Response to Surveys of High-Profile Topics: The Effects of Media Coverage and Public Engagement on Response to the National 2009 H1N1 Flu Survey

Nicholas D. Davis
NORC at the University of Chicago

James A. Singleton and Lina Balluz
Centers for Disease Control and Prevention

ABSTRACT:

The National 2009 H1N1 Flu Survey (NHFS) was conducted by NORC at the University of Chicago on behalf of the Centers for Disease Control and Prevention from October 2009-June 2010. The purpose of the survey was to monitor national and state-level influenza A (H1N1) 2009 monovalent and trivalent seasonal vaccination coverage rates during the 2009-2010 pandemic influenza season. The NHFS was conducted during a period of considerable media attention to influenza. The goal of this paper is to identify the impact of that attention – as well as corresponding shifts in public interest in influenza – on the Council of American Survey Research Organizations (CASRO) response rate and its components. A comparison of NHFS response rates over time to changes in the same rates for the Behavioral Risk Factor Surveillance System (BRFSS), another health-related but not influenza-specific survey, is examined to help quantify this effect. By separating long-term trends in response from shorter-term changes and comparing to aggregate indices of media coverage and public interest in the "flu," the paper determines that there is some evidence that changes in media attention may have both long-term and short-term effects on response propensity. We discuss the effects of changes in response related to potential non-response bias in survey estimates, identifying practical implications for conducting future pandemic influenza or other surveys during public health or other emergencies.

1. Introduction

In June of 2009, a novel strain of influenza known as 2009 pandemic influenza A (H1N1) virus was declared a pandemic by the World Health Organization. This influenza strain is also commonly known as the 2009 H1N1 flu or "swine flu". The development and rapid spread of illness from the 2009 H1N1 influenza virus became a major public health concern, and the influenza A (H1N1) 2009 monovalent vaccine (2009 H1N1 vaccine) was recommended initially for certain high-risk groups and later for the general public. The vaccine became available to the public in early October 2009. In order to monitor and evaluate influenza vaccination efforts among adults and children, the National 2009 H1N1 Flu Survey (NHFS) was implemented by the National Center for Immunization and Respiratory Diseases (NCIRD) and the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC).

The NHFS was a random digit dial (RDD) telephone survey conducted by NORC at the University of Chicago, operating from October 2009 through June 2010. The survey provided weekly monitoring of H1N1 and seasonal influenza vaccination coverage rates, nationally and at the state level, among all persons age six months and older. In addition to vaccination coverage, the survey obtained information regarding respondents' health behaviors and attitudes toward the flu, as well as a number of household demographics. The survey used a dual-frame sample of both landline and cellular telephone numbers and followed a five-week rolling sample design, in which sample lines were released each week and contact attempts continued for up to five-weeks or until a completed interview was obtained.

Due to increasing concern about a potential public health emergency, and also to the necessary public dissemination of updated information regarding the availability of influenza vaccines, the fall of 2009 saw elevated media attention to influenza and vaccine-related issues. This resulted in greater public awareness of and engagement in these issues than might otherwise have been the case. Both anecdotally and in the monitoring of key production metrics of the NHFS, survey administrators saw evidence that response rates to the NHFS declined over the course of the survey as media coverage and public engagement in the flu and flu vaccines waned.

A paper published in 2000 (Groves, Singer, Corning) established the Leverage-Saliency Theory of survey participation. In this theory, "leverage" describes the relevance of a particular survey attribute in the decision to participate, while "salience" refers to the level to which that attribute was made clear to the prospective participant. For example, considering a survey's topic as the attribute of interest, the leverage of this attribute is the extent to which the topic affects the decision to participate, and the salience of the attribute is the extent to which the survey's topic is clearly articulated. A subsequent article in 2004 (Groves, Presser, Dipko) showed that an individual's interest in the survey topic does affect the subject's decision to participate, and that this effect may lead to response bias in some situations.

