Modeling County-Level Vaccination Coverage Rates Across Time

N. Ganesh[1], Philip J. Smith[2], Kennon R. Copeland[1],
Kirk M. Wolter[3]
[1]NORC at the University of Chicago, 4350 East-West Highway, Bethesda, MD 20814
[2]Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333
[3]NORC at the University of Chicago, 55 E. Monroe, Chicago, IL 60603

Abstract
A small area model that accounts for trends and state effects is used to obtain county-level vaccination coverage rates for a multi-year period. Model-based county-level estimates - derived using a prediction approach - are a weighted combination of multiple estimates.

Key Words: EBLUP, NIS, Small Area Estimation

1. Introduction[†]

The National Immunization Survey (NIS) - a nationwide, list-assisted random digit dialing (RDD) survey conducted by the Centers for Disease Control and Prevention - monitors the vaccination rates of children between the ages of 19 and 35 months. Each year, the NIS conducts interviews with approximately 24,000 households across the United States and is designed to obtain estimates of vaccination coverage rates at the national and state level and for select local areas.

While the sample design does not support estimates meeting publication guidelines suggested by the National Center for Health Statistics for the vast majority of counties, there is an interest in monitoring vaccination rates at the sub-state level. Official estimates of county-level vaccination coverage rates for the two-year period 2004-2005 have been described by Smith and Singleton (2008).

In this paper, we discuss a small area model for county-level estimates to derive model-based estimates for the two-year periods from 1996 to 2003 based solely on direct estimates at the county level. Such methods can provide for expanded sets of estimates for examining trends over time.

The paper is organized as follows: the methodology for the county-level model is described in section 2; results from modeling county-level direct estimates for three or more doses of poliovirus vaccine are summarized in section 3; and concluding remarks are given in section 4.

[†] "The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of Centers for Disease Control and Prevention."

2. Methods

2.1 Small area models

Area-level models in small area estimation use auxiliary information along with random effects to account for small area/time variation in the true values of the parameters of interest. This can be expressed simply as:

$$y = X\beta + Zv + e$$

where y is a vector of small area level direct estimates for the variable of interest, β is a vector of fixed effects, X represents the auxiliary data, Z is a known matrix of coefficients, v is a vector of random effects, and e is a vector of sampling errors. The random effects thus capture the small area/time variation not explained by the auxiliary variables.

We describe how to develop small area estimates in the absence of auxiliary information on the areas. This is the situation when looking at historical estimates for which county-level characteristics for these periods are not available or too costly to obtain. Given the lack of predictor variables, a trend model (see section 2.3) making use of direct county-level estimates across time is considered.

2.2 Transformation of county-level direct estimates

County-level direct estimates, z_{it} (i.e., estimates based solely on the sample data and associated sample weights), for the proportion of 19-35 month-old children who were up-to-date for three or more doses of poliovirus vaccine were derived using 1996-2007 National Immunization Survey data. Due to sample size considerations, instead of modeling the county-level direct estimates for every county and year from 1996 to 2007, only counties and two-year periods with a sample size of at least 35 for the two-year period and at least 15 in each of the two years were modeled. That is, two-year county-level direct estimates for all available periods 1996-97, 1998-99, 2000-01, 2002-03, 2004-05 and 2006-07 were modeled to derive county-level model-based estimates for the periods of interest from 1996-97 to 2002-03.

Prior to fitting the model given in section 2.3, the data were transformed to log-odds: $y_{it} = \log[z_{it}/(1-z_{it})]$. An estimator ψ_{it} for the variance of y_{it} was derived using a Taylor series method. For values of z_{it} close to one, ψ_{it} will be large, and such direct estimates would receive an extremely small weight in the small area estimate. As a result, a maximum value for ψ_{it} was established for counties with values of z_{it} close to one:

$$\psi_{it} = \begin{cases} \dfrac{1}{z_{it}^2(1-z_{it})^2}[se(z_{it})]^2 & \text{if } z_{it} \le 0.9469 \\[2ex] \dfrac{0.9469(1-0.9469)}{n_{it}/deff} & \text{if } z_{it} > 0.9469, \end{cases} \qquad (1)$$

where $se(z_{it})$ is the standard error of z_{it}, 0.9469 is the 75th percentile of z_{it} for all counties and two-year periods, n_{it} is the sample size for county *i* and two-year period *t*, and $deff$ (=1.331) is the median value of the estimated design effect for all counties and two-year periods.

