	0.394	0.732	0.817	0.864	0.940	1.020	1.116	1.250	1.357	1.643	3.946
HPF in 2012 ACS	0.832	0.933	0.964	0.981	1.000	1.017	1.039	1.068	1.090	1.137	1.362
PPSF	0.163	0.486	0.652	0.744	0.896	1.038	1.255	1.543	1.772	2.361	7.240
PPSF in 2012 ACS	0.000	0.470	0.679	0.777	0.913	1.041	1.252	1.556	1.808	2.551	46.822

*HPF distribution excludes Alaska

5.1.2 Collapsing of Weighting Adjustment Cells

Two of the weighting adjustment factors, *NIF1* and *PPSF*, can require collapsing of adjustment cells so that all cells satisfy specified criteria, which is also true in the production ACS. The amount of collapsing that was required compared favorably to the production ACS. Among the initial *NIF* cells, only 0.2 percent of nearly 12,000 (across all states and months) had to be collapsed because of insufficient sample. All of the cells that required collapsing were for the multi-unit building type (recall that in each tabulation month, *NIF* cells are defined by state, sample month, and building type). In the 2012 ACS, the first and second noninterview adjustments required collapsing of 4.1 percent and 2.6 percent, respectively, of initial adjustment cells[13].

The collapsing of race and age/sex cells for the *PPSF* adjustment also compared favorably with the production ACS. Across all tabulation months, an average of six percent of race group cells cells did not meet the criteria from section 4.2.7 and required collapsing with another race group. After collapsing race groups, an average of 55 percent of age/sex cells required collapsing. In the 2012 1-year production ACS, 41 percent of race group cells, followed by 64 percent of age/sex cells did not meet the criteria. This indicates that, as long as we keep *PPSF* adjustments at the state level, we can use six race groups and cross them with the same 26 age/sex groups that are used in the production ACS, instead of using less detailed groups.

5.1.3 Coverage

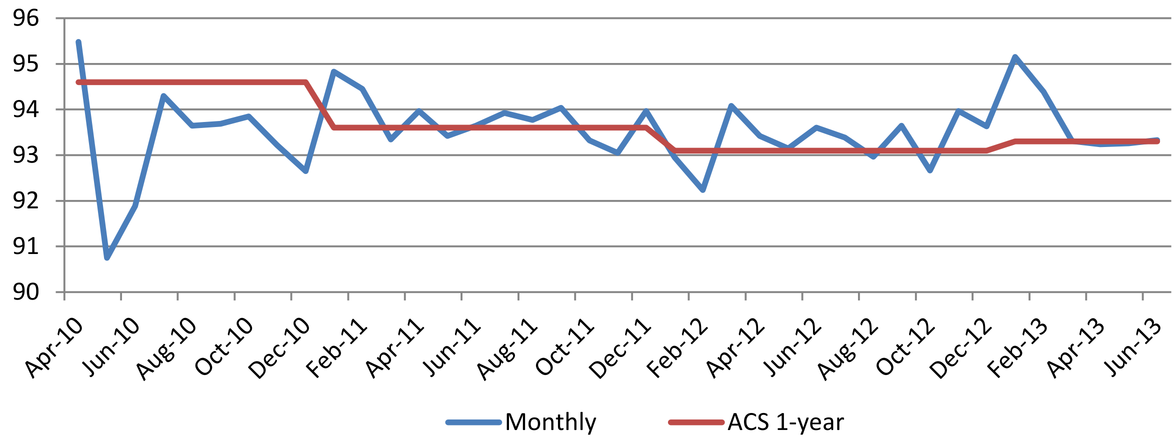
The population coverage rate is the ratio of the pre-controlled population estimate to the PEP population estimate. The pre-controlled population estimate is computed using the weight prior to post-stratification adjustments (*WNIF1*). Population coverage rates for the monthly samples are comparable to the annual coverage rates in the production ACS, though they have a larger range. Monthly coverage rates for the nation ranged from 90.7% to 95.5%, with a median of 93.5%. The production ACS has 1-year coverage rates ranging from 93.1% to 94.6% in the years 2010-2013. Figure 1 shows the monthly population coverage rates along with the rates from 2010-2013 ACS 1-year data[14].

---

[13] In the production ACS, cells for the first noninterview adjustment are formed by building type and census tract. Cells for the second noninterview adjustment are formed by county, building type, and tabulation date.

[14] The low value of 90.7%, in May 2010, is noticeably lower than the other months, which are much closer to the 1-year coverage rates from the production ACS. The total base weight of sampled housing units tabulated in May 2010 was also lower than in other months. In the 2010 production ACS, this anomaly was overcome through the noninterview adjustments. However, that was not the case with the weighting adjustments in this research, leading to a *WNIF1* based population estimate that was lower, compared to the independent estimate, than in other months. More investigation is required to determine why this was the case and how it could be compensated for.

Figure 1. Total Population Coverage Rates



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

5.2 Monthly Health Insurance Estimates

Monthly estimates of health insurance coverage for the nation, are shown on the graphs in Attachment A. Estimates for six selected states are shown in Attachments B through G. The estimates shown are for rates of uninsured and public/private coverage, in the categories of persons listed in Table 3. The figures also show 90% confidence intervals for the estimates.

Although we note trends that can be seen on these graphs, this is only for descriptive purposes and to draw attention to points of interest. The observations we make are not the results of statistical testing. It is not the purpose of this report to draw conclusions about the level of insurance coverage.

5.2.1 National-Level Estimates

A decreasing trend in the rate of uninsured persons can be seen in Figure A-1. It can also be seen in the estimates of uninsured for age groups 0-18 (Figure A-3) and 19-64 (Figure A-4). Among race groups, the only evident trend is for white non-Hispanics (Figure A-5). It could be argued that there are slight downward trends among black non-Hispanics (Figure A-6). Among poverty index categories, the 139-399 category (Figure A-9) shows a downward trend. The 0-138 category (Figure A-8) shows evidence of an overall downward trend, punctuated by in the middle of the series. The over 400 category (Figure A-10) shows a downward trend which then increases towards the end of the series.

Among persons with health insurance coverage, trends can be seen in the categories of private and public insurance, which are shown in Figures A-12 through A-17. In the age group 0-18, there is a downward trend in private insurance coverage (Figure A-12) and a corresponding upward trend in public insurance coverage (Figure A-15). In the 19-64 age group, there are upward trends in private and public insurance coverage (Figure A-13, Figure A-16). The trends noted in the private and public insurance coverage have one notable anomaly. In January 2013 there is a spike in private insurance coverage and a corresponding drop in public insurance coverage. Figure A-11 and Figure A-14 (0-64 age group) illustrate this. At this time, the cause of this phenomenon is not known. It's possible that it is 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. These characteristics are not the only ones where an unexplained jump occurred in that month. We also observed a spike in the number of persons with a bachelor degree or higher (not shown in this report). The monthly coverage rates described earlier also have a spike in that month.

The estimates for uninsured persons are presented another way in Figure A-2. Figure A-2 shows the rates of uninsured persons categorized by the lag between sampling and data collection (sample month and tabulation month). There is a clear separation between rates for the three groups. The lowest uninsured rates are among persons who were interviewed the same month they were sampled. The highest uninsured rates are among persons who were interviewed two months after they were sampled. Persons who were interviewed one month after they were sampled were in between, but closer to the rates of persons with no lag between sample month and interview month. All three series are mostly level across the 39 months and mostly free of large changes from one month to the next (we looked at these series for some states, but the larger variability made it difficult to draw any conclusions). These results provide additional support for taking the lag into account in the weighting methodology. However, there are isolated points that are cause for concern. For example, in June 2011, there is a bump in the rate for those with two months lag with no corresponding increase in the other two series. Figure A-2 also shows a sudden bump up in the rate of uninsured persons, suggesting that this temporary increase is being entirely driven by those sample persons with a two month lag between sample and interview month. Additional modifications to the weighting methodology may be able to address these kinds of anomalies.

5.2.2 State-Level Estimates

Trends that were observed at the national level largely disappear when looking at individual states. Most noticeable trends in state-level estimates are in California (Attachment B) and Texas (Attachment G), where trends are evident only in the total number of uninsured and the largest subgroups, and the trends in Texas are very slight with variability in the data making them hard to observe.

The absence of trends in state-level estimates may be because there are truly no trends for states included here or because high variances make the data too noisy. A discussion about variances of the estimates, both state and national level, is in section 5.2.3 (page 20).

Another issue with state-level estimates is with sample sizes (or no sample) for small domains. Small sample sizes lead to high variances for the estimates. Having no sample for a domain in some months results in estimates of zero for those months, reducing the usefulness of a time series for that domain.

5.2.3 Variance Estimates

Assessment of variance estimates was done using the coefficient of variation (CV). Table 5 shows median CVs of the monthly health insurance coverage estimates. These medians are given separately national and state-level estimates. States were grouped into four categories based on population size and median CVs were determined within each category. Generally, we consider CVs less than 10 percent as good, between 10 percent and 30 percent as moderate, and greater than 30 percent as large. The CVs for the national level estimates are generally very low, but moderate for the smallest population groups. The median CVs at the state-level fare worse. In the smallest states, the median CVs are good for only a few estimates. For the smallest groups, the median CVs are large even in large states.

Table 5. Median CVs (as a percentage) for Health Insurance Estimates

Uninsured	**Nation**	**States by Population Category**			
		< 2 million	2-5 million	5-10 million	> 10 million
Total Uninsured	0.68	11.65	5.86	4.72	3.03
Age 0-18	2.04	34.67	18.40	14.27	10.65
Age 19-64	0.64	11.32	5.56	4.36	2.89
White Non-Hispanic	0.99	13.99	7.91	6.54	4.49
Black Non-Hispanic	1.71	48.56	17.64	11.55	7.44
Hispanic	1.18	39.44	12.95	10.37	5.18
All Minorities	0.90	20.36	8.69	7.14	3.86
Poverty index 0-138	1.18	18.29	7.68	6.90	4.73
Poverty index 139-399	1.13	16.70	9.34	7.05	4.38
Poverty index 400+	2.16	31.14	19.37	14.51	8.58
With Private Insurance					
Age 0-64	0.25	2.56	1.64	1.12	0.83
Age 0-18	0.47	6.41	4.10	2.71	2.00
Age 19-64	0.24	1.91	1.20	0.82	0.60
With Public Insurance					
Age 0-64	0.62	9.90	5.38	4.34	2.70
Age 0-18	0.77	12.25	6.24	5.38	3.16
Age 19-64	0.72	11.38	6.67	4.94	3.24

Source: American 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

$$\text{Relative Difference} = \frac{(a-b)}{(a+b)/2} \quad (15)$$

Table 6 shows the relative differences using CVs for overall rate of uninsured persons at the national level CVs. For each year, the mean monthly CVs are actually smaller than what would be expected if the production ACS had 1/12 of its sample size. Table 7 gives a yearly summary of the relative differences among the states (N=50 for each year), including the results of a sign test. The sign test results show no evidence for the monthly CVs being systematically higher or lower than what would be expected.

The large variances for many state-level estimates indicate that monthly state-level estimates won't be useful. However, it may be worth considering subannual estimates based on larger periods, such as quarters. Sample sizes for quarterly samples are about three times that of monthly samples, so CVs would be reduced by roughly $(1/\sqrt{3})$.

Table 6. Comparing Mean Monthly CVs to CV from 1-year ACS Data (National Uninsured Rate)

Year	Production 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

Table 7. 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 (N=50)
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

5.3 Other Monthly Estimates

This section discusses monthly estimates for three characteristics, where it's expected to see gradual change over time or seasonal effects. These estimates are shown, alongside 1-year ACS estimates from the years 2010-2013, in Figures 2, 3, and 4. Figure 2 shows national monthly estimates for the total black/African American population. The pattern shows gradual increase through time, as expected. Comparisons of the monthly estimates and 1-year estimates from 2012 and 2013 are as expected. Monthly estimates in 2010 and 2011 are largely lower than the 1-year estimates. A likely cause for this is the use of different vintage PEP estimates in the weighting. Vintage 2013 PEP estimates were used for weighting the monthly samples. The ACS 1-year estimates that are shown were created from the production data of the corresponding years.

Figure 2. National Monthly Estimates of the Black/African American Population

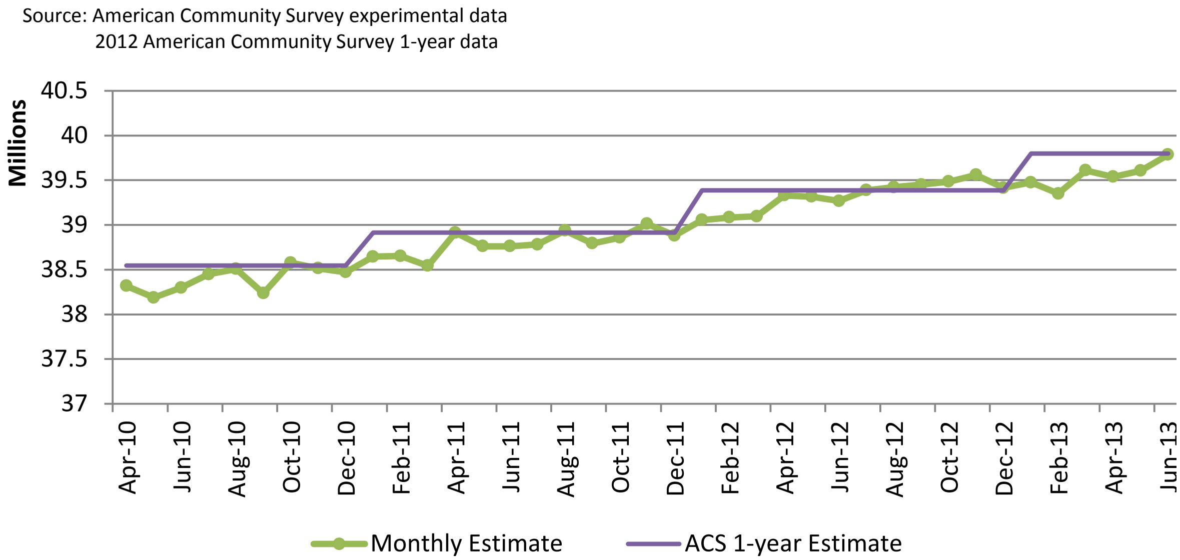


Figure 3. National Monthly Estimates of High School Graduates, Age 16+

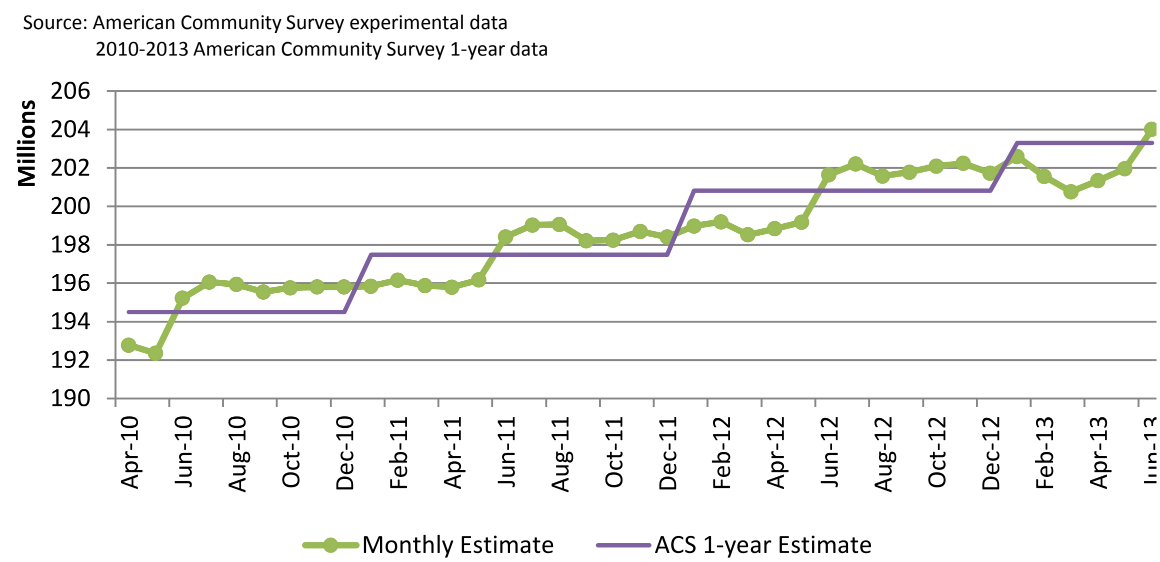
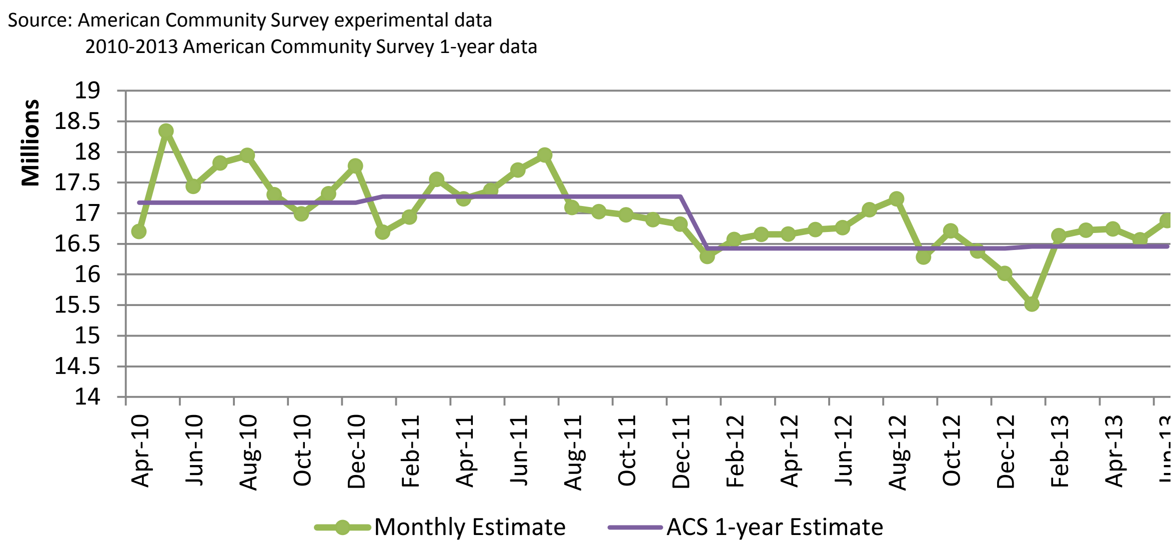


Figure 3 shows national monthly estimates of the number of persons, age 16+ with at least a high school diploma. This is an example of a characteristic where seasonal trends are expected, namely a jump in the estimate every June. The pattern shows clear jumps between May and June of each year, corresponding 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 will be too volatile to be reliable.

Figure 4 shows national monthly estimates of the number of vacant housing units, which is another characteristic where seasonal trends are expected. Evidence of a seasonal pattern that was expected can be seen, with estimates falling to their lowest points in the winter and peaking in the summer months.

Figure 4. National Monthly Estimates of Vacant Housing Units



6 Conclusions and Further Research

This research was intended to support a project where subannual estimates would only be produced after the production microdata for an entire year was collected and subject to the regular production processing (editing, imputation, etc.). Estimates produced in this fashion would be less useful than if they were released periodically throughout the year. However, there are significant operational and resource issues in producing this data on a flow basis. One of the most significant issues is how imputation will be done with samples that are only 1/12 as large as the annual samples. In the production ACS, editing and imputation is done after all data for a calendar year is collected. We had the benefit of using prior years' production data for research, but monthly estimates would not be very useful if we could only produce them on a yearly basis. Other issues include data capture and coding; tabulation; budget and personnel availability; and timing of data release.

As already noted, the group quarters, Alaska, and Puerto Rico samples were excluded from this project because of the nature of these samples. The group quarters sample is weighted separately from the housing unit sample, using an entirely different process. A separate research project will be needed to develop a methodology to weight monthly group quarters samples (or determine if it is even feasible). In addition, more research is needed to determine how the samples from remote parts of Alaska, which are only taken twice a year, can be treated so that monthly estimates for Alaska can be created.

Our assessment of the modified weighting methods for weighting the monthly samples was that they performed very well. The weighting adjustment factors were consistent with values in the production ACS and did not have the extreme values that are sometimes observed in the production ACS.

Variances for many monthly state-level estimates of health insurance coverage are too large for them to have practical use. However, variances of estimates for large states and the nation indicate that reliable state-level estimates can be produced for larger domains. 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. This approach has been successfully used for other products produced by the Census Bureau, such as the Section 203 determinations (Joyce, et al., 2014).

7 References

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.

Joyce, P. M., Malec, D., Little, R. J., Gilary, A., Navarro, A., & Asiala, M. E. (2014). Statistical Modeling Methodology for the Voting Rights Act Section 203 Language Assistance Determinations. *Journal of the American Statistical Association, 109*(505), 36-47.

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

King, K. (2009, October 1). Issues Related to Adding Sub-annual Estimates to the Data Products Available from the American Community Survey. *Federal Committee on Statistical Methodology: Proceedings of the 2009 Research Conference*.

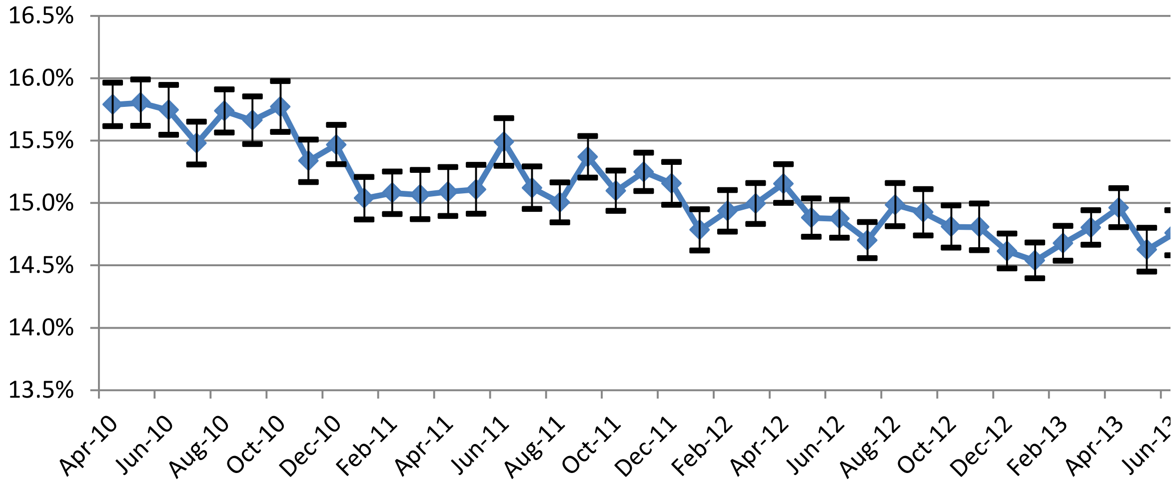
U.S. Census Bureau. (2014). *Design and Methodology: American Community Survey.* Washington, DC: U.S. Government Printing Office. Retrieved from http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology.pdf

Weidman, L., Alexander, C., Diffendal, G., & Love, S. (1995). Estimation Issues for the Continuous Measurement Survey. *Joint Statistical Meetings: Proceedings of the Section on Survey Research Methods* (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*(388), 781-790.

Attachment A.

