Modeling H1N1 Vaccination Rates

N. Ganesh, Kennon R. Copeland, Nicholas D. Davis,
National Opinion Research Center at the University of Chicago

James A. Singleton, Tammy Santibanez,
Centers for Disease Control and Prevention

ABSTRACT:

The National 2009 H1N1 Flu Survey, a RDD landline and cellular telephone survey that operated from October 2009 through June 2010 by NORC for the Centers for Disease Control and Prevention, tracked H1N1 (and seasonal influenza) vaccination coverage nationally on a weekly basis. National-level direct estimates for various socio-demographic groups and geographies were produced weekly based on a rolling sample of respondents. To reduce variability of estimates and trends, we estimated H1N1 vaccination rates using a survival analysis approach. In this paper, we consider parametric and non-parametric models with date of vaccination either interval censored or imputed, and we evaluate our models to select the "best" model that fits the data.

Introduction

In July 2009, the Advisory Committee on Immunization Practices (ACIP) issued recommendations for use of the influenza A (H1N1) 2009 monovalent vaccine (Centers for Disease Control and Prevention, 2009). Recognizing that the vaccine supply would not be ample immediately but would grow over time, ACIP identified: 1) initial target groups, consisting of approximately 159 million persons; and 2) a limited vaccine subset of the target groups, initially estimated at 42 million persons, to receive first priority while the H1N1 vaccine supply was limited (CDC, 2009). ACIP recommended expanding vaccination to the rest of the population as vaccine supplies increased.

To provide both timely estimates of vaccination coverage and reliable estimates of coverage in priority populations (e.g., the initial target groups and the limited vaccine subset) through the 2009–10 influenza season, CDC contracted with NORC at the University of Chicago to design and implement the 2009 National H1N1 Flu Survey (NHFS). The NHFS, conducted beginning the first week of October, 2009 and continuing through the last week of June, 2010, provided weekly and monthly estimates of H1N1 and seasonal flu vaccination coverage for the population along with information concerning knowledge, attitudes, and behaviors related to the H1N1 influenza virus and its prevention. Weekly national estimates were generated within the four to six days following the end of each survey week. Monthly estimates at both national and state levels were generated within the 11-13 days following the end of each survey month.

The weekly and monthly estimates were based upon relatively small samples and therefore subject to large sampling variability, resulting in unsmooth and unstable trends across time. This paper discusses modeling efforts undertaken to provide more stable and smooth vaccination coverage rate estimates.

Summary of the NHFS Sample Design

The NHFS consists of a national random-digit–dialed telephone survey based on a rolling weekly sample of landline and cellular telephones contacted to identify residential households (referred to as the NHFS sample). Within each contacted NHFS sample household, one adult and one child (if present) are randomly selected for interview. Monthly targets for the NHFS sample were established to achieve a total of approximately 4,889 completed adult interviews from landline households and 1,111 completed adult interviews from cellular-phone-only (CPO) or cellular-phone-mainly (CPM) households[1], or approximately 6,000 total adult interviews. The landline NHFS sample is augmented with a sample of children aged <18 years identified during screening for the National Immunization Survey (referred to as the NIS sample).

The target sample size of completed adult interviews within each telephone status group was allocated approximately equally across states. Deviations in state sample sizes were based upon the proportion minority (Hispanic, non-Hispanic Black) population in each state; states with higher minority populations were allocated a slightly larger sample.

Sample for the NHFS was released on a weekly basis, with each released panel remaining active for five weeks. Each sampled telephone number continued to be called across the five weeks until the number was resolved as non-residential, there was a confirmed refusal, or a completed interview was obtained. Completed interviews obtained within a survey week, defined as Sunday through Saturday, were then used in generating the estimates for that survey week. As shown in Figure 1, the estimates for a given survey week were thus based upon completed interviews from five panels. For example, estimates for week ending November 7 were constructed from the interviews completed during that week as a result of the panels (2-6) released in weeks ending October 10 through November 7.

[1] A household is "cell phone only" if there is a cell phone but no landline telephone in the home, and is "cell phone mainly" if there is a landline number available in addition to a cell phone but the respondent reports that it would be unlikely for the landline to be answered if it were to ring.

Figure 1:NHFS Rolling Sample Design

Figure 1

	Week End Date								
Panel	Oct 3	Oct 10	Oct 17	Oct 24	Oct 31	Nov 7	Nov 14	Nov 21	Nov 28
1	X	X	X	X	X				
2		X	X	X	X	X			
3			X	X	X	X	X		
4				X	X	X	X	X	
5					X	X	X	X	X
6						X	X	X	X
7							X	X	X
8								X	X
9									X

Estimates for a survey month were based upon all completed interviews from survey weeks with end dates within the survey month (e.g., November 2009 survey month consisted of survey weeks ending Nov 7, Nov 14, Nov 17, and Nov 24; sample completed Nov 25-30 were part of survey week ending Dec 1 and thus included in December 2009 survey month).

NHFS Data Collection

Between the first week of October, 2009 and the last 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.

Among the 670,841 landline telephones sampled, the percent that were identifiable as residential, non-residential, or non-working numbers, known as the resolution rate, was 77.5%. Among identified residential telephones, the percent completing the screener to determine the presence of an eligible adult was 99.6%, with 43.4% of sample adults in screened and eligible households being classified as completed adult interviews. The product of the resolution rate, the screener completion rate, and the interview completion rate, known as the CASRO response rate, was 34.0% for the landline sample. Of the 226,328 cell phone lines, 53.3% were resolved, 85.7% of personal- use lines completed the screener, and 55.9% of eligible adults completed the survey, leading to a CASRO rate of 25.5%.

