Estimating County-Level Vaccination Coverage Using Small Area Estimation with the National Immunization Survey-Child

Zachary H. Seeskin[1], Nadarajasundaram Ganesh[1], Poulami Maitra[1], Peter Herman[1], Kirk M. Wolter[1], Kennon R. Copeland[1], Ned English[1], Michael P. Chen[2], James A. Singleton[2], Tammy A. Santibanez[2], David Yankey[2], Laurie D. Elam-Evans[2], Natalie Sterrett[2], Chalanda S. Smith[2], Kevin Gipson[2], Seth Meador[2]
[1]NORC at the University of Chicago
[2]National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention

Abstract[1]
The National Immunization Survey-Child (NIS-Child) is a random digit dialing survey of parents and guardians of children age 19 to 35 months in the United States. The NIS-Child produces annual childhood vaccination coverage estimates at the national and state levels, as well as for select local areas and territories. We describe small area estimation methods using NIS-Child data to generate county-level vaccination coverage rates. Estimates for children by age two years are derived for children born 2007 through 2011 and 2012 through 2016 using 2008-2018 NIS-Child data, combining cohorts to increase sample size. The models use county-level predictors from the Area Health Resource File, Census Planning Database, natality birth records, and other sources. We describe our approach applying cross-sectional Lindley and Smith area-level models (also known as Fay-Herriot models), as well as our methods for selecting county-level predictors of vaccination coverage and limitations associated with these methods. County-level estimates are generated using the James-Stein approach, an empirical best linear unbiased prediction method. Further, we discuss an interactive mapping tool showing how the county-level vaccination coverage estimates vary across counties and how county-level coverage may be associated with county-level characteristics.

Key Words: Random digit dialing; Fay-Herriot model; James-Stein estimation; Data visualization

1. Introduction

The National Immunization Survey Child (NIS-Child) is a random digit dialing (RDD) survey conducted by NORC at the University of Chicago for the National Center for Immunization and Respiratory Diseases (NCIRD) at the Centers for Disease Control and Prevention (CDC) to estimate annual vaccination coverage rates for children age 19 to 35 months. Estimates are produced for the nation, each state, and select local areas and territories.

---

[1] The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Because local health authorities need data at the county-level for planning programs, interventions, and policy changes, we apply small area estimation methods with the NIS-Child to produce estimates of vaccination coverage rates by age two years for 3,137 counties in the United States. Estimates are produced separately for children born 2007 to 2011 and for children born 2012 to 2016, using 2008 to 2018 NIS-Child data. These small area estimates provide additional data beyond the NIS-Child estimates produced for larger geographic areas. The small area estimates can inform evaluations of the extent to which children in different counties are at risk for acquiring vaccine-preventable diseases and guide efforts to improve vaccination coverage.

Our methodology estimates vaccination coverage rates using a linear mixed model, specifically a modified version of the Lindley and Smith model or Fay-Herriot model (Lindley and Smith 1972, Fay and Herriot 1979). Direct survey estimates of vaccination coverage rates are produced using survey estimation based on NIS-Child data from the county and then applying Kaplan-Meier analysis to estimate vaccination coverage rates by age 24 months. In addition, a model-based estimate is obtained by using a linear model relating the NIS-Child direct survey estimates of vaccination coverage for large counties (counties with data for at least 35 children) and county groupings to county-level explanatory data sources listed in Section 3. The direct survey estimate and the model-based prediction for a given county are combined in a weighted average, where the weights given to these two components are proportional to their estimated precision.

This article describes our methodology and the county-level vaccination coverage estimates produced. Section 2 provides background on the NIS-Child and small area estimation, while Section 3 provides a detailed description of the methodology. Section 4 presents an analysis evaluating the estimates. Section 5 describes a mapping tool we developed to explore the estimates across geographic areas and examine the relationship between the estimates and county-level characteristics. Section 6 provides conclusions and describes potential uses and limitations of the estimates.

2. Background

2.1 NIS-Child

Conducted annually, the NIS-Child is designed to produce estimates of vaccination coverage for 56 or more geographic areas, including all U.S. states and selected local areas. NCIRD works with each of the 50 states, D.C., six metropolitan areas, and eight U.S. territories to increase early childhood vaccination coverage levels. The NIS-Child produces timely and recurrent data for 19 to 35 month old children in the United States to help evaluate the efficacy of these efforts. Trends observed in the NIS-Child estimates can be used as an effective surveillance tool for identifying groups at risk of low vaccination coverage, and therefore, can guide programs to increase vaccination coverage among these subpopulations. NIS-Child data may be used to estimate vaccination coverage in various domains defined by socio-demographic characteristics (e.g., income level or race/ethnicity category).

NIS-Child uses two phases of data collection to obtain vaccination information for a large national probability sample of young children: an RDD telephone survey designed to identify households with children aged 19-35 months, followed by a provider record check, which obtains provider-reported vaccination histories for these children. Prior to 2011, the RDD portion consisted only of landline telephone numbers. From 2011 onwards, the NIS-Child began producing estimates based on inclusion of a cell-phone component for the survey in order to capture the rising number of cell-phone-only households (Blumberg and Luke 2019). In 2018, the NIS-Child discontinued landline sampling and shifted to single-frame cell-phone RDD sampling to increase efficiency (Hill et al. 2019).

The NIS-Child estimates are produced using information from provider-reported immunization histories for these children. We use the term *adequate provider data* to refer to the set of households with vaccination histories provided, among which estimates are produced. The details of NIS-Child methodology are discussed in Wolter et al. (2017).

2.2 Small Area Estimation

Small area estimation can improve direct survey estimates for small geographic areas, particularly when the areas have modest sample sizes available. The goal of small area estimation is to not only use the direct survey estimates based on NIS-Child data available from the geographic area, but to also use a model-based estimate based upon the relationship between area-level estimates and covariates. The small area estimate is a weighted average of the direct survey estimate and the model-based prediction for an area, where the weights of the two components are proportional to their estimated precision. For detailed background on small area estimation, see Rao (2003) and Rao and Molina (2015).

The small area methods that were used to generate county-level vaccination coverage rates were similar to the methods that were previously used by the CDC to obtain county-level vaccination coverage rates for 19 to 35 month old children using 1995 to 2008 NIS-Child data (Smith and Singleton 2008, 2011) and 2007 to 2015 NIS-Child data (Ganesh et al. 2016). Unlike the estimation process in the previous years, vaccination coverage rates were produced for children by age 24 months (with some exceptions), instead of those produced for children age 19 to 35 months.

See Table 1 for a list of vaccines/doses for which we produced county-level estimates of vaccination coverage. Estimates were produced for two sets of annual birth cohorts using 2008 to 2018 NIS-Child data, for children born 2007 to 2011 and for children born 2012 to 2016. In 2008, the NIS-Child did not collect data for the vaccines Hib – Primary series, Hib – Full series, Rotavirus, and 4:3:1:3*:3:1:4 series. Therefore, these four series had estimates produced for the 2008 to 2011 birth cohort instead of 2007 to 2011.

Vaccination rates were produced for children by age 24 months with the exceptions of Birth dose HepB (by age 3 days), Rotavirus (by age 8 months), and HepA 2 or more doses (by age 35 months). Estimates were also produced for the percent of unvaccinated children by age 24 months.

Table 1: List of Modeled Vaccines, National Immunization Survey-Child, County-Level Estimates, 2008-2018

Vaccine Abbreviation	**Vaccine Description**
DTaP (≥ 3 doses)	Three or more doses of diphtheria, tetanus toxoids, and acellular pertussis vaccine (includes children who might have been vaccinated with diphtheria, tetanus toxoids, and pertussis vaccine (DTP) and diphtheria and tetanus toxoids vaccine (DT))
DTaP (≥ 4 doses)	Four or more doses of diphtheria, tetanus toxoids, and acellular pertussis vaccine (includes children who might have been vaccinated with diphtheria, tetanus toxoids, and pertussis vaccine (DTP) and diphtheria and tetanus toxoids vaccine (DT))
Polio (≥ 3 doses)	Three or more doses of poliovirus vaccine
MMR (≥1 dose)	One or more doses of measles, mumps, and rubella vaccine
Hib - Primary series	Either ≥ 2 or ≥ 3 doses, depending on product type received
Hib - Full series	Primary series and booster dose, which includes receipt of ≥3 or ≥4 doses, depending on product type received
HepB (≥3 doses)	Three or more doses of hepatitis B vaccine
Birth dose HepB	One dose hepatitis B vaccine administered between birth and age 3 days
Varicella (≥1 dose)	One or more doses of varicella vaccine
PCV (≥3 doses)	Three or more doses of pneumococcal conjugate vaccine
PCV (≥4 doses)	Four or more doses of pneumococcal conjugate vaccine.
HepA (≥1 dose)	One or more doses of hepatitis A vaccine.
HepA (≥2 doses by 35 months)	Two or more doses of hepatitis A vaccine
Rotavirus (by 8 months)	Three or more doses of RotaTeq vaccine or two or more doses of Rotarix vaccine. The maximum age for the final rotavirus dose is 8 months, 0 days
Influenza (≥2 doses)	Two or more doses of seasonal influenza vaccine (doses must be at least 24 days apart [4 weeks with a 4-day grace period])
4:3:1:3*:3:1:4	The combined seven-vaccine series (4:3:1:3*:3:1:4) includes ≥4 doses of DTaP, ≥3 doses of poliovirus vaccine, ≥1 dose of measles-containing vaccine, full series of Hib (3 or 4 doses, depending on product type), ≥3 doses of hepatitis B vaccine, ≥1 dose of varicella vaccine, and ≥4 doses of PCV
Unvaccinated children	Children for whom the household respondent reported that the child has zero vaccinations and either (1) has zero providers or (2) has one or more providers, all of whom reported zero vaccinations

3. Methodology

For each vaccine series and birth cohort range examined, we used a modified version of the Lindley and Smith model to model the county-level direct estimates for vaccination coverage rates with auxiliary data. The small area approach is outlined here for deriving vaccination coverage rates, using 1+ MMR (proportion with at least one dose of MMR) as an example.

For the purpose of developing geographic units for fitting the model, counties were grouped into areas with the goal of having 35 children with adequate provider data in each area. County groupings were based on definitions of public health districts provided by state immunization programs. Specifically, all counties with at least 35 children with adequate provider data served as their own geographic units in the model. Then, we grouped remaining counties within public health districts together. If the remaining area of a public health district had at least 35 children with adequate provider area, that area served as its own geographic unit in the model. Otherwise, the areas not achieving sufficient sample size were grouped into one geographic unit within the state as a unit in the model. The target of 35 children with adequate provider data in each area in the model was chosen so that the direct survey estimates of vaccination rates for the geographic areas were reasonably accurate. The 3,137 counties included in the analyses were grouped into 633 geographic units for estimates for the 2007-2011 birth cohort, 584 geographic units for analyses for estimates for the 2008-2011 birth cohort, and 751 geographic units for analyses of the 2012-2016 birth cohort.

In order to preserve the bounds of the proportions when modeling, the direct survey estimates are logit-transformed. The transformed direct survey estimate for a vaccine was defined as:

$$y_i = log\left(\frac{z_i}{1 - z_i}\right), \qquad (1)$$

where z_i is the direct survey estimate for the proportion of vaccinated children for 1+ MMR and i denotes the county or county groupings.

Since y_i is undefined when $z_i = 0$ or 1, the direct survey estimates for the county and county groupings were truncated to 0.005 if $z_i \leq 0.005$ or 0.995 if $z_i \geq 0.995$. The logit-transformed direct survey estimate for all county and county groupings was modeled as:

$$y_i = x_i'\beta + v_i + e_i, \qquad (2)$$

where x_i is a vector of covariates for county or county grouping i. Covariates were selected for the model using the Bayesian Information Criterion (Neath and Cavanaugh 2012). In the above model, the v_i's are random effects which capture the area-specific effect not captured by the regression component $x_i'\beta$, and e_i is the sampling error associated with the transformed direct survey estimate. Standard distributional assumptions of normality were made for the area-specific random effects, i.e., $v_i \sim N(0, \sigma_v^2)$, where σ_v^2 is an unknown variance parameter. Furthermore, the v_i's and the e_i's were assumed to be pairwise mutually independent with $e_i \sim N(0, \psi_i)$. As mentioned previously, since e_i is the sampling error, ψ_i is the sampling variance associated with the transformed direct estimate, and is estimated by

$$\psi_i = \frac{\mathrm{var}(z_i^{(s)}) \left[\frac{n_i^{(s)}}{n_i}\right]}{\left[z_i^{(s)}(1 - z_i^{(s)})\right]^2}, \tag{3}$$

where $z_i^{(s)}$ is the direct survey estimate for the vaccine type for the state that includes county or county grouping i, $n_i^{(s)}$ is the number of children with adequate provider data in the state that includes county or county grouping i, and n_i is the number of children with adequate provider data for county or county grouping i. The above estimate for ψ_i follows from a Taylor series approximation for the variance of y_i.

The model given by (2) can also be expressed as $y_i = \theta_i + e_i$, where θ_i is the true (but unknown) value for the logit-transformed proportion of vaccinated children for 1+ MMR in county or county grouping i, and $\theta_i = x_i'\beta + v_i$. Typically, θ_i is the parameter of interest in a given small area model. However, for the specified model, θ_i is the true logit-transformed proportion of vaccinated children for 1+ MMR in county i. Thus, after deriving the model-based estimate for θ_i, that estimate was inverse logit-transformed to obtain an estimate for the proportion of vaccinated children for 1+ MMR. Finally, σ_v^2 was estimated using a restricted maximum likelihood estimator, and β was estimated using a weighted least squares estimator.

For counties with at least 15 children with adequate provider data, which indicated sufficient precision to utilize the direct survey estimate for the ultimate estimator, the small area estimation approach combined the direct survey estimate and model-based estimate. The county-level estimates were produced using the James-Stein estimator or Empirical Best Linear Unbiased Predictor (EBLUP) estimator for θ_i, given by:

$$\hat{\theta}_i = \frac{\psi_i}{\psi_i + \hat{\sigma}_v^2} x_i'\hat{\beta} + \frac{\hat{\sigma}_v^2}{\psi_i + \hat{\sigma}_v^2} y_i, \tag{4}$$

where $\hat{\beta}$ is the weighted least squares estimator of β, and $\hat{\sigma}_v^2$ is the restricted maximum likelihood estimator of σ_v^2. Note that the James-Stein estimator is a weighted linear combination of the regression estimate ($x_i'\hat{\beta}$) and the direct estimate (y_i). For counties with fewer than 15 children with adequate provider data, θ_i was estimated by a regression estimator:

$$\hat{\theta}_i = x_i'\hat{\beta}. \tag{5}$$

The estimator $\hat{\theta}_i$ is still in the logit scale. Thus, the estimate for the proportion of vaccinated children for 1+ MMR in each county was obtained by transforming from logit to a proportion. That is,

$$\hat{p}_i = \left(\frac{\exp(\hat{\theta}_i)}{1 + \exp(\hat{\theta}_i)}\right), \tag{6}$$

An estimate for the mean squared error (MSE) of the estimator $\hat{\theta}_i$ is given in Rao and Molina (2015). Since we modeled aggregated geographic areas, we transformed the model-based MSE to get the MSE of all the counties. Now, since the final model-based estimate $\hat{p}_i$ involved transforming to a proportion, the initial estimate for the MSE was adjusted to take into account the transformation. Using a Taylor series approximation, MSE of $\hat{p}_i$ can be approximated as follows:

$$\hat{v}(\hat{p}_i) = [p_i(1-p)]^2 \hat{v}(\hat{\theta}_i), \tag{7}$$

where $\hat{v}(\hat{\theta}_i)$ is the MSE estimator of $\hat{\theta}_i$ as given in Rao and Molina (2015). Finally, for each state, the county-level model-based estimates were ratio-adjusted so as to have the weighted (by population counts) sum of the county-level model-based estimates agree with the state-level vaccination coverage rate for 1+ MMR obtained from the NIS-Child.

Auxiliary data sources were used to provide the model covariates x_i. The covariates included demographic, health, and economic characteristics along with election results from the following data sources: 2018-19 Area Health Resource File (AHRF), 2019 Census Planning Database, 2012-2016 Vital Statistics Natality data, 2019 Robert Wood Johnson Foundation (RWJF) County Health Rankings Analytic data, 2012-2018 Vaccine Tracking System (VTrckS) data, and 2016 Presidential Election data.

These data sources were reviewed for candidate covariates that were plausibly related to vaccination coverage rates. The candidate covariates are presented in Appendix Table A1. Before conducting the estimation, missing values of potential predictors were imputed where necessary by applying sequential regression single imputation approach with predictive mean matching (Harrell 2015, Harrell 2020). Among 105 candidate covariates, 8 or 7.6% had missing data rates of 2.0% or more across the 3,137 counties, and the maximum missing data rate was 4.4%.

4. Evaluation of Estimates

For each model, the number of covariates selected by data source are presented in Appendix Table A2, along with pseudo-R^2 measures of goodness of fit. The pseudo-R^2 is defined as:

$$1 - \frac{LL(M^{full})}{LL(M^{Intercept})}, \tag{8}$$

where LL denotes the log-likelihood of the model, M^{full} denotes the full model and $M^{Intercept}$ is the intercept-only model. The pseudo-R^2's range between 0.11 and 0.99, with a median of 0.47. The lowest pseudo-R^2 measures are attributed to the models for the unvaccinated children rates whereas the models for Flu (≥ 2 doses) and HepB (Birth dose) had the largest pseudo-R^2's. This indicates that there was variation across the models as to what extent the covariate set explained differences in vaccination coverage among counties; i.e., available predictors were strongly predictive of county- or county grouping-level vaccination coverage for Flu (≥ 2 doses) and HepB (Birth dose), but were not very predictive of rates of unvaccinated children.

We further examined the precision of the estimates, specifically reviewing standard errors. Across 34 sets of estimates produced for difference combinations of vaccine series and birth cohort ranges, the median standard error of the small area estimates across 3,137 counties ranged from 0.2 to 10.9 percentage points, with a median of 3.6 percentage points.

The estimates are reported in the NIS Mapping Tool discussed in Section 5. After fitting the models, we also evaluated the small area methodology and resulting estimates using two methods. Our methods rely on the fact that, for areas with sufficient sample size, the direct survey estimate should be precise enough to provide a good benchmark for the small area estimate. The methods we employed were to:

1. Aggregate the county estimates to the state-level without ratio-adjustment and compare this aggregated estimate to the state direct survey estimate.

2. Select predictors for the small area models with a process excluding all counties and country groupings with 250 or more children with adequate provider data and then compare the resulting EBLUP small area estimates for these areas to the direct survey estimates. We conducted separate analysis of the results for large counties, where large here refers to counties with 250 or more children with adequate provider data.

We ran the evaluation across the different vaccination status estimates. For the state estimates, we examined the root mean square errors (RMSE's) of the aggregated small area vaccination estimates across the 50 states and D.C., where the RMSE for each estimated vaccination coverage rate is defined as

$$\sqrt{\frac{1}{51}\sum_{i=1}^{51}(\hat{p}_i - z_i)^2}, \tag{9}$$

where $\hat{p}_i$ for state i is the aggregate of the county small area estimates to the state-level without ratio-adjustment to agree with the state direct estimate and z_i is the direct state estimate. We also examined the magnitude of the bias

$$\left|\frac{1}{51}\sum_{i=1}^{51}(\hat{p}_i - z_i)\right|. \tag{10}$$

For the second analysis, we examined the RMSE's of the large counties employing variable selection excluding such geographies. For this analysis, the RMSE for each estimated vaccination coverage rate is

$$\sqrt{\frac{1}{m}\sum_{j=1}^{n}\left(\tilde{p}_j - z_j\right)^2}, \tag{11}$$

where m is the number of counties with 250 children with adequate provider data, $\tilde{p}_j$ for county j is the estimate with adjustment based on the state direct estimate, and z_j is the direct county estimate. The corresponding bias magnitude is

$$\left| \frac{1}{m} \sum_{j=1}^{n} (\tilde{p}_j - z_j) \right|, \tag{12}$$

The results from the evaluation are presented here in Table 2. We find that the median RMSE across models varied between 1.5 and 1.7 percentage points across the two evaluations, with a maximum RMSE across the two of 4.1 percentage points.

Table 2: Evaluation of Small Area Estimates: RMSE's Based on Comparison of Estimates to Direct Survey Estimates for Large Geographic Areas

RMSE Distribution Across Vaccination Models	State	Large Counties (With Variable Selection Excl. Large Geographies)
Minimum	0.011	0.008
25th Percentile	0.013	0.011
Median	0.017	0.015
75th Percentile	0.024	0.022
Maximum	0.041	0.034
Number of Areas per Model	51	30 to 47

The bias magnitudes are presented in Table 3. The results indicate overall low bias in the areas examined, with median bias magnitudes of 0.7 percentage points for the state analysis and 0.3 percentage points for the analysis of large counties. The maximum bias magnitude for any model examined is 1.0 percentage point. This indicates that most of the error in the estimates is random rather than fixed. In fact, we found that when examining bias-squared as a percentage of the mean square error, the median values were 15.6% for the state analysis and 3.1% for the county analysis.

Table 3: Evaluation of Small Area Estimates: Bias Magnitudes Based on Comparison of Estimates to Direct Survey Estimates for Large Geographic Areas

Bias Magnitude Distribution Across Vaccination Models	State	Large Counties (With Variable Selection Excl. Large Geographies)
Minimum	0.001	0.000
25th Percentile	0.004	0.002
Median	0.007	0.003
75th Percentile	0.008	0.004
Maximum	0.010	0.009
Number of Areas per Model	51	30 to 47

5. NIS Mapping Tool

To facilitate visualization and use of the small area vaccination coverage estimates, a mapping tool that builds upon *The National Opioid Misuse Community Assessment Tool* (https://opioidmisusetool.norc.org/) was developed. The purpose of the *NIS Vaccination Coverage Interactive Mapping Tool* (http://nis-mappingtool.norc.org/) is to allow the user to visualize vaccination coverage estimates while simultaneously seeing underlying demographic characteristics. The mapping tool includes small area estimates and indicators for which counties were included in model fitting (without grouping).

Estimates are presented for U.S. counties and 419 public health districts defined by groupings of counties, where the public health district estimates are based on aggregating the county-level estimates within the district and weighting based on the estimated size of the eligible population. Note that not every county in the U.S. is assigned to a public health district.

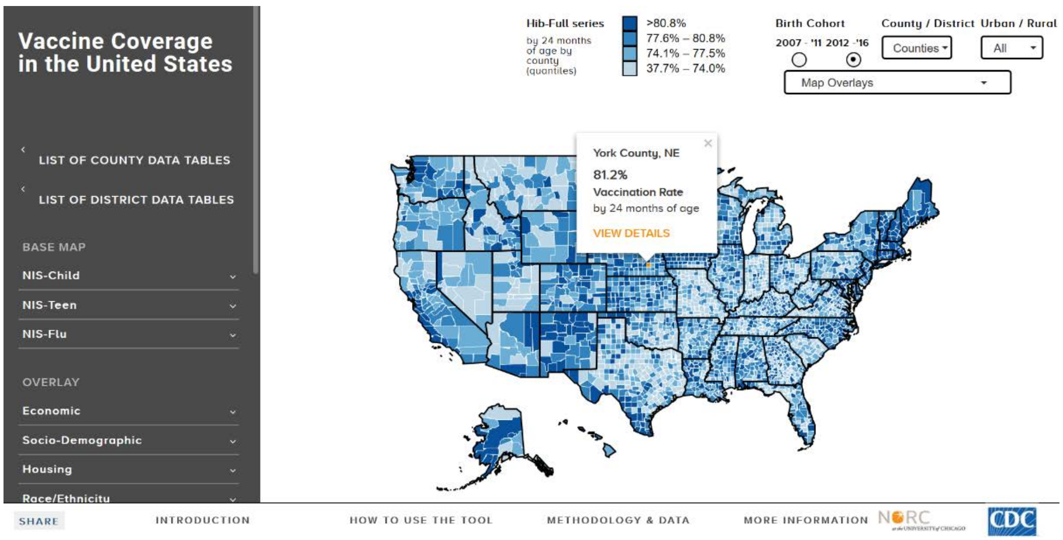

Figure 1: Image of NIS Mapping Tool

The tool has been populated with demographic and economic information, including race/ethnicity, age distribution, poverty rate, and other population characteristics that may be related to vaccination uptake, and can create printable fact sheets by public health district or county. In addition, address-level data are included regarding different vaccine provider types. Utilizing data visualization principles and mapping technology, this information could be used to help policymakers and immunization programs to quickly identify areas of low vaccination coverage and direct limited resources towards activities to help these communities. The tool also includes small area estimates from the National Immunization Survey-Teen (NIS-Teen) and the National Immunization Survey-Flu (NIS-Flu).

6. Conclusions

We produced county-level estimates of vaccination coverage in the United States using a small area estimation methodology with NIS-Child data provided by CDC. This will provide valuable insights for future immunization program planning, interventions, and policy changes. In addition, the available NIS Mapping Tool allows exploration of estimates across geographic areas and association of vaccination coverage with key covariates. Overall, we have demonstrated the value of our small area estimation approach for informing planning of immunization programs.

We note limitations associated with the estimates. Estimates for small counties should be used with caution due to strong reliance on model-based estimates rather than direct survey estimates, as there were not sufficient numbers of sample children with adequate provider data to generate accurate direct estimates of vaccination coverage. Information is included in the NIS Mapping Tool regarding which counties had fewer than 15 children with adequate provider data, for which estimates were produced solely based on model predictions rather than direct survey estimates. Continuing to identify data sources that may improve the fit of small area models is an area for continued investigation that may further improve the quality of small area estimates for vaccination coverage in the future.

References

Blumberg, S.J. and Luke, J.V. 2019. Wireless substitution: Early release of estimates from the National Health Interview Survey, July-December 2018. National Center for Health Statistics. June 2019. Available from: http://www.cdc.gov/nchs/nhis.htm

Fay, R.E. and Herriot, R.A. 1979. Estimates of income for small places: An application of James-Stein procedures to Census data. Journal of the American Statistical Association, 74, 269–277.

Ganesh, N., Chakraborty, A., Copeland, K., Wolter, K., Santos, K., Liu, L., Smith, P. J., Yankey, D., Jeyarajah, J., Santibanez, T., Singleton, J., Greby, S., Elam-Evans, L., and Smith, C. 2016. Small-area estimation methods for county-level vaccination coverage rates using the NIS. Presentation at the 2016 Joint Statistical Meetings.

Hill, H.A., Singleton, J.A., Elam-Evans, L.D., Nguyen, K., Pingali, S., Walker, T., Yankey, D., Zhao, Z., Johnson, C., Skalland, B., Wolter, K., and Fredua, B. 2019. Transition from a dual-frame (cell-phone and landline) to a single-frame (cell-phone) sample design: Impact on vaccination coverage estimates, National Immunization Survey-Child, 2014-2018. Available at: https://www.cdc.gov/vaccines/imz-managers/coverage/childvaxview/pubs-presentations/NIS-child-vac-coverage-estimates-2014-2018.html

Harrell Jr., F.E., 2015. Regression modeling strategies: With applications to linear models, logistic and ordinal regression, and survival analysis. Springer.

Harrell Jr., F. E. 2020. HMisc: Harrell miscellaneous. R package version 4.4-0. https://cran.r-project.org/web/packages/Hmisc/Hmisc.pdf

Lindley, D.V. and Smith, A.F.M. 1972. Bayes estimates for the linear model. Journal of the Royal Statistical Society Series B, 34(1), 1-41.

Neath, A.A. and Cavanaugh, J.E. 2012. The Bayesian information criterion: background, derivation, and applications. Wiley Interdisciplinary Reviews: Computational Statistics, 4(2), 199-203.

Rao, J.N.K. 2003. Small area estimation. New York: Wiley.

Rao, J .N .K. and Molina, I. 2015. Small area estimation. New York: Wiley.

Smith, P.J. and Singleton, J.A. 2008. Vaccination coverage for selected counties: Attainment of the Healthy People 2010 goals and association with county-level indices of access to care, economic conditions, and demographic composition. Public Health Reports, 123, 155-172.

Smith, P.J. and Singleton, J.A. 2011. County-level trends in vaccination coverage for selected counties in the U.S., children aged 19-35 months, United States, 1995-2008: Attainment of the healthy people 2010 objectives and association with county-level indices of access to care, economic conditions, and demographic composition. Morbidity and Mortality Weekly Report. Surveillance Summaries, 60(4), 1-86.

Wolter, K. M., Smith, P. J., Khare, M., Welch, B., Copeland, K. R., Pineau, V. J., and Davis, N. 2017. Statistical methodology of the National Immunization Survey, 2005-2014. National Center for Health Statistics. Vital and Health Statistics 1(61).

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
	Economic Dependence Typology	
Farming dependent county	2018-19 Area Health Resource File	25 percent or more of the county's average annual labor and proprietor's earnings were derived from farming, or 16 percent or more of jobs were in farming, 2010-2012
Mining dependent county	2018-19 Area Health Resource File	13 percent or more of the county's average annual labor and proprietors' earnings were derived from mining, or 8 percent or more of jobs were in mining, 2010-2012
Manufacturing dependent county	2018-19 Area Health Resource File	23 percent or more of the county's average annual average annual labor and proprietors' earnings were derived from manufacturing, or 16 percent or more of jobs were in manufacturing, 2010-2012
Federal/state government dependent county	2018-19 Area Health Resource File	14 percent or more of the county's average annual labor and proprietors' earnings were derived from Federal/State government during or 9 percent or more jobs were in Federal/State government, 2010-2012
Recreation dependent county	2018-19 Area Health Resource File	Computed using three data sources: 1) Percentage of wage and salary employment in entertainment and recreation, accommodations, eating and drinking places, and real estate as a percentage of all employment reported by the Bureau of Economic Analysis; 2) Percentage of total personal income reported for these same categories by the Bureau of Economic Analysis; and 3) Percentage of vacant housing units intended for seasonal or occasional use reported in the 2010 Census
Nonspecialized dependent county	2018-19 Area Health Resource File	County did not meet the economic dependence threshold for any one of the other above types
	Policy Types	
Low education typology county	2018-19 Area Health Resource File	20 percent or more of county residents age 25-64 did not have a high school diploma or equivalent, determined by the American Community Survey 5-year average data for 2008-12
Low employment typology county	2018-19 Area Health Resource File	Less than 65 percent of county residents age 25-64 were employed, determined by the American Community Survey 5-year average data for 2008-12
High poverty typology county	2018-19 Area Health Resource File	20 percent or more of its residents were poor as measured by the American Community Survey five-year estimates for 2008-12
Persistent poverty typology county	2018-19 Area Health Resource File	20 percent or more of county residents were poor, measured by the 1980, 1990, 2000 censuses, and the American Community Survey 5-year average data for 2007-11
Persistent child poverty typology county	2018-19 Area Health Resource File	20 percent or more of county related children under 18 were poor, measured in the 1980, 1990, 2000 censuses, and the American Community Survey 5-year average data for 2007-11
Population loss typology county	2018-19 Area Health Resource File	Number of county residents declined between the 1990 and 2000 censuses and also between the 2000 and 2010 censuses
Retirement destination	2018-19 Area Health Resource File	Number of residents age 60 and older grew by 15 percent or more between 2000 and 2010 censuses due to net migration

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
	Other Socio-Economic Indicators	
Percent white	2018-19 Area Health Resource File	The percentage of the total population that is White non-Hispanic/Latino (2017 Census County Characteristics File)
Percent black	2018-19 Area Health Resource File	The percentage of the total population that is Black non-Hispanic/Latino (2017 Census County Characteristics File)
Percent Asian	2018-19 Area Health Resource File	The percentage of the total population that is Asian (2017 Census County Characteristics File)
Percent Hispanic	2018-19 Area Health Resource File	The percentage of the total population that is Hispanic/Latino (2017 Census County Characteristics File)
Population under 5 years	2018-19 Area Health Resource File	The percentage of the total population that is under 5 years old
Population 10 to 19 years	2018-19 Area Health Resource File	The percentage of the total population that is 10 to 19 years old
Population under 20 years	2018-19 Area Health Resource File	The percentage of the total population that is under 20 years old
Percentage of population female among population under 5	2018-19 Area Health Resource File	The percentage of the population under 5 years old that is female
Median age	2018-19 Area Health Resource File	Median age based on 2010 Census.
Per capita income	2018-19 Area Health Resource File	Per capita personal income (REIS, 2017)
Household income	2018-19 Area Health Resource File	Median household income (2013-17 ACS)
Children in deep poverty	2018-19 Area Health Resource File	The percentage of 0-17 year old children in deep poverty (ACS 2013-2017)
Children in poverty	2018-19 Area Health Resource File	The percentage of 0-17 year old children in poverty (Census SAIPE 2017)
Divorced females	2018-19 Area Health Resource File	Percentage of divorced female, based on the ACS 2013-2017
Food stamp recipients	2018-19 Area Health Resource File	Percent of population that receives food stamps/SNAP (Census SNAP file, 2010)
Disabled Children	2018-19 Area Health Resource File	Percentage of non-institutionalized children (< 18 years) who are disabled (2013-17 ACS)
Children without health insurance	2018-19 Area Health Resource File	The percentage of children less than 19 years old without health insurance (Census SAHIE)
Unemployment rate	2018-19 Area Health Resource File	The percentage of civilians ages 16 years and over in the labor force that are unemployed (BLS, 2018)
Housing units with more than 1 person per room	2018-19 Area Health Resource File	Percent of housing units with more than 1 person per room (ACS 2013-2017)
No telephone service	2018-19 Area Health Resource File	The percentage of occupied housing units that do not have a working telephone and available service
Home value	2018-19 Area Health Resource File	Median home value (ACS 2013-2017)
Rent	2018-19 Area Health Resource File	Median gross rent (ACS 2013-2017)

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
Urban housing units	2018-19 Area Health Resource File	Percent of housing units classified as urban (Census 2010)
Population density	2018-19 Area Health Resource File	Population density per square mile (2010 Census Redistrict)
Housing density	2018-19 Area Health Resource File	Housing density per square mile (2010 Census Redistrict)
Agriculture, forestry, fishing, mining, hunting employment	2018-19 Area Health Resource File	Percentage of 16+ civilian employed population working in agriculture, forestry, fishing, mining, or hunting professions (ACS 2013-2017)
Construction employment	2018-19 Area Health Resource File	Percentage of 16+ civilian employed population working in the construction profession (ACS 2013-2017)
Education, health care and social assistance employment	2018-19 Area Health Resource File	Percentage of people aged 16 and older who are in the Percentage of 16+ civilian employed population working in education, health care and social assistance professions (ACS 2013-2017)
Manufacturing employment	2018-19 Area Health Resource File	Percentage of people aged 16 and older who are in the manufacturing profession (ACS 2013-2017)
Other industry employment	2018-19 Area Health Resource File	Percentage of people aged 16 and older who are in other industries (ACS 2013-2017)
Rural population	Census Planning Database	The percentage of 2010 population that lives outside of an Urbanized Area or Urban Cluster (2010 Census)
No high school diploma	Census Planning Database	The percentage of people aged 25 years and over at time of interview who are not high school graduates and have not received a diploma or the equivalent in the ACS population (2013-17 5 year ACS)
College degree or more	Census Planning Database	The percentage of people aged 25 years and over at the time of interview with a college degree or higher in the ACS population (2013-17 5 year ACS)
Receive public assistance	Census Planning Database	Percentage ACS occupied housing units that receive public assistance income (general assistance and Temporary Assistance to Needy Families) (2013-17 5 year ACS)
Non-owner occupied housing units	Census Planning Database	Percentage of 2010 Census occupied housing units that are not owner occupied, whether they are rented or occupied without payment of rent (2013-17 5 year ACS)
No health insurance	Census Planning Database	Percentage of people who have no health insurance coverage, public or private, in the ACS population (2013-17 5 year ACS)
School enrolled toddlers	Census Planning Database	Percentage of children age 3 and 4 that are enrolled in school in the ACS (2013-17 5 year ACS)
Census return rate	Census Planning Database	Percentage of completed 2010 Census mail forms received from addresses
Primary care physicians	Robert Wood Johnson County Health Rankings (2019)	Ratio of population to primary care physicians (2017 Area Health Resource File/American Medical Association)
Preventable hospital stays	Robert Wood Johnson County Health Rankings (2019)	Rate of hospital stays for ambulatory-care sensitive conditions per 100,000 Medicare enrollees 2016 Mapping Medicare Disparities Tool)

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
Flu vaccinations	Robert Wood Johnson County Health Rankings (2019)	Percentage of fee-for-service (FFS) Medicare enrollees that had an annual flu vaccination (2016 Mapping Medicare Disparities Tool)
Income inequality	Robert Wood Johnson County Health Rankings (2019)	Ratio of household income at the 80th percentile to income at the 20th percentile (2013-17 ACS 5 year estimates)
Social associations	Robert Wood Johnson County Health Rankings (2019)	Number of membership associations per 10,000 population (2016 County Business Patterns)
Life expectancy	Robert Wood Johnson County Health Rankings (2019)	Average number of years a person can expect to live (2015-17 NCHS Mortality File)
Premature age-adjusted mortality	Robert Wood Johnson County Health Rankings (2019)	Number of deaths among residents under age 75 per 100,000 population (2015-17 CDC WONDER Mortality data)
Diabetes prevalence	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults aged 20 and above with diagnosed diabetes (2016 CDC Diabetes Interactive Atlas)
Food insecurity	Robert Wood Johnson County Health Rankings (2019)	Percentage of population who lack adequate access to food (2016 Map the Meal Gap)
Limited access to healthy foods	Robert Wood Johnson County Health Rankings (2019)	Percentage of population who are low-income and do not live close to a grocery store (2015 USDA Food Environment Atlas)
Severe housing cost burden	Robert Wood Johnson County Health Rankings (2019)	Percentage of households that spend 50% or more of their household income on housing (2013-17 ACS 5 year estimates)
	Natality Measures	
Children with low birth weight	2012-16 Natality data	Percentage of children born with birth weight < 2500 grams
Children with very low birth weight	2012-16 Natality data	Percentage of children born with birth weight < 1500 grams
Births to young mothers	2012-16 Natality data	Percentage of children born to mothers less than 20 years old
Births with gestations less than 37 weeks	2012-16 Natality data	Percentage of children with gestations less than 37 weeks
Births to unmarried mothers	2012-16 Natality data	Percentage of children born to unmarried mothers
Births in hospitals	2018-19 Area Health Resource File	Percentage of births in hospitals (2017 AHA Survey Database)
	Household and Family Characteristics	
Single parent household	2018-19 Area Health Resource File	Percentage of single parent households (2010 Census)
Average family size	2018-19 Area Health Resource File	Average family size as in 2010 Census.
Husband wife household	2018-19 Area Health Resource File	Percentage of husband-wife households (2010 Census)
Family with female heads	2018-19 Area Health Resource File	Percentage of families with a female head (2010 Census)

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
Speaks a language other than English at home	Census Planning Database	The percentage of the population aged 5 years and over that speaks a language other than English at home (ACS 2013-2017)
Spanish (or Spanish Creole)-speaking people	Census Planning Database	Percentage of people ages 5 years and over who speak English less than "very well" and speak Spanish or Spanish Creole at home in the ACS. Examples include Ladino and Pachuco
Moved from another residence within the last year	Census Planning Database	The percentage of the population aged 1 year and over that moved from another residence in the U.S. or Puerto Rico within the last year
Citizen at birth	Census Planning Database	The percentage people who are citizens of the United States at birth in the ACS. This includes respondents who said that they were born in the United States, Puerto Rico, a US Island Area (such as Guam), or abroad of American (US citizen) parent or parents. (2013-17 5 year ACS)
Spanish(or Spanish Creole)-speaking household	Census Planning Database	The percentage of all ACS occupied housing units where a Spanish or Spanish Creole language was assigned as the household language and no one ages 14 years and over speaks English only or speaks English "very well".
	Behavioral Characteristics	
Adults with fair or poor health	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults who are in poor or fair health based on the 2016 Behavioral Risk Factor Surveillance System
Poor physical health days	Robert Wood Johnson County Health Rankings (2019)	Average number of physically unhealthy days reported in past 30 days based on the 2016 Behavioral Risk Factor Surveillance System
Poor mental health days	Robert Wood Johnson County Health Rankings (2019)	Average number of mentally unhealthy days reported in past 30 days based on the 2016 Behavioral Risk Factor Surveillance System
Adult smoking	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults who are current smokers based on the 2016 Behavioral Risk Factor Surveillance System
Adult Obesity	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults who are currently obese, based on the 2015 CDC Diabetes Interactive Atlas
Food environment index	Robert Wood Johnson County Health Rankings (2019)	Index of factors that contribute to a healthy food environment, from 0 (worst) to 10 (best) (2015 & 2016 USDA Food Environment Atlas, Map the Meal Gap from Feeding America)
Physical inactivity	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults age 20 and over reporting no leisure-time physical activity (2015 CDC Diabetes Interactive Atlas)
Access to exercise opportunities	Robert Wood Johnson County Health Rankings (2019)	Percentage of population with adequate access to locations for physical activity (2010 & 2018 Business Analyst, Delorme map data, ESRI, & US Census Tigerline Files)
Excessive drinking	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults reporting binge or heavy drinking (2016 Behavioral Risk Factor Surveillance System)

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
Driving alone to work	Robert Wood Johnson County Health Rankings (2019)	Percentage of the workforce that drives alone to work (2013-17 ACS 5 year estimates)
Long commute - driving alone	Robert Wood Johnson County Health Rankings (2019)	Among workers who commute in their car alone, the percentage that commute more than 30 minutes (2013-17 ACS 5 year estimates)
Frequent physical distress	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults reporting 14 or more days of poor physical health per month (2016 Behavioral Risk Factor Surveillance System)
Frequent mental distress	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults reporting 14 or more days of poor mental health per month (2016 Behavioral Risk Factor Surveillance System)
Insufficient sleep	Robert Wood Johnson County Health Rankings (2019)	Percentage of adults who report fewer than 7 hours of sleep on average (2016 Behavior Risk Factor Surveillance System)
	2016 Presidential Election Data	
Votes for Donald Trump	2016 Presidential Election Returns	Percentage of all voters voting for Donald Trump
Votes for Donald Trump or Hillary Clinton	2016 Presidential Election Returns	Percentage of all voters voting for either Donald Trump or Hillary Clinton
Voters	2016 Presidential Election Returns	Percentage voting among the estimated voting age population, based on 2014-18 ACS 5 year estimates
	Vaccine-Specific Information	
DTaP	2012-18 VTrckS	DTaP doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
DTaP+Polio	2012-18 VTrckS	DTaP + Poliovirus doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
DTaP+Polio+HepB	2012-18 VTrckS	DTaP + Poliovirus + Hepatitis B doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
HepA	2012-18 VTrckS	Hepatitis A doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
HepB	2012-18 VTrckS	Hepatitis B doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
Hib	2012-18 VTrckS	Hib doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
MMR	2012-18 VTrckS	Measles, mumps, rubella doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
Polio	2012-18 VTrckS	Poliovirus doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data

Appendix Table A1: Candidate Predictors for Small Area Estimation Models

Potential Predictors	Data Source	Description
PCV	2012-18 VTrckS	Pneumococcal conjugate vaccine doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
Rotavirus	2012-18 VTrckS	Rotavirus doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data
Varicella	2012-18 VTrckS	Varicella doses ordered through VTrckS relative to the estimated number of 1 to 2 year old children based on 2014-18 American Community Survey data

Appendix Table A2: Pseudo-R^2 and Number of Variables Selected for Models by Data Source

Vaccine Type	Birth Cohort Years	Number of Variables Selected							Pseudo R^2
		Total	AHRF (2018-19)	Census Planning Database	Election Returns (2016)	Natality (2012-16)	RWJF County Health Rankings (2019)	VTrckS (2012-18)	
Number of candidate variables		105	48	13	3	5	25	11	
Varicella (≥1 dose)	2012-16	10	4	2	0	1	3	0	0.38
4:3:1:3*:3:1:4	2012-16	19	10	4	0	0	4	1	0.41
HepB (≥3 doses)	2012-16	16	5	3	1	1	4	2	0.50
HepA (≥1 dose)	2012-16	16	7	1	0	0	5	3	0.75
HepA (≥2 doses)	2012-16	23	11	2	0	2	5	3	0.67
Hib-Full series	2012-16	13	5	2	1	0	4	1	0.47
Hib-Primary series	2012-16	13	6	2	1	1	3	0	0.45
MMR (≥1 dose)	2012-16	6	2	1	0	1	1	1	0.34
PCV (≥3 doses)	2012-16	10	7	2	0	0	1	0	0.47
PCV (≥4 doses)	2012-16	11	3	3	0	0	4	1	0.53
Pol (≥ 3 doses)	2012-16	10	3	3	0	1	3	0	0.49
DTaP (≥ 3 doses)	2012-16	15	5	3	0	1	5	1	0.50
DTaP (≥ 4 doses)	2012-16	16	6	2	1	0	4	3	0.50
Unvaccinated children	2012-16	13	8	2	0	2	0	1	0.11
Flu (≥ 2 doses)	2012-16	21	9	4	1	1	5	1	0.98
HepB (Birth dose)	2012-16	13	3	2	0	1	3	4	0.87
Rotavirus	2012-16	14	9	4	0	0	0	1	0.61
Varicella (≥1 dose)	2007-11	14	3	3	0	0	6	2	0.36
4:3:1:3*:3:1:4	2008-11	9	5	1	1	0	2	0	0.36
HepB (≥3 doses)	2007-11	10	3	3	0	1	3	0	0.47
HepA (≥1 dose)	2007-11	15	4	2	0	1	8	0	0.94
HepA (≥2 doses)	2007-11	14	4	1	0	1	8	0	0.74
Hib-Full series	2008-11	10	4	1	1	0	4	0	0.41
Hib-Primary series	2008-11	12	7	0	1	0	4	0	0.35
MMR (≥1 dose)	2007-11	7	5	1	0	1	0	0	0.26
PCV (≥3 doses)	2007-11	11	3	1	0	0	6	1	0.35
PCV (≥4 doses)	2007-11	12	6	0	0	0	5	1	0.54
Pol (≥ 3 doses)	2007-11	9	3	3	0	0	2	1	0.31
DTaP (≥ 3 doses)	2007-11	12	9	2	0	0	1	0	0.37
DTaP (≥ 4 doses)	2007-11	17	9	3	1	0	4	0	0.41
Unvaccinated children	2007-11	23	11	4	0	2	5	1	0.11
Flu (≥ 2 doses)	2007-11	14	5	4	0	2	3	0	0.99
HepB (Birth dose)	2007-11	19	6	4	1	1	4	3	0.99
Rotavirus	2008-11	22	7	2	1	0	8	4	0.76