Monitoring Key Estimates from the National Health Interview Survey throughout the Realignment of Census Bureau Regional Offices[1]

Norilsa Toribio, Reid Rottach
US Census Bureau, 4700 Silver Hill Road, Washington DC 20233

Abstract

The NHIS is one of six household surveys that are currently part of the Demographic Data Monitoring System (DDMS), which captures survey responses from interviewer laptops on a daily basis and will frequently load survey cost data as well. Within the DDMS is a component to model survey outcomes and display tabular and graphical output. This paper presents details of the models used to study the effect of the Census Bureau's Regional Office realignment on key estimates of the NHIS, as well as issues encountered in developing the system and interpreting the results.

1. Introduction

The NHIS conducts household interviews to monitor the health of the United States population. The survey is sponsored by the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control and Prevention (CDC). The NHIS provides information annually on the health status of the civilian, non-institutionalized population residing in one of the 50 states and the District of Columbia, and it is the primary source of information on the health of its target population. The NHIS collects data on topics such as health insurance coverage, injuries, health status, use of health services, immunizations, etc. In addition to the survey's core questions, the NHIS contains supplements that are designed to address any changes in public health data needs (NCHS, 2012).

The NHIS has a stratified multistage sample design. The first stage of sample selection stratifies and selects the Primary Sampling Units (PSUs). A PSU consists of a county or a group of contiguous counties.

After PSUs are selected, the second stage of sample selection identifies the sample units within the selected PSUs. A new sample design is implemented after each decennial census. The U.S. Census Bureau carries out the data collection for the NHIS. A Census Bureau Field Representative (FR) conducts face-to-face interviews for the survey with Computer Assisted Personal Interviewing (CAPI).

---

[1] *Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.*

On January 1, 2012, the U.S. Census Bureau began phasing in a new Regional Office (RO) structure that redefined the U.S. Census Bureau's field operations. The next section provides an overview of this change in managing structure.

2. Background

The Census Bureau Regional Office structure contained twelve regional offices since 1961. The boundaries of these twelve regions are shown below.

Figure 1: Map of the twelve regional offices as of December 31, 2011.

The changes in the management structure were deemed necessary due to:

- Budget cuts
- Survey sponsors' demands for lower costs, improved efficiency, and increased responsiveness
- An increase in the use of real time administrative data to create leaner infrastructures

As a result, in January 2012 the U.S. Census Bureau began a transition that would consolidate its RO operations into six Regional Offices. This shift took place over seven waves, with the implementation of the first wave in January, and the last wave in November. The reorganization was complete by December 2012.

One eighth of the country changed over to the new management structure in each of the first six waves, then the remaining one fourth of the country converted to the new structure in the seventh wave. The map below shows the boundaries of the six remaining Regional Offices.