While these studies related to individual decisions to participate in a survey, the NHFS provides a unique opportunity to explore whether there is evidence that large-scale, nationwide shifts in public interest in and media attention to a survey's topic can affect response rates. Due to the timing of survey administration, the relevance or leverage of the survey's topic was initially heightened but declined over time. The salience of the survey topic was rather high, with the questionnaire's introductory script reading "…we are conducting a nationwide study of the swine flu pandemic, also known as the 2009 H1N1 influenza virus." While it is a common goal for a survey's topic to be relevant to the target population, problems may arise when the relevance changes over the duration of the survey, especially if this results in changes to the structure of response propensity.

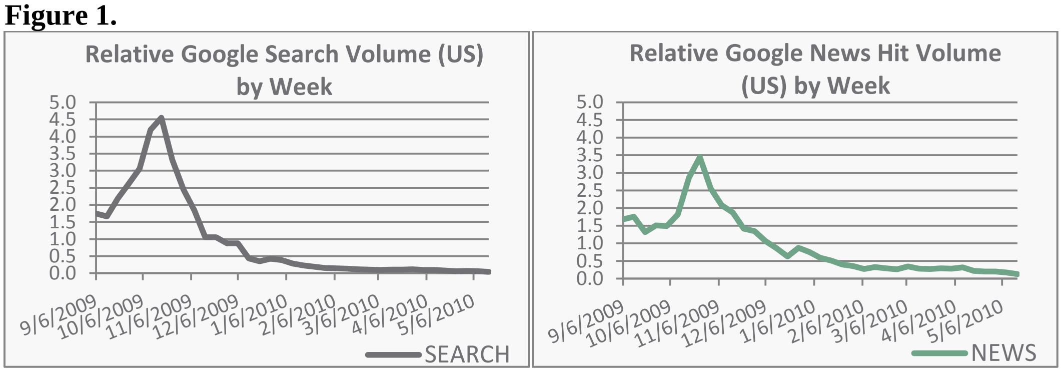
2. Methods

The data available to analyze this problem include the NHFS CASRO and component response rates, including the resolution rate, screener completion rate and interview completion rate, by week of sample release. For completed interviews, responses to select survey items are also available. For the purposes of this paper, we did not use sample cases released during the last five weeks of data collection, as these cases did not have the full five week period to respond.

For the purposes of comparison, similar data from the Behavioral Risk Factor Surveillance System (BRFSS) were also collected, including response rates at the monthly level and responses to select survey items. BRFSS data is convenient for comparison because, like the NHFS, it is a nationwide, health-related survey targeting adults. Unlike the NHFS, however, BRFSS is not H1N1 or influenza-specific and thus did not present this topic as saliently to potential respondents. Another difference between the NHFS and BRFSS is the inclusion of cell phone sample cases in the NHFS but not in the BRFSS sample.

Finally, we obtained aggregate measures of public engagement and media coverage. Gathered from Google's free internet services, we obtained Google Search volume figures by week, using the search terms "H1N1", "flu shot", "flu vaccine", and "swine flu", as a proxy for general public engagement or interest in these topics. Similarly, the number of Google US News stories retrieved (hit volume) by week, using the same search terms, was used as a representation of media attention to the influenza pandemic and vaccination-related issues. Figure 1 shows the national-level relative Google Search volume and relative Google News hits by week from September 2009 through May 2010. The results are shown relative to a baseline volume. That is, each weekly value is expressed as a percentage of the average weekly volume over the entire time period presented. Due to this multiplicative nature, we use the natural logarithm transformation to obtain data that is more linear over time.

Figure 1.



2.1 NHFS Response Rate Comparison

Between the first week of October, 2009 and the fourth week of May, 2010, a total of 897,169 telephone lines were drawn from the NHFS sample frame. Of these, 670,841 were landlines and 226,328 were cell phones. For these sample cases, we examined changes over time in the CASRO and component response rates, by week of sample release. In standard terminology, the resolution rate is the percentage of sampled lines that were resolved as either residential, non-working or out-of-scope, and the screener completion rate is the percentage of residential households completing the questionnaire's screening section. For landline telephones, screening consisted merely of confirming that an adult (18+) was on the line. For cell phone cases, screening consisted of confirming that an adult was on the line and that the household would be considered a "cell only" or "cell mainly" household. Finally, the interview completion rate is the percentage of screened and eligible household respondents that completed the survey by providing at least their H1N1 and seasonal influenza vaccination coverage status. The

CASRO response rate can be thought of as the product of the resolution rate, screener completion rate, and interview completion rate.