Figure A-1. Monthly Estimates of Uninsured Persons in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

Figure A-2. Monthly Estimates of Uninsured Rate by Lag between Sample Date and Data Collection

Source: American Community Survey experimental data

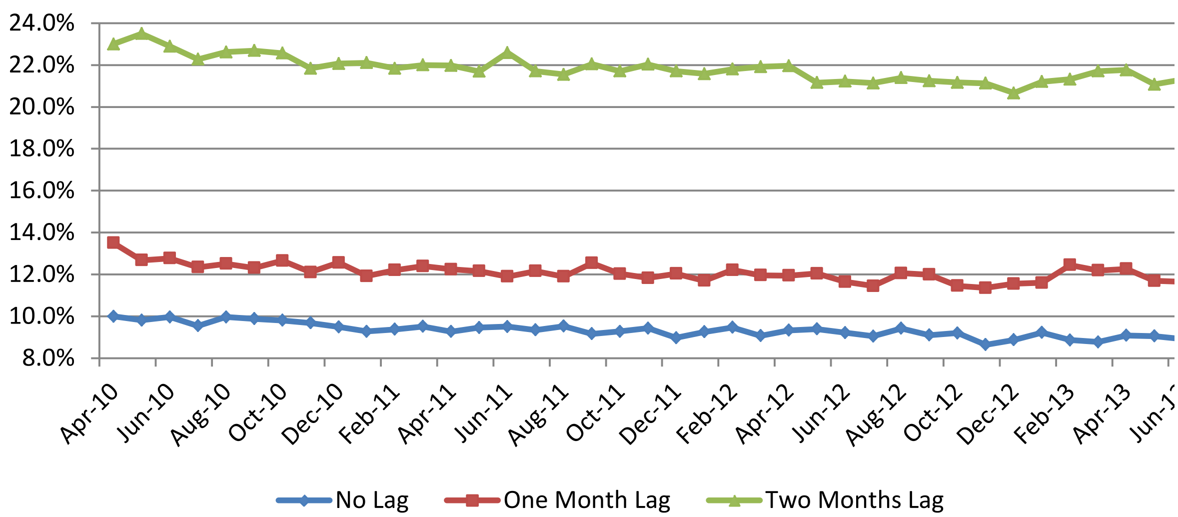
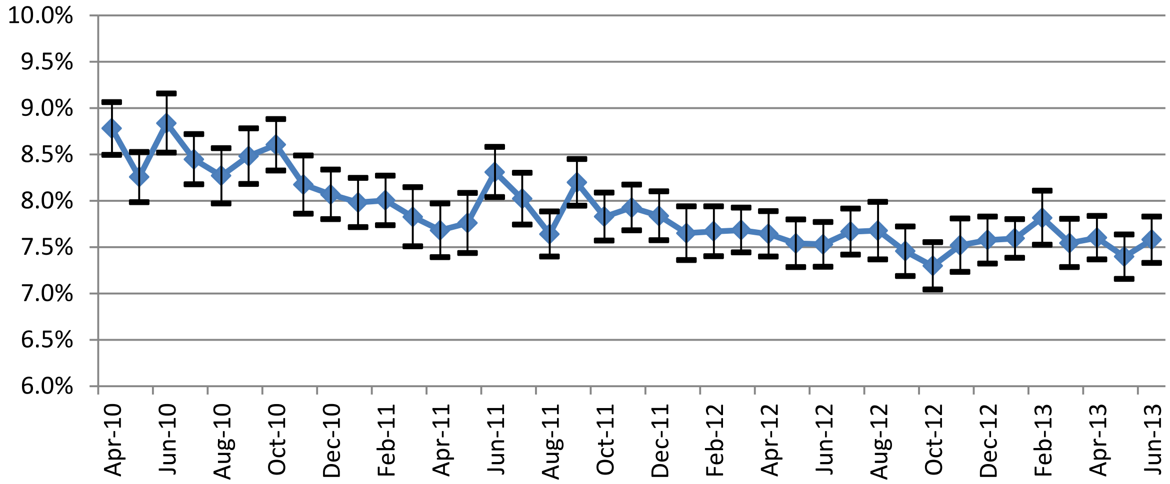


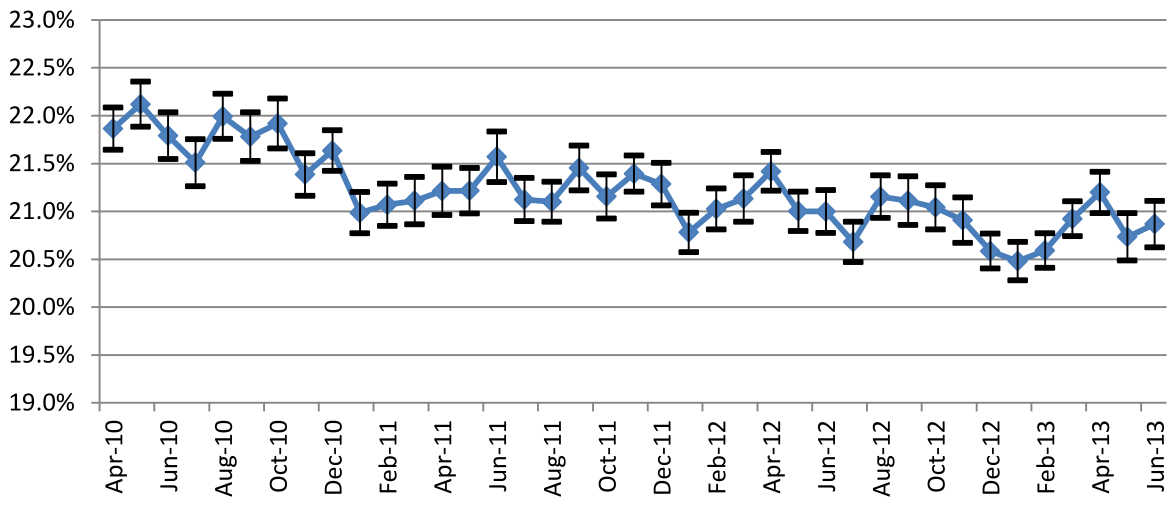
Figure A-3. Monthly Estimates of Uninsured Rate for Age 0-18 in the United States



Source: American Community Survey experiemntal data

Excludes Alaska
Error bars show 90% confidence intervals

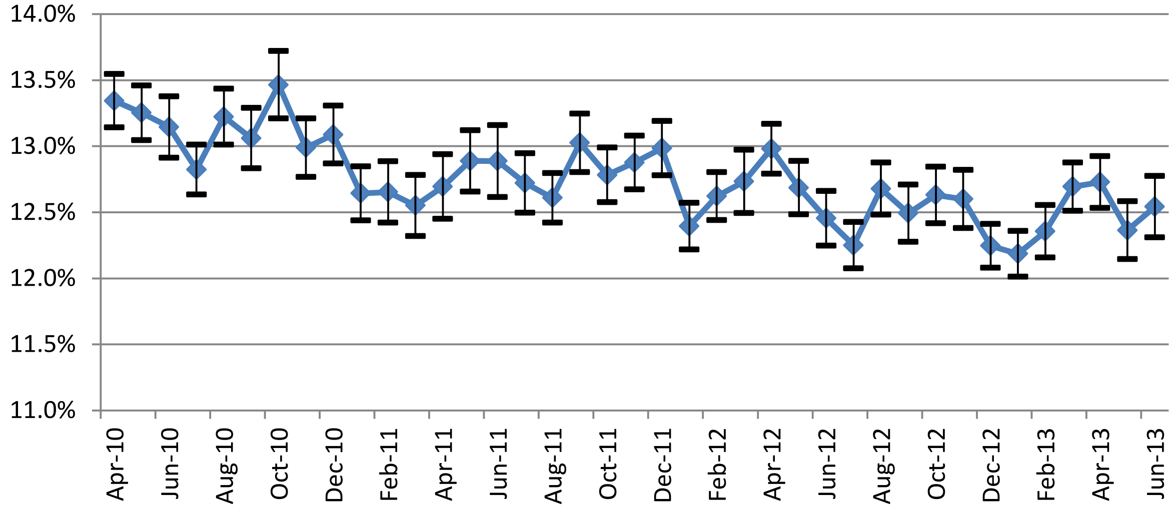
Figure A-4. Monthly Estimates of Uninsured Rate for Age 19-64 in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

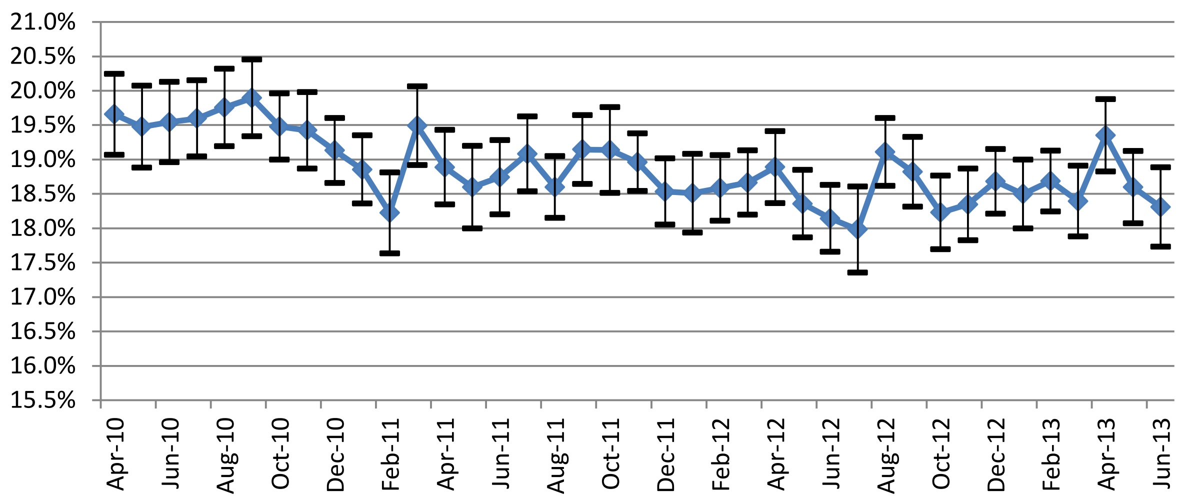
Figure A-5. Monthly Estimates of Uninsured Rate for White Non-Hispanics, Age 0-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

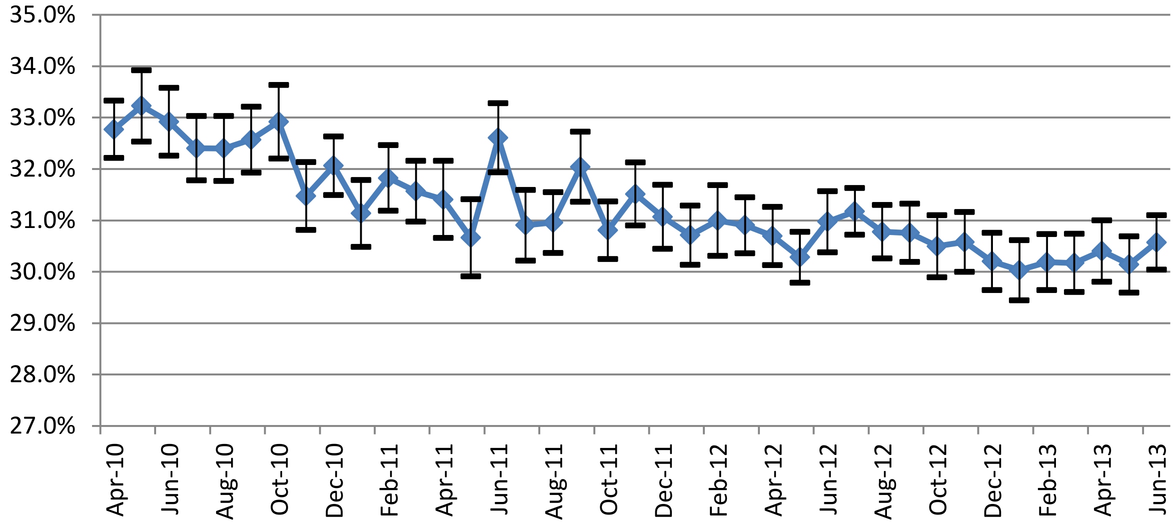
Figure A-6. Monthly Estimates of Uninsured Rate for Black Non-Hispanics, Age 0-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

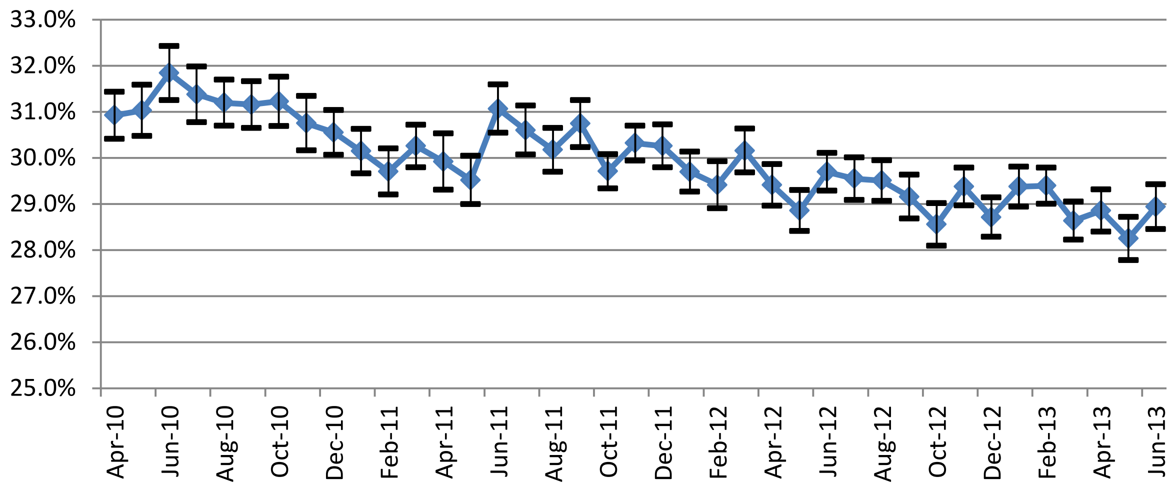
Figure A-7. Monthly Estimates of Uninsured Rate for Hispanics Age 0-64 in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

Figure A-8. Monthly Estimates of Uninsured Rate for Poverty Index 0-138, Age 0-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

Figure A-9. Monthly Estimates of Uninsured Rate for Poverty Index 139-399, Age 0-64, in the United States

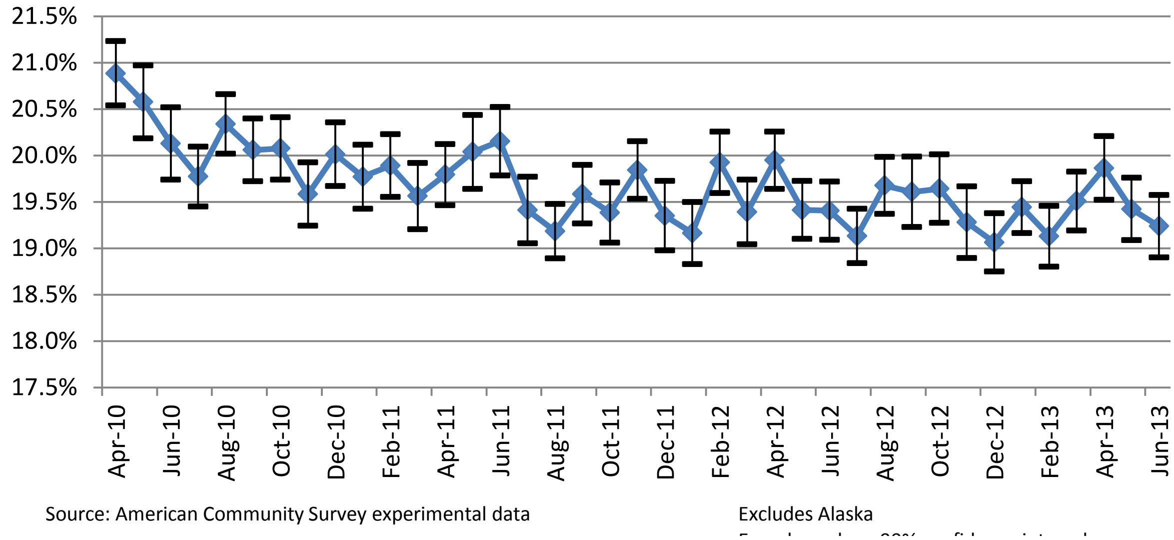


Figure A-10. Monthly Estimates of Uninsured Rate for Poverty Index over 400, Age 0-64, in the United States

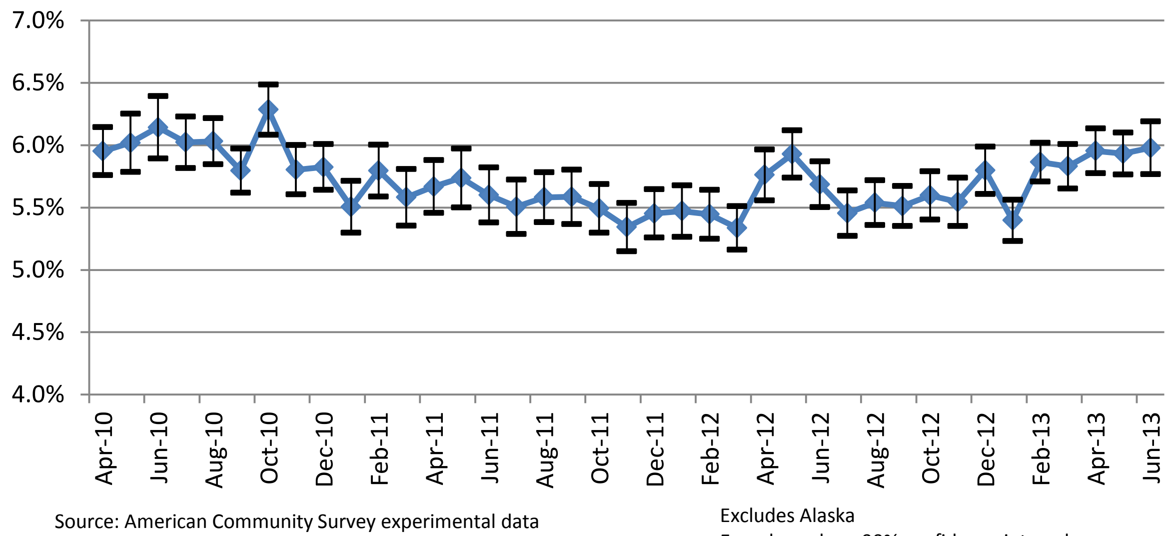
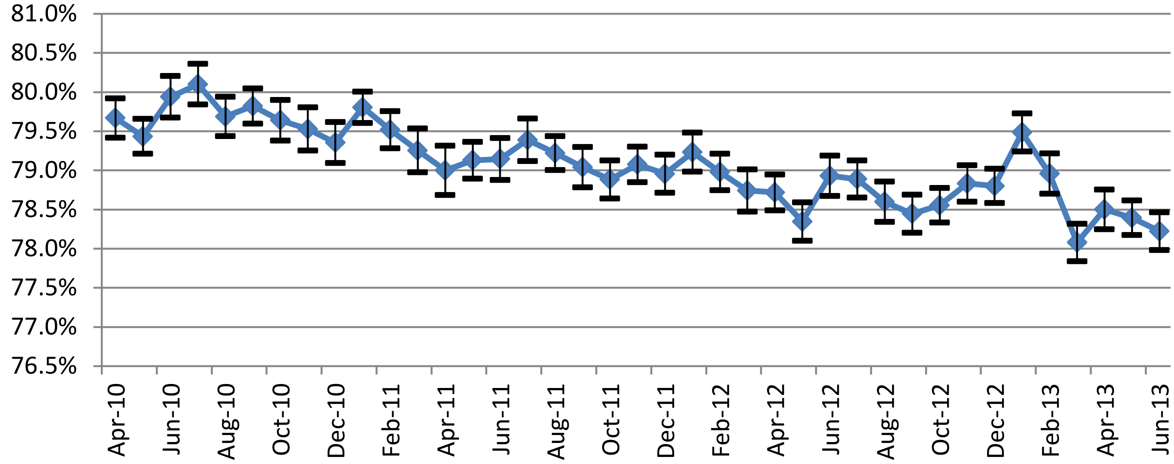


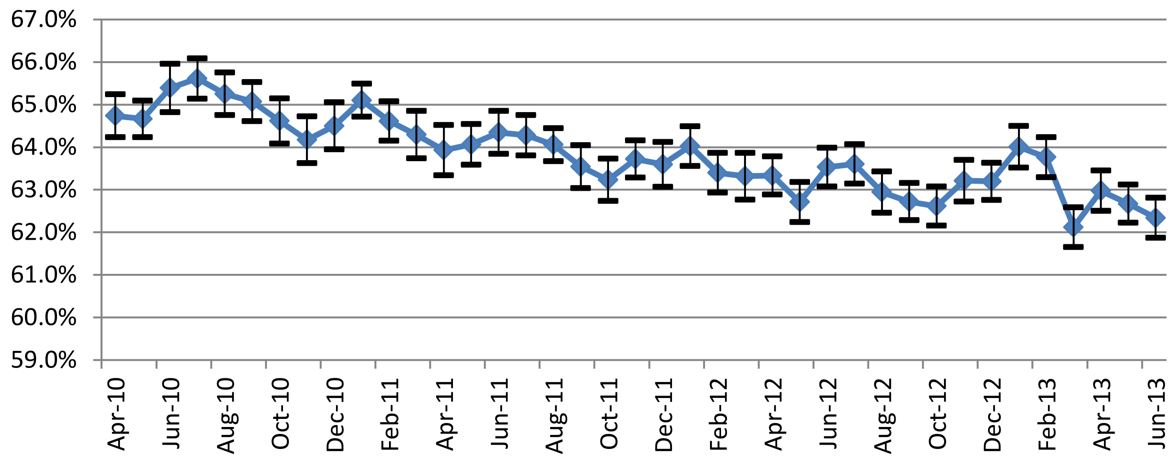
Figure A-11. Monthly Estimates for Percent of Insured That Have Private Insurance, Age 0-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

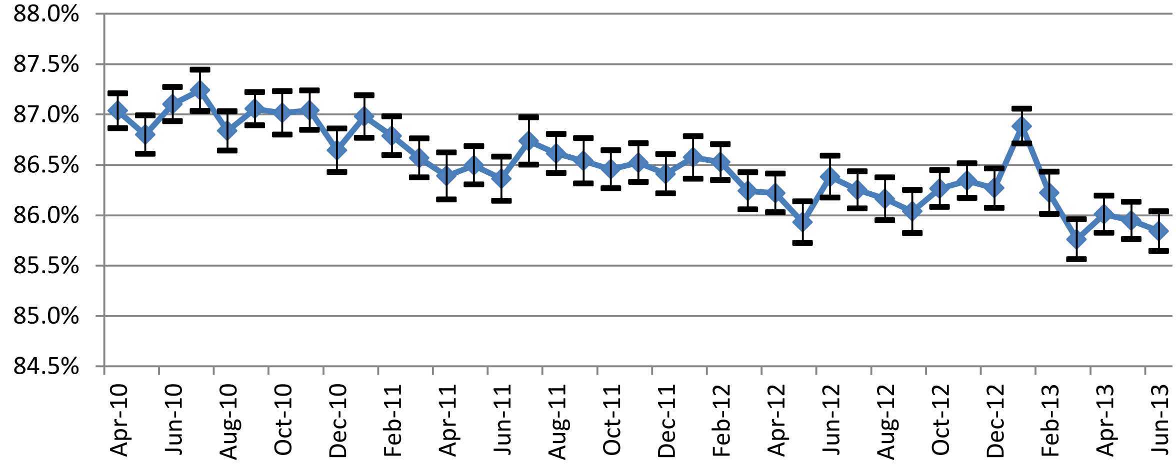
Figure A-12. Monthly Estimates for Percent of Insured That Have Private Insurance, Age 0-18, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

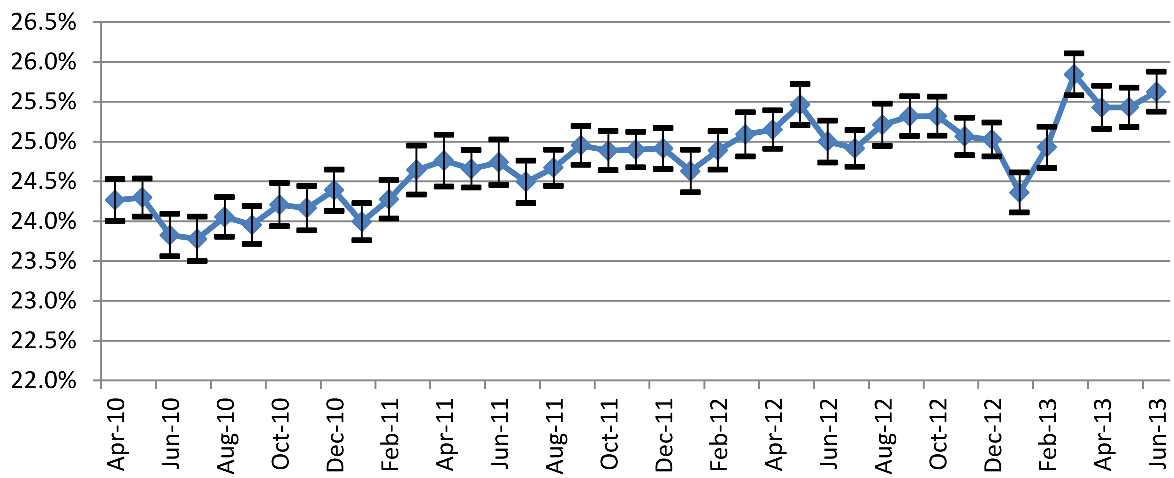
Figure A-13. Monthly Estimates for Percent of Insured That Have Private Insurance, Age 19-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