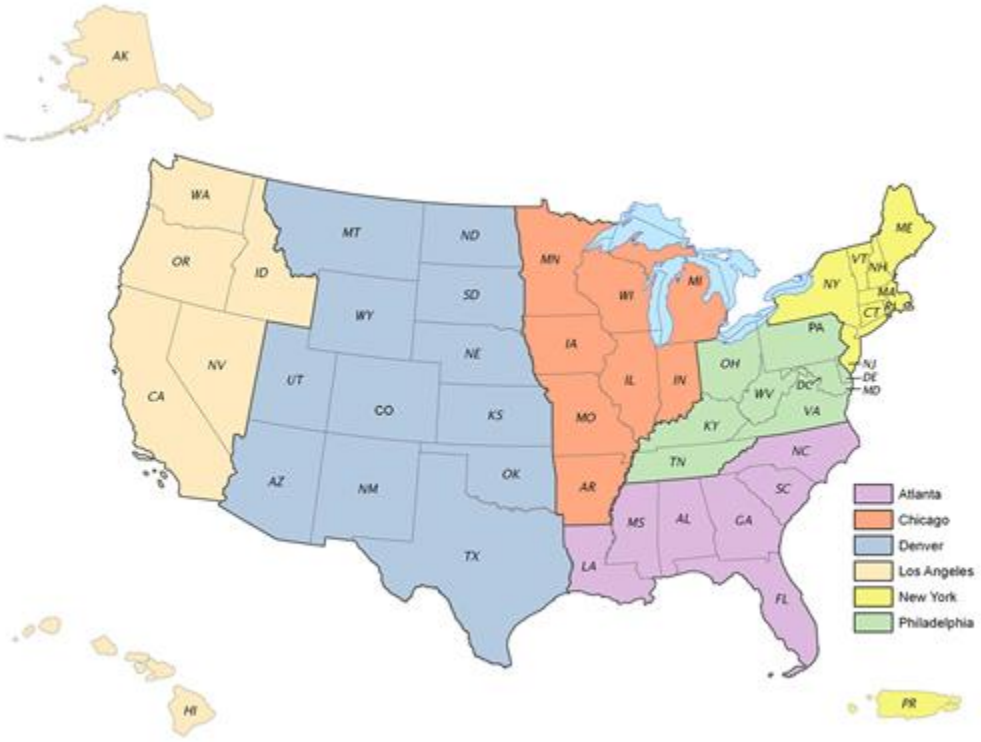

Figure 2: Map of the six regional offices as of January 2013.

As this changeover occurred, we monitored the survey responses daily in areas under the new management as well as areas that had not yet transitioned. Our goal was to determine whether the change in the U.S. Census Bureau's field operations was having an effect on the estimates produced. To do our analysis, we used unedited CAPI data as well as edited data that was used during the testing stages.

3. Key Variables

We modeled twelve key variables from three components of the NHIS questionnaire, namely, the Adult, Child, and Family questionnaires to determine whether the change in management was affecting the estimates produced by them. They are:

- **Private health insurance:** This item is asked of all family members and it provides the percent of respondents who have private health insurance.
- **Medicaid health insurance:** This item is asked of all family members and it provides the percent of respondents who have Medicaid, a type of health insurance provided by the federal government for those requiring financial assistance.

- **No health insurance:** This item is also asked of all members in the family. It indicates the percent of respondents who are uninsured.
- **Did not get needed medical care due to cost in the past twelve months**: This item indicates whether the family member did not receive needed medical care because of cost in the past year.
- **Needed help with personal care needs:** This item is asked of all family members and it indicates whether the member needed any help with activities of daily living.
- **Usual place of medical care (adult):** this item is asked of the sample adult and it indicates whether the adult has a place where they go when they are sick or need advice about their health.
- **Flu vaccination in the past twelve months (adult):** this item is asked of the sample adult and it indicates whether the respondent has had a vaccine within the last year that protects against influenza.
- **Current asthma (adult):** this item applies to sample adults who have indicated that they have had asthma in the past. The item is asked to assess whether the sample adult still suffers from the disease.
- **Smoking status:** this item applies to sample adults who have smoked at least 100 cigarettes in their entire life. The question addresses whether the sample adult still smokes.
- **Usual place of medical care (child):** this question applies to the sample child and it indicates whether there is a place where the child is usually taken for health-related reasons.
- **Flu vaccination in the past twelve months (child):** this question is asked of the sample child and it indicates whether the child has had a vaccine against influenza in the past twelve months.
- **Current asthma (child):** this question is asked of the sample child if they have had asthma. This item addresses whether the child still suffers from the disease.

4. Data

We used edited survey data to test and develop potential models. We were then able to use unedited CAPI data for the last half of 2011, which we used to select our twelve models. In addition, we made use of Census Bureau geography to construct variables to indicate when a particular area would change over to the new management structure.

