Monitoring County-Level Vaccination Coverage During the 2004–2005 Influenza Season

Haomiao Jia, PhD, Michael Link, PhD, James Holt, PhD, Ali H. Mokdad, PhD, Lei Li, PhD, Paul S. Levy, ScD

Background: During the 2004–2005 influenza season, the United States faced a sudden shortage of influenza vaccine. In response, the Centers for Disease Control and Prevention (CDC) and the Advisory Committee on Immunization Practices (ACIP) recommended prioritizing vaccination for persons aged 65 and older and others at high risk. To monitor subsequent vaccination coverage, several questions about influenza vaccination were added to the ongoing Behavioral Risk Factor Surveillance System (BRFSS). This study provided real-time county-level estimates of influenza vaccination coverage from the BRFSS each month from October 2004 through January 2005.

Method: The methods used a variation of small area estimation procedures suitable for situations in which most small areas have few or no survey respondents, and rapid assessment is essential. Both model-based methods and nonparametric spatial-smoothing methods were used in a three-step procedure.

Results: The highest vaccination rates during the 2004–2005 influenza season were seen in the upper Midwest and the Southeast. Areas with the lowest vaccination rates were the intermountain West, southern California, portions of Washington and Oregon, and various areas across the Eastern United States, often coinciding with urban areas. Intrastate variations were especially pronounced in the Eastern United States, particularly in Georgia, Florida, Tennessee, Kentucky, North Carolina, Virginia, and New York. These states all had areas with low immunization rates as well as areas with high rates.

Conclusions: The results showed that vaccination coverage varied significantly across states and substate regions. Our findings show that this methodology can provide estimates with reasonable reliability for planning during public health emergencies.

Introduction

Growing concern about a future influenza pandemic, underscored by the U.S. government’s multibillion dollar funding proposal for pandemic planning, highlights the need for valid and reliable local-level health data to inform program planning and evaluation, resource allocation, and real-time action steps. Influenza vaccination statistics are usually drawn from national surveys, vaccination programs, or Medicare beneficiary data. These sources provide updated data on influenza vaccination for the United States as a whole and for certain subpopulations. However, none of these sources can provide timely vaccination statistics for most U.S. counties. Moreover, few if any of these sources are designed to facilitate the rapid estimation of coverage that is essential during an unfolding public health event. Small-area estimation (SAE) techniques, when used in conjunction with the appropriate health surveillance data, can, however, be powerful tools both for addressing an unfolding public health event and for planning future public health action.

A case in point is the sudden and unexpected shortage in the availability of influenza vaccine during the 2004–2005 influenza season in the United States. In response, the Advisory Committee on Immunization Practices (ACIP) and Centers for Disease Control and Prevention (CDC) recommended vaccination only for those who were at high risk of influenza-related complications and for certain categories of high-risk individuals, and asked healthier adults to defer or forgo vaccination. Because of the high vulnerability of elderly adults and others in the priority groups to complications of influenza, it was important to establish a rapid-response surveillance system whereby this vacci-
nation strategy could be monitored as the influenza season progressed and modified if necessary.7–9

Delivery of health services such as immunization is often a highly local phenomenon, and it was considered important to obtain estimates of vaccination coverage for areas as small as counties and to provide maps showing temporal–spatial patterns of vaccination rates.10 By examining counties or clusters of counties with low estimated vaccination rates, the officials managing vaccine redistribution could identify areas with low coverage rates and then reallocate available vaccine or other resources to these areas. Here we illustrate how SAE techniques were applied to an ongoing health surveillance system, the Behavioral Risk Factor Surveillance System (BRFSS), to provide as close to real-time local-level estimates as possible during the 2004–2005 influenza season. This effort has application for future rapid response monitoring, such as during a pandemic, as well as applicability for using past data to help identify areas of need for future planning.