A total of 63,659 completed interviews from the combined NHFS samples had been collected as of the last week of May, 2010. The cell telephone sample accounted for 12,662 (19.9%) of the NHFS completes.

Statistical Methods for Obtaining Estimates for the NHFS

Persons with completed interviews from a given survey week (month) were weighted to reflect: 1) sample selection (random selection of telephone numbers, selection of adult/child within household, number of telephone lines on which the sampled person could be reached); 2) combination of panels, NHFS and NIS samples (for children), and landline/cell telephone samples; and 3) ratio adjustment to population controls. Ratio adjustment was carried out through a raking approach using race/ethnicity, age, gender, and region (state for monthly weighting) as the raking dimensions. Estimated H1N1 and seasonal vaccination coverage were derived using the sample weights and data on reported vaccination status.

Estimated H1N1 and seasonal vaccination coverage were derived using the sample weights and data on reported vaccination status. To summarize the statistical methodology by which vaccination coverage rates and their standard errors are estimated from the sample, let Y_{vhij} be an indicator of vaccination status for the v^{th} influenza vaccine , for the j^{th} person in the i^{th} sampled household in the h^{th} stratum of the NHFS sampling design, which is equal to 1 if the person reports a vaccination for the v^{th} influenza vaccine (H1N1, seasonal) and 0 otherwise. Also, let W_{hij} denote the final sampling weight for this sampled person. Letting

$$\hat{Y}_{vh} = \sum_{i,j} W_{hij} Y_{vhij}$$

and

$$\hat{T}_h = \sum_{i,j} W_{hij}$$

Then the national estimator of vaccination coverage rate for vaccine v may be expressed as

$$\hat{\theta}_v = \frac{\sum_h \hat{Y}_{vh}}{\sum_h \hat{T}_h} .$$

Letting

$$Z_{vhij} = \frac{W_{hij}\left(Y_{vhij} - \hat{\theta}_v\right)}{\hat{T}_h},$$

$$Z_{vhi} = \sum_j Z_{vhij} ,$$

and

$$\overline{Z}_{vh} = \frac{\sum_i Z_{vhi}}{m_h},$$

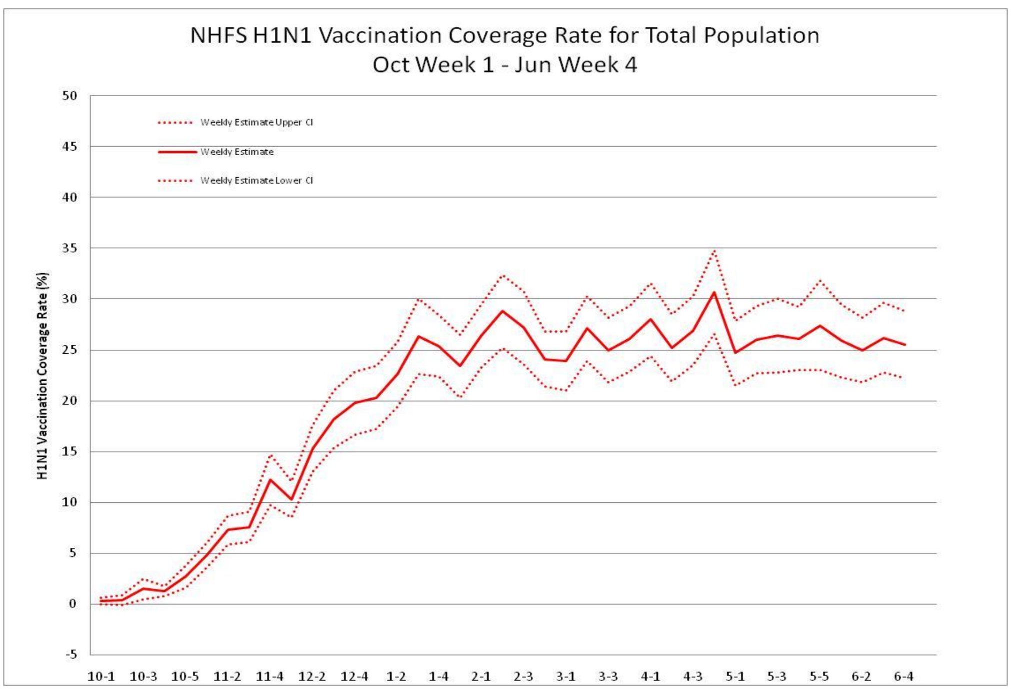
an estimator of the variance of the estimated vaccination coverage rate can be expressed as