2.3 Trend model

Our model for the logit of the county-level direct estimate, y_{it}, for county *i* and two-year period *t* is given by (2), where *m* is the total number of counties with at least one two-year period with sufficient sample size.

$$y_{it} = \mu + \alpha_t + \beta_s + v_i + u_{it} + e_{it},\ i = 1,...,m,\ t = 1,...,n_i, \qquad (2)$$

where $t = 1,...,n_i$ indexes the two-year periods for county *i*. Note that logits may not be defined for all two-year periods for a given county (the total number of two-year periods possible is *T*); the logit would not be defined for any two-year period in which the county sample size was less than 35 and/or less than 15 in either of the two years.

For the county-level model, we assume: 1) a mean, μ, for the base two-year period (1996-1997) for all counties in states without a state-effect β_s; 2) a state effect , β_s, for all counties in state *s* (the state effect β_s is included only for states with multiple counties); 3) and a period effect, α_t, for two-year period *t* relative to the base two-year period ($\alpha_1 \equiv 0$), defined for all counties with defined logits in a given two-year period *t*. The v_i's are random effects which capture the county specific effect not captured by $\mu + \alpha_t + \beta_s$, u_{it} is a county-by-time random effect which explains the additional variability not captured by either $\mu + \alpha_t + \beta_s$ or v_i , and e_{it} is the sampling error associated with y_{it}.

It is assumed that the v_i's, the u_{it}'s and the e_{it}'s are pairwise mutually independent, and $v_i \sim N(0,\sigma_v^2)$, $u_{it} \sim N(0,\sigma_u^2)$, $e_{it} \sim N(0,\psi_{it})$. The sampling variance ψ_{it} is assumed known without error. The assumption of known sampling variance is frequently used in area-level models in small area estimation (e.g., Fay and Herriot, 1979).

Since the state effect β_s is included only for states with multiple counties, all β_s's in the model are estimable. For example, District of Columbia does not have a state effect, only a county effect.

The parameter of interest in model (2) is

$$\theta_{it} = \mu + \alpha_t + \beta_s + v_i + u_{it}.$$

Thus, the true but unknown value for the logit of the proportion of 19-35 month-old children who were up-to-date for three or more doses of poliovirus vaccine for a given county/two-year period can be expressed as the sum of a mean for the base two-year period, the period mean, the state mean, the county effect, and the county-by-time effect.

2.4 Model-based estimates

The model-based estimate of θ_{it} used in estimating county-level vaccination rates is derived using the best linear unbiased predictor [BLUP, see Rao (2003)] approach. The BLUP is the model-based estimate $\tilde{\theta}_{it}$ that has the smallest mean squared error (i.e., minimum variance for $(\tilde{\theta}_{it} - \theta_{it})$), subject to the constraint that $\tilde{\theta}_{it}$ is unbiased (i.e.,

$E(\tilde{\theta}_{it}) = E(\theta_{it})$). The BLUP for θ_{it} can be expressed as (see Appendix for details on the derivation of $\tilde{\theta}_{it}$):

$$\tilde{\theta}_{it} = (1 - w_{it}^{(1)}) y_{it} + \left(\frac{w_{it}^{(1)}}{1 + w_i^*} \right) (\tilde{\mu} + \tilde{\alpha}_t + \tilde{\beta}_s) + \left(\frac{w_{it}^{(1)} w_i^*}{1 + w_i^*} \right) \sum_{j=1}^{n_i} \frac{w_{ij}^{(2)}}{w_i^*} (y_{ij} + \tilde{\alpha}_t - \tilde{\alpha}_j), \qquad (3)$$

where by definition $\tilde{\alpha}_1 \equiv 0$, $(\tilde{\mu}, \tilde{\alpha}_2, ..., \tilde{\alpha}_T, \tilde{\beta}_1, ..., \tilde{\beta}_S)'$ is the best linear unbiased estimator of $(\mu, \alpha_2, ..., \alpha_T, \beta_1, ..., \beta_S)'$, and

$$w_{it}^{(1)} = \frac{\psi_{it}}{\sigma_u^2 + \psi_{it}}, \ w_{ij}^{(2)} = \frac{\sigma_v^2}{\sigma_u^2 + \psi_{ij}}, \ w_i^* = \sum_{j=1}^{n_i} w_{ij}^{(2)} .$$

Thus, the model-based estimate for the logit of the proportion of 19-35 month-old children who were up-to-date for three or more doses of poliovirus vaccine for a given county/two-year period, $\tilde{\theta}_{it}$, is thus a weighted combination of three unbiased estimates of θ_{it}:

1) the logit of the direct estimate;
2) an estimate for the main effect model for θ_{it} (v_i and u_{it}, being random effects with zero expected values, are thus not included in this estimate); and
3) an estimate which itself is a weighted combination of multiple unbiased estimates of θ_{it}, each of which specifically accounts for differences between the period effect for period t and the period effect for other periods.

The three unbiased estimates of θ_{it} are differentially weighted, with the weights, $(1 - w_{it}^{(1)})$, $\left(\frac{w_{it}^{(1)}}{1 + w_i^*} \right)$, and $\left(\frac{w_{it}^{(1)} w_i^*}{1 + w_i^*} \right)$, summing to one. The weights are determined by the variances associated with each estimate and can be expressed in terms of the variances of the logit of the direct estimates, ψ_{it}, the variance of the county effect (σ_v^2), and the variance of the county-by-time effect (σ_u^2), as indicated above.

Since the model-based estimate $\tilde{\theta}_{it}$ depends on unknown variance components (σ_v^2, σ_u^2), an **empirical** model-based estimate, $\hat{\theta}_{it}$, is obtained by substituting an estimate $(\hat{\sigma}_v^2, \hat{\sigma}_u^2)$ for (σ_v^2, σ_u^2) in (3). After deriving the empirical model-based estimate $\hat{\theta}_{it}$, the model-based estimate, $\hat{p}_{it}$, of the proportion of 19-35 month-old children in county *i* and two-year period *t* with three or more doses of poliovirus vaccine was obtained by back-transforming the empirical model-based estimate $\hat{\theta}_{it}$. That is,

$$\hat{p}_{it} = \frac{\exp(\hat{\theta}_{it})}{1 + \exp(\hat{\theta}_{it})} .$$

The variance of the empirical model-based estimate $\hat{\theta}_{it}$ was derived along the lines given in Rao (2003), and the variance for the model-based estimate $\hat{p}_{it}$ was derived using a Taylor series method.

3. Results

The model given by (2) was fitted using y_{it} and ψ_{it} given in section 2.2. There were 281 counties with sufficient sample size for at least one two-year period (m=281 and T=6), with 49 states having two or more counties with sufficient sample size. Among the 49 possible state effect parameters, β_s, only 17 were significant at the 0.1 level. That is, in the fitted model, only 17 states had a state effect parameter β_s. In other words, for 32 of the states, there was no discernible difference in their poliovirus vaccination coverage and the coverage at the national level.

Table 1 provides parameter estimates for the final model. The negative estimates for α_2 and α_3, and $\alpha_2 > \alpha_3$ indicate an overall decline in the logit of the vaccination rates for three or more doses of poliovirus vaccine for the periods 1998-99 and 2000-01 compared to the base period 1996-97. The positive estimates for α_4 , α_5 and α_6 , and $\alpha_6 > \alpha_5 > \alpha_4$ indicate an increase in the logit of the vaccination rates for three or more doses of poliovirus vaccine for the periods 2002-03, 2004-05 and 2006-07 compared to the base period 1996-97.

A Q-Q plot of the standardized residuals (Figure 1) indicates that beyond the second standard deviation, the sample quantiles deviate significantly from the normal quantiles.

For each one of the two-year periods of interest (1996-97 to 2002-03), Table 2 gives the distribution of the difference between the county-level direct estimates and model-based estimates for percent of 19-35 month-old children with three or more doses of poliovirus vaccine. Approximately 50% (80%) of the county-level direct estimates for each of the two-year periods are within $\pm$2.5 ($\pm$5.0) percentage points of the model-based estimates. As the median standard error for the counties is greater than three percentage points, this suggests that for most counties/periods, the model-based estimate is not significantly different from the direct estimate.

Figure 2 maps the model-based estimates for the period 2002-03. As seen in the map, most of the counties with high vaccination rates for three or more doses of poliovirus vaccine are counties in east coast states.

For Pulaski County, AR, Figure 3 gives a plot of the direct estimate and the model-based estimate, along with associated 95% confidence intervals for the two-year periods 1996-97 to 2002-03. The model-based estimate is similar to the direct estimate but has confidence intervals that are more narrow.

Table 3 gives the median length of the 95% confidence intervals for the county estimates for percent of 19-35 month-old children with three or more doses of poliovirus vaccine, for each one of the two-year periods from 1996-97 to 2002-03. The confidence intervals associated with the model-based estimates are more than 40 percent narrower than the confidence intervals associated with the direct estimates.

As a consequence of the model-based estimates having tighter confidence intervals, the period-to-period variability in the estimates is also reduced. Table 4 provides distributional statistics for the direct and model-based county level estimates over the four two-year periods from 1996-97 to 2002-03. Two statistics examined were the standard deviation of the estimates across time and the difference between the maximum and minimum estimates across time. The national level estimates for poliovirus vaccinations rates were relatively stable between 1996 and 2003 (varying between 89.4 and 91.6, but neither increasing nor decreasing consistently across the period). As seen in Table 4, the model-based estimates are much more stable across time than are the direct estimates.

4. Conclusion

Use of a trend model allows for model-based estimates of county-level vaccination coverage rates with confidence intervals more than 40 percent narrower than those for the corresponding direct estimates. Thus, the model-based estimates offer improved precision over the county-level direct estimates. One benefit of the model-based estimates is improved precision of estimates across time, which could allow for improved monitoring of vaccination coverage rates.

The model-based estimates may also offer an improvement over adhoc approaches such as pooling NIS county-level direct estimates over long multi-year periods (e.g., 3 to 5 years) using simple weighted proportions. The model-based approach presented in this paper 'borrows strength' by using a weighted combination of a direct estimate and a model estimate, allowing estimates for narrower periods of time. This will also serve to minimize potential bias that may be incurred when there is a trend over time.

The model developed here makes good use of the available data, affording the opportunity to derive efficient estimates for a county/period with nothing more than the direct county estimates across periods. While there may be some difficulty in interpreting the various terms in the model-based estimate, it is unbiased and minimizes variance.

There remains the question of model fit. Some of the model assumptions may not hold. For example, the normal Q-Q plot indicates that the transformed data may have a non-normal distribution. This seems to suggest that in order to protect against an incorrectly specified distribution for the data, a non-parametric method should be used to estimate the unknown variance components (σ_v^2, σ_u^2) for use in the empirical model-based estimate $\hat{\theta}_{it}$. In this research, we used maximum likelihood to estimate the variance components; however, alternate methods such as nonparametric bootstrap or estimating equations could also be considered.

Another concern is that the estimate for the design effect used in (1) may be too small. The design effect was used when computing an estimated variance of y_{it} only for counties with z_{it} close to one. Typically for the National Immunization Survey, design effects are much larger. A larger design effect will result in model-based estimates having a smaller weight associated with the logit of the direct estimate y_{it}.

The assumption of equal period effect α_t for all counties in a given period is unlikely to hold. Future work should involve developing models to allow for unequal period effects for counties in a given period. Finally, data analysts may use the model-based estimates over time for each county to fit time-trend curves. The danger in such methods is that by not taking into account the variability associated with the model-based estimates, the fitted time-trend curves will likely be over-smoothed. Also, the fitted time trend curves will likely not be sensitive to picking up recent departures from the prior year time trend curves.

References

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. Volume 74, pp. 269-277.

Rao, J.N.K. (2003). *Small Area Estimation*. Wiley, New York.

Smith, P.J, and Singleton, J.A. (2008). *Vaccination Coverage Estimates for Selected Counties: Achievement of Healthy People 2010 Goals and Association with Indices of Access to Care, Economic Conditions, and Demographic Composition*. Public Health Reports 123, pp. 155-172.

Table 1: For parameters in trend model (2), the estimate, the associated standard error, and the p-value for a hypothesis test that the parameter is equal to zero [Data source: National Immunization Survey 1996-2007].

Description	Parameter	Estimate	Standard error	p-value
Mean for the base year 1996-97 for counties in reference states*	μ	2.2679	0.0368	0.0000
Period effect for 1998-99	α_2	-0.0748	0.0403	0.0634
Period effect for 2000-01	α_3	-0.1410	0.0392	0.0003
Period effect for 2002-03	α_4	0.0702	0.0418	0.0930
Period effect for 2004-05	α_5	0.3458	0.0447	0.0000
Period effect for 2006-07	α_6	0.4595	0.0465	0.0000
State effect for AK	β_{AK}	-0.3953	0.1499	0.0084
State effect for AZ	β_{AZ}	-0.3176	0.1209	0.0086
State effect for CT	β_{CT}	0.5964	0.1496	0.0001
State effect for ID	β_{ID}	-0.3863	0.1391	0.0055
State effect for LA	β_{LA}	-0.4307	0.1341	0.0013
State effect for ME	β_{ME}	0.5105	0.1429	0.0004
State effect for MA	β_{MA}	0.6104	0.1178	0.0000
State effect for MN	β_{MN}	0.5230	0.1811	0.0039
State effect for MT	β_{MT}	-0.3316	0.1300	0.0107
State effect for NH	β_{NH}	0.4847	0.1426	0.0007
State effect for NM	β_{NM}	-0.4519	0.1497	0.0025
State effect for NC	β_{NC}	0.5303	0.2187	0.0153
State effect for OR	β_{OR}	-0.3676	0.1371	0.0073
State effect for RI	β_{RI}	0.8136	0.1608	0.0000
State effect for TX	β_{TX}	-0.3109	0.1116	0.0053
State effect for UT	β_{UT}	-0.2622	0.1343	0.0509
State effect for VT	β_{VT}	0.6811	0.1269	0.0000
Variance of county effect	σ_v^2	0.0511	0.0089	0.0000
Variance of county-by-time effect	σ_u^2	0.0308	0.0063	0.0000

*The reference states are all states with no significant state effect (i.e., all states excluding AK, AZ, CT, ID, LA, ME, MA, MN, MT, NH, NM, NC, OR, RI, TX, UT, and VT).

Figure 1: Normal Q-Q plot of the standardized residuals from trend model (2) [Data source: National Immunization Survey 1996-2007].

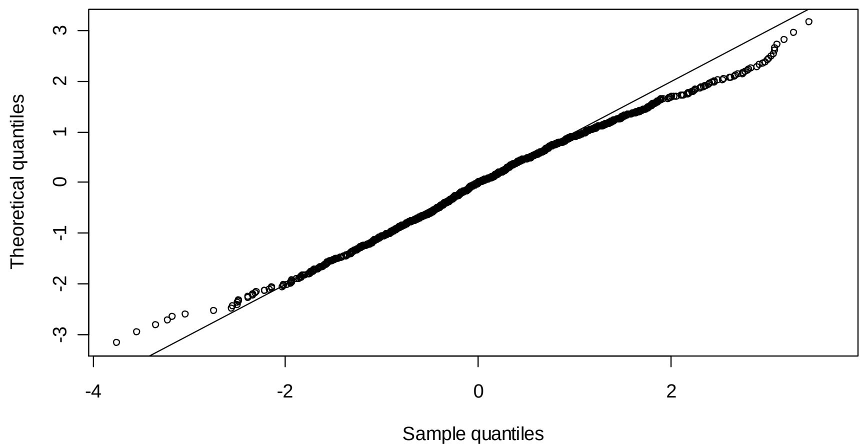

Table 2: Distribution of the difference between county-level direct estimates and model-based estimates of poliovirus vaccination coverage rates (%) [Data source: National Immunization Survey 1996-2007].

Two-year period	Min	P2.5	P10	P25	Median	P75	P90	P97.5	Max
1996-97	-8.6	-6.6	-3.9	-1.7	0.1	1.9	3.9	5.5	9.2
1998-99	-15.8	-8.4	-5.0	-1.5	0.1	2.3	4.0	6.8	8.0
2000-01	-16.6	-8.6	-4.6	-1.3	0.2	2.6	4.7	6.9	9.9
2002-03	-13.4	-7.9	-3.8	-1.7	0.0	1.8	4.2	6.2	7.3

Table 3: Median length of the 95% confidence interval (CI) for county-level estimates of poliovirus vaccination coverage rate [Data source: National Immunization Survey 1996-2007].

Two-year period	Median length of 95% CI for direct estimate	Median length of 95% CI for model-based estimate
1996-97	12.7	6.8
1998-99	13.8	7.1
2000-01	15.1	7.9
2002-03	12.9	6.5

Figure 2: County-level model-based estimates of poliovirus vaccination coverage rate for counties with sufficient sample for the two-year period 2002-03 [Data source: National Immunization Survey 1996-2007].

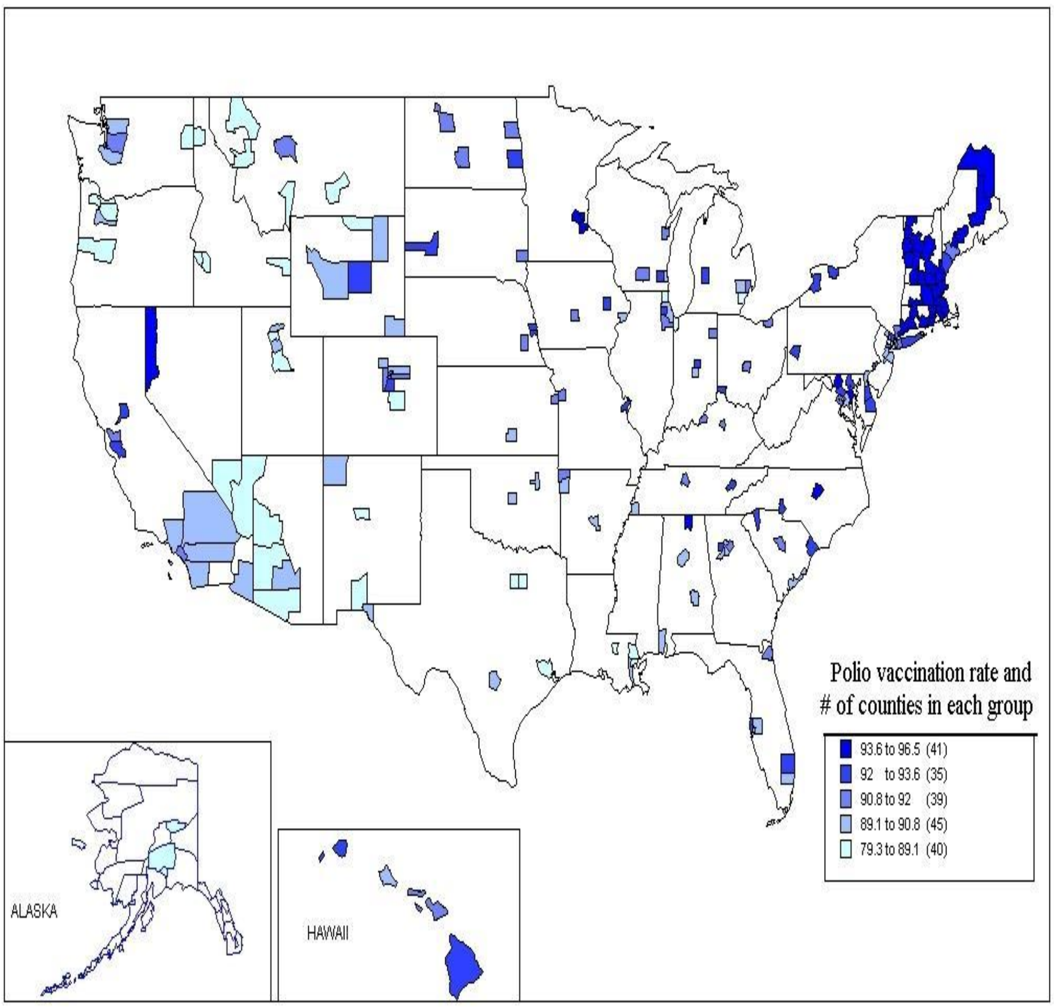

Figure 3: Direct and model-based estimates of poliovirus coverage vaccination rate, and associated 95% confidence intervals (CI), for Pulaski county [Data source: National Immunization Survey 1996-2007].

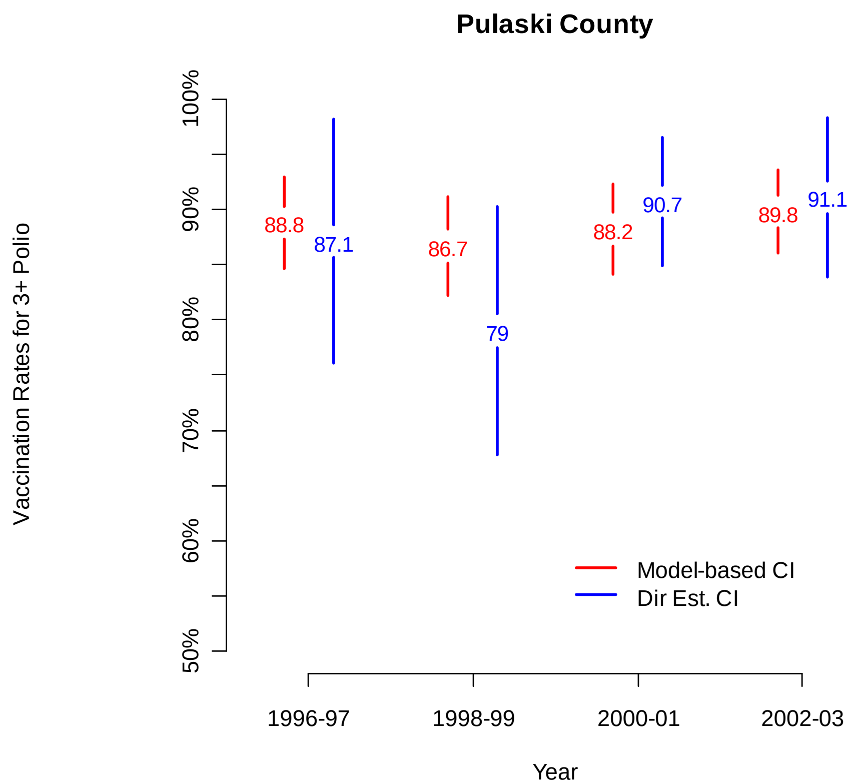

Table 4: Distributional statistics for county-level two-year estimates over time (184 counties with estimates for all two year periods from 1996-97 to 2002-03) [Data source: National Immunization Survey 1996-2007].

	Standard Deviation of the estimates across time (%)		Difference between the maximum and the minimum of the estimates across time (%)	
Percentile	Direct	Model-based	Direct	Model-based
5th	1.2	0.4	2.6	0.8
10th	1.5	0.4	3.2	0.9
25th	2.1	0.6	4.7	1.4
Median	3.3	0.9	7.2	2.0
75th	4.4	1.2	9.7	2.6
90th	5.9	1.6	12.7	3.5
95th	7.0	1.7	15.1	3.9

Appendix

Derivation of (3) in section 2.4.

In this section, we derive the best linear unbiased predictor for the parameter of interest, $\theta_{it} = \mu + \alpha_t + \beta_s + v_i + u_{it}$. First, we define some notation that is needed in the derivation of the BLUP.

Let $\gamma = (\mu, \alpha_2, ..., \alpha_T, \beta_1, ..., \beta_S)'$, where S is the number of states with two or more counties that have sufficient sample size, and T is the number of periods. Moreover,
$X' = (x_{11}, ..., x_{1n_1}, ..., x_{m1}, ..., x_{mn_m})$, where x_{it} is such that $x_{it}'\gamma = \mu + \alpha_t + \beta_s$,
$y = (y_{11}, ..., y_{1n_1}, ..., y_{m1}, ..., y_{mn_m})'$, $v = (v_1, ..., v_m)'$, $u = (u_{11}, ..., u_{1n_1}, ..., u_{m1}, ..., u_{mn_m})'$,
$\mathrm{var}(y) = V = diag(V_1, ..., V_m)$, where $V_i = \sigma_v^2 J_{n_i} + \sigma_u^2 I_{n_i} + \Psi_i$, $\Psi_i = diag(\psi_{i1}, ..., \psi_{in_i})$,
J_{n_i} is a $n_i \times n_i$ matrix of 1's, and I_{n_i} is a $n_i \times n_i$ identity matrix.

We derive the BLUP for a general $\theta = l'\gamma + k'v + q'u$, where $l \in \Re^p, k \in \Re^m$ and $q \in \Re^n$, where $p = \dim(\gamma) = S + T$ and $n = \sum_{i=1}^{m} n_i$. By appropriately choosing *l*, *k* and *q*, θ_{it} can be obtained as a special case of θ . Let the BLUP of θ be given by

$$\hat{\theta} = a'y + b,$$

where $a \in \Re^n$ and *b* is a scalar. Since $\hat{\theta}$ is unbiased for θ ,

$$l'\gamma = E(\theta) = E(a'y + b) = a'X\gamma + b,$$
$$\Rightarrow b = 0,\ l' = a'X.$$

Moreover,

$$\mathrm{var}(\hat{\theta} - \theta) = \mathrm{var}(a'y - l'\gamma - k'v - q'u)$$
$$= a'Va + \sigma_v^2 k'k + \sigma_u^2 q'q - 2\sigma_v^2 a'J_* k - 2\sigma_u^2 a'q,$$

where $J_* = diag(1_{n_1}, ..., 1_{n_m})$, 1_{n_i} is an $n_i \times 1$ vector of 1's.

Lagrange multipliers are used to minimize $\mathrm{var}(\hat{\theta} - \theta)$ subject to the constraint $l' = a'X$. That is, let

$$f = a'Va + \sigma_v^2 k'k + \sigma_u^2 q'q - 2\sigma_v^2 a'J_* k - 2\sigma_u^2 a'q + 2(a'X - l')\lambda.$$

Computing the gradient of f with respect to *a* and λ ,

$$\frac{\partial f}{\partial a} = 2Va - 2\sigma_v^2 J_* k - 2\sigma_u^2 q + 2X\lambda = 0$$
$$\frac{\partial f}{\partial \lambda} = 2(X'a - l) = 0$$

Solving the above equations, we obtain

$$\lambda = (X'V^{-1}X)^{-1}(\sigma_v^2 X'V^{-1}J_* k + \sigma_u^2 X'V^{-1}q - l)$$

$$a = -V^{-1}X(X'V^{-1}X)^{-1}(\sigma_v^2 X'V^{-1}J_* k + \sigma_u^2 X'V^{-1}q - l) + \sigma_v^2 V^{-1}J_* k + \sigma_u^2 V^{-1}q$$

Hence,

$$\tilde{\theta} = a'y = l'\tilde{\gamma} + \sigma_v^2 k' J_*' V^{-1}(y - X\tilde{\gamma}) + \sigma_u^2 q' V^{-1}(y - X\tilde{\gamma}),$$

where $\tilde{\gamma}$ is the best linear unbiased estimator of $\gamma = (\mu, \alpha_2, ..., \alpha_T, \beta_1, ..., \beta_S)'$, that is,

$$\tilde{\gamma} = (\tilde{\mu}, \tilde{\alpha}_2, ..., \tilde{\alpha}_T, \tilde{\beta}_1, ..., \tilde{\beta}_S)' = (X'V^{-1}X)^{-1}X'V^{-1}y.$$

By appropriately choosing l, k and q, and noting that $V^{-1} = diag(V_1^{-1}, ..., V_m^{-1})$, it follows that the BLUP of θ_{it} is given by

$$\tilde{\theta}_{it} = x_{it}'\tilde{\gamma} + \sigma_v^2 1_{n_i}' V_i^{-1}(y_i - X_i\tilde{\gamma}) + \sigma_u^2 q_{(t)}' V_i^{-1}(y_i - X_i\tilde{\gamma}),$$

where $X_i' = (x_{i1}, ..., x_{in_i})$, $y_i = (y_{i1}, ..., y_{in_i})'$ and $q_{(t)}$ is a $n_i \times 1$ vector with t^{th} entry equal to 1 and all other entries equal to 0. Furthermore, since

$$V_i^{-1} = D_i^{-1} - \sigma_v^2 \left[1 + \sigma_v^2 \sum_{j=1}^{n_i} (\sigma_u^2 + \psi_{ij})^{-1} \right]^{-1} D_i^{-1} J_{n_i} D_i^{-1},$$

$$D_i = diag(\sigma_u^2 + \psi_{i1}, ..., \sigma_u^2 + \psi_{in_i}),$$

after some simplification, it follows that

$$\tilde{\theta}_{it} = x_{it}'\tilde{\gamma} + \sigma_v^2 \left[1 + \sigma_v^2 \sum_{j=1}^{n_i} (\sigma_u^2 + \psi_{ij})^{-1} \right]^{-1} \sum_{j=1}^{n_i} \left(\frac{y_{ij} - x_{ij}'\tilde{\gamma}}{\sigma_u^2 + \psi_{ij}} \right) + \frac{\sigma_u^2}{(\sigma_u^2 + \psi_{it})}(y_{it} - x_{it}'\tilde{\gamma})$$

$$- \frac{\sigma_u^2}{(\sigma_u^2 + \psi_{it})} \sigma_v^2 \left[1 + \sigma_v^2 \sum_{j=1}^{n_i} (\sigma_u^2 + \psi_{ij})^{-1} \right]^{-1} \sum_{j=1}^{n_i} \left(\frac{y_{ij} - x_{ij}'\tilde{\gamma}}{\sigma_u^2 + \psi_{ij}} \right).$$

Further simplification and recombination of terms yields formula (3) in section 2.4.