Materials and Methods

Data sources

The BRFSS is a large, ongoing telephone interview survey conducted by state health departments with assistance from the CDC.11 The survey uses random-digit-dialing methods to generate probability-based samples for the 50 states and District of Columbia. More than 300,000 adult interviews are completed annually. The BRFSS data contain weighting factors to adjust for nonresponse and post-stratification, as well as unequal sampling probabilities.12,13

From November 1, 2004 to February 28, 2005, a new set of influenza vaccination–related questions was added to the BRFSS to determine whether respondents had been vaccinated for influenza recently and to identify those in ACIP-defined priority groups (text of these questions is provided in Appendix A, found online at www.ajpm-online.net). Priority groups of adults included the following: all individuals aged ≥65 years; individuals reporting that they had asthma, other lung problems, heart problems, diabetes, kidney problems, and weakened immune system; healthcare workers with direct patient contact; and women who were pregnant at the time of the interview.8

To facilitate rapid processing and reporting of these data, states submitted their completed interviews to the CDC on a weekly rather than monthly basis, which is normally the case. Automated processing and weighting systems provided analysts with clean data sets within 3 business days, facilitating rapid dissemination of results.

From November 2004 through February 2005, a total of 102,890 adults were interviewed. For all adults, the average number of respondents per county was 32. The sample sizes for counties within states were approximately proportional to the county populations; therefore, a large number of small counties had few or no respondents and only 15% of counties had ≥50 respondents. For adults aged ≥65 years and the priority subgroups, the average number of observations available per county per month was much smaller, at <10.

For the SAE analysis, cumulative vaccination coverage was estimated at the end of each month for the period between October 2004 and the end of each month through January 2005. Because interviews were spread out over the entire month each month, the data collected during a given month was used to report on vaccination coverage for the previous month(s) rather than the month in which the interview was conducted. For example, seasonal coverage through October 31, 2005 was initially estimated using data collected in November 2005, and then updated each month by adding data collected from subsequent months. The final (season-end) estimate for October was then based on four months of data (November through February). In contrast, seasonal coverage through January 31, 2005 was estimated using data collected in February 2005. Estimates were provided for adults as a whole, for all adults in vaccine-priority groups, and for adults aged ≥65 years.

In addition to the BRFSS data, county-specific ecologic data were collected from various sources and merged with the BRFSS survey data by matching the county of the survey respondents’ residence. These county-level variables were predictors in the SAE models to link areas that shared similar characteristics. Examples of county-specific data elements included unemployment rate and influenza-related mortality rate. Ecologic data were derived from the census,14 vital statistics data,15 and the Area Resource File.16

Estimation procedure

Although the BRFSS has a very large, geographically dispersed sample and provides estimates for selected counties and metropolitan areas, it was not designed to produce county-level estimates in all areas. As a result, a significant percentage of counties lack sufficient data to provide accurate design-based estimates of vaccination rates.4,6,11 Alternative methods that “borrow” information from related counties by either adding data from similar areas or using explicit area-linking modes can improve the validity of local estimates for those small counties.4,6,17 Such procedures, or SAE techniques, have been used to provide reliable estimates from the BRFSS data for counties with small or nonexistent samples.3,6,17 Using a variation of these techniques, this study produced cumulative monthly estimates of vaccination coverage for all adults and for adults in ACIP-defined priority groups for all 3117 U.S. counties for the months of October 2004 through January 2005.

To facilitate this, both model-based SAE methods and nonparametric spatial-smoothing methods were used in a three-step procedure:

Step 1. A generalized linear mixed model (GLMM) was used to estimate county vaccination rates.5,18 The probability that a person in the county received a vaccination during each month was a function of both individual-level variables (age, race/ethnicity, and gender) and county-level ecologic variables. The selection of individual- and county-level covariates was similar to the process reported earlier.6 County-specific variables thought to be related to vaccination were selected first, and then variables with a statistically significant relationship with vaccination were identified through regression. Because some of the county-specific ecologic variables were moderately to highly correlated, a factor analysis process was applied to identify four factors
from the variables: (1) socioeconomic characteristics (unemployment rate, poverty rate, per capita income, median household income, and proportion of adults with high school education); (2) mortality (age-adjusted mortality by gender and race/ethnicity); (3) demographic characteristics (proportion living in rural areas and proportion of adults who are married); and (4) healthcare resources and utilization (ratios of hospitals, hospital beds, and hospital admissions to population). Two influenza-related variables were also added to the model: the influenza- and pneumonia-related death rates and the percentage of adults who are healthcare workers. The regression analysis was conducted separately for the nine census divisions: New England, Mid-Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, and Pacific.

Step 2. Cumulative vaccination rates from cross-sectional data were calculated based on the estimated percentages of respondents who were vaccinated during the month and the estimated vaccination rates in the previous month. Therefore, estimated cumulative vaccination rates were nondecreasing over the time.