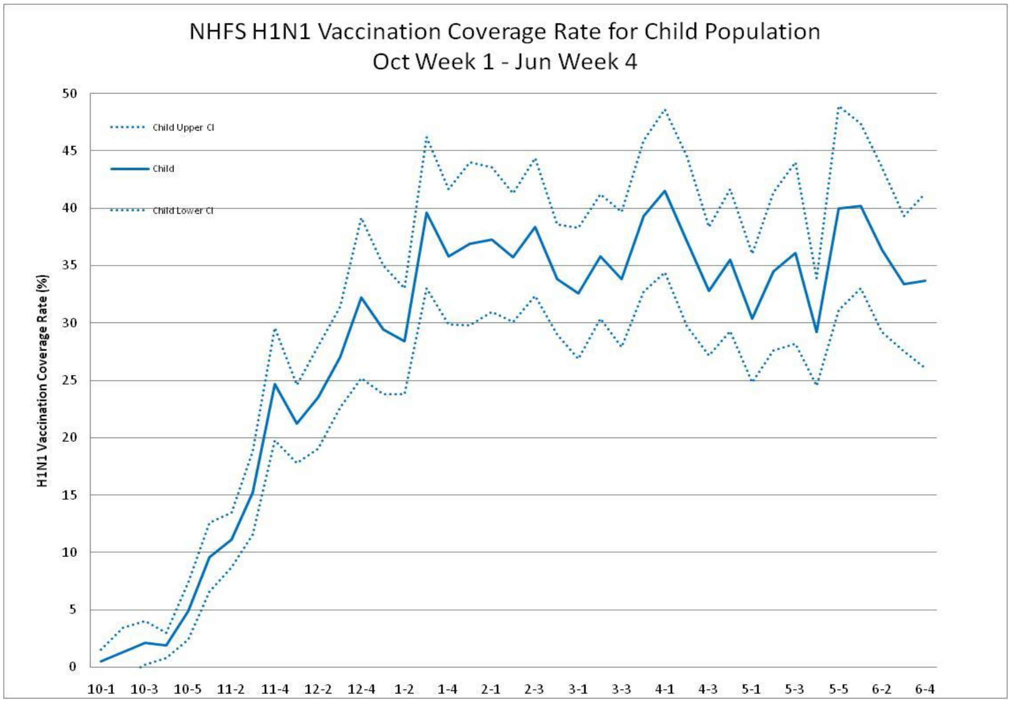
$$v(\hat{\theta}_v) = \sum_h \frac{m_h}{m_h - 1}\left(Z_{vhi} - \overline{Z}_{vh}\right)^2.$$

In these equations, m_h denotes the number of sampled households containing persons with completed interviews in the h^{th} stratum. Note that these formulae extend to estimates for any subgroup, s, of the population. A similar approach is taken for derivation of vaccination coverage estimates for a survey month.

Weekly H1N1 vaccination coverage estimates and associated 95% CIs for the total population and for children between the ages of 6 months and 17 years for October Week 1 through June Week 4 are shown in Figures 2 and 3. As can be seen, the vaccination coverage estimate trend is not monotonically increasing, which must be the case for the underlying population vaccination coverage, due to independent samples from week-to-week and sampling variability. In addition, the trend is very unstable.

Figures 2 and 3





Smoothed Estimates of Vaccination Coverage

In an attempt to provide more stable estimates of the vaccination coverage rates, both for individual time periods and across time, two approaches were undertaken to combine data across week. The challenge in combining data across weeks is that the reference period varies, as respondents were asked their vaccination status as of the date of the interview. However, respondents reporting having received a vaccination were also asked month of vaccination, which is used in the two approaches.

Composite Estimation

The first approach entailed generation of composite estimates (Schaible, 1978, Wolter, 1979) of monthly vaccination coverage for each month prior to the current survey week using data from the current survey week along with all other survey weeks following the end of each individual month, and incorporating those monthly coverage estimates with weekly estimates of the interim vaccination coverage to a specific survey week to generate what was referred to as an “enhanced” weekly vaccination coverage estimate.

Extending the notation provided for the official weekly estimates, let $\hat{\theta}_{vmw}$ be the estimated vaccination coverage for a prior month, m, based upon sample data from the w^{th} survey week, $\hat{\theta}_{vw}$ be the estimated vaccination coverage for the current month based upon the sample data from the w^{th} survey week, and n_w be the number of completed interviews from the w^{th} survey week. Then the composite vaccination coverage estimate for the m^{th} month can be expressed as

$$\hat{\theta}_{vm} = \sum_{w \in W(m)} \frac{n_w}{\sum_{w' \in W(m)} n_{w'}} \hat{\theta}_{vmw},$$

where $W(m)$ represents the set of survey weeks following the m^{th} month, and the estimated variance of the composite estimate for m^{th} month can be expressed as

$$v(\hat{\theta}_{vm}) = \sum_{w \in W(m)} \left(\frac{n_w}{\sum_{w' \in W(m)} n_{w'}} \right)^2 v(\hat{\theta}_{vmw}).$$

The "enhanced" vaccination coverage estimate for the w^{th} survey week can then be expressed as

$$\hat{\theta}_{vw} = \sum_{m \in M(w)} \hat{\theta}_{vm} + \hat{\theta}_{vw},$$

where $M(w)$ represents the set of months prior to the w^{th} survey week, and the estimated variance of the vaccination coverage can be expressed as (under the assumption of zero correlation between monthly estimates)

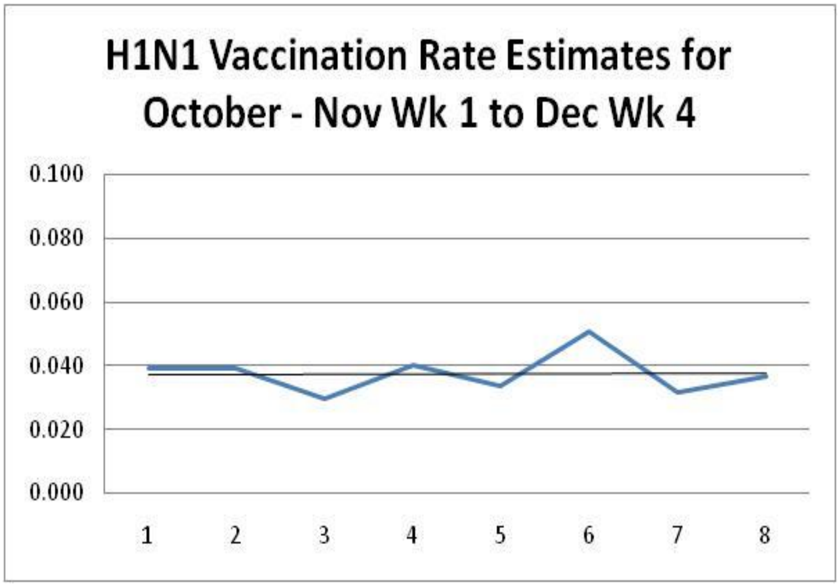
$$v(\hat{\theta}_{vw}) = \sum_{m \in M(w)} v(\hat{\theta}_{vm}) + v(\hat{\theta}_{vw}).$$