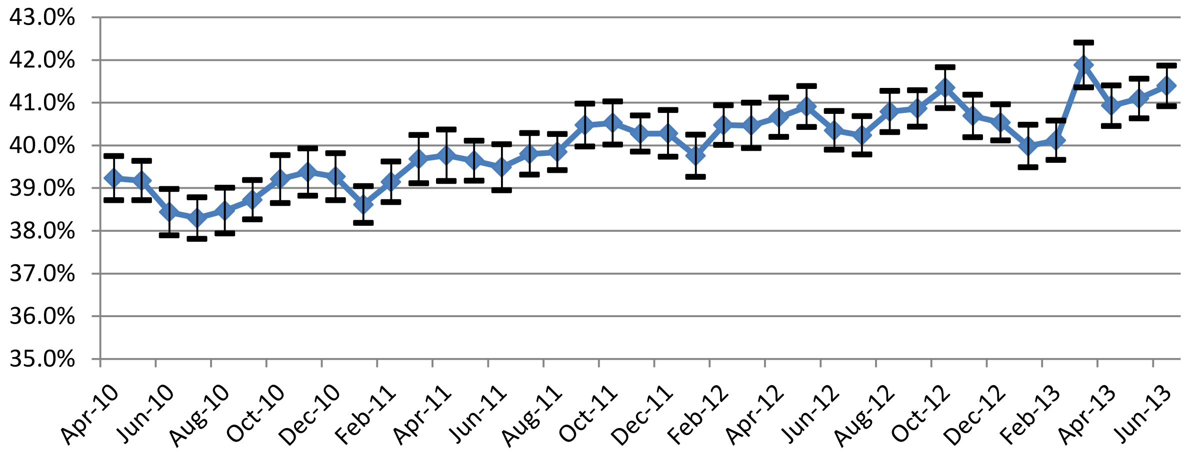
Figure A-14. Monthly Estimates for Percent of Insured That Have Public Insurance, Age 0-64, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

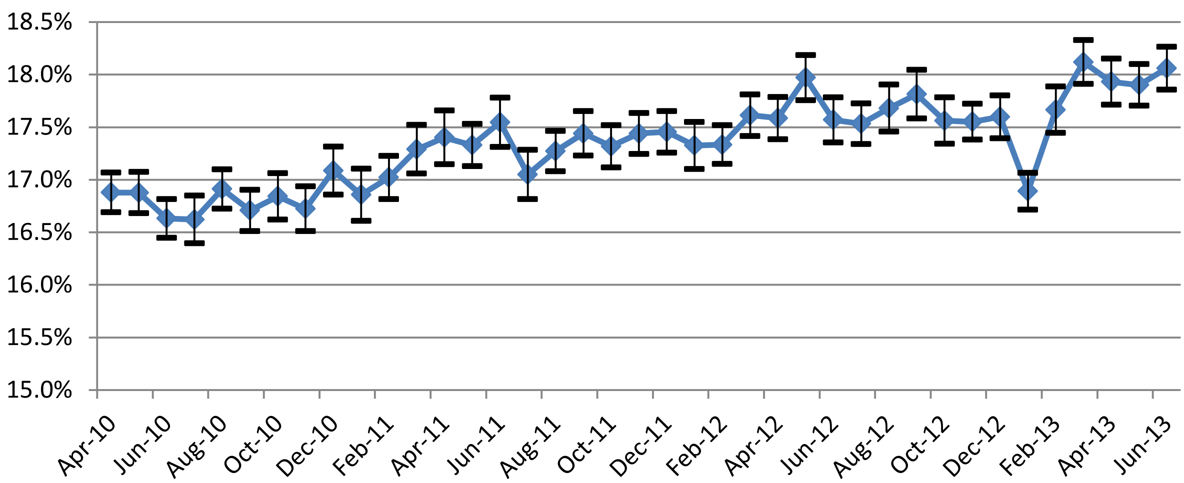
Figure A-15. Monthly Estimates for Percent of Insured That Have Public Insurance, Age 0-18, in the United States



Source: American Community Survey experimental data

Excludes Alaska
Error bars show 90% confidence intervals

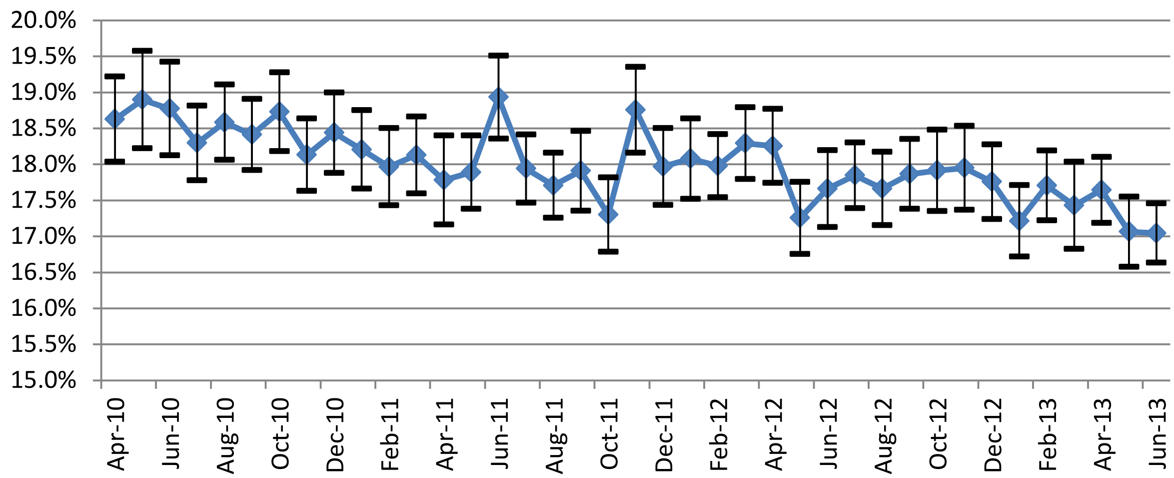
Figure A-16. Monthly Estimates for Percent of Insured That Have Public Insurance, Age 19-64, in the United States



Source: American Community Survey experimental data

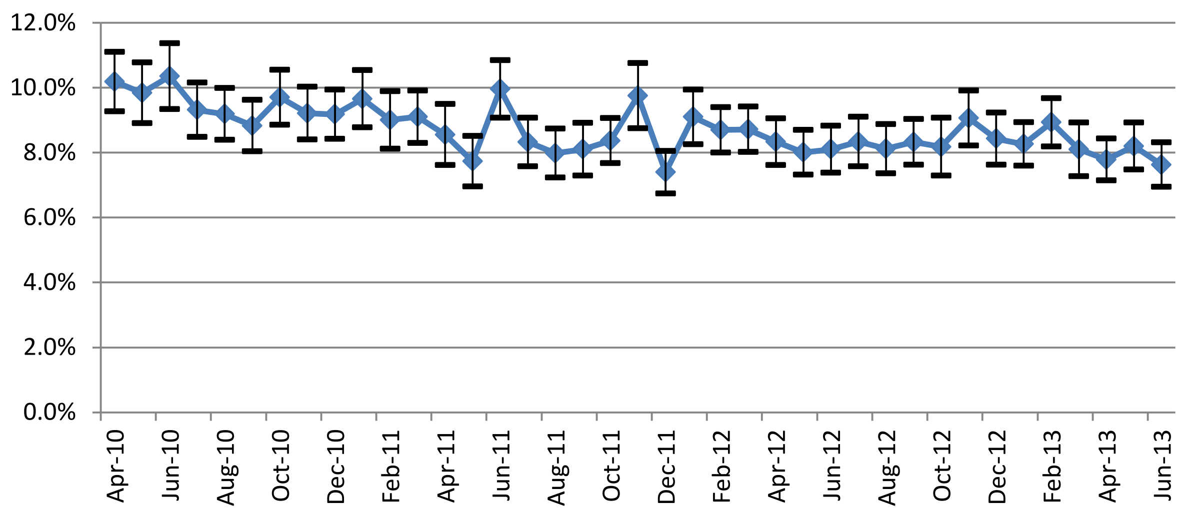
Excludes Alaska
Error bars show 90% confidence intervals

Figure B-1. Monthly Estimates of Uninsured Persons in California



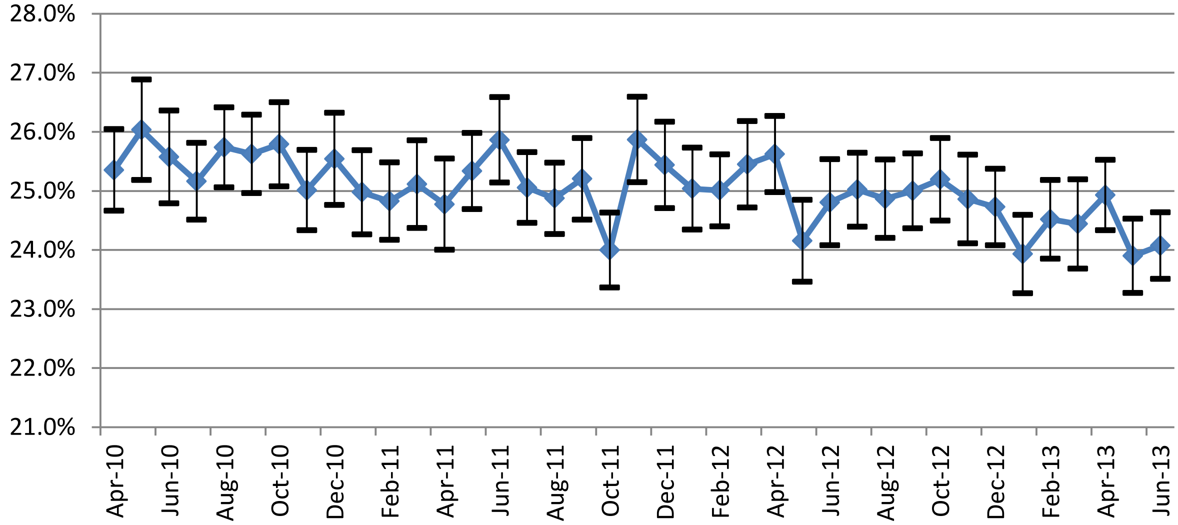
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-2. Monthly Estimates of Uninsured Age 0-18 in California



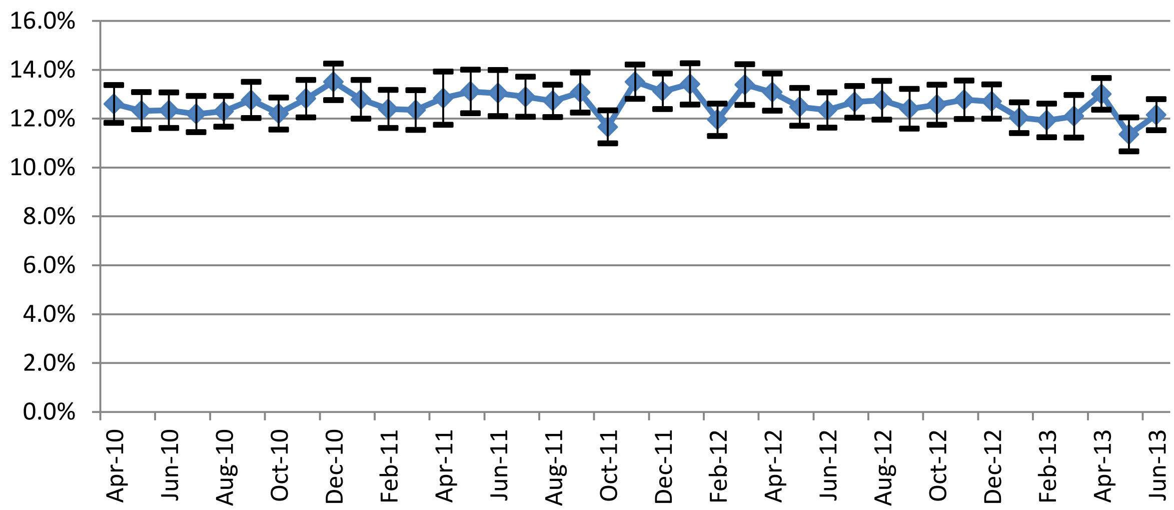
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-3. Monthly Estimates of Uninsured Age 19-64 in California



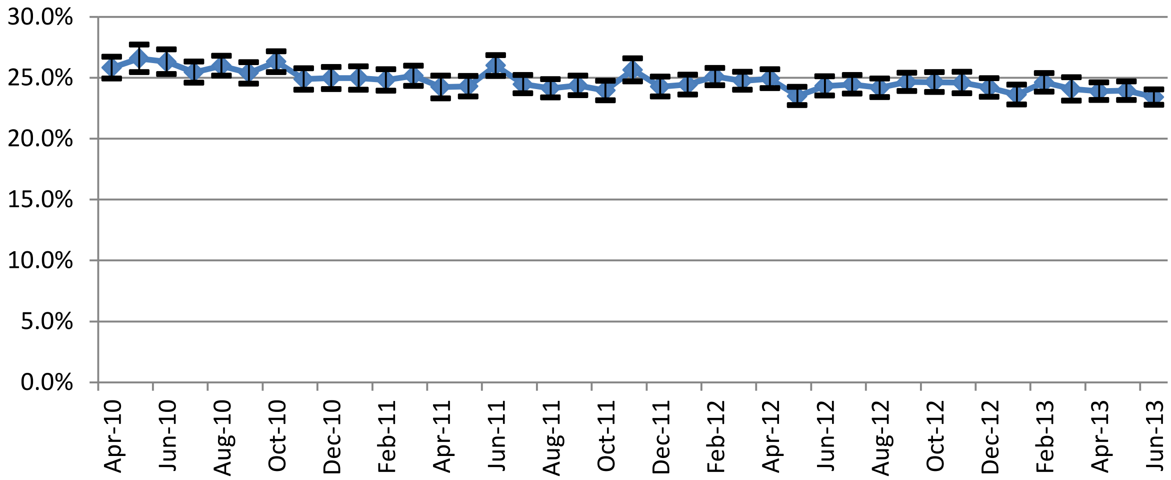
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-4. Monthly Estimates of Uninsured White Non-Hispanics, Age 0-64, in California



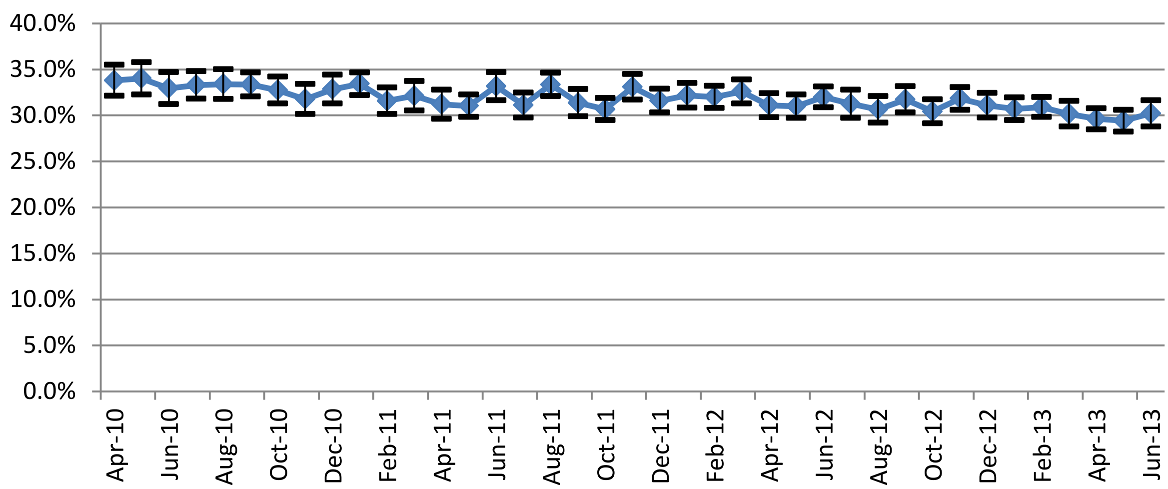
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-5. Monthly Estimates of Uninsured Minorities, Age 0-64, in California



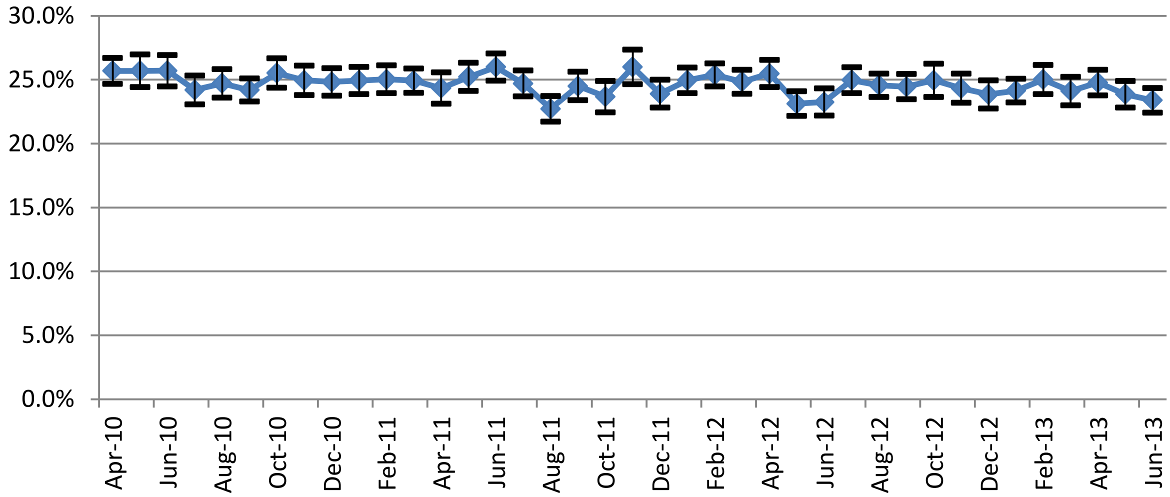
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in California



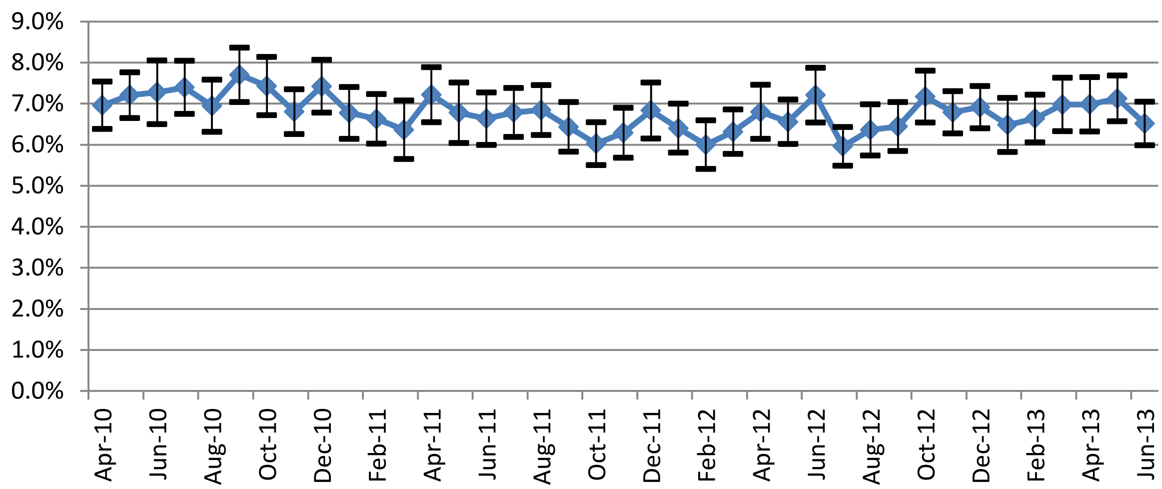
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in California



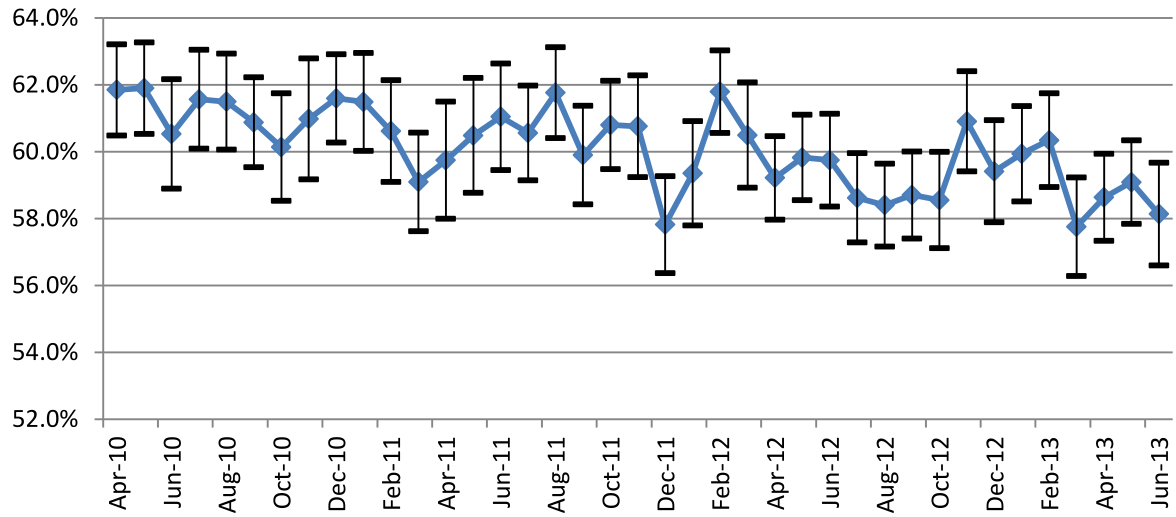
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in California



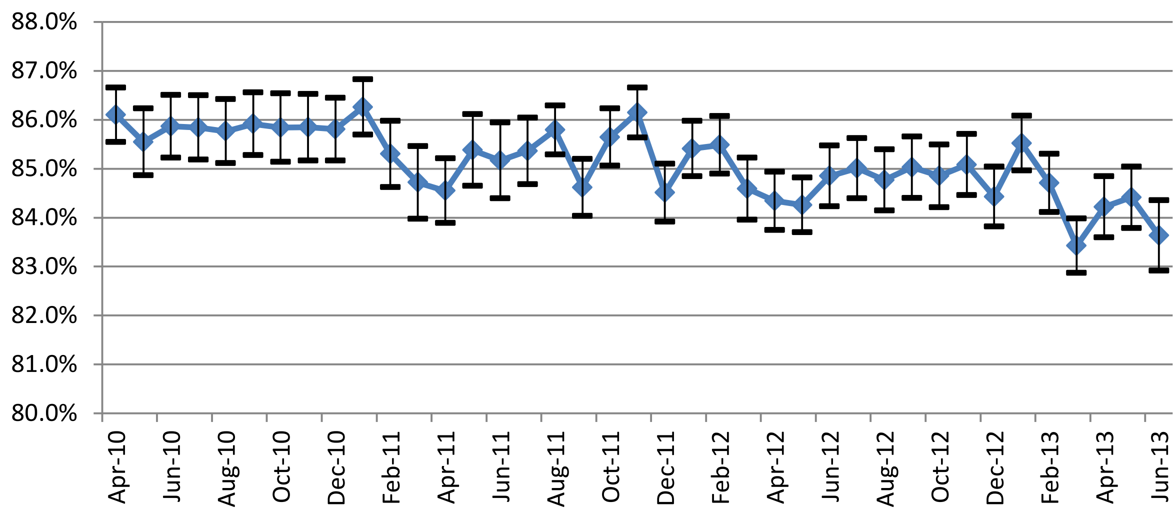
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in California