We constructed the variable CURRENT_WAVE to indicate which wave of the transition corresponded to the current production month. Therefore, this is a categorical variable holding values 1-7 for the 12 months of the year 2012, and having a value of 0 for the historical data. We made use of Census Bureau tracts to determine when a particular area would change over to the new management structure. With this tract information, we constructed the variable WAVE_RO_CHANGE, which we used to indicate the wave during which a particular area was scheduled to undergo the change in the management structure.

Using the variables CURRENT_WAVE and WAVE_RO_CHANGE, we created the variable that we used as the indicator of whether the case came from an FR that was being managed under the new system; we call this variable the Z Variable. The Z Variable was constructed to have values 0 or 1 depending on whether the household interviewed was located in an area that was under the new structure. The Z Variable had a value of 1 for any particular interview only if the scheduled wave of change for the area had passed or was currently in progress when the FR conducted the interview. For instance, if a Census Bureau tract was scheduled to convert to the new management structure during Wave 1, and an interview in this tract was taking place during Wave 4 then the Z Variable would have had a value of 1 for this interview case since the area had already changed over to the new structure several waves ago. Hence, we constructed the Z Variable to equal one whenever CURRENT_WAVE $\geq$ WAVE_RO_CHANGE and to equal zero otherwise. We examined the significance of the Z Variable in each of the models to make inferences on whether the change in management structure was having an effect on the data being collected.

The Z Variable, CURRENT_WAVE and WAVE_RO_CHANGE were used as predictors in each of our twelve models. Also, the twelve-month lagged mean of the response variable was used in each of the models to account for seasonality. In addition to these variables, the models included covariates to improve model fit and to account for confounding factors, such as geographic differences in the transitioned areas.

We considered many variables to be included as predictors in our models. However, we were careful in the final number of covariates chosen for each one of the models keeping in mind the potential for model over specification. We considered the difficulties regarding small sample sizes, missing values and the correlation of covariates. Below we describe some of the variables we considered, although not all of the following were included in our models.

- **Demographic:**
  - Age of the respondent
  - Age of the respondent squared
  - Highest level of education
  - Hispanic origin of the respondent
  - Born in the United States
  - Race of the respondent

- **Household:**
  - Receive cash assistance from a government program
  - Housing unit owned/bought/rented

- **Health:**
  - Certain key variables such as the health insurance variables were used as predictors in other models
  - Health condition compared to twelve months ago (for sample child and

sample adult)
- Difficulty of walking a quarter of a mile by oneself
- An indicator of whether the child has an impairment that requires the use of special equipment
- Health condition
- Paid sick leave

- **Tract-Level Variables from the 2000 Census Bureau Planning Database:**
  - Percent American/Indian Native
  - Percent Asian
  - Percent Black
  - Percent Hispanic
  - Percent White
  - Percent of persons that did not graduate from high school
  - Percent of persons aged 65 years or older
  - Percent of persons below poverty
  - Percent of persons with public assistance
  - Percent of persons unemployed

- **Lagged Means:**
  - Twelve-month lagged mean of the response variable

5. The Models

All of our key variables were coded as binary having values of 0 or 1, where 1 represents the response of interest. As a result, logistic regression using PROC LOGISTIC was used to model all of our key measures. Our twelve models share the general form:

$$Logit(p_i) = \beta_0 + \beta_z z_i + \tilde{\beta}_c' \tilde{c}_i$$

where z_i is an indicator variable; $z_i = 1$ if the case was collected in an area under the new management system and $z_i = 0$ otherwise; p_i is the probability that the key measure is equal to one, and $\tilde{c}_i$ is the vector of covariates.

As a first step, we used the forward selection algorithm to narrow down the number of covariates to consider. We then reduced the number of possible covariates based on their significance using the Wald Chi-Square test statistic as a basis for the decision. The model fit was evaluated using the Hosmer and Lemeshow goodness-of-fit statistic and Pearson's R-squared.