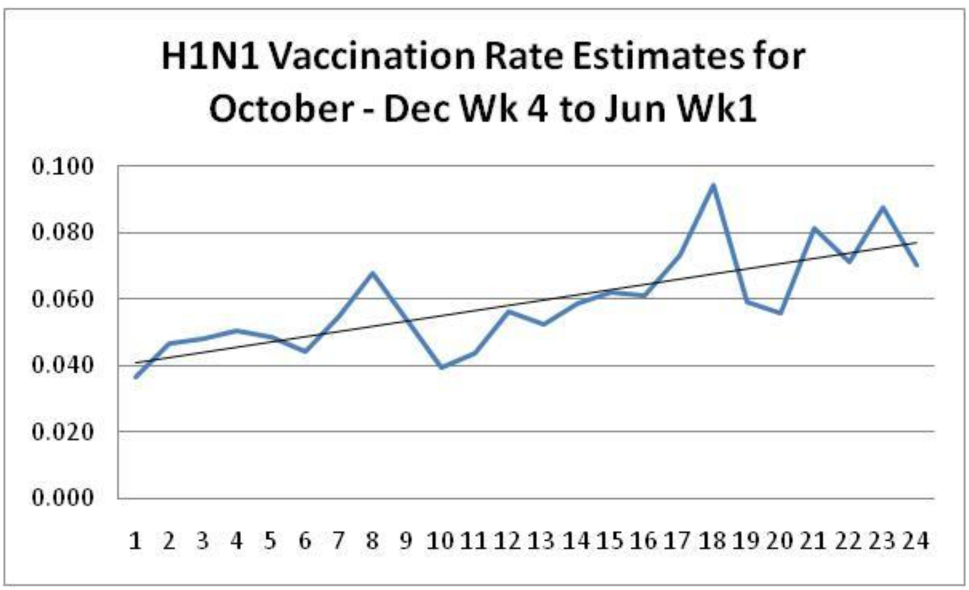
Potential Recall Error

Reviewing the values, $\hat{\theta}_{vmw}$, across weeks, it was observed that the values increased across time for October, and decreased for November, December, and January. Based on this finding, it is theorized that respondents were experiencing difficulty recalling month of vaccination the longer from the vaccination date the interview occurred. As the interview asks "Since September, have you received an H1N1 vaccination?" and then follows with "In which month did your receive your vaccination?", if the respondent has difficulty recalling the month but knew it was in the past, they may have used the initial question as an anchor from which to answer the month question.

Figures 4 and 5 show the estimated vaccination coverage rate for October for Nov week 1 to Dec week 4, and for Dec week 4 to Jun week 4, respectively. As can be seen, the estimates were roughly constant through Dec week 4, and have been generally increasing since Dec week 4.