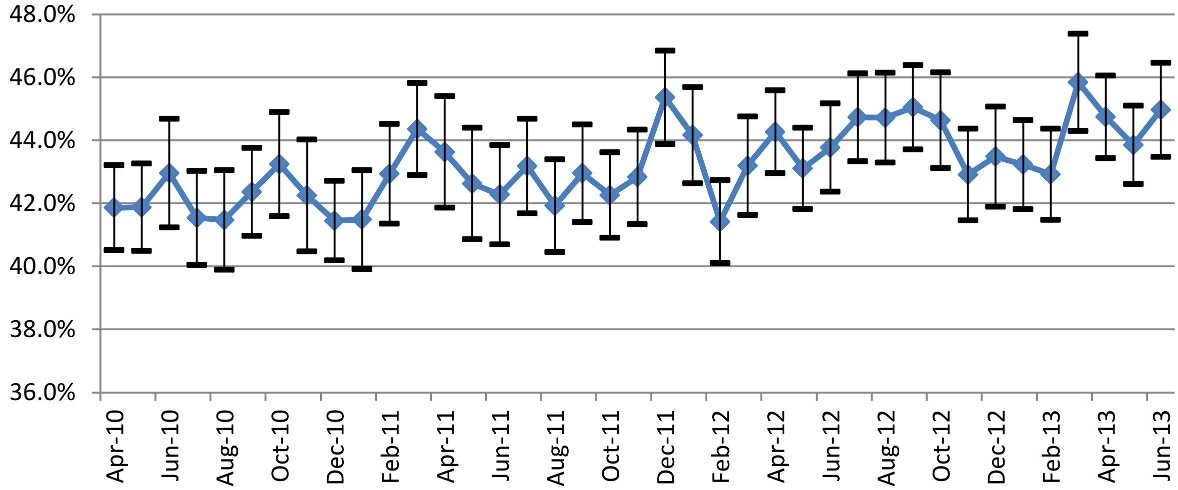
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in California



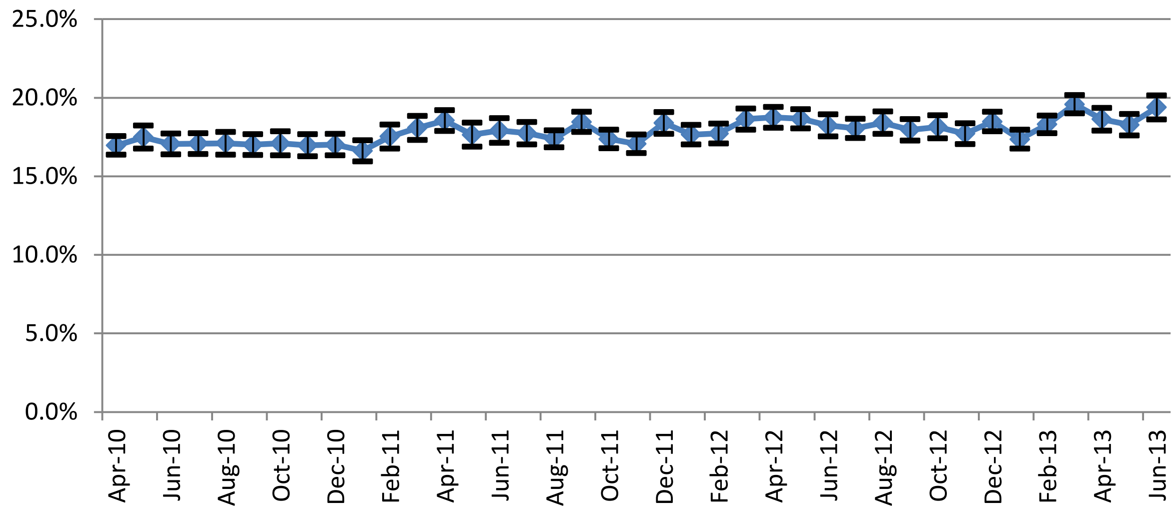
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in California



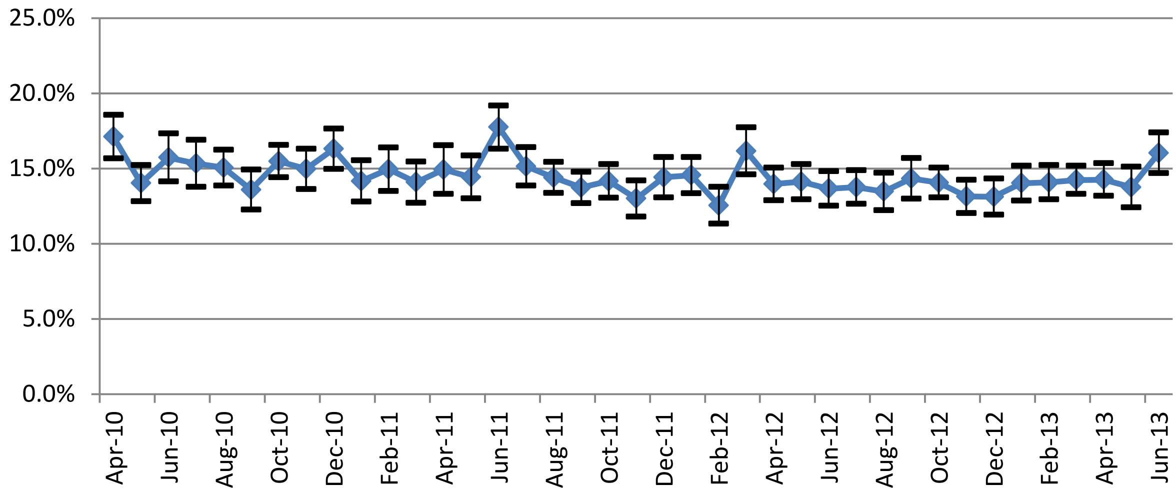
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure B-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in California



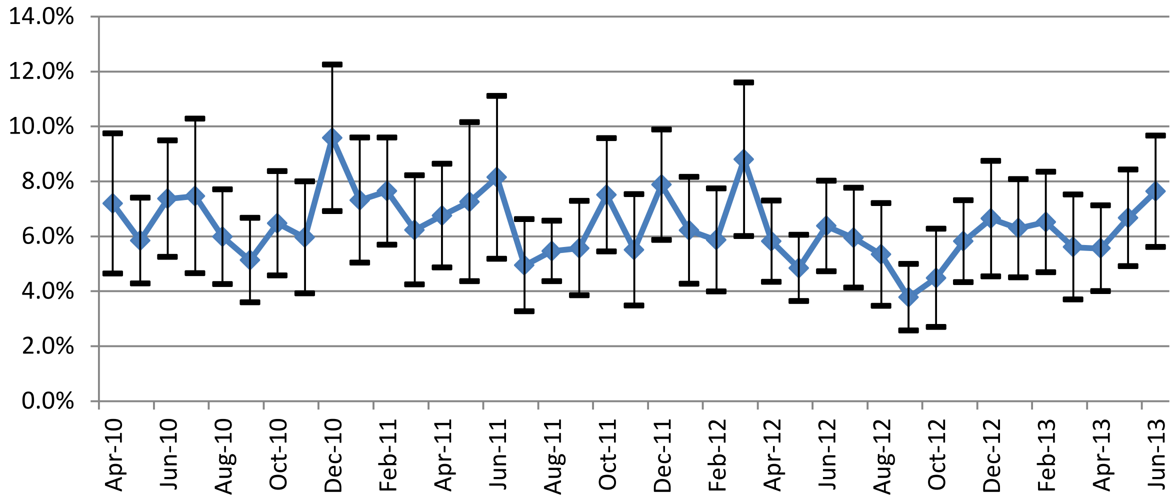
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-1. Monthly Estimates of Uninsured Persons in Kentucky



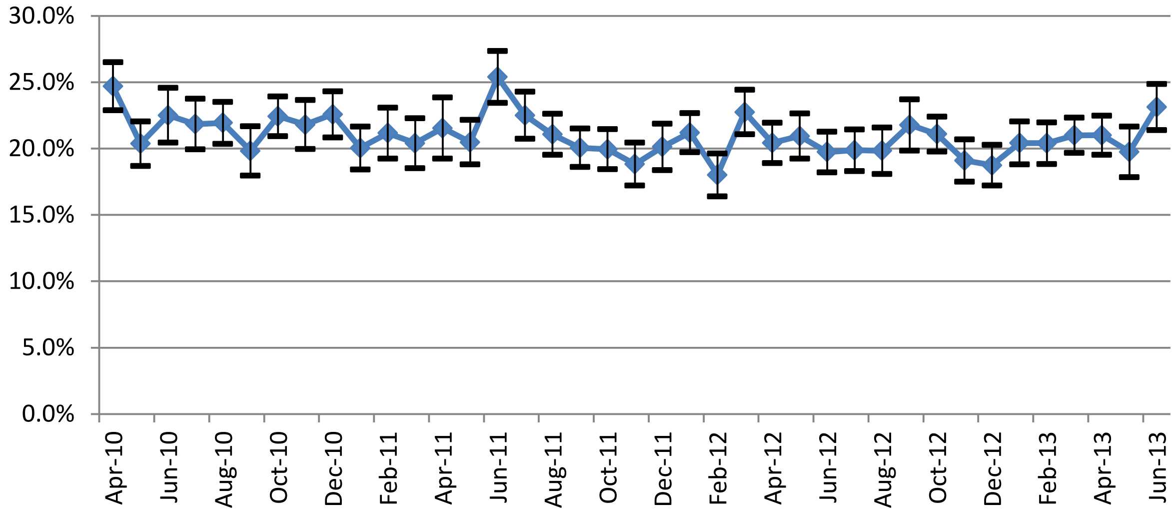
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-2. Monthly Estimates of Uninsured Age 0-18 in Kentucky



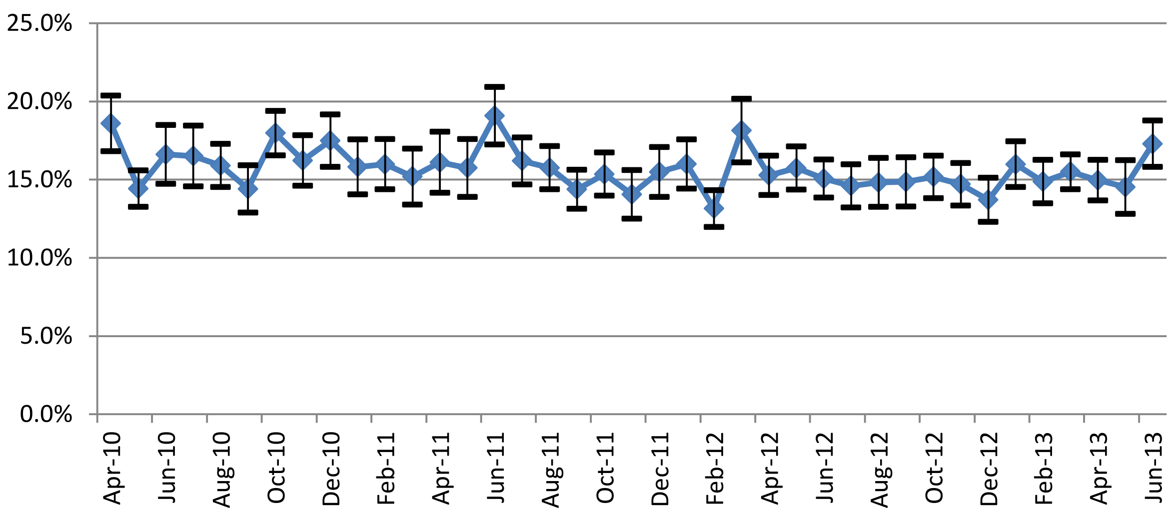
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-3. Monthly Estimates of Uninsured Age 19-64 in Kentucky



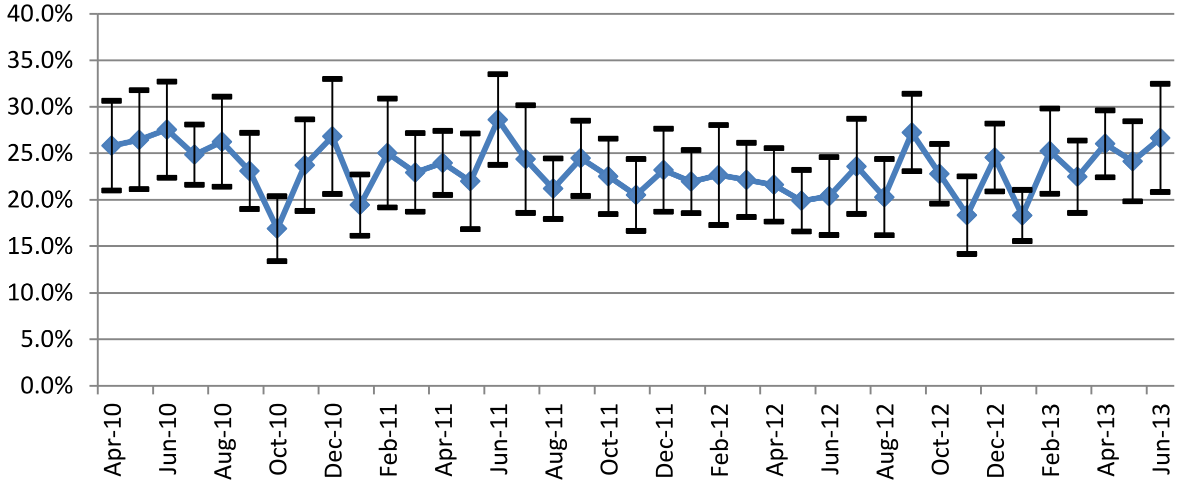
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-4. Monthly Estimates of Uninsured White Non-Hispanics, Age 0-64, in Kentucky



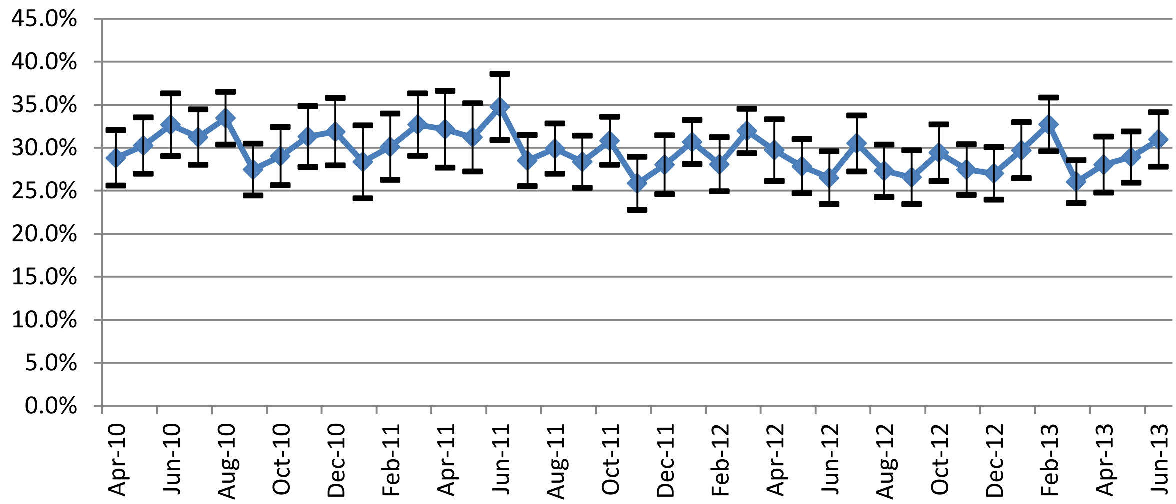
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-5. Monthly Estimates of Uninsured Minorities, Age 0-64, in Kentucky



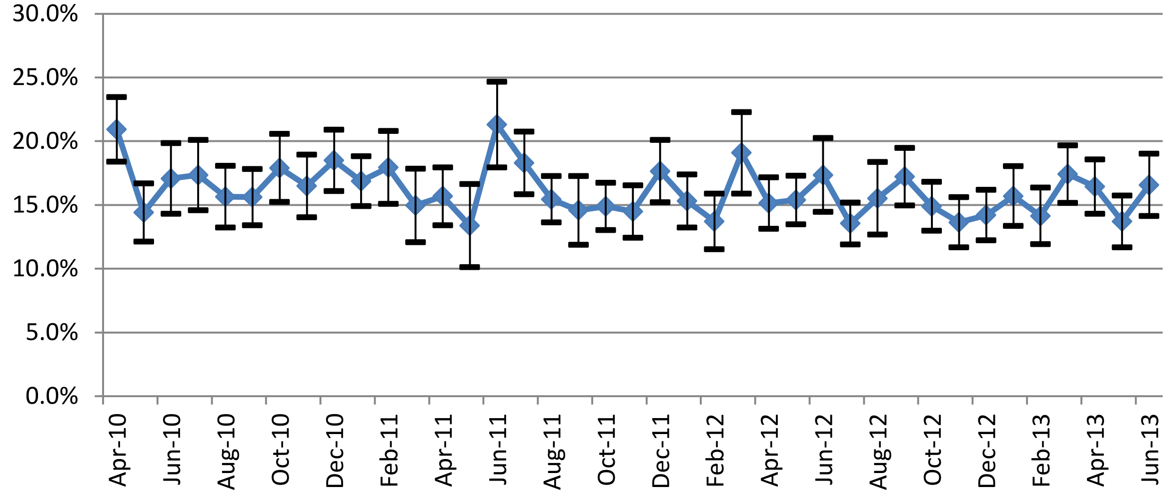
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in Kentucky



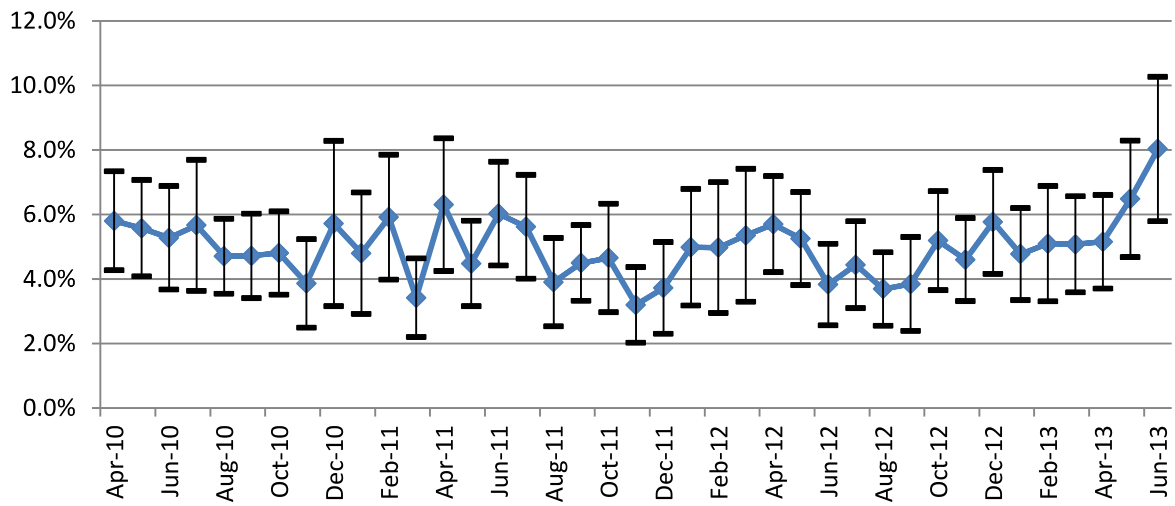
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in Kentucky



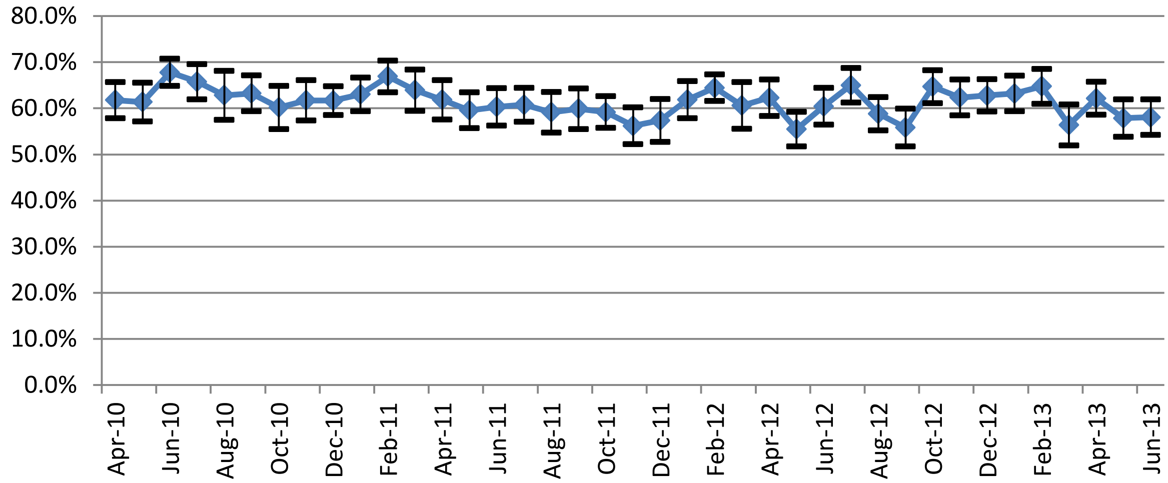
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in Kentucky



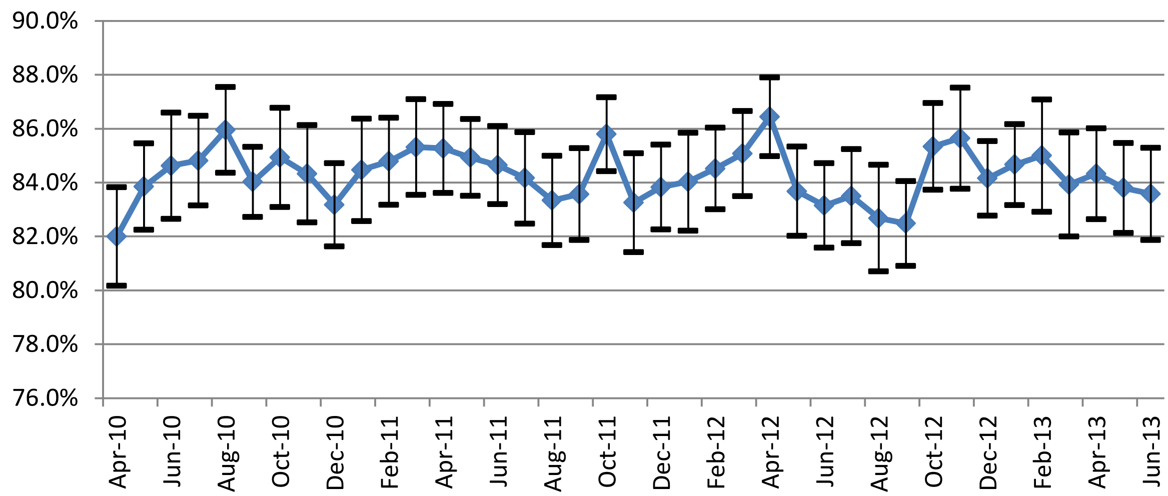
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in Kentucky



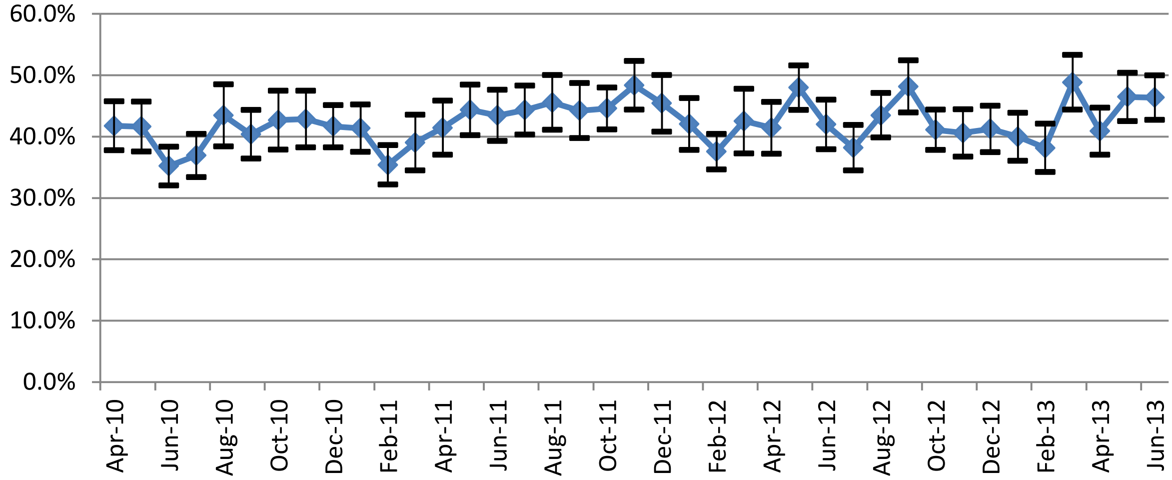
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in Kentucky