We were careful in the selection of covariates because not all of the potential covariates could be applied to every model. The variables from the Family Questionnaire were the only ones that could be used as covariates across all twelve models. The remaining covariates were questionnaire-specific. Therefore, when we modeled sample adult and sample child key measures, only the variables from their respective questionnaires and the Family Questionnaire were used as covariates. Similarly, covariates from the Sample

Adult and Sample Child questionnaires could not be used when modeling key variables from the Family Questionnaire. For instance, we could not use a covariate indicating whether the respondent has paid sick leave to model health insurance because the covariate only applied to the sample adult.

Because our goal was to monitor the changes as they were happening, we could not wait until the changes began to take place to start developing our models. For this reason, our models were selected in the year 2011, a few months prior to the start of Wave 1 of the change. When testing the models, we developed a "mock" Z Variable that simulated the change in management structure on the 2011 NHIS data. Since the change in management had not yet occurred in 2011, our goal was to be able to construct models that, while highly significant, had Z values that were insignificant.

Once the change in management began to take place on January 1, 2012, we began to run the models we had developed during our testing phase. We ran our models every day using the unedited data collected in the field the previous day along with data that had accumulated up to that point. In addition to the daily data, our models also contained data from the year 2011, which we had used during our testing phase.

The Data Monitoring Systems Team, the IT team working on this project, turned the outputs produced by our daily models into the graphical outputs that we examined to help answer the research question of whether the change in management was having an effect on the data being collected. Each day, the Systems Team turned the CAPI data into a SAS dataset in order for the models to be run. Then, after our models produced the statistics of interest, the Systems Team used Oracle Application Express (APEX) to produce graphs and summary tables displaying the Beta-Z coefficient for each key measure and their corresponding confidence boundaries.

Each day, we examined the significance of the Z Variable in each one of our models. We were interested in the value of the Beta-Z coefficient in relation to its upper and lower bounds with a 95% confidence interval. We monitored this Beta-Z coefficient daily, testing the hypothesis that Beta-Z was equal to zero, looking for patterns and keeping track of any that were significant or very close to the boundaries.

6. Limitations

We faced certain challenges when selecting our models. Our main concern was that the NHIS sample size is small, especially when compared to that of other surveys that were part of the project. With no sample cuts or augmentations, the expected number of completed interviews is approximately 35,000 households per year (NCHS, 2012). The small sample size made it quite challenging to select models for the sample adult and sample child key measures because, especially at the beginning of data collection in early 2012, the data gathered was not sufficient to find a proper fit for the sample child and sample adult models.

Another concern regarding the models for the sample adult and sample child key measures was that due to skip patterns, these key variables were often blank. For instance, the key measure "Current asthma" for the sample adult (the situation is analogous to the sample child key measure). The question generating this key measure is only asked if the sample adult indicates that they have been told by a health professional that they had asthma. If the answer to this preliminary question is "Yes" then they key

measure "Current asthma" is asked to determine whether the respondent still suffers from the disease. Consequently, the question is never asked of respondents who have never been told they had asthma and, as a result, in many cases this key measure could not be modeled.

7. Changes to the Models

As Wave 1 progressed and more data was accumulated, it became apparent that some of our models could be revised in order to improve the fit. Before committing to making any changes, we ran the proposed revised models separately every day, keeping track of the model statistics to make sure that the improvements we were attempting to make would remain stable over time. We implemented changes that affected the sample adult key measures, two of the sample child key measures as well as the Medicaid Health Insurance key measure.

The changes made to the aforementioned models consisted of revisions to their list of covariates to remove existing covariates and replace them with new ones or to replace an existing covariate with a recode of itself. In order to ease the implementation of the changes, we chose not to update the estimates that had already been produced by the models with the estimates produced by the revised models. The changes were implemented as Wave 1 was phasing out at the end of March 2012.