Step 3. Moving weighted averages of neighboring counties were applied to smooth the estimates obtained from the previous step with use of the head-banging method.5,19

Further details of these three steps are described in Appendix B online at www.ajpm-online.net.

Evaluation of estimates

In the absence of a gold standard for county-level estimates, this study examined the consistency of estimates by using county-level estimates as “building blocks” to construct an aggregated state average that was compared to the state composite estimate.5,17,20 As the size of the area for aggregation increases, the sample size becomes larger and the aggregated state average should approach the state composite estimate.5,17,20 As the size of the area for aggregation increases, the sample size becomes larger and the aggregated state average should approach the state composite estimate, which more accurately reflects the true state rate.20 To ensure maximum sample sizes, the estimates for October 2004 were used for evaluation. Data from Alaska and Washington DC were excluded from this analysis because each of these areas was treated as a single county in the estimation.

Results

For the United States as a whole, the estimated cumulative percentage of individuals vaccinated by the end of each month of the 2004–2005 influenza season is shown in Figure 1 for all adults, for those aged ≥65 years, and for all high-priority adult groups. At the end of September 2004, 1.2% of all adults had received influenza vaccine. Vaccination coverage increased as the season progressed, and by the end of January 2005, the overall coverage rate was 24.4%. More importantly, estimated coverage increased among adults aged ≥65 years to 68.0% in January. Among all people considered high priority for receiving the vaccine, estimated vaccination coverage increased to 41.4% during the same timeframe.

Figure 1. Cumulative influenza vaccination coverage among U.S. adults by month during the 2004–2005 influenza season.

County-Level Coverage

The geographic pattern of influenza vaccination coverage rates was examined by mapping vaccination coverage at the end of each month for the three groups (Figure 2 and Figures A1 and A2 in the online appendix at www.ajpm-online.net). As shown on the maps, this study found wide geographic variation in influenza coverage. For all adults by the end of January 2005, estimated coverage rates ranged from 12.6% to 62.3%. Among all adults, the standard deviation of the distribution of rates was 6.0% and the rate in 90% of counties (from the 5th to the 95th percentile) fell between 18% and 35%. For adults from priority groups, the coverage rate in 90% of counties fell between 23% and 44%; and for adults aged ≥65 years, the rate in 90% of counties fell between 48% and 64%.

By examining county-level maps, we observed that the areas with the highest vaccination rates for all people during the 2004–2005 influenza season were the upper Midwest and the Southeast. Areas with the lowest vaccination rates were the intermountain West, southern California, portions of Washington and Oregon, and various areas across the Eastern United States, often coinciding with urban areas. Other areas with above-average coverage rates included the central Plains and Great Lakes states, the Rocky Mountains, and broadly scattered areas in the Eastern United States. Intrastate variations were especially pronounced in the Eastern United States, particularly in Georgia, Florida, Tennessee, Kentucky, North Carolina, Virginia, and New York. These states all had areas with low immunization rates as well as areas with very high rates. This variability within states highlights the usefulness of conducting surveillance at the substate level. Areas with the greatest ratios of immunization improvement (defined as the ratio of the January 2005 immunization rate to the October 2004 immunization rate) were portions of Mississippi, Georgia, Florida, and Alabama. Other areas with high levels of improvement in immunization coverage for the influenza season included New England, the Mid-Atlantic, and the southern Plains states.
Figure 2. Influenza vaccination rates by counties for all adults during the 2004–2005 influenza seasons.
errors was differences were