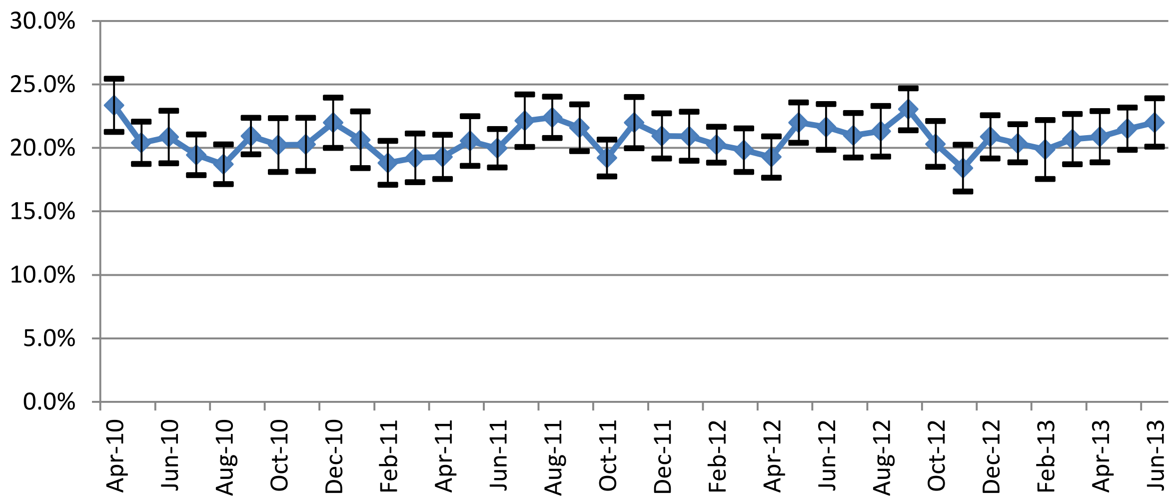
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in Kentucky



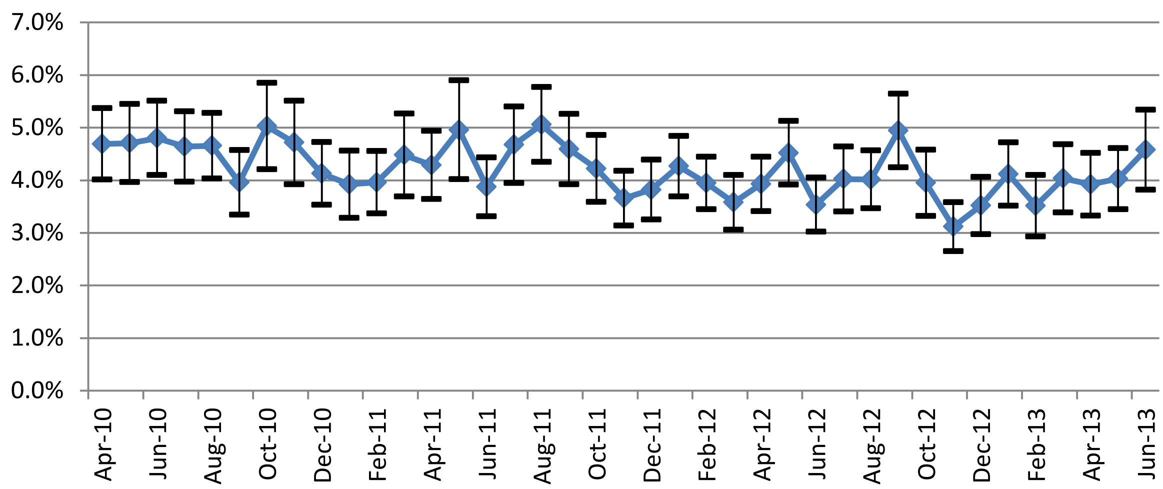
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure C-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in Kentucky



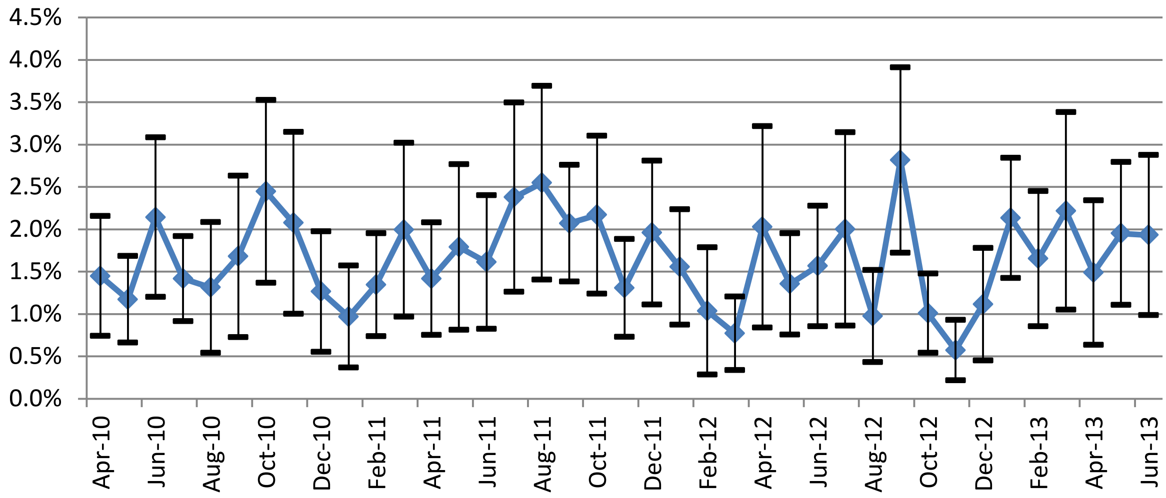
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-1. Monthly Estimates of Uninsured Persons in Massachusetts



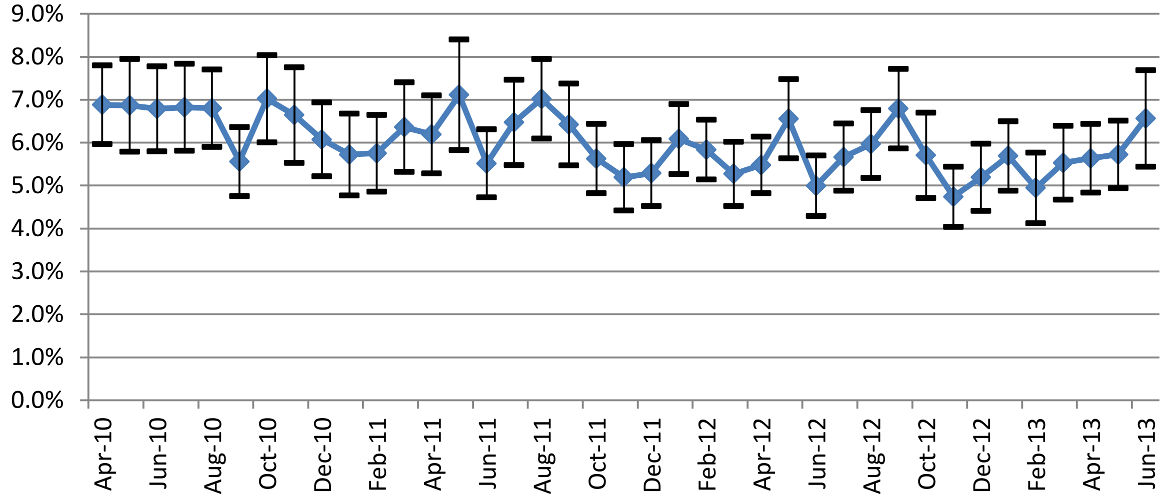
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-2. Monthly Estimates of Uninsured Age 0-18 in Massachusetts



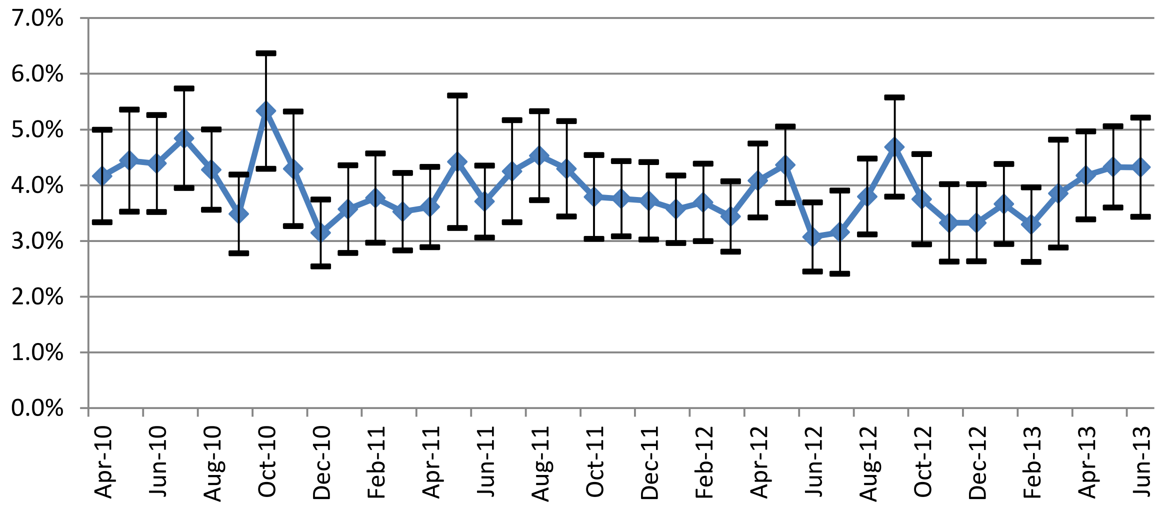
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-3. Monthly Estimates of Uninsured Age 19-64 in Massachusetts



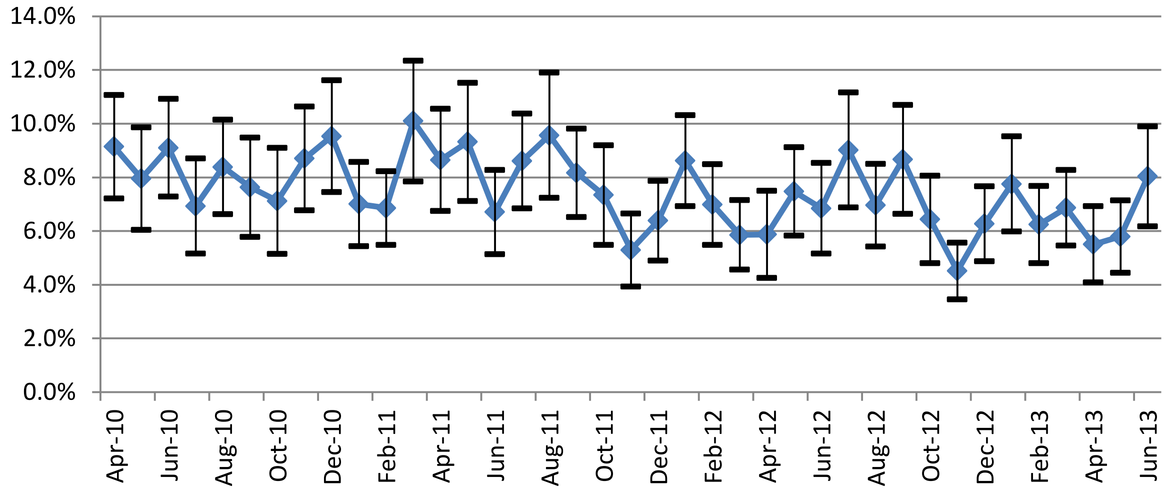
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-4. Monthly Estimates of Uninsured White Non-Hispanics, Age 0-64, in Massachusetts



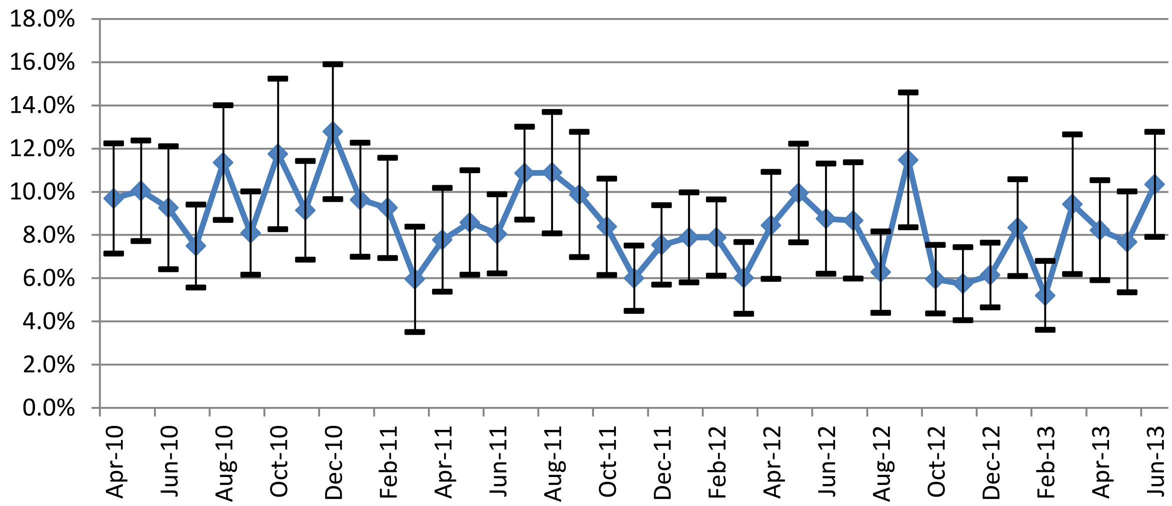
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-5. Monthly Estimates of Uninsured Minorities, Age 0-64, in Massachusetts



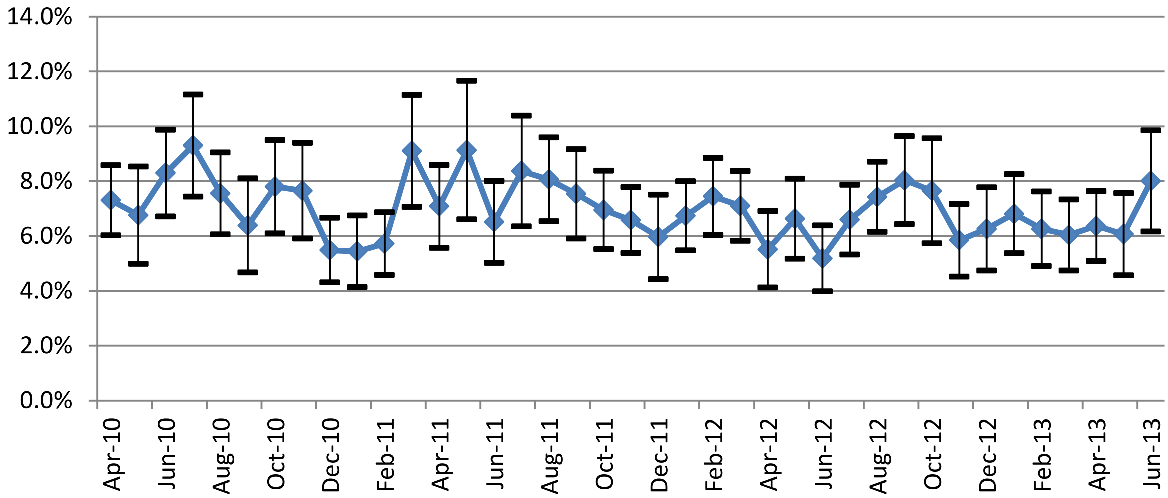
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in Massachusetts



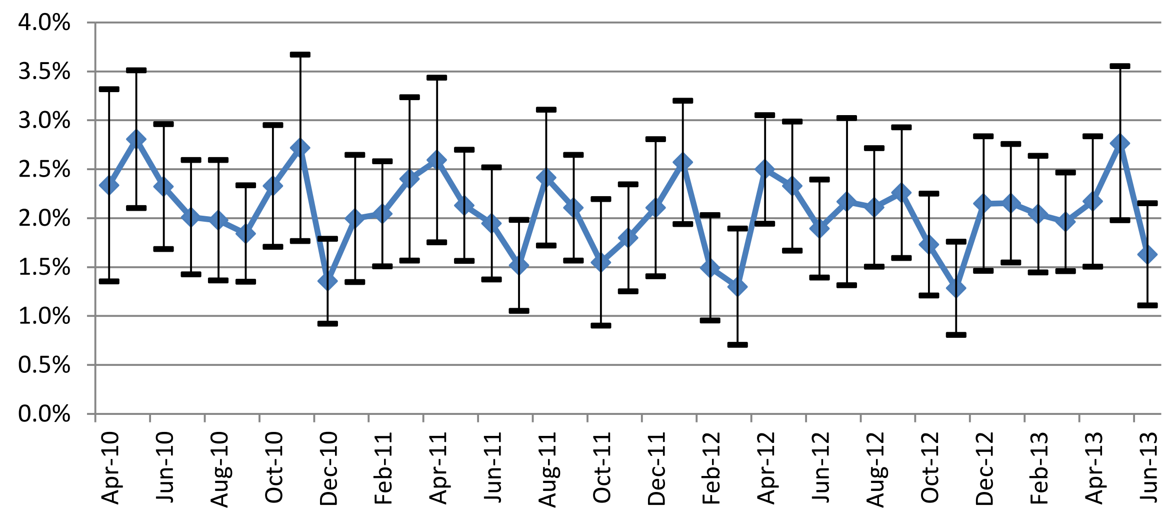
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in Massachusetts



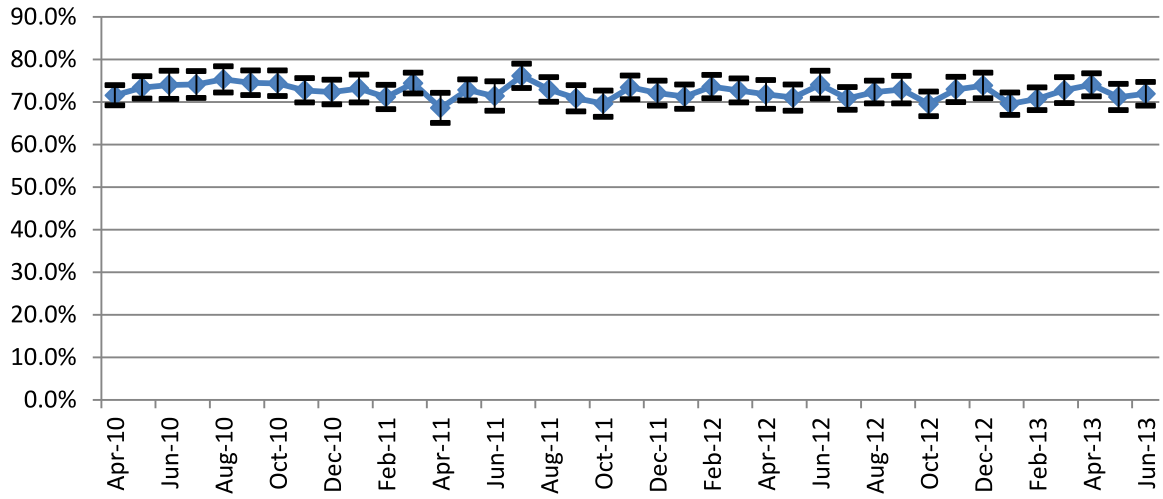
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in Massachusetts



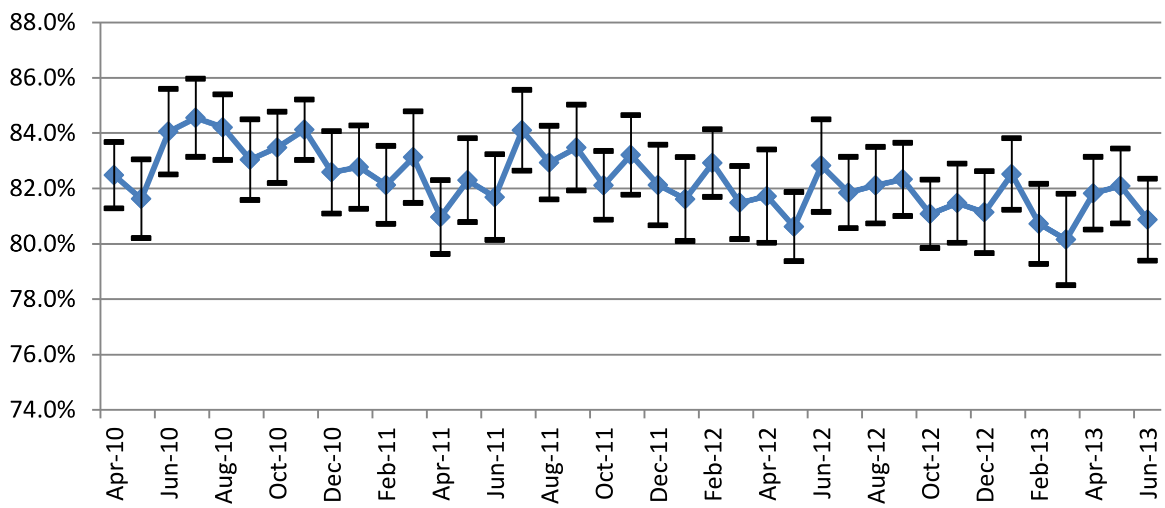
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in Massachusetts



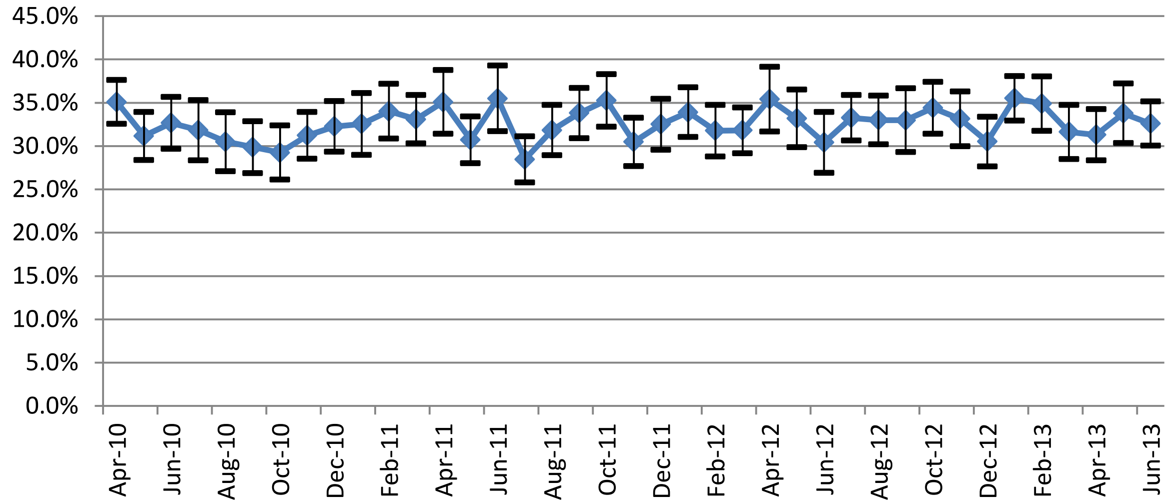
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in Massachusetts