If it were true that topic relevance were the predominant factor driving changes in the CASRO response rate, then we would expect to see changes in the CASRO response rate manifest primarily in the interview completion rate. The resolution rate would not be expected to change at all relative to topic interest, and as noted earlier, the screening section of the NHFS was very short, in the case of landline telephone cases requiring only the confirmation of an adult on the line. Noting that the CASRO response rate declined over the course of the survey, seeing this decline primarily in the interview completion rate would provide evidence that waning topic relevance was a factor.

As mentioned earlier, NHFS response rates were also compared to similar rates from BRFSS. BRFSS response data includes landline sample cases only, and was available at only the monthly level because BRFSS sample is released on a monthly basis. The response data includes the CASRO response rate (also available for the NHFS), the resolution rate, and the survey cooperation rate, which is equal to the product of the screener completion rate and the interview completion rate (unlike the NHFS, separate screener completion rates and interview completion rates were not available for BRFSS). If the declining trend in the NHFS rates were not seen in the BRFSS response rates, this would further support our hypothesis that changing interest in the flu was a factor, because changing interest in the flu would be expected to have a greater effect on the response rates of flu-specific surveys such as the NHFS than on other surveys that are not flu-specific.

2.2 Modeling the Effects of Public Engagement on the Interview Completion Rate

If trends in the NHFS CASRO and component response rates, and comparison of these rates to BRFSS rates and Google aggregates, suggest a long-term correlation between public engagement and response propensity, these results would not necessarily imply that changes in media coverage or public engagement *cause* changes in interview completion rates. Our second analysis attempts to identify and quantify whether there is evidence of a shorter term relationship between these covariates.

To remove the long-term linear trend in the interview completion rates and examine short-term variations, the weekly interview completion rate data was “de-trended”. This was done by subtracting the linear trend from each weekly interview completion rate. The linear trend was calculated as the equation of the least-squares linear trend line through the data. In other words, we subtracted from each weekly interview completion rate the expected interview completion rate under smooth and uniform change, leaving only fluctuations about this line. This yields residual fluctuations centered at zero, absent a linear trend. This detrending was done at the state level.

A similar transformation was made for the aggregates of Google News and Google Search data. For each week of sample release, the log transformation of the average relative hit volume during the week of sample release and the prior week, was plotted over the duration of the survey. The linear trend in these averaged and transformed results was then subtracted from each data point as described above. Figure 2 shows the residual (i.e., “de-trended) log average Google Search volume and Google News hits corresponding to each week of sample release, once again at the national level.

Figure 2.

As can be readily seen in Figure 2 above, there was a strong temporal correlation between Google Search volume and Google News hit volume over the time period considered. To avoid multicolinearity in our model, we therefore chose just one as a covariate. Because Google Search volume was available at the state level, as was the interview completion rate, we chose Google Search volume as a model covariate.

To help quantify the relationship between Google Search volume and NHFS interview completion rates, and to check whether any national-level relationship holds at the state level, a linear regression model was built with the dependent or response variable being the within-state residual interview completion rate by week of sample release. There were 51 x 34 = 1734 data points in the model (the 50 states plus DC, by 34 weeks of sample release, recalling that the last five weeks of sample release were removed). The predicting covariate was the within-state residual log Google Search volume, averaged over the week of sample release and the prior week. Errors were assumed to be random with a normal distribution and mean zero. The analysis was performed using Statistical Analysis Software (SAS).

Fitting the model at the state level was important. Doing so afforded a larger sample size for analysis, and also controlled for geographical variations in public engagement in flu-related topics. For example, the spike in Google Search volume occurred somewhat earlier in the American southwest than in the remainder of the country.

2.3 Assessment of Potential Bias

Previous study of data from the NHFS has indicated a potential for non-response bias inherent in the survey. Specifically, responders who completed the survey in the first two weeks after sample release reported influenza vaccination at a higher rate than late responders who completed the survey in weeks 3-5 of dialing (Singleton et al, 2010). Assuming that late responders are similar to non-responders, this suggests that there may have been non-response bias in the NHFS if the data collection had been less than five weeks.