8. Output from the Models

We ran our models each day examining the Beta-Z estimate in relation to its confidence bounds with a 95% confidence interval, which we used as the basis to help answer the research question of whether the change in management was having an effect on the estimates being produced. Because of this, we wanted to enhance our output and produce graphs that resembled quality control charts so that we could determine quickly whether the Beta-Z coefficient was outside its confidence bounds on any given day. As a result, we centered our confidence intervals around zero, making our graphs resemble quality control charts. Our confidence intervals were defined as follows:

$$LCI = -(StdErr * 1.96)$$
$$UCI = (StdErr * 1.96)$$

Whenever the Beta-Z coefficient was outside the confidence bounds, we monitored the estimate to determine whether any alarming trends were developing and to identify any tendencies that could warrant further investigation. We expected any trends or patterns to develop slowly because the models were cumulative, in that the data used any given day was the same as data used the previous day with one more day's worth of data collection.

The model outputs from Private Health Insurance, Flu Vaccination in the Past 12 Months (Adult) and Usual Place of Medical Care (Child) follow. We consider these model outputs to be representative of the NHIS models for this project.

Model: Private Health Insurance
Covariates: Lagged Mean of Private Health Insurance, Z Variable, Current Wave, Wave of RO Change, Race, Tenure, Education, Hispanic Origin, Percent of Persons Below Poverty Level

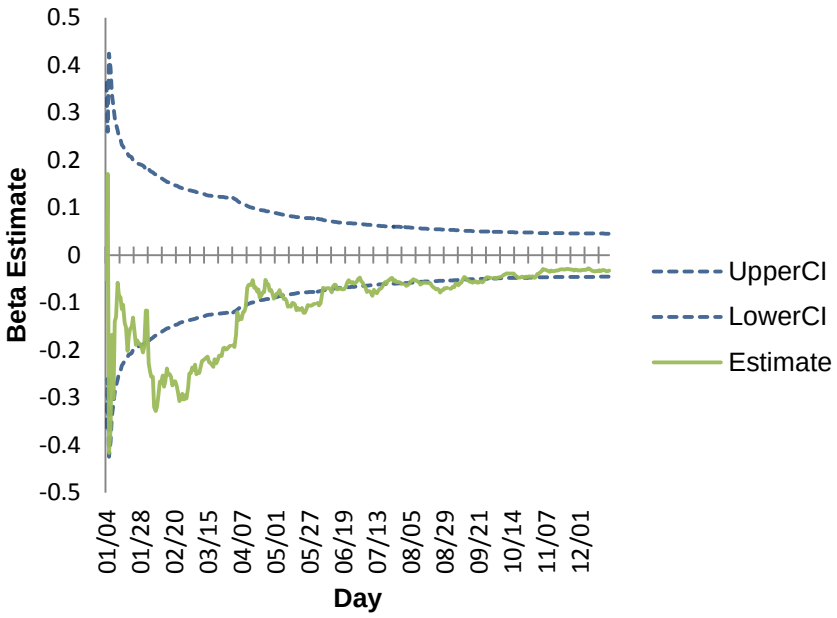

Figure 3: National NHIS Beta-Z for Private Health Insurance by day (95 % CI).

As Figure 3 shows, the Beta-Z estimate for the Private Health Insurance key measure became out of bounds not too long after data collection under the new management structure started on January 1, 2012. As Wave 2 began on April 1, the beta value was slowly trending back within the boundaries and it stayed within bounds from mid-April until the end of the month.

Waves 3, 4 and 5 had beta values that were out of bounds for most of the data collection period. However, when out of bounds, the betas were very close to the lower confidence bound. By the start of Wave 6 in October, the Beta-Z estimate was within the confidence bounds, remaining within the limits for the rest of the year 2012.

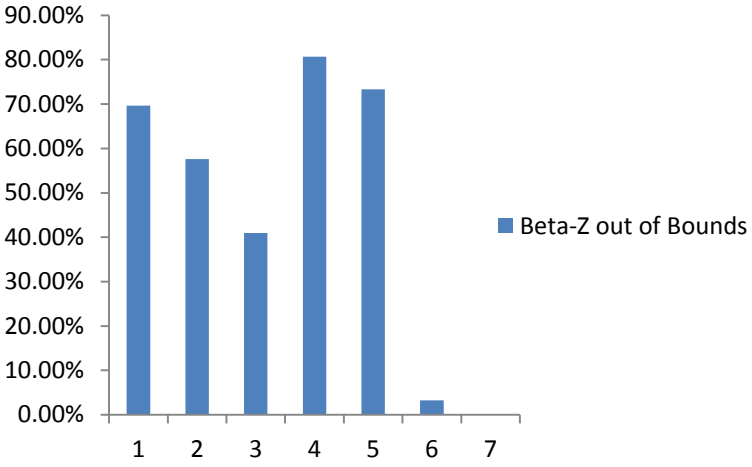

Figure 4: Percentage of Beta-Z values out of bounds for Private Health Insurance by Wave.

As Figure 4 shows, Wave 4 of the change (in August 2012) experienced the highest percent of Beta Z's that were out of bounds. By Wave 6, there were almost no betas out of bounds for the rest of the year.

Tracking the Beta-Z for Private Health Insurance was very interesting. The Survey of Income and Program Participation (SIPP), a survey that was also part of the project, was also tracking their Private Health Insurance key measure. As we monitored the daily outputs, we started to notice that our Beta-Z estimate was beginning to trend outside the confidence intervals while something completely different was happening with SIPP. As Figure 5 shows, SIPP's Beta-Z for the Private Health Insurance key measure had been well within bounds and rather stable. Seeing how different the outputs were for the same key measure triggered an investigation.

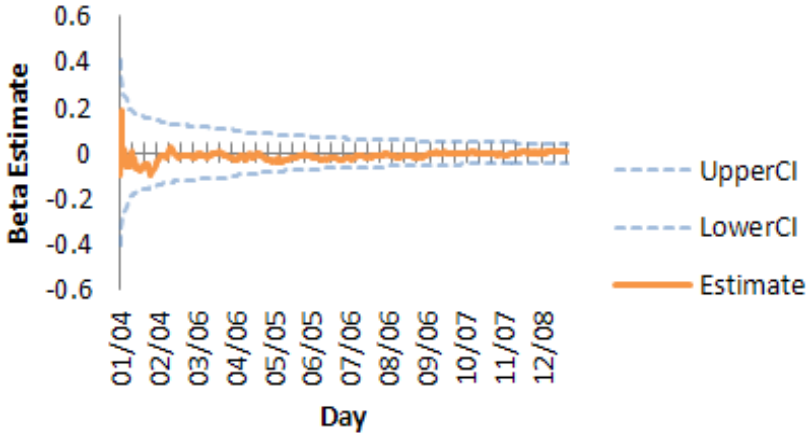

Figure 5: National SIPP Beta-Z for Private Health Insurance Coverage by day (95% CI).

Why were NHIS and SIPP so different on the Private Health Insurance variable? We worked towards answering this question during Wave 1. We evaluated the covariates being used in both models and we ran separate models for NHIS that, to the extent possible, incorporated the same covariates being used by SIPP. Doing so did not change the significance of the Beta-Z value, and our output did not seem to be greatly affected by the change in covariates.

SIPP's Health Insurance model included household members of age 15 and older, while ours included every person in the household. SIPP's model included marital status as a covariate but we could not include this variable as a covariate since children were included in our models and marital status was not part of the sample child questionnaire.

One of the biggest differences between our models had to do with how the question is asked in both surveys. For SIPP, the questionnaire asks whether the household member is covered by private health insurance "in this month" (U.S. Census Bureau 2008). The NHIS questionnaire has a preliminary question asking if a person is covered by any kind of health insurance. If the answer is "Yes", "Don't Know" or "Refused" then a question is asked to enter all the types of health insurance that apply; this is where a person identifies if they have private health insurance.