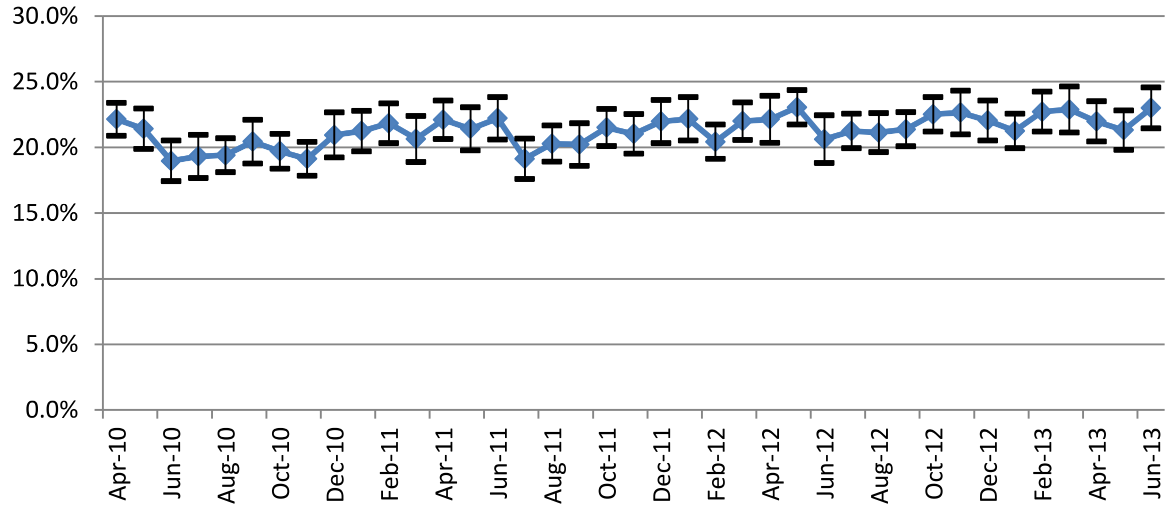
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in Massachusetts



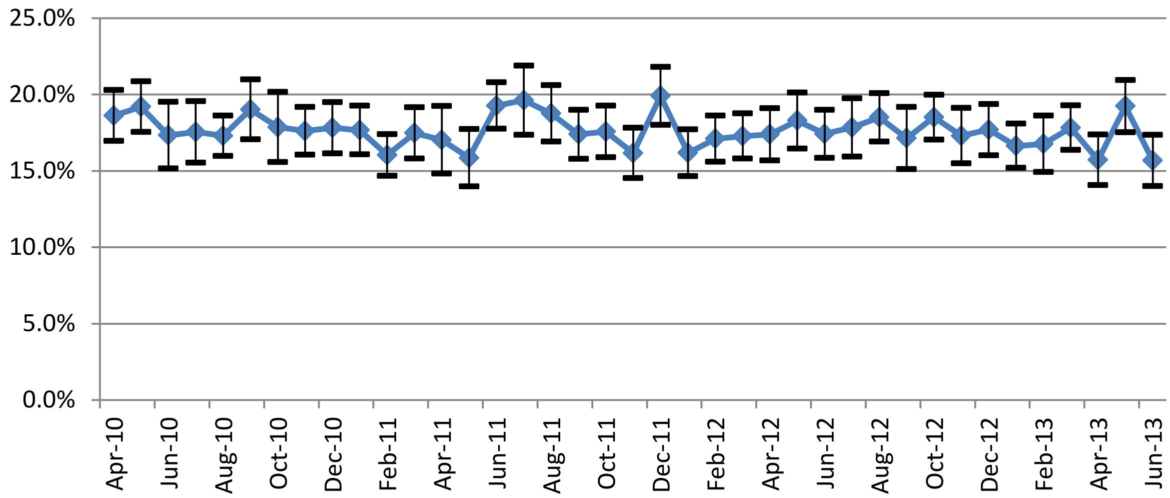
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure D-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in Massachusetts



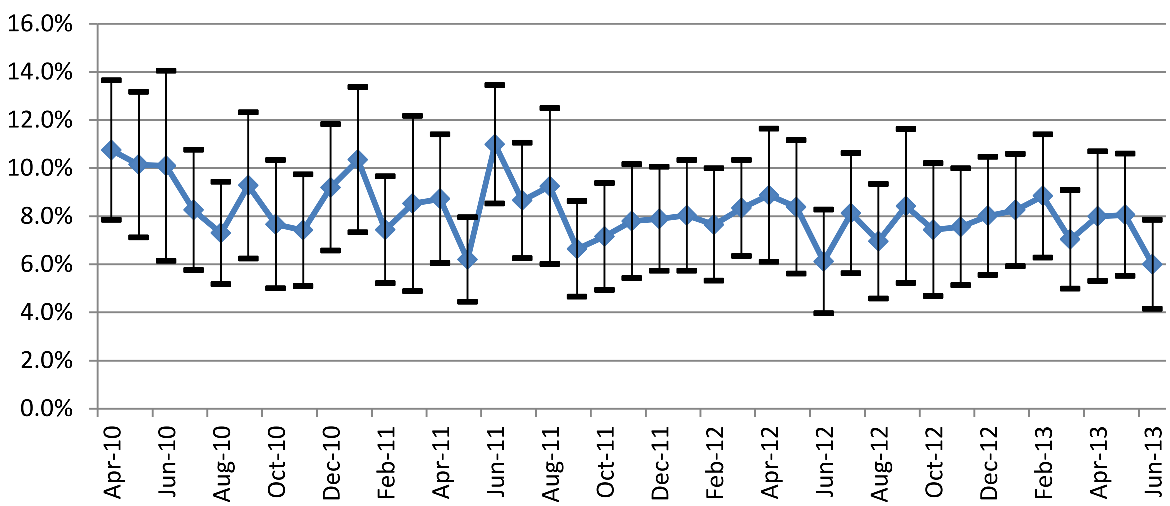
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-1. Monthly Estimates of Uninsured Persons in Mississippi



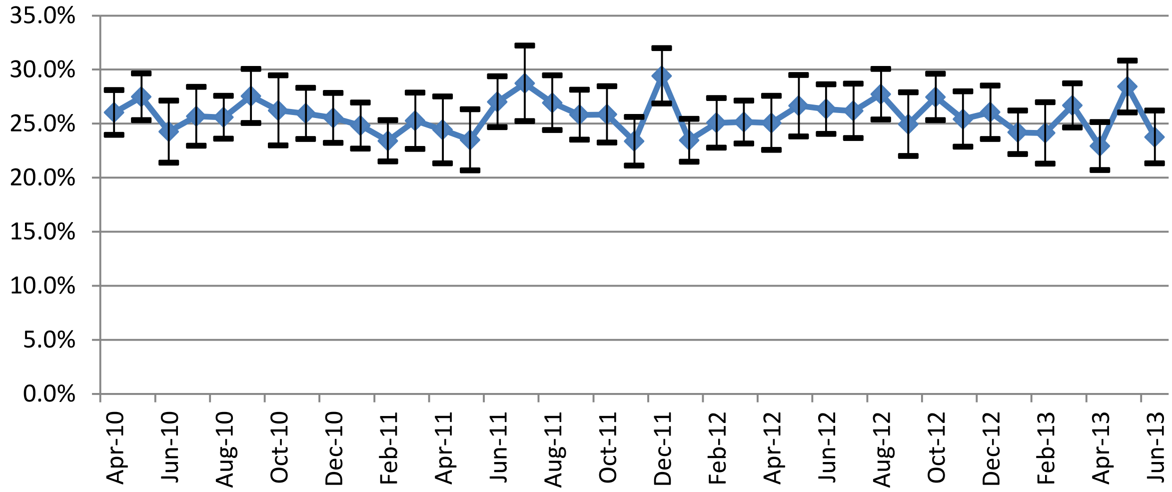
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-2. Monthly Estimates of Uninsured Age 0-18 in Mississippi



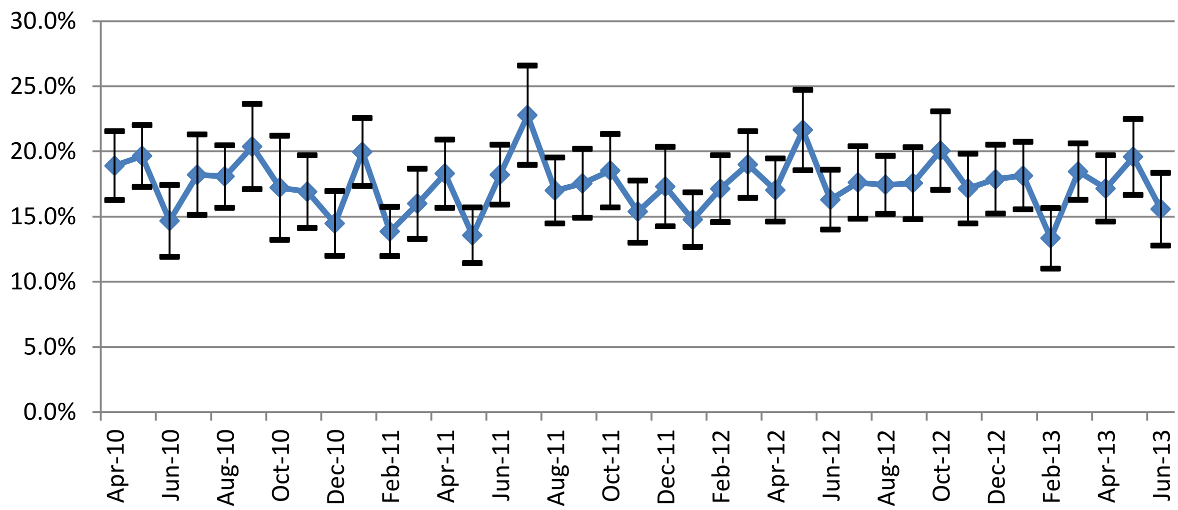
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-3. Monthly Estimates of Uninsured Age 19-64 in Mississippi



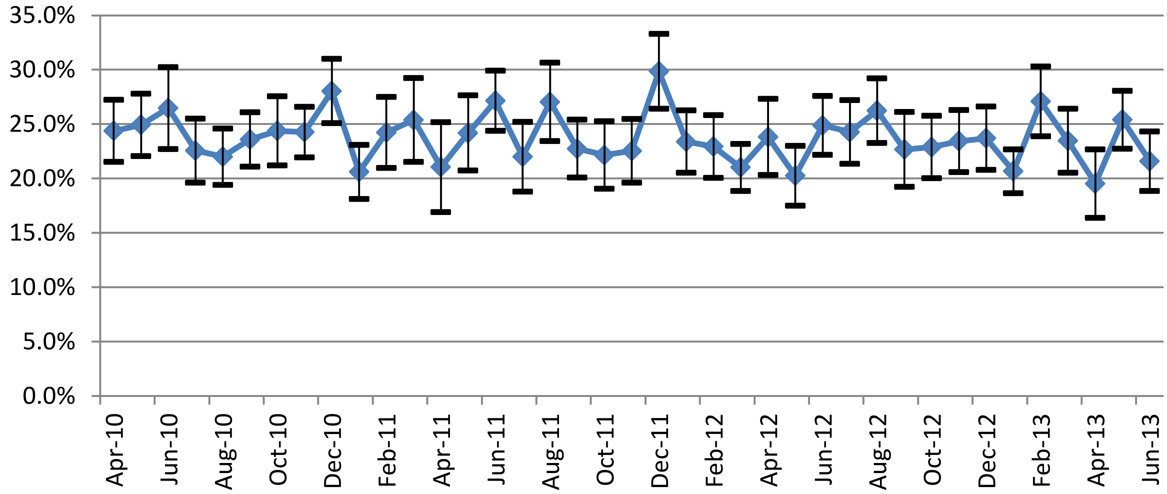
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-4. Monthly Estimates of Uninsured White Non-Hispanics, Age 0-64, in Mississippi



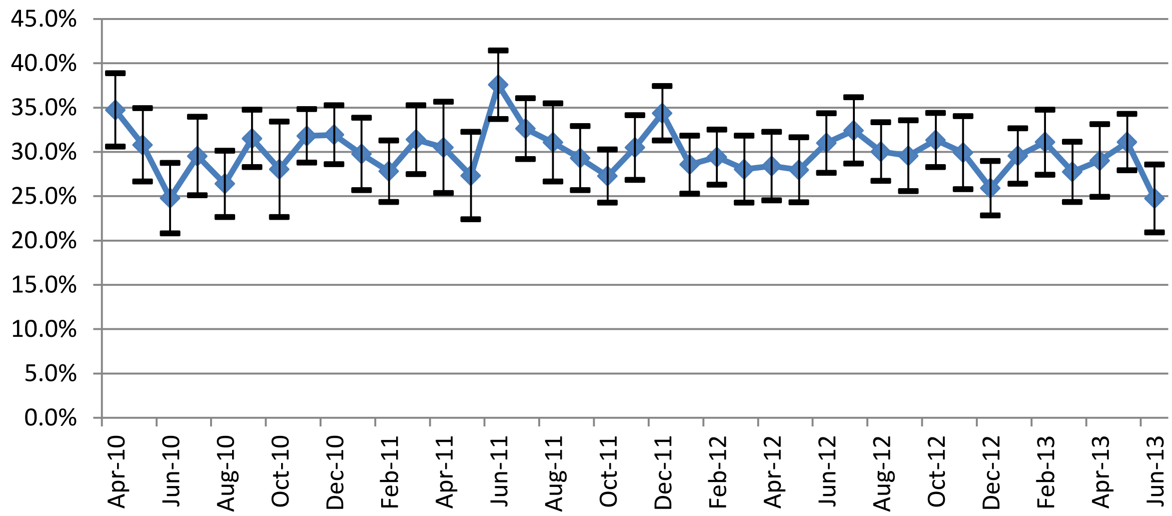
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-5. Monthly Estimates of Minorities, Age 0-64, in Mississippi



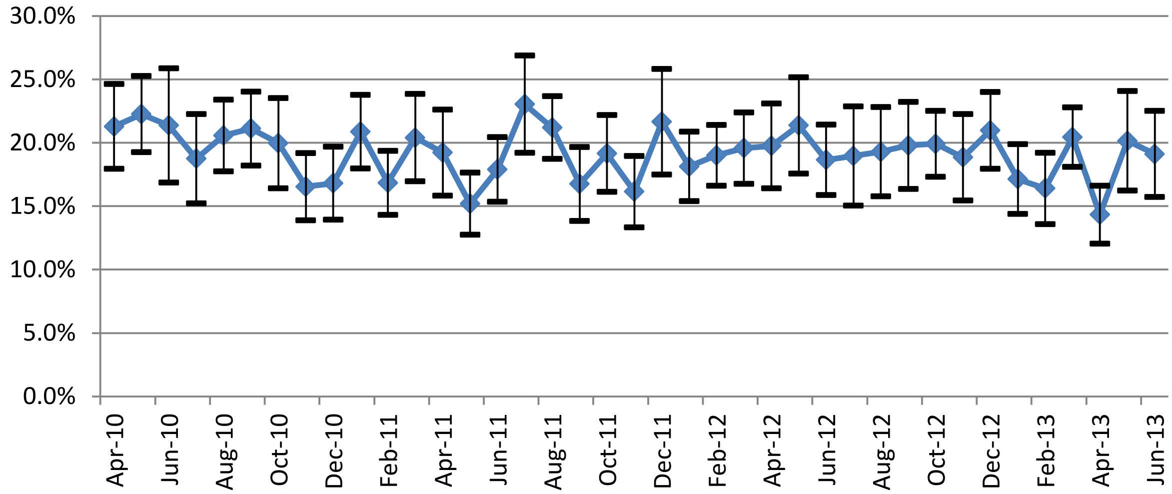
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in Mississippi



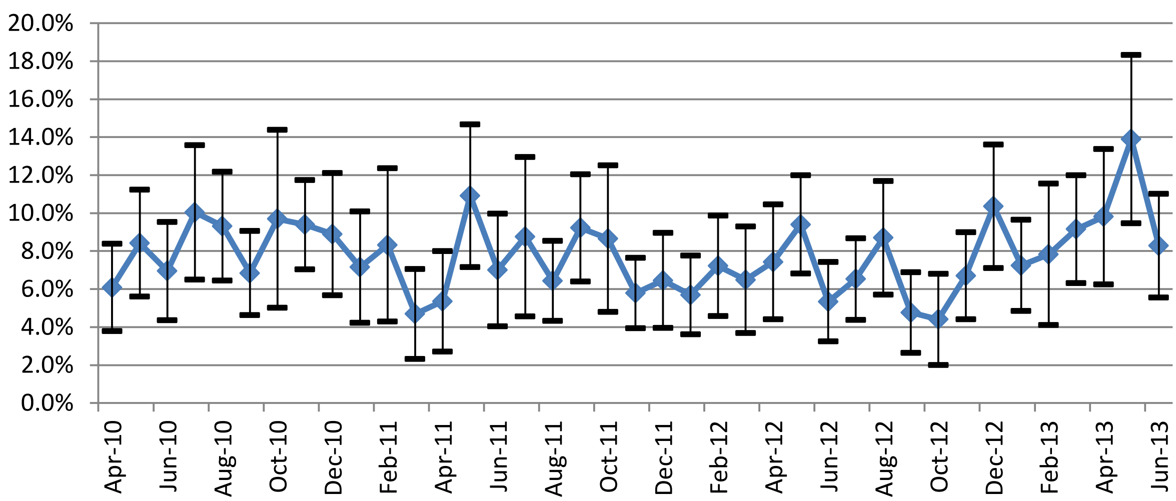
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in Mississippi



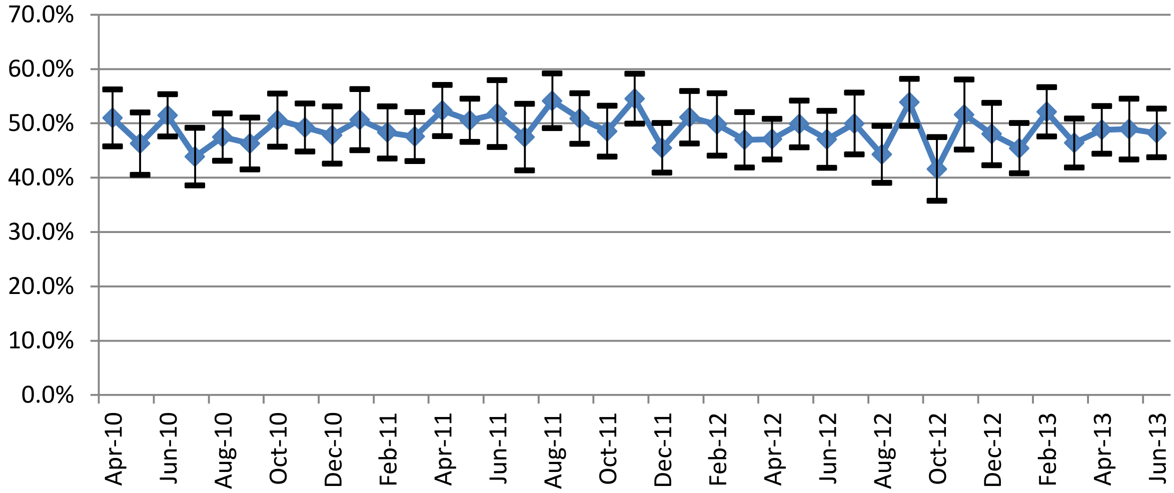
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in Mississippi



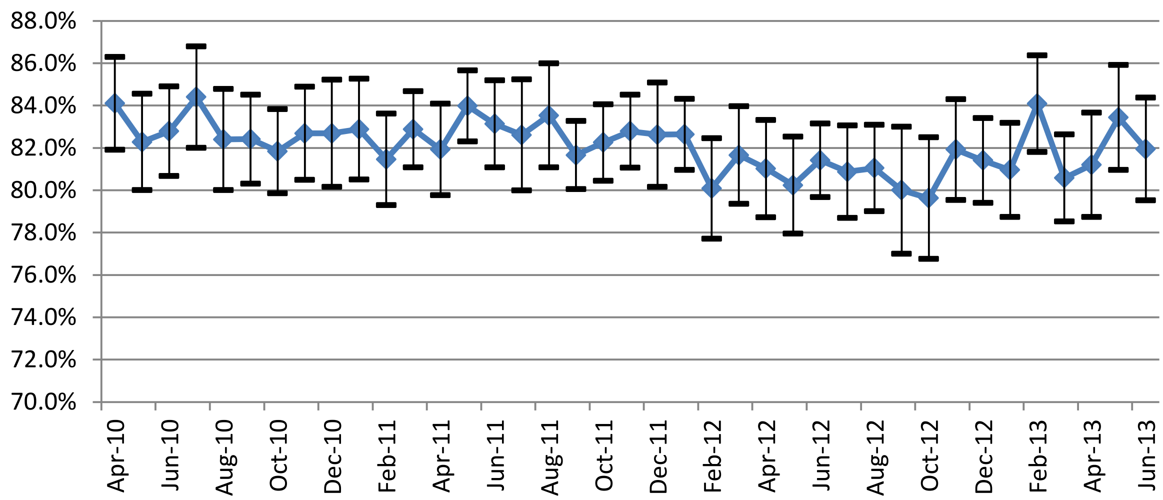
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in Mississippi



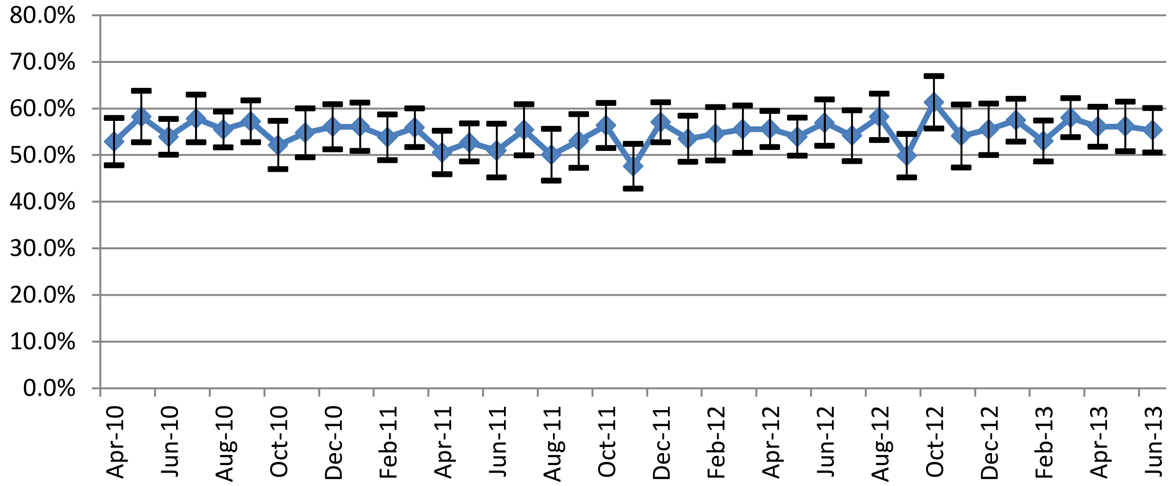
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in Mississippi



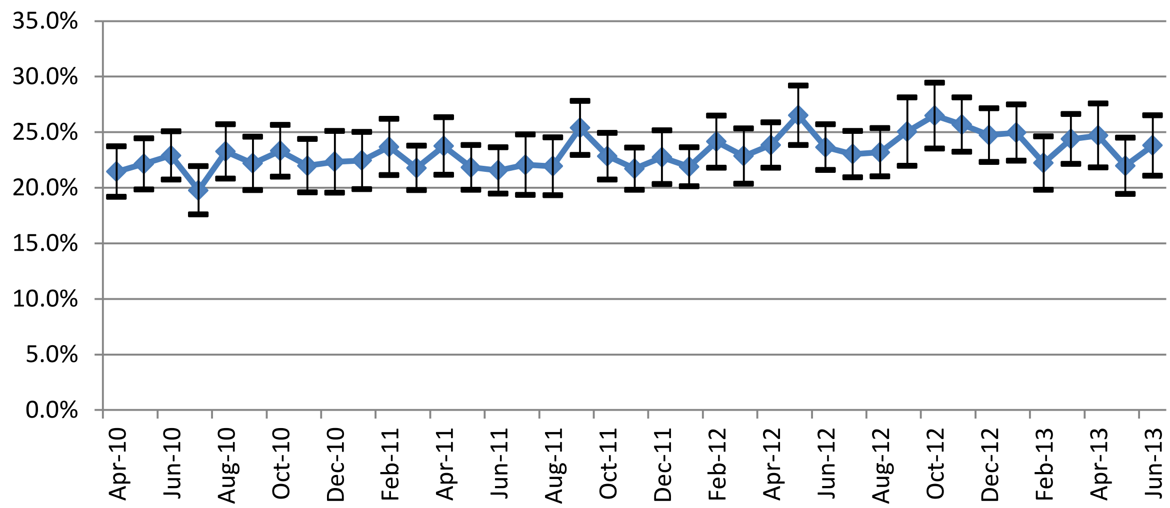
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure E-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in Mississippi



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

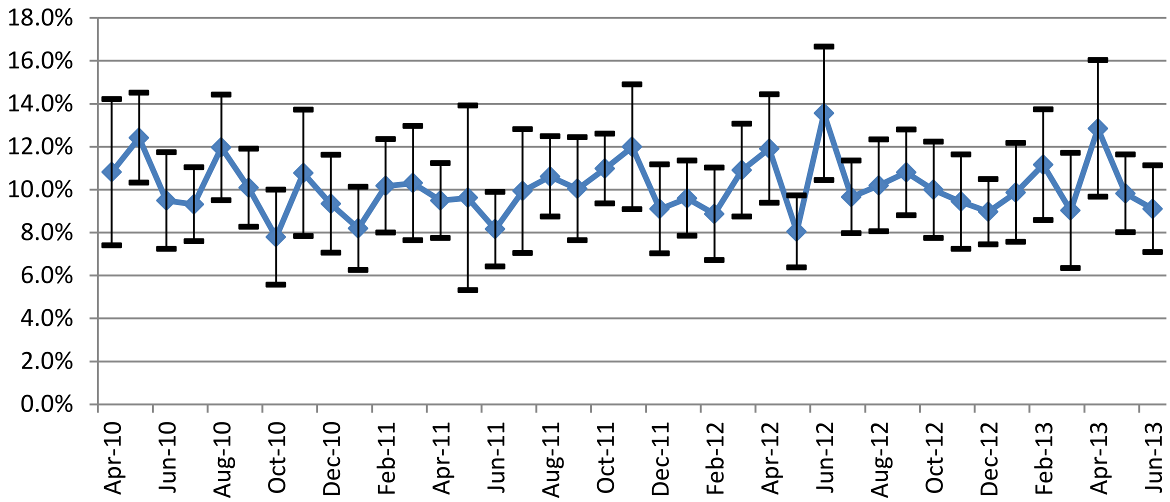
Figure E-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in Mississippi



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

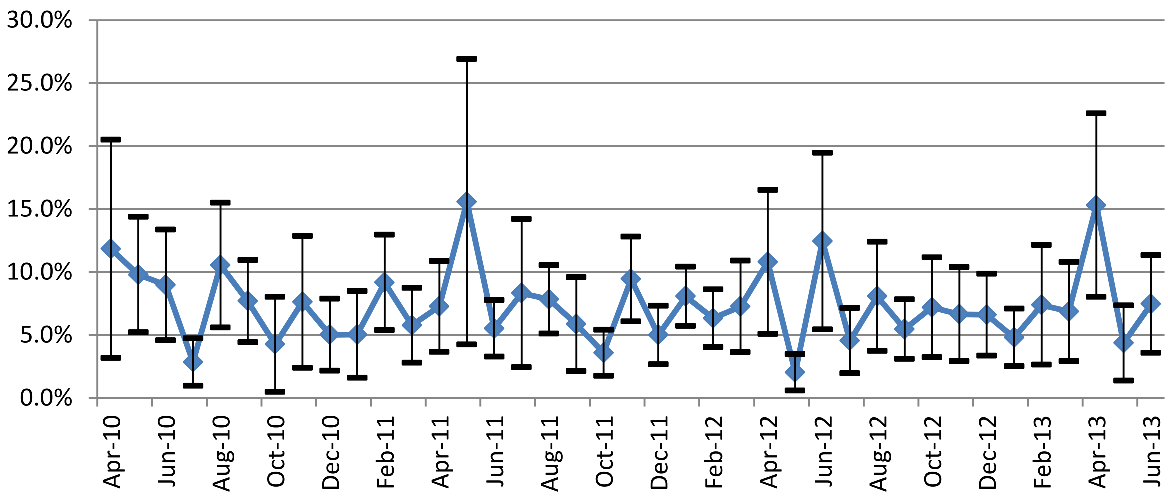
Attachment F.