As the interview completion rate of the survey changed over time, it is possible that the structure of response changed resulting in temporal shifts in the potential bias of vaccination coverage estimates. To further gauge the potential for bias resulting in the changing structure of response to the NHFS over time, we compared monthly H1N1 and seasonal flu vaccination coverage estimates for adults as of the end of each month to similar end-of-month estimates obtained from BRFSS.

To do so, estimates of end-of-month H1N1 and seasonal influenza vaccination coverage rates were derived by employing the Kaplan-Meier (K-M) estimator (Kaplan and Meier, 1958). Vaccinations reported with unknown month of uptake were included in the analysis with the month of vaccination imputed. The estimates were derived from both NHFS data and BRFSS data collected from adults 18 years or older at the time of interview. We compare the NHFS K-M estimates to the BRFSS K-M estimates by absolute difference in estimates over time and identify whether there is any temporal change.

3. Results

3.1 NHFS Response Rate Comparison

Figure 3 shows the NHFS CASRO response rate by week of sample release, from the 9/27/09 sample release through the 5/16/10 sample release date. The data show a clear spike in the CASRO response rate in late October, followed by a steady decline, with an overall decreasing trend over the entire length of the survey as seen in the linear trend line. Figures 4, 5, and 6 below show that the long-term changes in the CASRO response rate can be explained primarily by changes in the interview completion rate.

Figure 3. Council of American Survey Research Organizations (CASRO) Response Rate by Week of Sample Release, National 2009 H1N1 Flu Survey

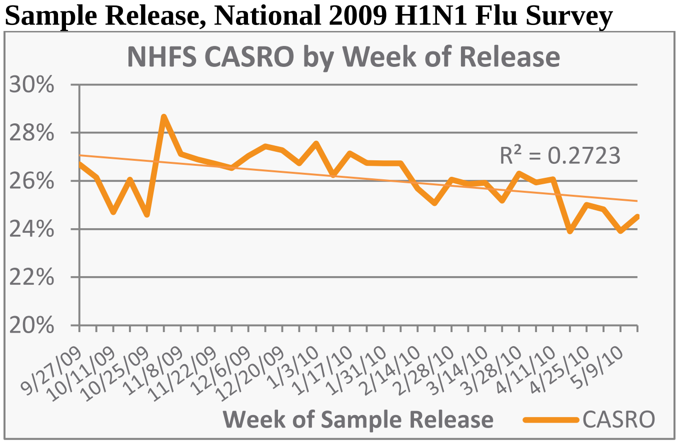


Figure 4. Resolution Rate by Week of Sample Release, National 2009 H1N1 Flu Survey

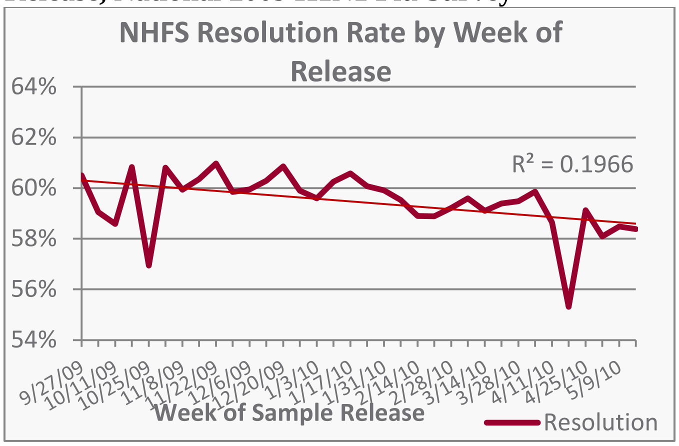


Figure 5. Screener Completion Rate by Week of Sample Release, National 2009 H1N1 Flu Survey

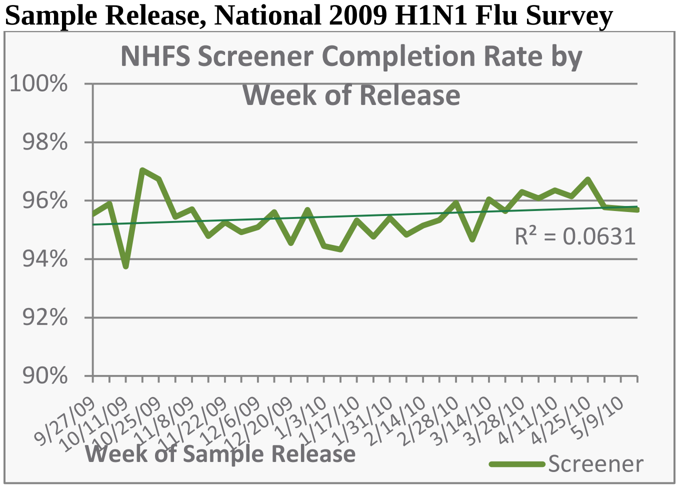


Figure 6. Interview Completion Rate by Week of Sample Release, National 2009 H1N1 Flu Survey

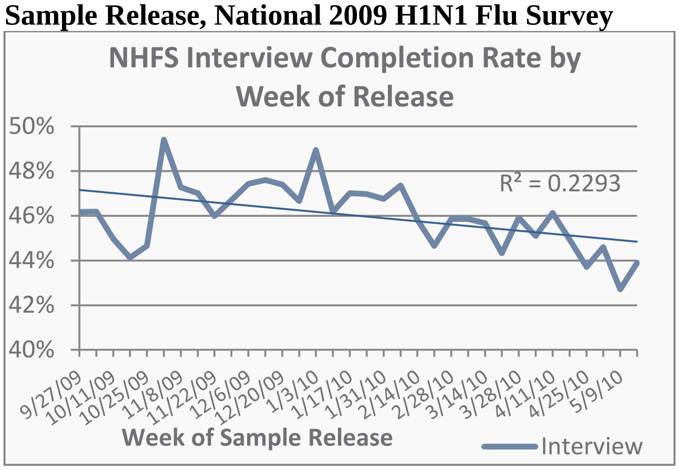


Figure 7 shows the monthly BRFSS landline CASRO response rate from October 2009 through May 2010. The peak in the BRFSS CASRO rate occurred in January, much later than the late October peak observed for NHFS. These changes in the BRFSS CASRO rate were primarily the result of changes in the BRFSS resolution rate (Figure 8), as Figure 9 shows that there was no discernible trend in the cooperation rate over the same time period.

Figure 7. Council of American Survey Research Organizations (CASRO) Response Rate by Month of Sample Release, BRFSS

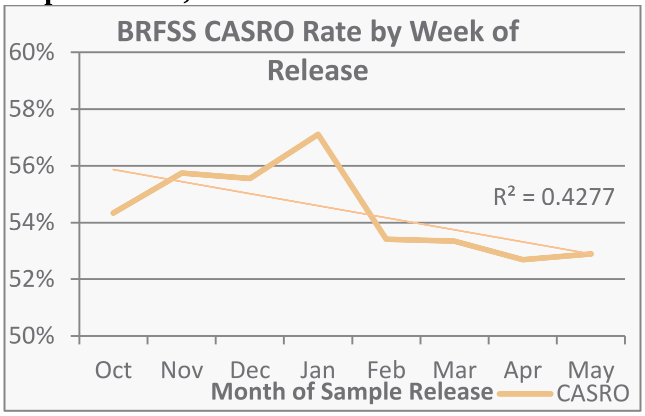


Figure 8. Resolution Rate by Month of Sample Release, BRFSS

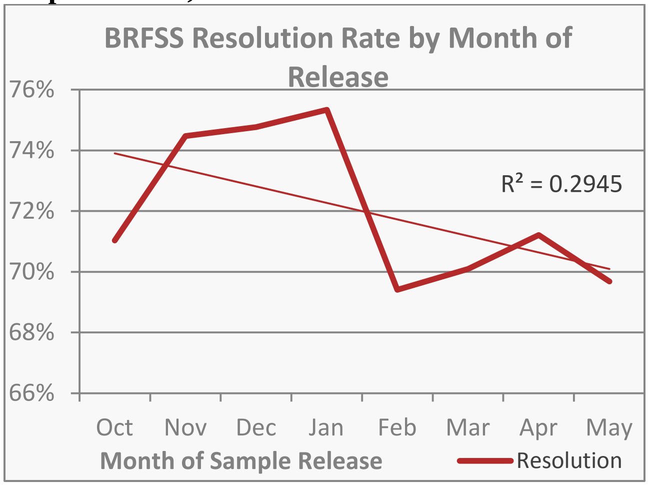
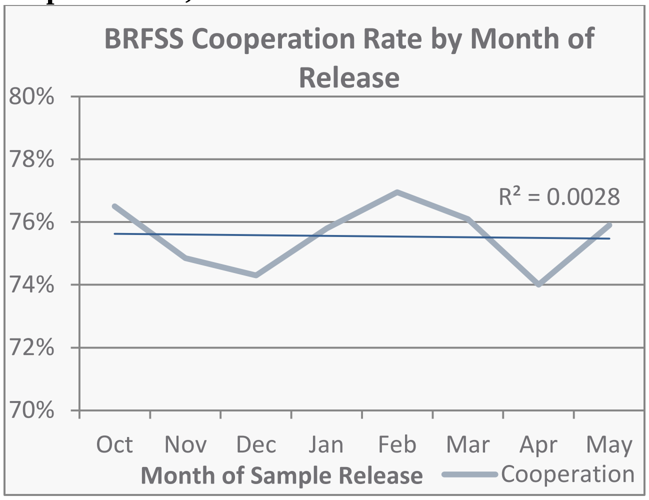


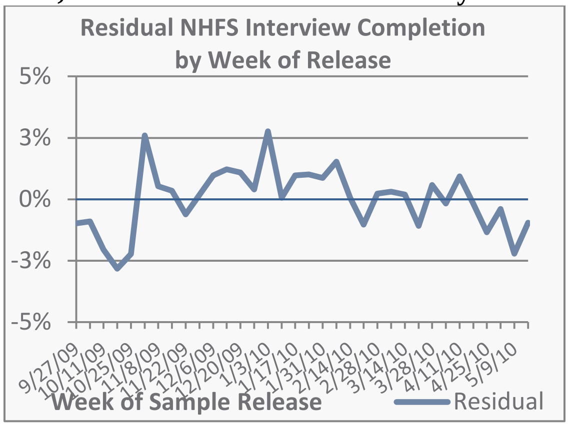
Figure 9. Cooperation Rate by Month of Sample Release, BRFSS



3.2 Modeling the Effects of Public Engagement on the Interview Completion Rate

As seen in Figure 6 above, the interview completion rate showed a clear decreasing trend over the course of the survey. Any analysis of changes in the completion rate would be dominated by this long-term trend. Therefore, as mentioned earlier, the data was "de-trended" by subtracting from each weekly interview completion rate the linear predictor. The residual fluctuations centered at zero, absent a linear trend, are depicted in Figure 10. While the chart shown here depicts the weekly interview completion rates at the national level, this "de-trending" was actually carried out individually for each state.

Figure 10. "De-Trended" Interview Completion Rate, National 2009 H1N1 Flu Survey



The results of our linear regression analysis, shown in Table 1, indicate that, nationally and at the state level, there was a significant association between fluctuations in Google search volume and fluctuations in the interview completion rate. The term 'ResSearch', which is the within-state residual log Google Search volume, averaged over the week of sample release and the prior week, entered the model with a positive and statistically significant parameter estimate. The positive parameter estimate indicates a positive relationship between Google Search volume and the interview completion rate.

Table 1.

Model Selection Order	Covariate	Parameter Estimate	R-Square	P-Value
1	Intercept	-0.00048		
2	ResSearch	0.0254	0.1320	0.0065

Because each observation was at the state by sample-release date level, these results imply that the positive correlation between Google Search volume and interview completion rate hold at the state level. This is true overall, but it is possible that for some individual states the relationship may not hold, as a linear regression model was not fit separately for each state.

3.3 Assessment of Potential Bias

Table 2 presents the NHFS Kaplan-Meier estimates alongside the BRFSS K-M estimates, all of which are censored Kaplan-Meier estimates of vaccination coverage among adults as of the end of each month. Figure 11 is a plot of the difference between the NHFS and BRFSS K-M estimates by month with confidence intervals for the difference.

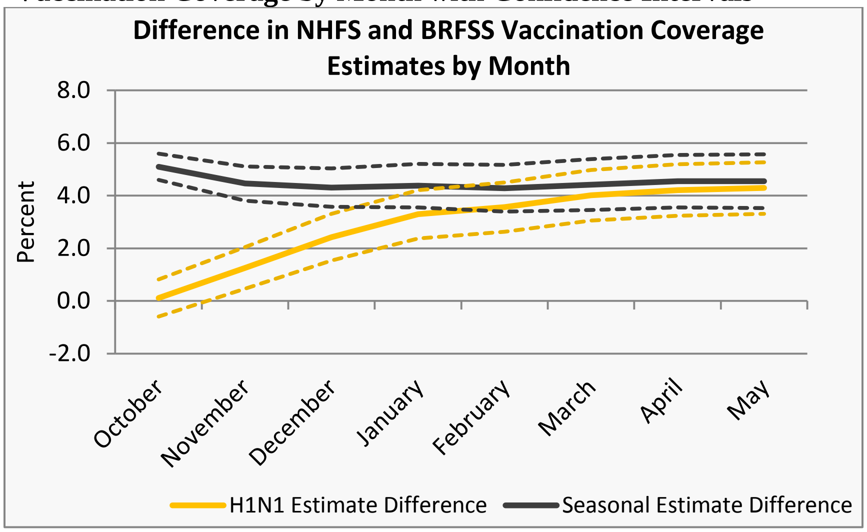
The results show that the NHFS estimate of H1N1 vaccination coverage for October was very close to the corresponding BRFSS estimate. However, over time the difference in estimates steadily increased and by January the NHFS estimate was 21.6 percent, 3.3 percentage points larger than the BRFSS estimate. In May 2010, the difference in estimates was 4.3 percentage points. The plot of the difference in seasonal influenza vaccination coverage estimates shows relatively NHFS larger estimates throughout the survey period.

Table 2. Comparison of NHFS and BRFSS Kaplan-Meier Estimates of Influenza Vaccination Status Among Adults 18+, 2009-2010 Influenza Season[1]

Month	NHFS K-M Estimates		BRFSS K-M Estimates	
	H1N1	Seasonal	H1N1	Seasonal
October 2009	4.6	28.5	4.5	23.4
November	11.7	37.1	10.4	32.6
December	17.0	40.1	14.6	35.8
January 2010	21.6	42.2	18.3	37.8
February	23.9	43.1	20.3	38.8
March	25.2	43.7	21.2	39.3
April	25.7	44.0	21.5	39.5
May	26.0	44.1	21.7	39.6

[1] Estimates presented are cumulative, end-of-month estimates

Figure 11. NHFS Minus BRFSS Estimate of Vaccination Coverage by Month with Confidence Intervals



4. Discussion

These results show a decline in the interview completion rate of the NHFS over the course of the survey, and this corresponds to a similar downward trend in both public engagement in and media coverage of flu-related topics as measured by Google aggregates. Furthermore, no such trend was seen in the BRFSS cooperation rate, which was used as a control for comparison. These results suggest that there is a general, long-term correlation between public engagement in a survey's topic and average propensity to respond as measured by the interview completion rate. Furthermore, the results of our logistic regression analysis of detrended short-term changes show an association between volume of influenza-related searches on the internet and NHFS interview completion rates.

The results of section 3.3 further suggest that the NHFS estimates of 2009 H1N1 vaccination coverage were close to the corresponding BRFSS estimates early in the survey period, but were later moderately larger than the estimates derived from BRFSS data. One possible explanation for this effect is that media coverage and public engagement in flu-related topics were at a peak in late October 2009, increasing the overall likelihood of survey participation even among those less likely to have obtained the 2009 H1N1 vaccine by the time of the interview. Then, as public awareness returned to lower levels in later months, response rates decreased among those who were unlikely to have been vaccinated but remained somewhat elevated among segments of the population that were more likely to have received the 2009 H1N1 vaccine. However, our investigation of the demographic makeup of completed interviews over time yielded no discernible patterns. For example, the percentage of respondents that reported working in a health care setting remained roughly constant over the course of the survey, and the percentage of completed interviews from adults ≥65 years did not change appreciably either. Furthermore, it is not clear why this trend would be visible in the 2009 H1N1 influenza estimates but not in the seasonal influenza estimates, where a similar trend in the differences is not seen.

These analyses have several limitations. Notably, the data were only available at the state level, while there were likely geographical variations in Google activity at the metropolitan area or media market level. Furthermore, public engagement and media coverage of influenza related topics may also have been affected by the availability of the 2009 H1N1 vaccine in each particular area. Ideally, the regression analysis would have been carried out at the media market or metropolitan area level. Also, while there appears to be an association between topic interest and response propensity, topic relevance may be a function of more than simply interest in the survey's topic. For example, a potential respondent that has already been vaccinated for the flu may find the flu-related survey less relevant to him/her, even if the respondent finds the topic interesting. Additional limitations result from BRFSS being only a landline survey while the NHFS was a dual-frame survey, and BRFSS vaccination coverage data was not available for adults in some states. Finally, the regression analysis was conducted with a relative small sample, and the results are from only one study and should be independently confirmed.

Surveys of high-profile topics may be subject to temporal changes in media coverage of and public interest in the topic of interest. These changes may differentially affect response propensity in the population, and this carries important implications for the administration of surveys of timely or important topics. First, the survey design should balance the need for timely estimates with the need to allow an adequate time period for response, as shortened response periods may result in additional nonresponse bias. As mentioned before, the NHFS allowed up to five weeks for respondents to complete the interview, and if this period had been shorter there is evidence that some survey estimates might have been biased. However, further study showed that the apparent bias would not be significant enough to override the advantages of timeliness gained by a shorter follow-up period (Singleton et al., 2011).

The potential for bias resulting from differential response propensity to surveys of high-profile topics, as predicted by Leverage-Saliency Theory, may be mitigated by including the relevant questionnaire items in other established, more general surveys, or by strategic phrasing of the introductory text. Including relevant items in already established surveys may also result in added efficiency and cost-effectiveness, if the existing survey meets the researchers' needs for timeliness and comprehensiveness. In any case, it is recommended that serial, cross-sectional surveys should monitor response rates, sample characteristics, key outcomes, and external measures such as media coverage and public engagement, over the course of survey administration. This is particularly important when conducting pandemic influenza or other surveys during public health or other emergencies.

References

Groves, Robert M., Eleanor Signer, and Amy Corning. 2000. "Leverage-Saliency Theory of Survey Participation: Description and an Illustration.“ *The Public Opinion Quarterly* 64.3:299-308.

Groves, Robert M., Stanley Presser, and Sarah Dipko. 2004. "The Role of Topic Interest in Survey Participation Decisions.“ *The Public Opinion Quarterly* 68.1:2-31.

Kaplan, E. L., and Paul Meier. 1958. “Nonparametric Estimation from Incomplete Observations.” *Journal of the American Statistical Association* 53.282:457-473.

Singleton, James A., Kennon R. Copeland, Nicholas Davis, N. Ganesh, Kirk M. Wolter, and Gary Euler. 2010. “The National 2009 H1N1 Flu Survey: Rapid Data Collection and Early Responder Analysis.” Paper presented at the AAPOR 65th Annual Conference 2010.

Singleton, James A., Nicholas Davis, Kennon Copeland, Tammy Santibanez, N. Ganesh, Carey Drews-Botsch, and Kirk Wolter. 2011. “Design of Health Surveys for Public Health Emergencies: Early Responder Bias in the National 2009 H1N1 Flu Survey (NHFS).” Paper presented at the 10th Conference on Health Survey Research Methods 2011.