When using such an approach, interpreta-

don reasonable reliability as judged by the validation assess-

2005 influenza season. Although the estimates had

vaccination rates for all U.S. counties during the 2004–

health surveillance data provided estimates of influenza

Discussion and Conclusions

The BRFSS, which is designed to provide reliable

health data for each state, is an excellent source for

substate data because it is more likely than most other

nationwide surveys to include samples from each of the

3117 counties.5,6,13 Applying SAE techniques to these

health data for each state, is an excellent source for

SAE model in many applications.6,17,18 This is because

one of the major aims was to make the estimation

procedure computationally simple, easy to maintain,

flexible in the inclusion of individual-level and ecologic

covariates, and less restrictive to sample size require-

ments. Some computationally intensive SAE methods

such as empirical Bayes and hierarchical Bayes,17,21–23

sophisticated model inference techniques,18,24 and

models with spatial depending structures25 require

large mean sample sizes for calculation. Previous stud-

ies have shown that all of these SAE methods require at

least several hundred observations per area to converge

quickly and to show meaningful improvement over

simpler models.18,22,23 In addition, spatial data-

smoothing methods require relatively stable local data

to start with and are efficient only when large numbers

of observations were available in the majority of small

areas.5,6

Because models of monthly vaccination rates should

incorporate possible correlations between different

months, a model with autoregressive random effects was

appropriate here.17,20 However, besides the larger sample

size requirements, inference and prediction for these

time-series models and spatial-survival models are diffi-

cult and require special software and programming.25,26

Spatial data are typically smoothed to better reveal

spatial trends. Although a complicated pattern of spa-

tial variation can be modeled with SAE techniques, this

study used a nonparametric “head-banging” procedure

for spatial data smoothing.19 This procedure is effective

in removing random noise and it is easy to imple-

ment.5,6 The smoothed values were set within 25% from

their regression estimates to preserve the credible

spikes and edges in spatial data. This smoothing “bor-

rowes” information only from bordering counties and is

thus truly local in character. Estimates from smoothed

models have several other advantages. First, because of

small mean sample sizes, model convergence could be

a problem. Regression estimates for counties with

strong individual effects could be heavily biased, result-

ing in more-scattered patterns. Spatial smoothing can

remove additional background noise and further stabi-

Consistency of Estimates

Table 1 presents some discrepancy statistics between

aggregated state average and state composite estimates.

In general, the aggregated average of SAEs matched

with state composite estimates perfectly. The Pearson

correlation coefficients were >0.90, mean absolute
differences were <1.3%, and the root of mean square
errors was <2.0% for all three groups.

Discussion and Conclusions

The BRFSS, which is designed to provide reliable

health data for each state, is an excellent source for

substate data because it is more likely than most other

nationwide surveys to include samples from each of the

3117 counties.5,6,13 Applying SAE techniques to these

health data for each state, is an excellent source for

SAE model in many applications.6,17,18 This is because

one of the major aims was to make the estimation

procedure computationally simple, easy to maintain,

flexible in the inclusion of individual-level and ecologic

covariates, and less restrictive to sample size require-

ments. Some computationally intensive SAE methods

such as empirical Bayes and hierarchical Bayes,17,21–23

sophisticated model inference techniques,18,24 and

models with spatial depending structures25 require

large mean sample sizes for calculation. Previous stud-

ies have shown that all of these SAE methods require at

least several hundred observations per area to converge

quickly and to show meaningful improvement over

simpler models.18,22,23 In addition, spatial data-

smoothing methods require relatively stable local data

to start with and are efficient only when large numbers

of observations were available in the majority of small

areas.5,6

Because models of monthly vaccination rates should

incorporate possible correlations between different

months, a model with autoregressive random effects was

appropriate here.17,20 However, besides the larger sample

size requirements, inference and prediction for these

time-series models and spatial-survival models are diffi-

cult and require special software and programming.25,26

Spatial data are typically smoothed to better reveal

spatial trends. Although a complicated pattern of spa-

tial variation can be modeled with SAE techniques, this

study used a nonparametric “head-banging” procedure

for spatial data smoothing.19 This procedure is effective

in removing random noise and it is easy to imple-

ment.5,6 The smoothed values were set within 25% from

their regression estimates to preserve the credible

spikes and edges in spatial data. This smoothing “bor-

rowes” information only from bordering counties and is

thus truly local in character. Estimates from smoothed

models have several other advantages. First, because of

small mean sample sizes, model convergence could be

a problem. Regression estimates for counties with

strong individual effects could be heavily biased, result-

ing in more-scattered patterns. Spatial smoothing can

remove additional background noise and further stabi-

Table 1. Discrepancy statistics between aggregated state average from county estimates and state composite estimates