Sample design is another key difference between SIPP and NHIS that may have influenced how different our Private Health Insurance models were behaving. SIPP's

design consists of rotating sample panels where a household is visited more than once during the life of the design (U.S. Census Bureau 2006). For NHIS, a household can only be interviewed once during the course of the design. Due to these differences, we concluded that the nature of the question in both surveys was different enough to produce outputs that were not similar.

Model: Flu Vaccination in the Past 12 Months (Adult)
Covariates: Lagged Mean of No Health Insurance, Paid Sick Leave, Difficulty of Walking a Quarter of a Mile by Oneself, Current Wave, Wave of RO Change, Z Variable, Lagged Mean of Flu Vaccination in the Past 12 Months (Adult)

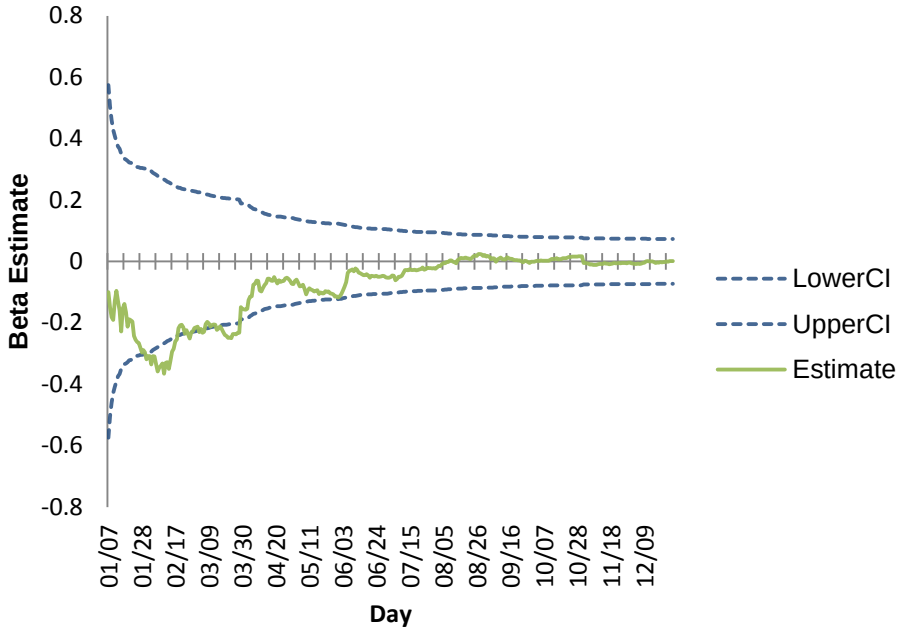

Figure 6: National NHIS Beta-Z for Flu Vaccination in the Past 12 Months- Adult by day (95 % CI).

The output for this key measure reflects what we expected to obtain from our models. We can see that during the beginning of the data collection period, when we did not yet have much data, our Beta-Z estimate was not very stable. However, as time progressed and the amount of data increased, the Beta-Z estimate started to stabilize and settle around zero.

Flu Vaccination in the Past 12 Months (Adult) is one of the models we changed during the end of Wave 1 in order to improve model validity. As Figure 6 shows, by the beginning of Wave 2 all of the Beta-Z values were within the confidence bounds. Forty-two betas were outside the confidence bounds, all being out of bounds during the first wave of the change.