Figure F-1. Monthly Estimates of Uninsured Persons in North Dakota



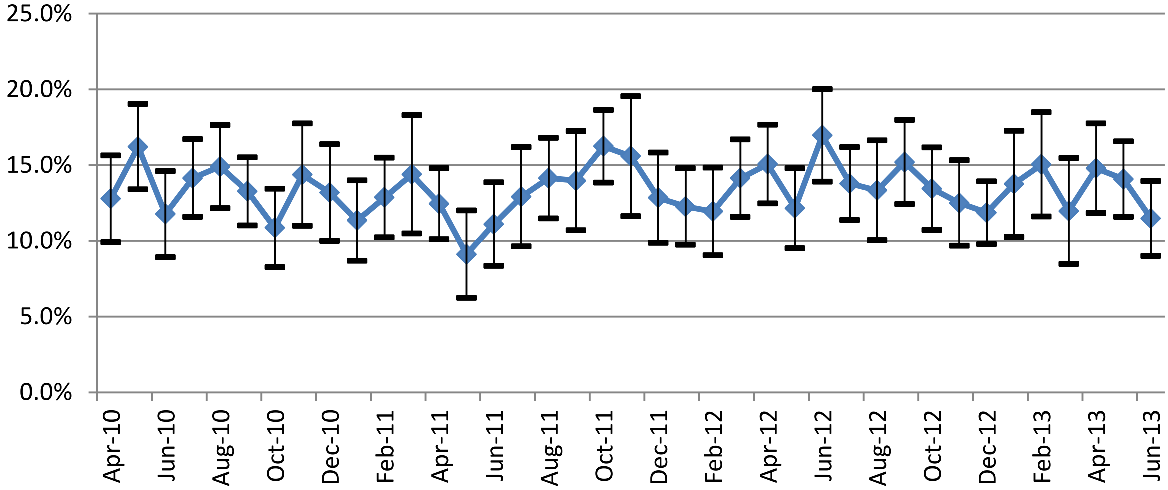
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-2. Monthly Estimates of Uninsured Age 0-18 in North Dakota



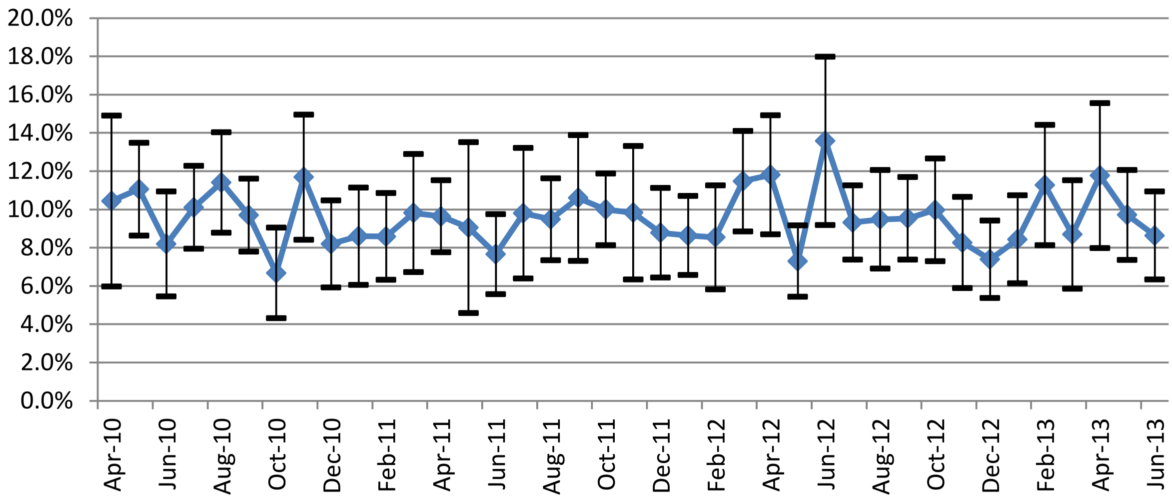
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-3. Monthly Estimates of Uninsured Age 19-64 in North Dakota



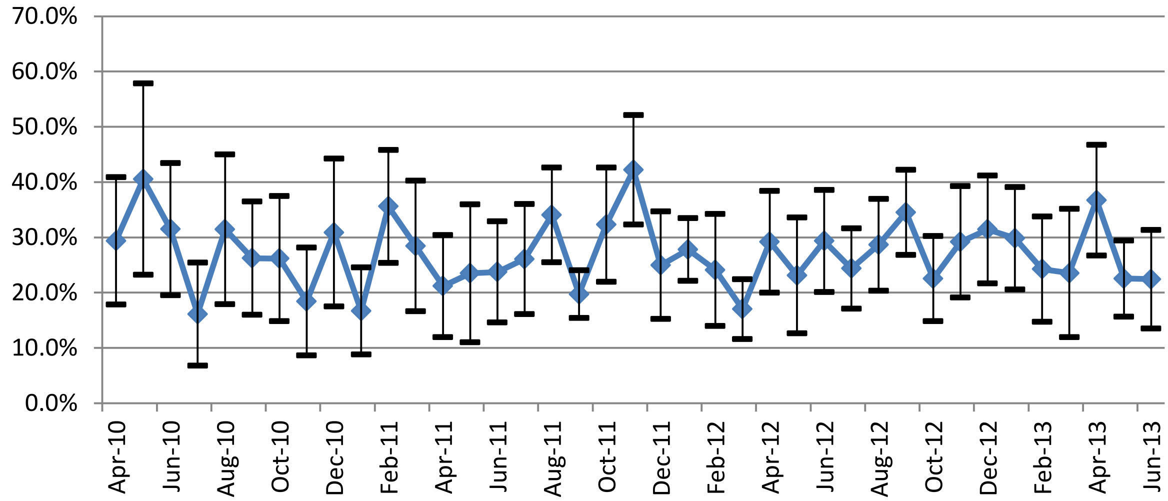
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-4. Monthly Estimates of Uninsured White Non-Hispanics, Age 0-64, in North Dakota



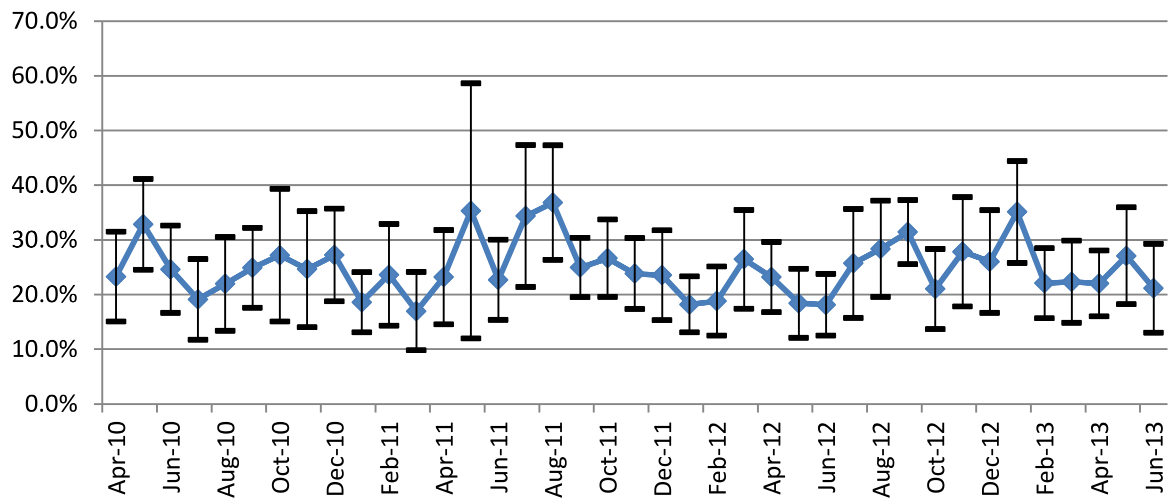
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-5. Monthly Estimates of Uninsured Minorities, Age 0-64, in North Dakota



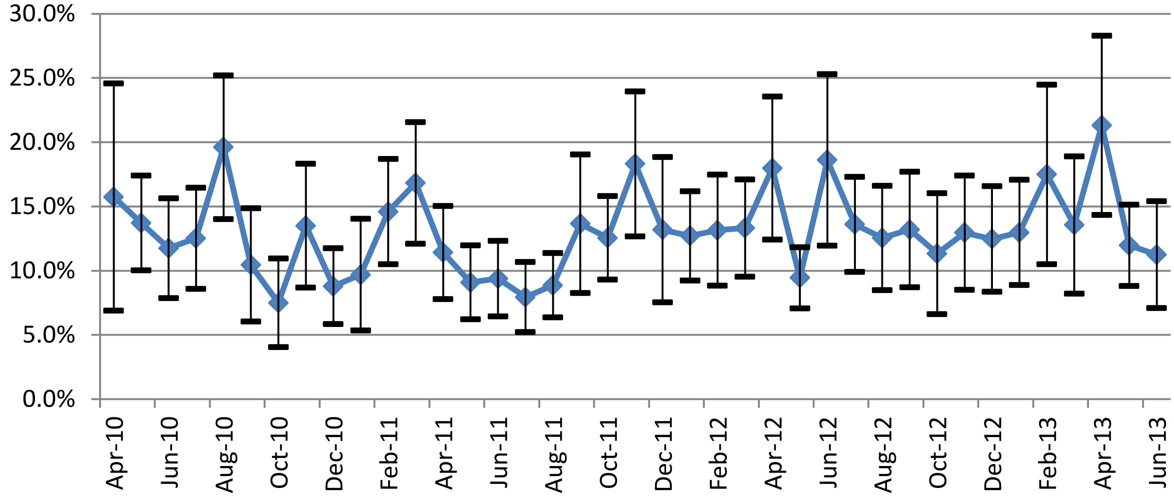
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in North Dakota



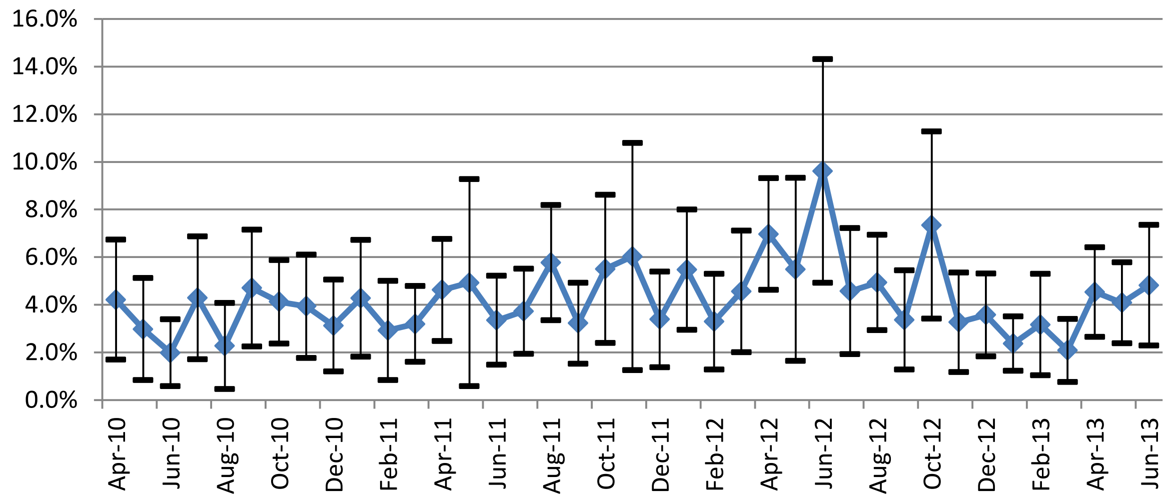
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in North Dakota



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in North Dakota



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

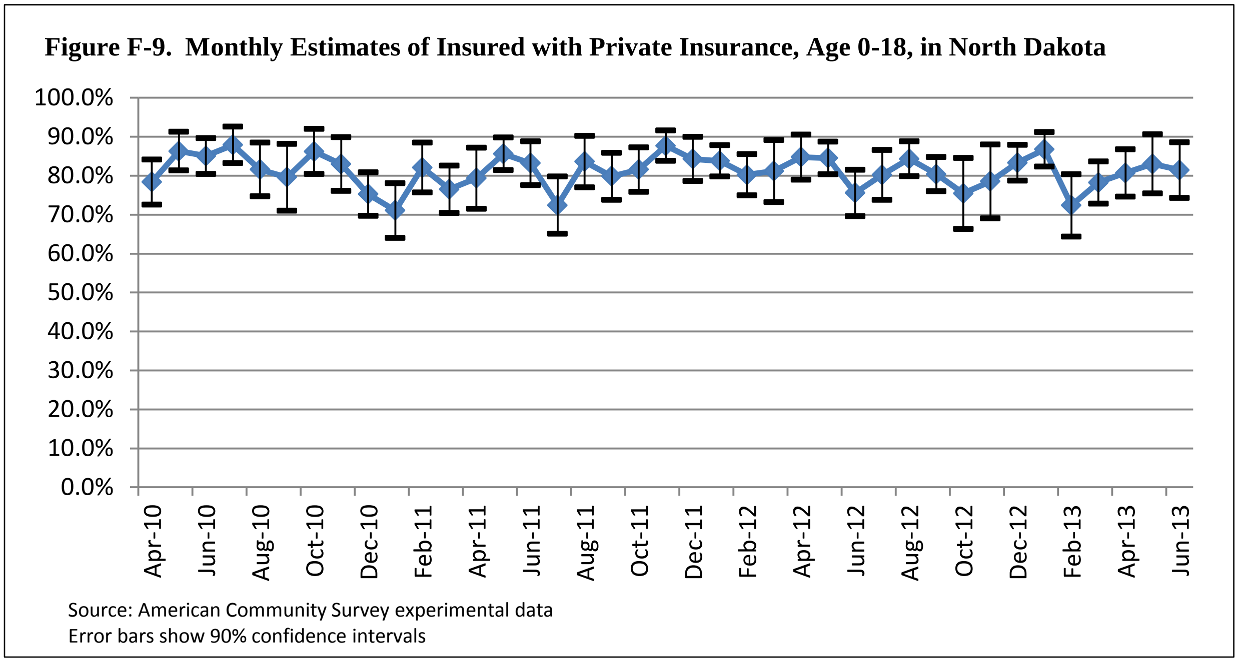


Figure F-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in North Dakota
100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
Apr-10
Jun-10
Aug-10
Oct-10
Dec-10
Feb-11
Apr-11
Jun-11
Aug-11
Oct-11
Dec-11
Feb-12
Apr-12
Jun-12
Aug-12
Oct-12
Dec-12
Feb-13
Apr-13
Jun-13
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

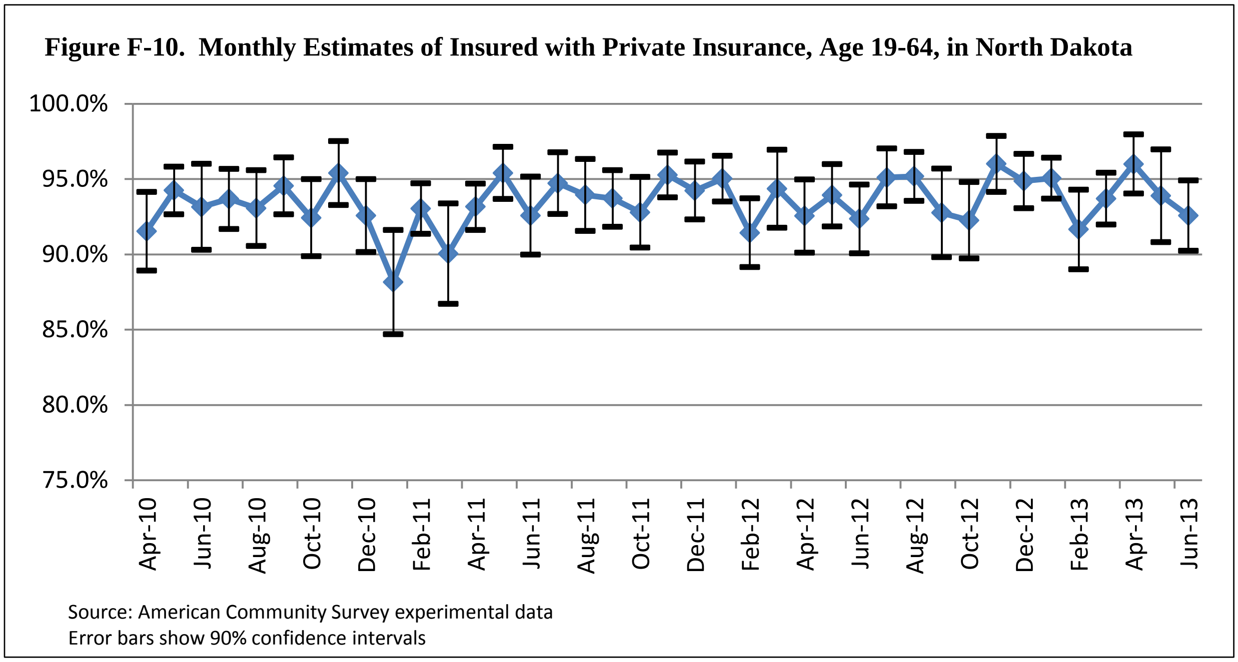
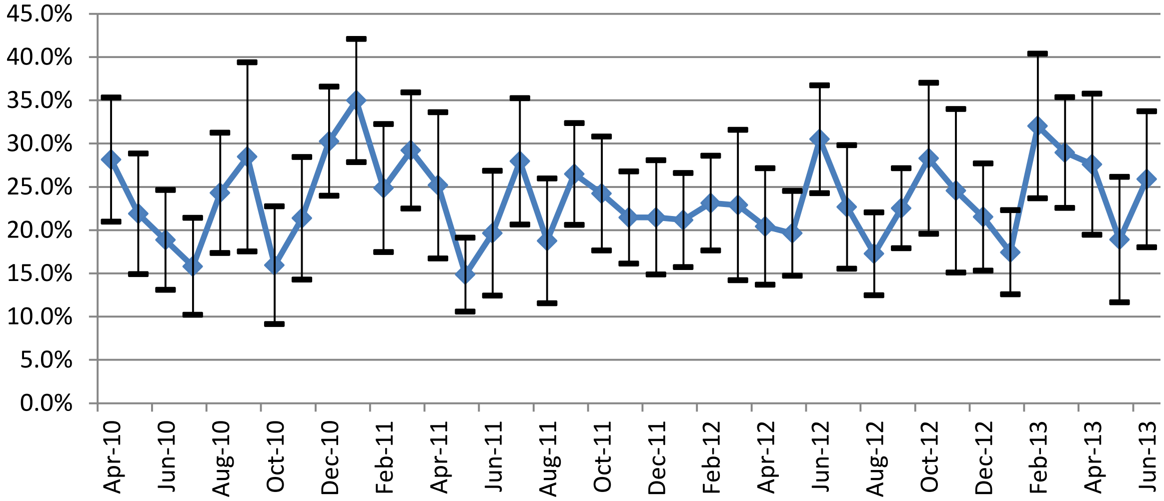


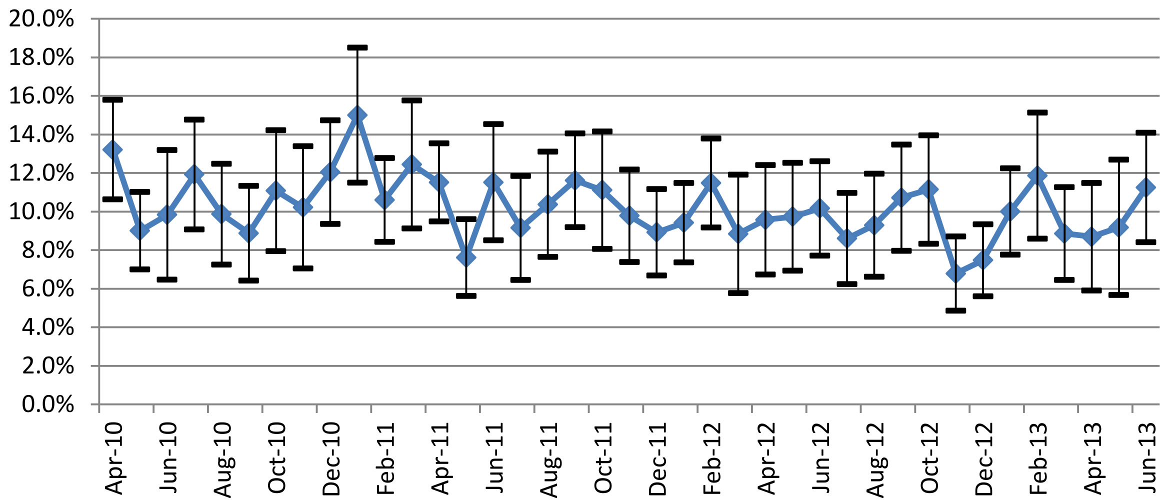
Figure F-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in North Dakota
100.0%
95.0%
90.0%
85.0%
80.0%
75.0%
Apr-10
Jun-10
Aug-10
Oct-10
Dec-10
Feb-11
Apr-11
Jun-11
Aug-11
Oct-11
Dec-11
Feb-12
Apr-12
Jun-12
Aug-12
Oct-12
Dec-12
Feb-13
Apr-13
Jun-13
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure F-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in North Dakota