<table>
<thead>
<tr>
<th>Discrepancy statistics</th>
<th>All adults</th>
<th>Aged ≥65</th>
<th>Priority group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficient</td>
<td>0.954</td>
<td>0.965</td>
<td>0.904</td>
</tr>
<tr>
<td>Spearman correlation coefficient</td>
<td>0.931</td>
<td>0.948</td>
<td>0.896</td>
</tr>
<tr>
<td>Mean absolute difference</td>
<td>0.382%</td>
<td>1.279%</td>
<td>1.163%</td>
</tr>
<tr>
<td>$1/49\sum_{i=1}^{49}</td>
<td>r_i - \bar{p}</td>
<td>$</td>
<td></td>
</tr>
<tr>
<td>Mean relative absolute difference</td>
<td>4.822%</td>
<td>5.272%</td>
<td>8.547%</td>
</tr>
<tr>
<td>$1/49\sum_{i=1}^{49}</td>
<td>r_i - \bar{p}</td>
<td>/\bar{r}_i \times 100%$</td>
<td></td>
</tr>
<tr>
<td>Root of mean square error</td>
<td>0.633%</td>
<td>1.760%</td>
<td>1.943%</td>
</tr>
</tbody>
</table>

$\bar{p}_i$, aggregated average of county-level estimates for the state; $r_i$, composite estimates for the state.
lize the estimates. Second, the hierarchical dependent structure among counties was not modeled in the GLMM used here. Spatial correlation of data must be adjusted because neighboring counties are likely to have similar rates. Spatial data smoothing can incorporate correlations of neighboring counties by calculating moving averages of neighboring counties without using autoregressive techniques. Third, because this study did not estimate individual values of the random-effect term, estimates for a small number of counties could be poor and result in outliers.\textsuperscript{27} Smoothing can normalize extreme values for those counties with much higher or lower rates than their neighboring counties; however, it may also, at times, have the negative effect of smoothing real differences between neighboring counties. The results also showed that the estimates had sufficient sensitivity to show substate variations. The mean county-level estimates for all adults and for two population subgroups were close to the national estimates. The ranges of the values of the county-level estimates were slightly wider than those of the state-level composite estimates. The standard deviations of the estimates were larger among counties with 0 to 50 respondents.

A limitation of the regression-estimation approach is that it requires area-level data, which might not be available for all counties. The accuracy of regression estimates may also be biased in some counties by poor prediction. Further, the regression model does not take into account the spatial hierarchical structure of the data. This study included a spatial data-smoothing procedure to minimize regression-estimation bias and to adjust for the correlations among neighboring counties. However, the efficiency of this effort requires further evaluation.

Development of SAEs can be critical during an emerging public health event, especially one distributed over a wide geographic area such as in the case of a pandemic influenza outbreak. The need for such data to monitor vaccine coverage based on prioritization recommendations was highlighted during the 2004–2005 influenza season. The technique can also be applied to data collected during noncrisis periods to provide information for future planning. For instance, data collected during a nonshortage year could be used to guide planning for vaccine distribution in a subsequent year, highlighting areas of particular need within a state. When combined with mapping techniques and temporal–spatial analysis of local data, these techniques can be powerful tools for the management and response to emerging as well as to ongoing public health situations.

No financial conflict of interest was reported by the authors of this paper.

References

Percent Vaccinated

- **< 21.5%** (over -2 standard deviations)
- **21.6 - 33.8%** (-1 to -2 standard deviations)
- **33.9 - 46.1%** (-1 standard deviation to mean)
- **46.2 - 58.4%** (mean to +1 standard deviation)
- **58.5 - 70.7%** (+1 to +2 standard deviations)
- **70.8% and greater** (over +2 standard deviations)

Source: Behavioral Risk Factor Surveillance System

Figure A1. Influenza vaccination rates by counties for population aged 65 years during the 2004–2005 influenza season.
Figure A2. Influenza vaccination rates by counties for adults in priority groups during 2004–2005 influenza season.
Monitoring County-Level Vaccination Coverage During the 2004–2005 Influenza Season

Haomiao Jia, PhD, Michael Link, PhD, James Holt, PhD, Ali H. Mokdad, PhD, Lei Li, PhD, Paul S. Levy, ScD

Appendix A: New Questions Added to BRFSS to Monitor Influenza Vaccination Coverage, November 1, 2004–February 2005

Q1. During the past 12 months, have you had a flu shot?
   Read if necessary: We want to know if you had a flu shot injected in your arm.
   1 Yes
   2 No
   7 Don’t know/Not sure
   9 Refused

Q2. During the past 12 months, have you had a flu vaccine that was sprayed in your nose?
   1 Yes – go to Q4
   2 No – If Q1 is “Yes” go to Q4, otherwise go to Q6
   7 Don’t know/Not sure No – If Q1 is “Yes” go to Q4; if Q1 is “No” go to Q6, otherwise go to Q7
   9 Refused