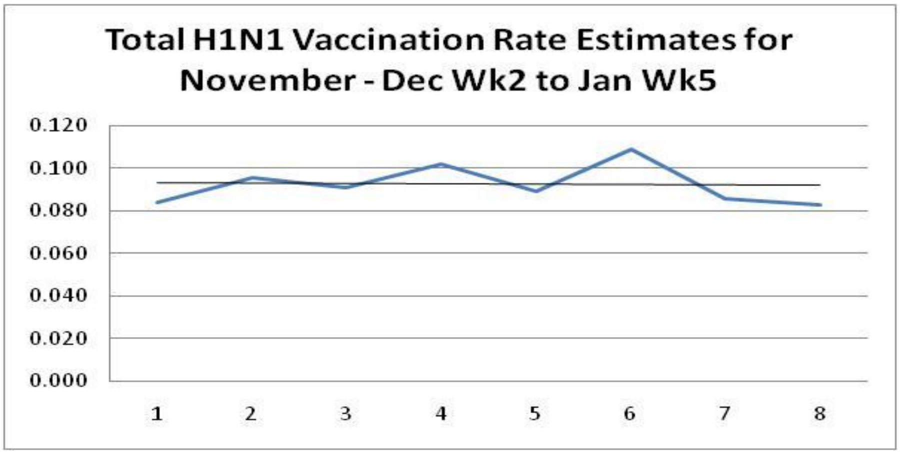
Figures 4 and 5

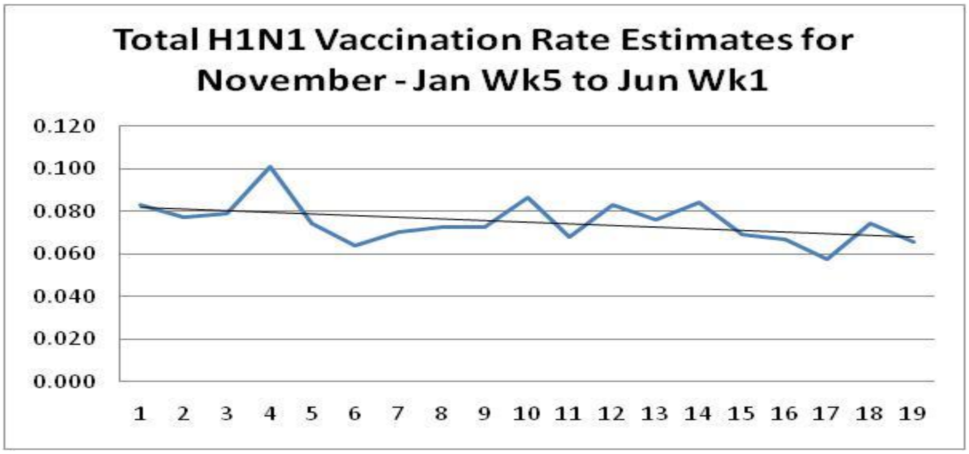




Figures 6 and 7 show the estimated vaccination coverage rate for November for Dec week 2 to Jan week 5, and for Jan week 5 to Jun week 4, respectively. As can be seen, the estimates were roughly constant through Jan week 5, and have been generally decreasing since Jan week 5. Similar results were observed for December and January.

Figures 6 and 7



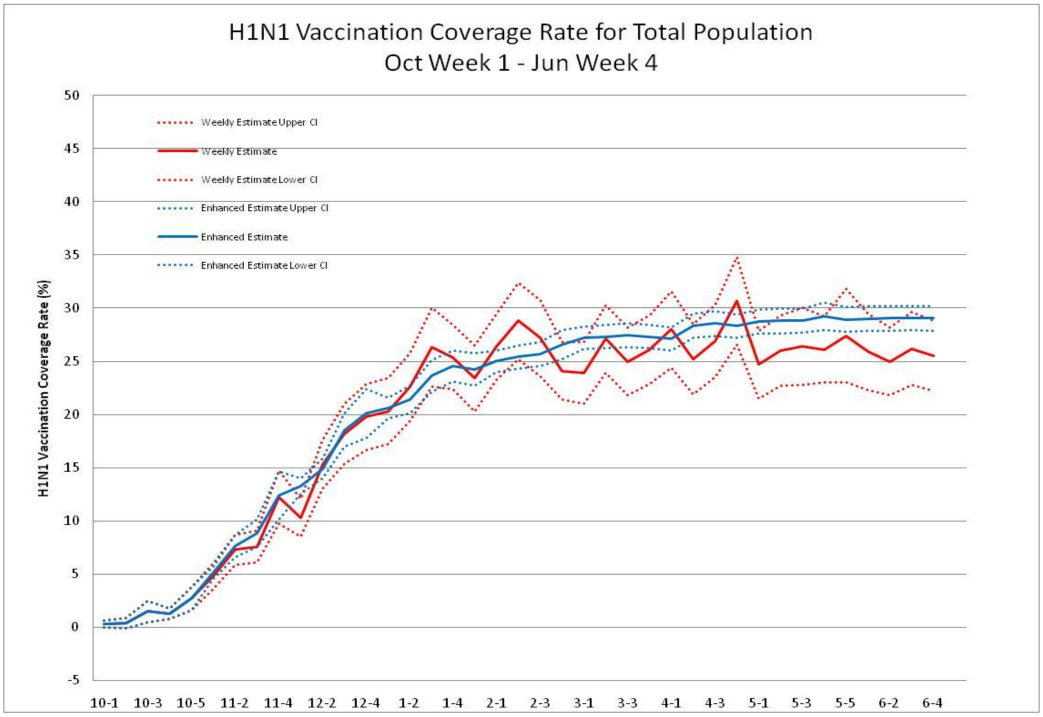


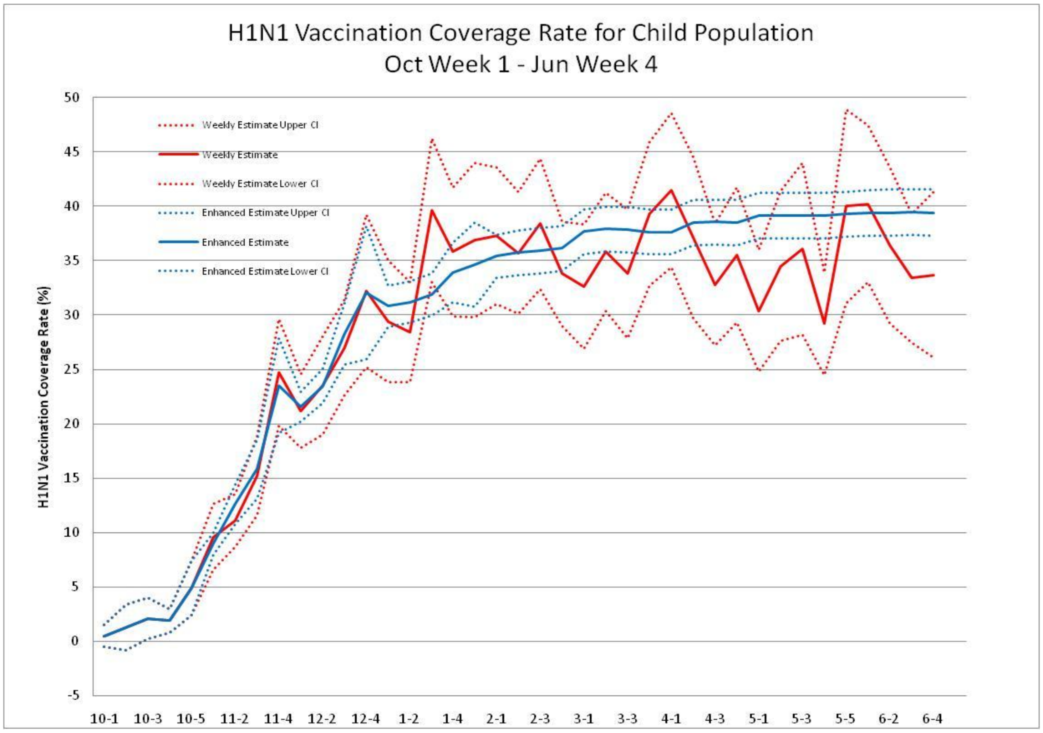
Based on the findings relative to potential recall error, the enhanced estimation was modified to utilize data from only those survey weeks determined to have provided "accurate" recall of the vaccination rate for the months of October through January.

Enhanced Weekly Estimates

Enhanced weekly H1N1 vaccination coverage estimates and associated 95% CIs for the total and child populations for October Week 1 through June Week 4 are shown in Figures 8 and 9, respectively, along with the official weekly estimates. As can be seen, the enhanced vaccination coverage estimate trend is smoother and the individual estimates have narrower CIs than the official estimates. However, the estimates are not constrained to be non-decreasing.

Figures 8 and 9





Survival Estimation

A second approach to generating monthly estimates of vaccination coverage entailed application of a weighted Kaplan-Meier survival approach (Klein and Moeschberger, 1997). For the Kaplan-Meier approach, the "censor" and "time-to-event" variables were defined as follows:

1. If a person was vaccinated in a month prior to the month of interview, then the "censor" variable was set to "not censored" and the "time-to-event" variable was set to the month of vaccination.
2. Instead, if a person was vaccinated in the same month as the month of interview, then the "censor" variable was set to "censored" and the "time-to-event" variable was set to the month preceding the month of interview.
3. If a person was unvaccinated as of the day of the interview, then the "censor" variable was set to "censored" and the "time-to-event" variable was set to the month preceding the month of interview.

For persons that were unvaccinated as of the day of the interview, their "time-to-event" variable was set to the month preceding the month of interview as it is possible for these cases to be vaccinated after the day of the interview but prior to the end of the month of interview. To be consistent with this definition, persons that were vaccinated on the month of interview were also defined to be unvaccinated (i.e., "censored") as of the month preceding the month of interview. This monthly survival analysis approach was used by the CDC to estimate influenza vaccination coverage by state and selected population groups (Centers for Disease Control and Prevention, 2010a; Centers for Disease Control and Prevention, 2010b).

In order to obtain estimates for the most recent interview month (i.e., month of June), for cases that were interviewed in the month of June, the censor and time variables were defined as follows:

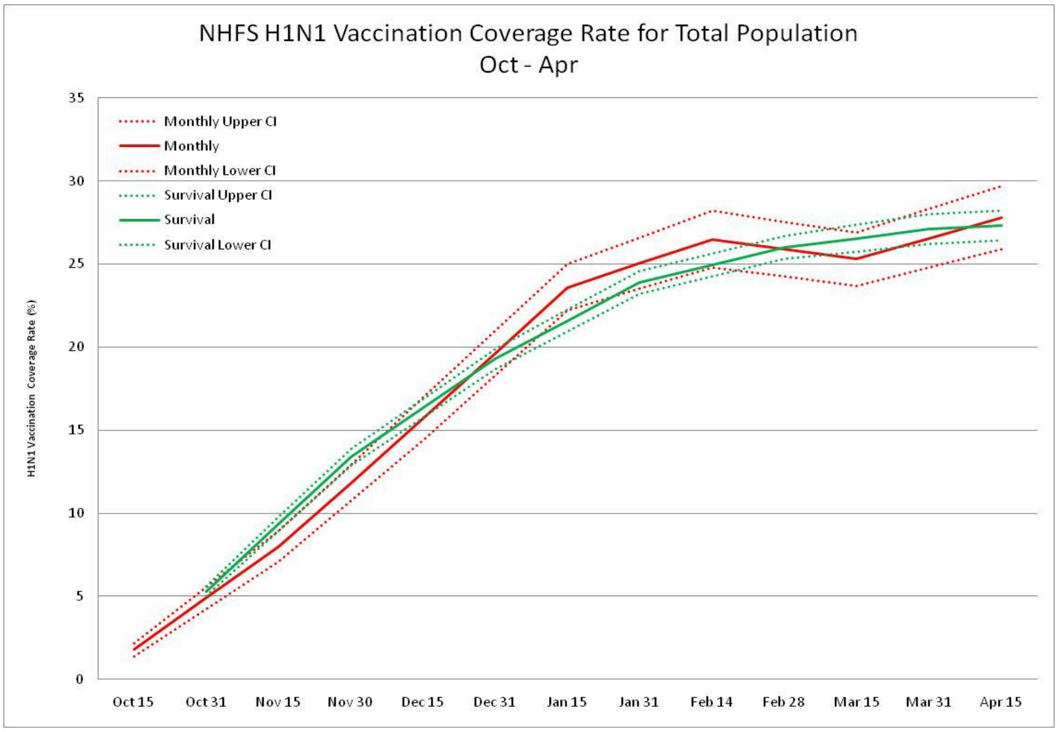
1. If a person was vaccinated in the month of June, then the "censor" variable was set to "not censored" and the "time-to-event" variable was set to month of June.
2. If a person was unvaccinated as of the day of the interview, then the "censor" variable was set to "censored" and the "time-to-event" variable was set to the month of June.

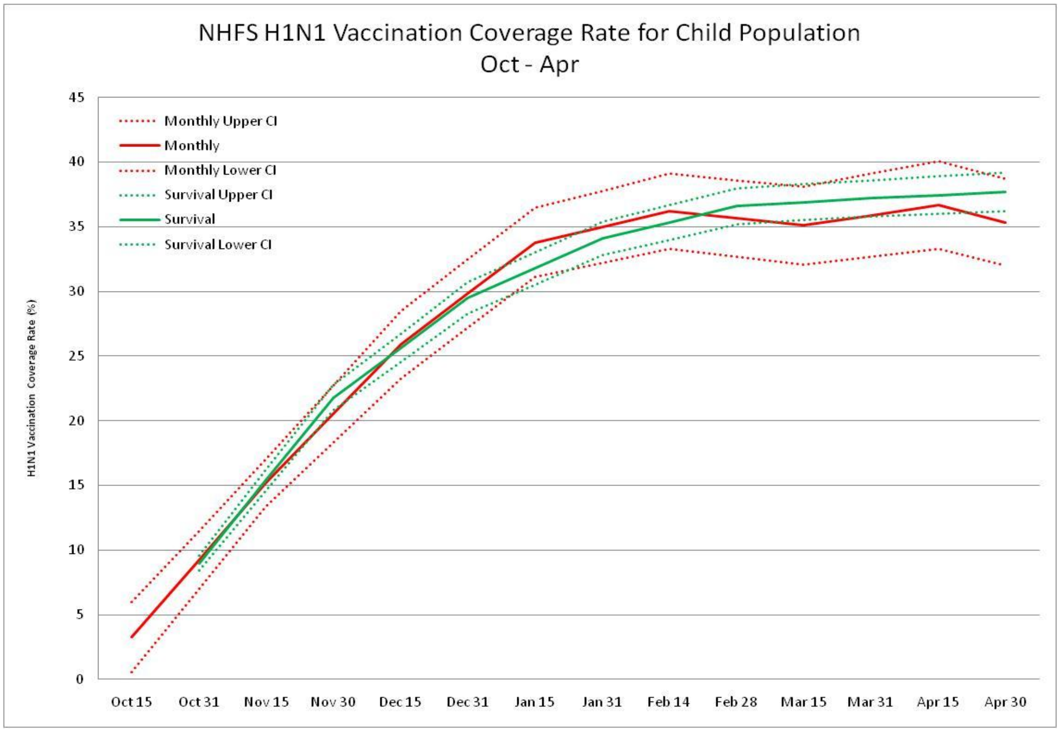
This approach underestimates cumulative vaccination coverage as of the end of the most recent interview month, because it assumes all persons unvaccinated as of the date of their interview remained unvaccinated by the end of the month.

As mentioned previously, survey weights were taken into account when using the Kaplan-Meier approach in SUDAAN. In particular, the monthly weights from each survey month were appropriately normalized using the number of interviewed adults/children in each state and month. These normalized weights, and the previously defined "censor" and "time-to-event" variables were used in SUDAAN to produce the monthly estimates of vaccination coverage.

KM monthly H1N1 vaccination coverage estimates and associated 95% CIs for the total and child populations for October through April are shown in Figures 10 and 11, respectively, along with the official monthly estimates. As can be seen, the KM monthly vaccination coverage estimates are greater than the official monthly vaccination coverage estimates, as is expected given the official monthly estimates represent the vaccination coverage as of approximately the mid-point of the survey month whereas the KM and enhanced monthly vaccination coverage estimates represent the vaccination coverage as of the last day of the calendar month. It can also be seen that the KM monthly estimates yield smaller CI widths than does the official monthly estimates.