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

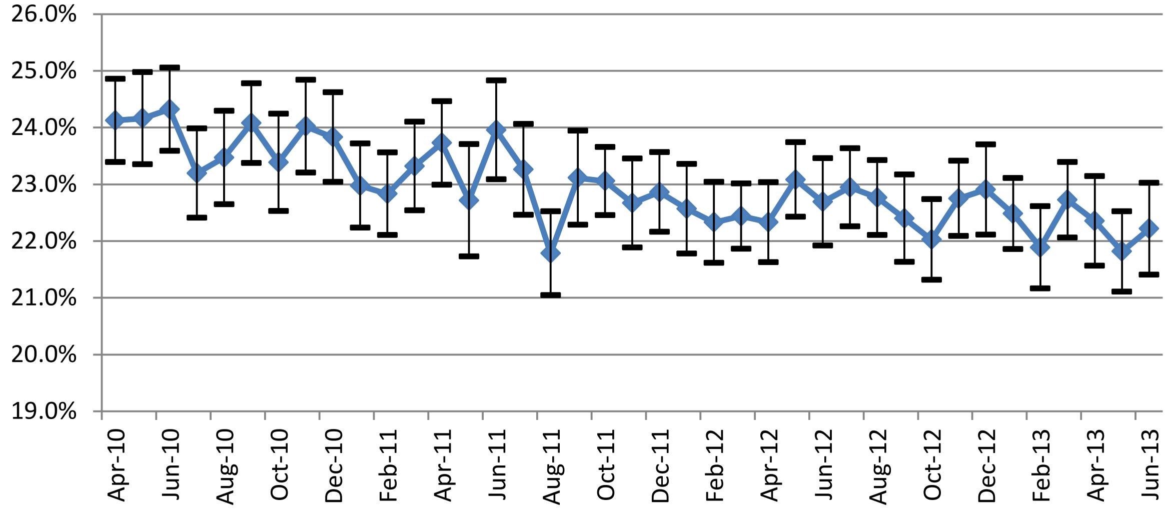
Figure F-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in North Dakota



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

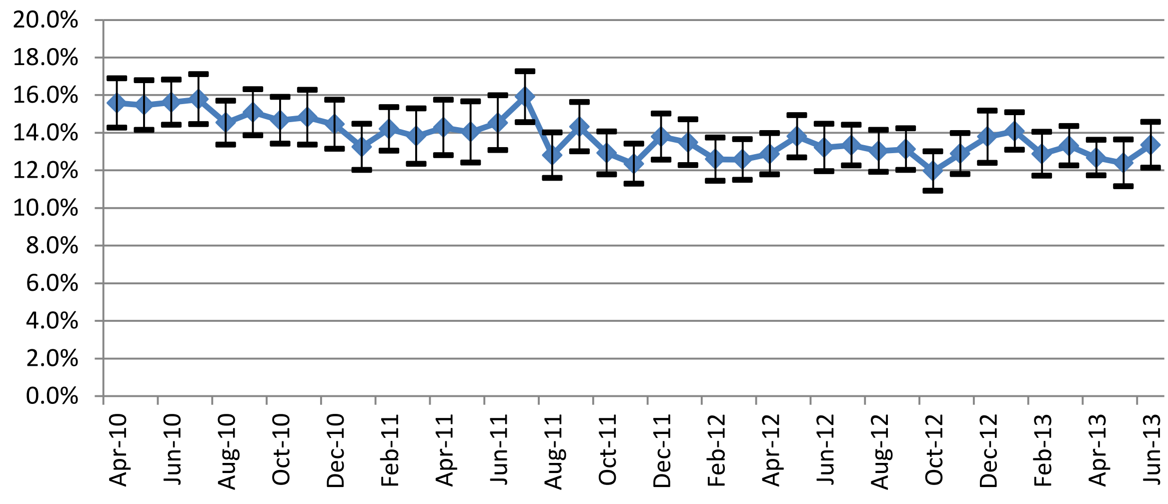
Attachment G.

Figure G-1. Monthly Estimates of Uninsured Persons in Texas



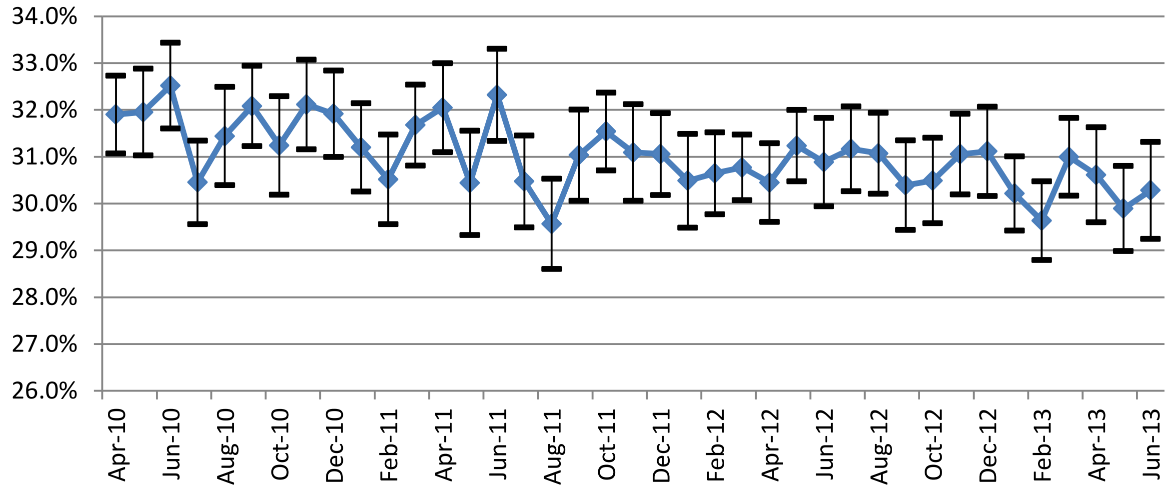
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-2. Monthly Estimates of Uninsured Age 0-18 in Texas



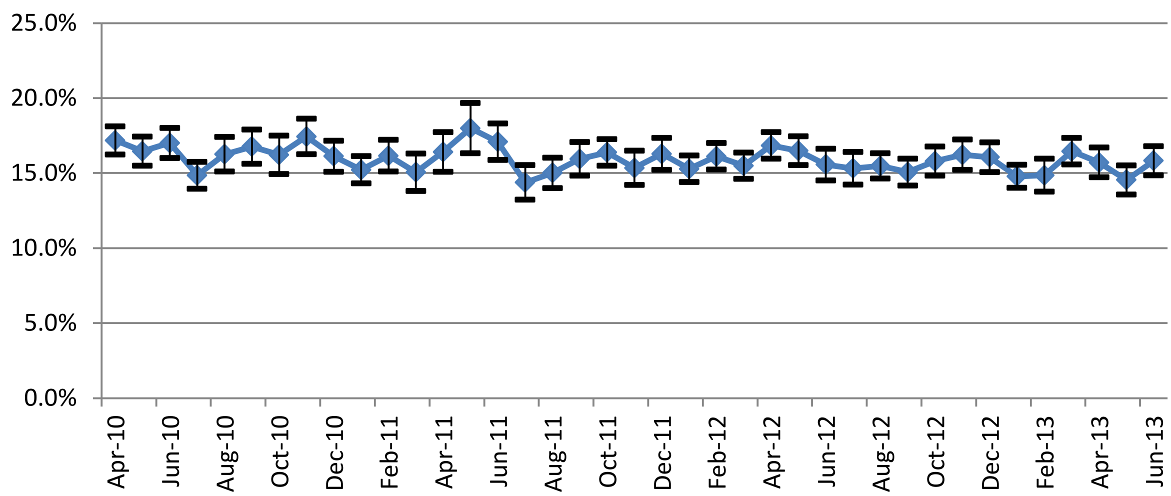
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-3. Monthly Estimates of Uninsured Age 19-64 in Texas



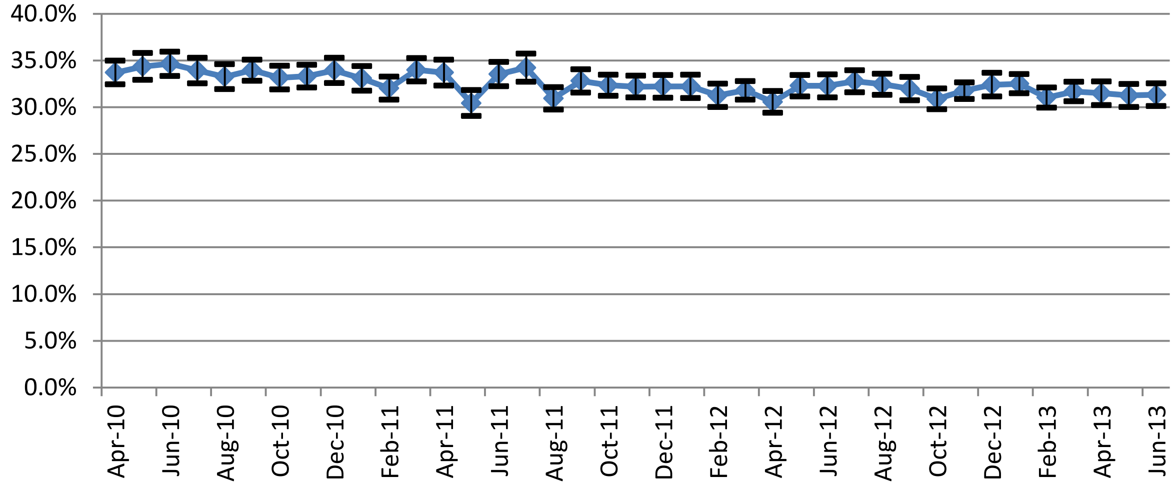
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-4. Monthly Estimates of White Non-Hispanics, Age 0-64, in Texas



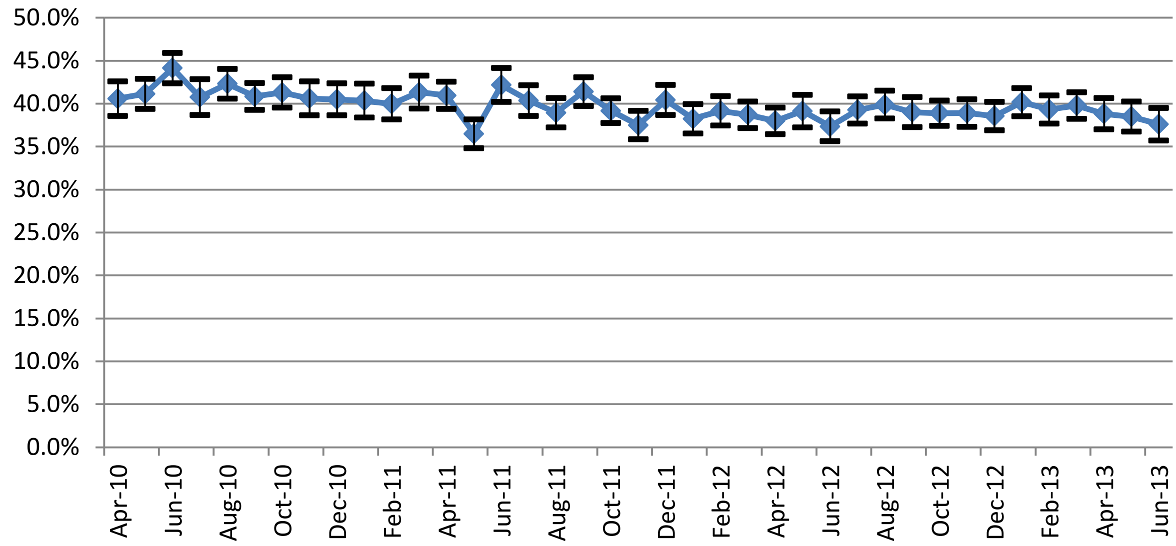
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-5. Monthly Estimates of Minorities, Age 0-64, in Texas



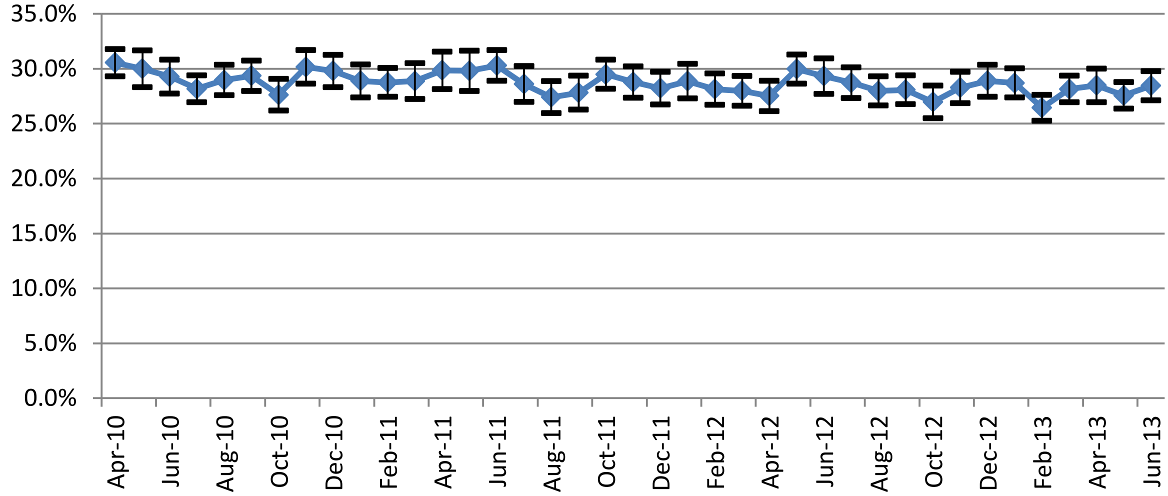
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-6. Monthly Estimates of Uninsured with Poverty Index 0-138, Age 0-64, in Texas



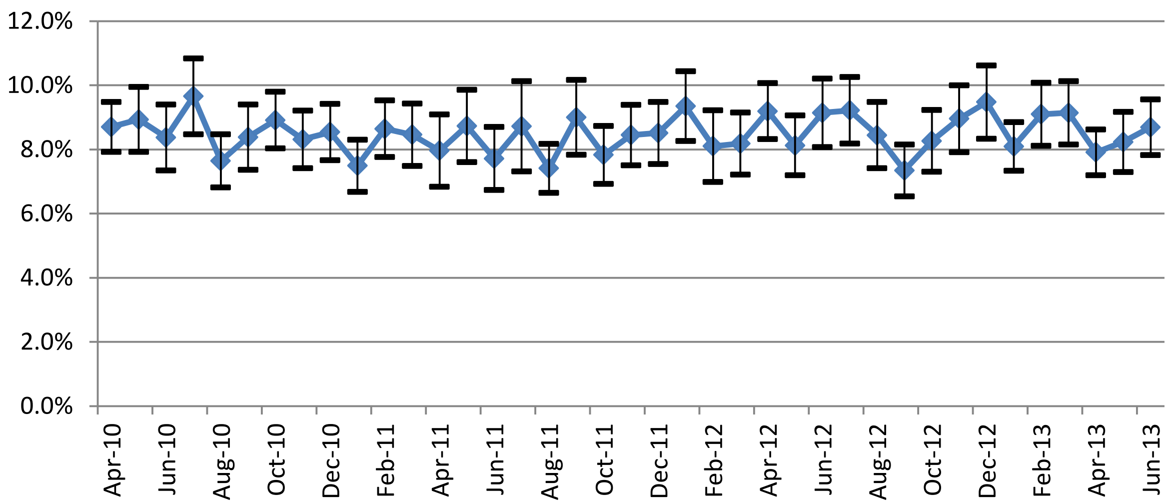
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-7. Monthly Estimates of Uninsured with Poverty Index 139-399, Age 0-64, in Texas



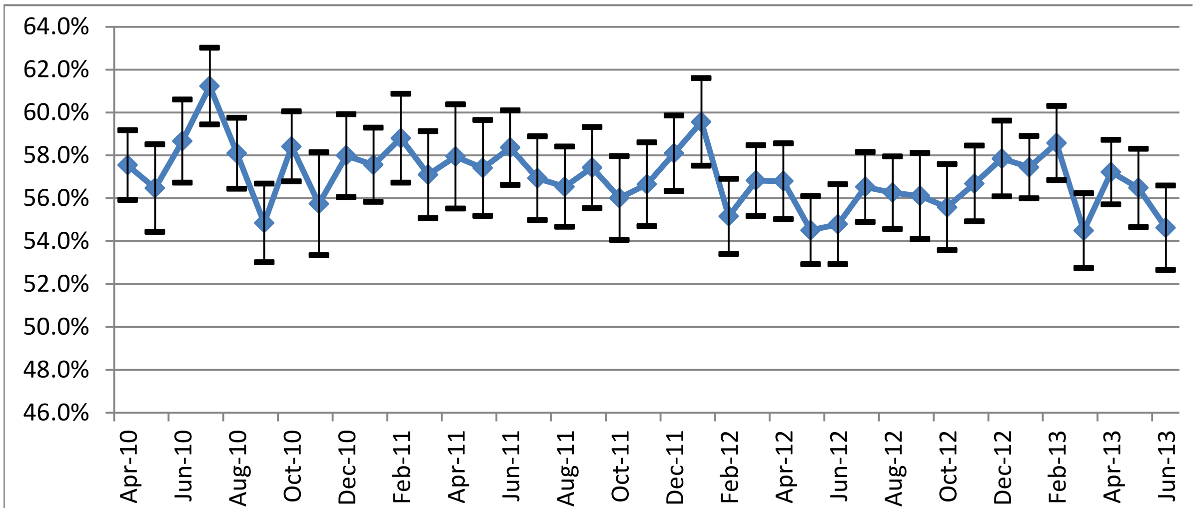
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-8. Monthly Estimates of Uninsured with Poverty Index 400+, Age 0-64, in Texas



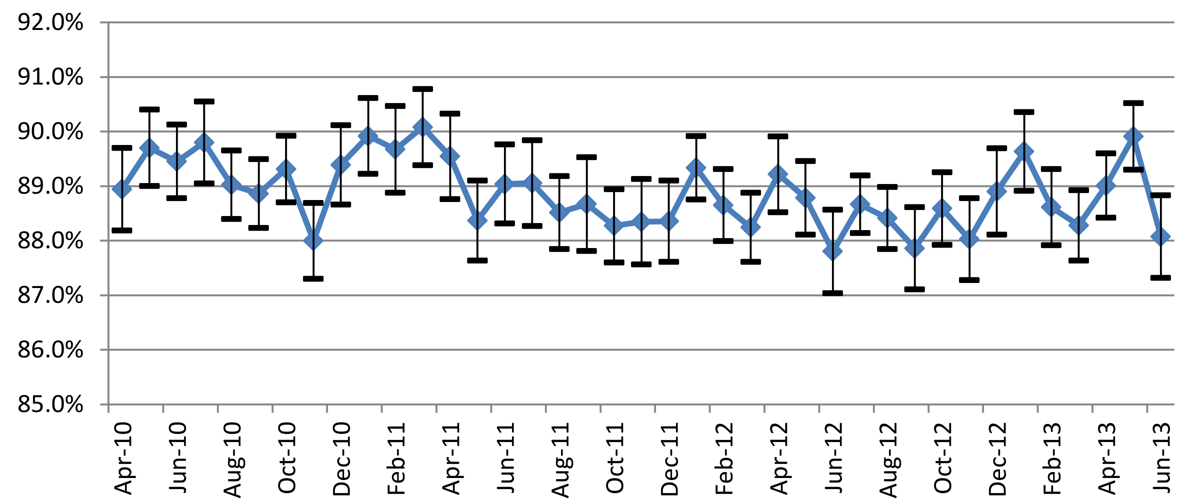
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-9. Monthly Estimates of Insured with Private Insurance, Age 0-18, in Texas



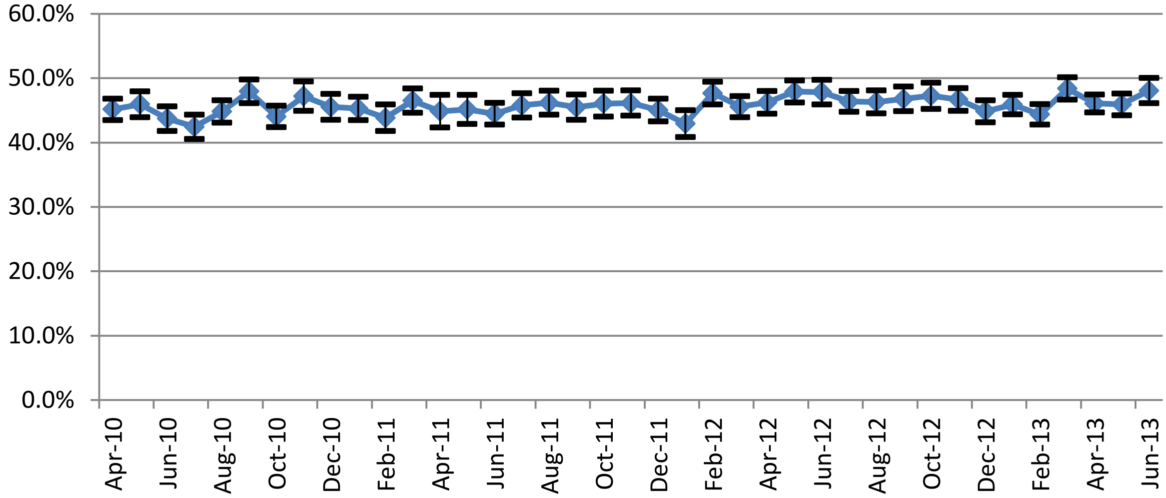
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-10. Monthly Estimates of Insured with Private Insurance, Age 19-64, in Texas



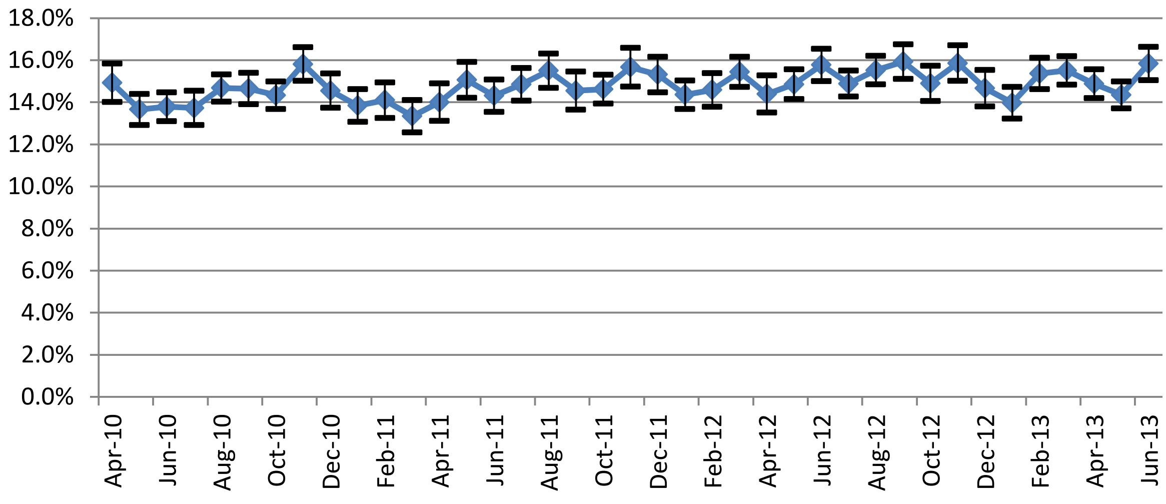
Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-11. Monthly Estimates of Insured with Public Insurance, Age 0-18, in Texas



Source: American Community Survey experimental data
Error bars show 90% confidence intervals

Figure G-12. Monthly Estimates of Insured with Public Insurance, Age 19-64, in Texas



Source: American Community Survey experimental data
Error bars show 90% confidence intervals