Q3. During what month and year did you receive your most recent flu vaccination?
   If “Yes” to both Q1 and Q2, also say: “The most recent flu vaccination may have been either the flu shot or the flu spray.”
   __/________ Month / Year
   77/7777 Don’t know/Not Sure (Probe: “Was it before or after September 2004?” Code approximate month and year)
   99/9999 Refused

If Q3 is DK or RF go to Q6

Q4. Where did you go to get your most recent [FILL: flu shot/vaccine that was sprayed in your nose]? CATI fill in appropriate response from Q1 and Q2.
   Read only if necessary:
   01 A doctor’s office or health maintenance organization (HMO)
   02 A health department
   03 Another type of clinic or health center
      [Example: a community health center]
   04 A senior, recreation, or community center
   05 A store [Examples: supermarket, drug store]
   06 A hospital [Example: in-patient]
   07 An emergency room
   08 Workplace
   or
   09 Some other kind of place
   77 Don’t know/Not sure (Probe: How would you describe the place where you went to get your most recent flu vaccine?)
   99 Refused

If Q5 is before 9/2004 go to Q5, if Q5 is DK or RF, go to Q5, otherwise go to Q6

Q5. What is the MAIN reason you have NOT received a flu vaccination for this current flu season? [Interviewer note: The current flu season = Sept. ’04 – Mar. ’05] Do not read answer choices below. Select category that best matches response.
   01 Need: Do not need it
   02 Need: Doctor did not recommend it
   03 Need: Did not know that I should be vaccinated
   04 Need: Flu is not that serious
   05 Need: Had the flu already this flu season
   06 Concern about vaccine: side effects/can cause flu
   07 Concern about vaccine: does not work
   08 Access: Plan to get vaccinated later this flu season
   09 Access: Flu vaccination costs too much
   10 Access: Inconvenient to get vaccinated
   11 Vaccine shortage: saving vaccine for people who need it more
   12 Vaccine shortage: tried to find vaccine, but could not get it
   13 Vaccine shortage: not eligible to receive vaccine
   14 Some other reason
   77 Don’t know/Not sure (Probe: “What was the main reason?”)
   99 Refused

Q6. If Q3 is 09/2003 through 03/2004 go to Q7, otherwise ask Q6
   Did you get a flu vaccination during the ‘last flu season’ in other words during the months of September 2003 through March 2004?
   1 Yes
   2 No
   7 Don’t know/Not sure (Do not probe)
   9 Refused

Q7. Have you ever had a pneumonia shot? This shot is usually given only once or twice in a person’s lifetime and is different from the flu shot. It is also called the pneumococcal vaccine.
   1 Yes
   2 No
   7 Don’t know/Not sure (Do not probe)
   9 Refused

Q8. Has a doctor, nurse, or other health professional ever said that you have any of the following health problems? Read each problem listed below:
   Asthma
   Lung problems, other than asthma
   Heart problems
   Diabetes
   Kidney problems
   Weakened immune system caused by a chronic illness, such as cancer or HIV/AIDS, or medicines, such as steroids
   -or-
   Sickle cell anemia or other anemia
   1 Yes – Go to Q9
   2 No – Go to Q10
   7 Don’t know/Not sure (Probe by repeating question) – Go to Q10
   9 Refused – Go to Q10

Q9. Do you still have (this/any of these) problem(s)?
   1 Yes
   2 No
   7 Don’t know/Not sure (Do not probe)
   9 Refused

Q10. Do you currently work in a health care facility, such as a medical clinic, hospital, or nursing home? If necessary say: This includes part-time and volunteer work.
   1 Yes – Go to Q11
   2 No – Go to Next Section
Appendix B: Details of the Estimation Procedure

This study used the following three-step procedure employing both multilevel regression and spatial smoothing. Because data were collected for four months and each month furnished an estimate of the proportion vaccinated for previous months but not for the data collection month or following months, the number of eligible respondents decreased with each month.

Step 1: Multiple logistic regression estimation

This study used a random-effects logistic regression model to estimate for each of the 3,117 U.S. counties the probability that a person in the county received an influenza immunization during each month as a function of personal level variables and ecological variables specific to the county.