Model: Usual Place of Medical Care (Child)
Covariates: Lagged Mean of Private Health Insurance, Current Wave, Wave of RO Change, Z Variable, Lagged Mean of Usual Place of Medical Care (Child)

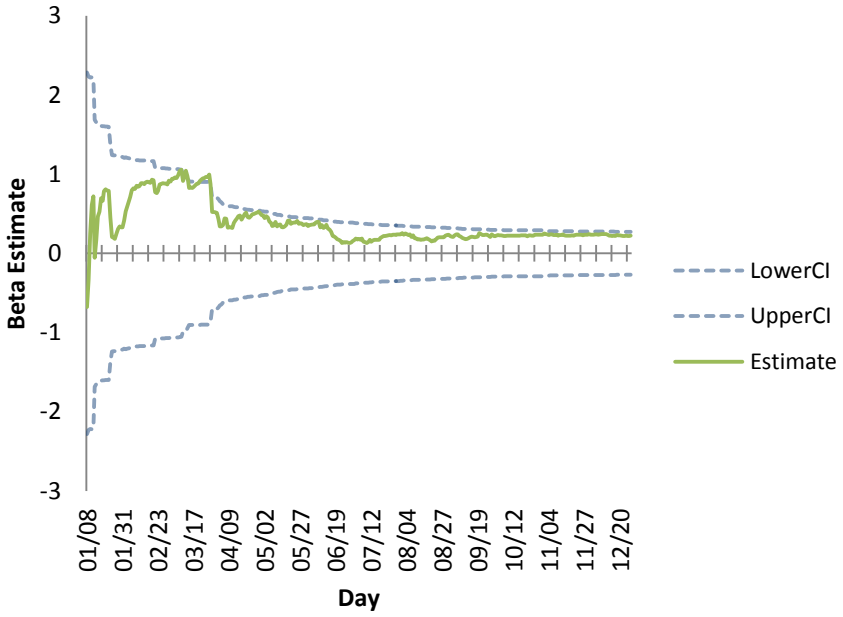

Figure 7: National NHIS Beta-Z for Usual Place of Medical Care –Child by day (95 % CI).

The key measure "Usual Place of Medical Care (Child)" is also one of the models that were modified at the end of Wave 1 to improve model validity. Eleven betas were outside the bounds for this key measure, and they all became out of bounds during the first wave of the change. This was an interesting model to keep track of because, as Figure 7 shows, although the betas were within bounds after the first wave, the Beta-Z estimate for "Usual Place of Medical Care (Child)" stayed relatively close to the upper confidence bound throughout the year 2012.

9. Conclusion

Our goal was to determine whether the change in the Regional Office management structure was having an effect on the survey data collected. The changes were phased in across seven waves in 2012, with the first in January and the seventh in November. The measure we chose for this study was the regression coefficient on an indicator of the new management structure, which we called the Beta-Z coefficient.

In predicting survey responses, we tested whether the Beta-Z coefficient was significant, controlling for other factors. However, we could not control for everything. Hence, this project began with the understanding that greater than 5% of the coefficients would likely end up outside the 95% confidence bounds even if there were no management effect. The final output produced by the team showed that approximately 10% of the estimates were outside the confidence bounds. The judgment is subjective, but this data alone would be weak evidence to suggest that the management change itself had an effect on the quality of our data.

10. Acknowledgements

We would like to thank Kimball Jonas and Robyn Sirkis who provided much appreciated support throughout the course of this project and also provided valuable assistance during the writing of the research summarized here.

We would also like to thank Anne Kearney for the valuable input she provided throughout the project.

References

Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985-94. National Center for Health Statistics. Vital Health Stat 2(110). 1989.Series 2 no. 110

NCHS (National Center for Health Statistics) (2012).“About the National Health Interview Survey.” < http://www.cdc.gov/nchs/nhis/about_nhis.htm>. Posted June 2012, accessed March 2013.

Office of Management and Budget (2006).“Standards and Guidelines for Statistical Surveys.”<http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf. Posted September 2006, accessed July 2013

U.S. Census Bureau (2008). “Survey of Income and Program Participation” <http://www.census.gov/sipp/core_content/2008/2008.html> Posted March 2008, accessed March 2013.

U.S. Census Bureau (2006). “Survey of Income and Program Participation”
< http://www.census.gov/sipp/overview.html> Posted May 2006, accessed March 2013.