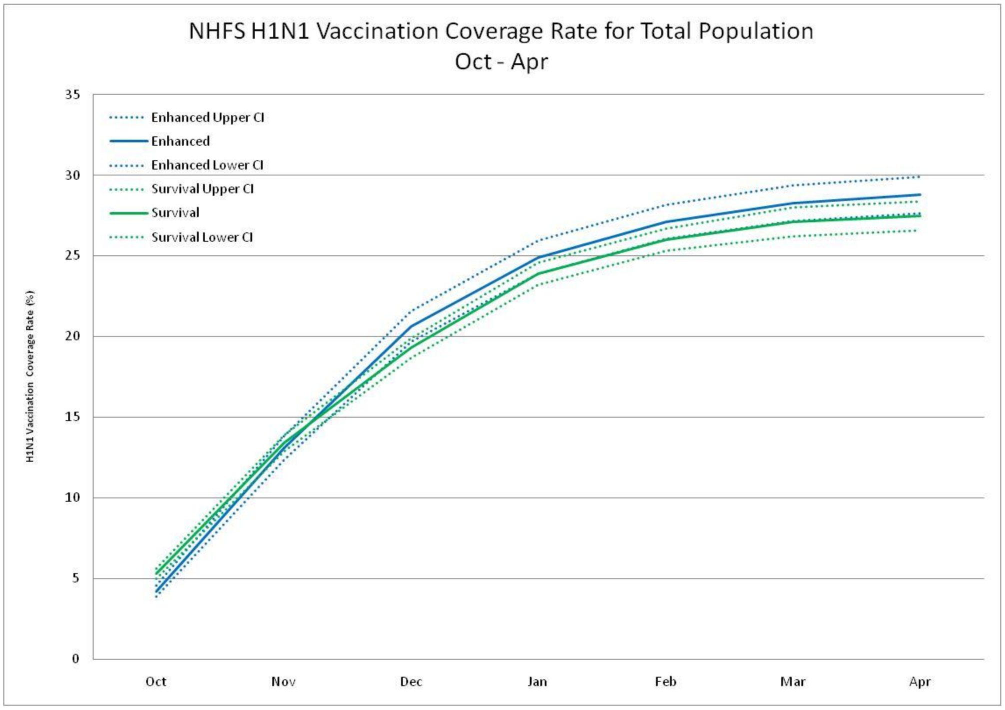
Figures 10 and 11

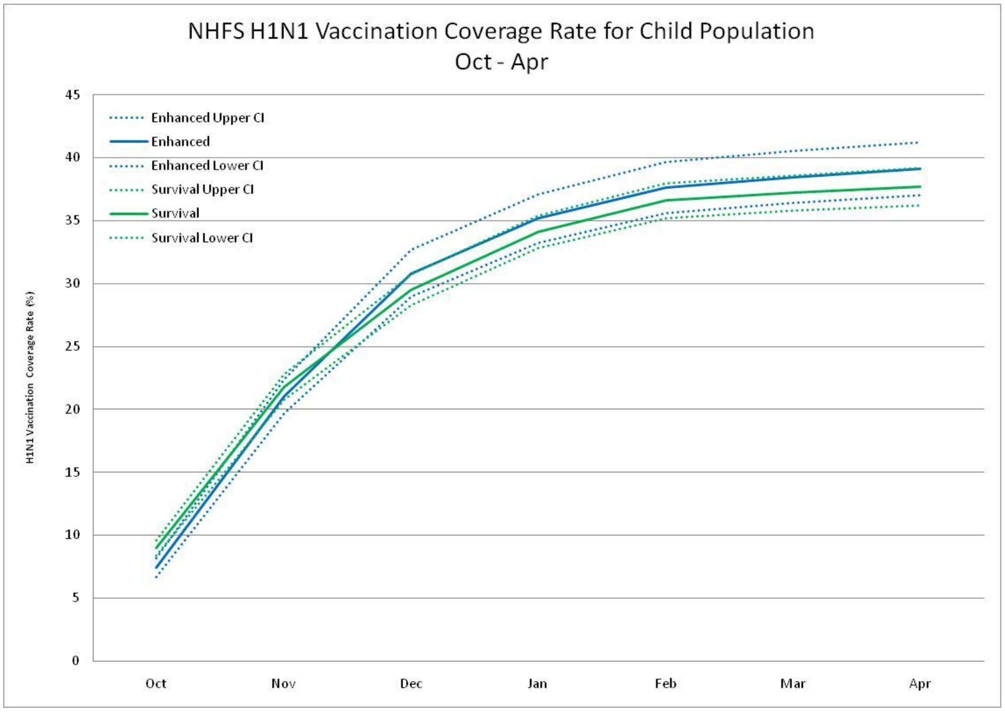




KM monthly H1N1 vaccination coverage estimates and associated 95% CIs for the total and child populations for October through April are shown in Figures 12 and 13, respectively, along with the enhanced monthly estimates. As can be seen, the enhanced monthly vaccination coverage estimates are greater than the KM monthly vaccination coverage estimates, as is expected given the enhanced monthly estimates account for the observed recall error in reporting month of vaccination. It can also be seen that the enhanced and KM monthly estimates yield similar CI widths.

Figures 12 and 13





Discussion

Survival estimates provide the opportunity to reduce variability and stabilize trends for survey estimates over time. In addition, a given generation of survival estimates yield non-decreasing estimates for a given trend, which meets with users' expectations[2]. Monthly survival estimates for vaccination rates are easily derived using standard statistical software packages, such as SUDAAN. Survival estimates are not, however, appropriate for dynamic cohorts of the population, such as women pregnant during the vaccination period.

To increase the usability of the data, finer granularity is needed, such as weekly or daily survival estimates. Such granularity is possible with survival approaches, assuming survey data on week/date of vaccination are available or can be reliably modeled.

As to application for monitoring of vaccination rates, the composite estimates appear preferable to the weekly estimates based solely on completed interviews during the survey week. For monthly analysis, survival estimates appear preferable.

Further work is needed on the application of survival methods in deriving weekly and daily estimates, with particular emphasis on the problem of accounting for recall error in the survival estimates.

References

Centers for Disease Control and Prevention. "Use of influenza A (H1N1) 2009 monovalent vaccine: recommendations of the Advisory Committee on Immunization Practices (ACIP)," 2009. MMWR 2009;58(No. RR-10).

Centers for Disease Control and Prevention (2010a). "Interim results: state-specific influenza A (H1N1) 2009 monovalent vaccination coverage – United States, October 2009-January 2010." MMWR 2010;59:363-368.

Centers for Disease Control and Prevention (2010b). "Interim results: state-specific seasonal influenza vaccination coverage – United States, August 2009-January 2010." MMWR 2010;59:477-484.

Klein, J.P. and Moeschberger, M.L. (1997), Survival Analysis: Techniques for Censored and Truncated Data, Springer: New York.

Schaible, W.L. (1978) "Choosing Weights for Composite Estimators for Small Area Statistics." Proceedings of the Section on Survey Research Methods, American Statistical Association, 741-746.

Wolter, K.M. (1979). "Composite Estimation in Finite Populations," *Journal of the American Statistical Association,* 4:604-613.

[2] It should be noted that survival estimates are only non-decreasing within a given estimation run. As new survival estimates are created across time, the user sees and compares estimates from different estimation runs, and the revealed series is not necessarily non-decreasing.