Considering the notation $y$ for the dichotomous measure of presence or absence of influenza vaccination, suppose that $y_{ij}$ is a Bernoulli variable with probability $p_{ij}$ for observations in the demographic category $j$ and county $i$ for any given month. Considering the cross-sectional BRFSS data, we used a multilevel logistic regression model with a county-level random effect $\mu_i$: \[
\logit(p_{ij}) = x_{ij}' \beta + \mu_i \tag{1}
\]
where $x_{ij} = (x_{ij1}, \ldots, x_{ijk})$ was a vector of $k$ personal level (i.e., age, race, and sex) and county level covariates for each respondent and $\beta = (\beta_1, \ldots, \beta_k)$ were corresponding fixed effects. The random effect $\mu_i$ was assumed to be independent and normally distributed with a mean of zero and a variance equal to $\sigma^2_{\mu}$. Given that $n_{ij}$, the population in demographic category $j$ and county $i$, and the total population $n_i = \sum_{j} n_{ij}$ in county $i$ were known from the census, we obtained the county-level rate for:

$$
\hat{p}_i = \sum_j (n_{ij} / n_i) \hat{p}_j
$$

Parameters for the GLMM were estimated by approximation in terms of Taylor series also known as the linearization method.\textsuperscript{1,2} To simplify analysis, individual values of random effects were not used when calculating $\hat{p}_i$. However, we included the random-effect term to obtain accurate estimations for fixed effects and variances.\textsuperscript{3,4}

The process of estimation for respondents aged 65 and older was similar to the process of estimation for all adults. We first calculated the total population aged 65 years and older using $n_{ij65+} = \sum_{\text{age} \geq 65} n_{ij}$ for county $i$ and then obtained the monthly county-level rate as $\hat{p}_i = \sum_{\text{age} \geq 65}(n_{ij} / n_{ij65+}) \hat{p}_j$. To estimate the rate for adults from the priority group, we examined the data in two separate groups: age 18-64 years and age 65 and older. Let $\hat{p}_{i1}$ and $\hat{p}_{i2}$ be the estimated rates for age 18-64 years from the priority group and age 65 and older, respectively, for county $i$. Because $a_r$, the proportion of adults aged 18-64 years, is known, $\hat{p}_r$, the estimated rate for the priority group, is as

$$
\hat{p}_i = a_r \hat{p}_r + (1 - a_r) \hat{p}_{i2}.
$$

Step 2: Estimation of cumulative rates

Because the objective was to estimate the vaccination rates at the end of each month or the cumulative rates, we calculated cumulative rates from cross-sectional data based on the estimated percentages of respondents who received vaccine during the month and the estimated vaccination rates in the previous month.

Let $t$ be the indicator of the month ($t = 1, 2, \cdots$ for October, November, etc.). The rate by the end of month $t$ was given by

$$
\hat{p}_d = 1 - (1 - \hat{p}_{i,t-1})(1 - \hat{p}_r) = \hat{p}_{i,t-1} + (1 - \hat{p}_{i,t-1})\hat{p}_r, \tag{2}
$$

where $\hat{p}_{i,t-1}$ is the vaccination rate by the end of month $t-1$ and $\hat{p}_r$ is the probability of receiving flu vaccine during month $t$. This step was designed to ensure that estimated vaccination rates obtained from regression were non-decreasing over the time.

Step 3: Spatial smoothing (head banging)

To further improve our estimation, we proposed a moving weighted average of neighboring counties to smooth estimates obtained from the previous step. We used the head-banging method described by Mungiole et al.\textsuperscript{5} as follows. Let $\hat{p}_i$ be the estimated prevalence from previous steps for county $i$. Each county had up to 27 neighboring counties. The median estimated rate was calculated for neighboring counties ($m_i$). These neighboring counties were then grouped according to whether their estimated rates fell above or below $m_i$. The following two quantities were defined as:

- high screen for county $i = \text{weighted median prevalence of neighboring counties} \geq m_i$
- low screen for county $i = \text{weighted median prevalence of neighboring counties} < m_i$

Weights were based on the county population. If $\hat{p}_i$ was between the high and low screen, its value was not changed. If $\hat{p}_i$ was greater (or less) than its high screen, then its value was changed to the high (or low) screen. The smoothing method was applied to all counties, and then the above process was repeated 10 times to generate a “moving” average.\textsuperscript{5,6}

Because estimates for some small suburban counties neighboring one or several large metropolitan counties would be changed to the values of those metropolitan counties, we decided that a change to high screen or low screen should represent no more than a 25% difference from the original values.

References: