1

2 **Vaccination Coverage in India in 2011: A Small Area Estimation Approach**

3

4

Santanu Pramanik^a

5

Nithiyananthan Muthusamy^a

6

Rajeev Gera^a

7

Ramanan Laxminarayan^{a,b,c, §}

8

9 ^aPublic Health Foundation of India, ISID Campus, 4 Institutional Area, Vasant Kunj, New
10 Delhi – 110070, India.

11 [§] Corresponding author, ramanan@phfi.org

12 ^b Princeton University, Princeton NJ 08544.

13 ^c Center for Disease Dynamics, Economics & Policy, 1616 P St NW, Suite 430, Washington
14 DC 20036.

15

16

17

18 **Manuscript:** 3,500 words **Abstract:** 260 words

19

20 **Keywords:** National rural health mission; Linear mixed model; Shrinkage estimates; Full

21 immunization; Measles; Vaccination coverage inequality

22 **Abstract**

23 Information on population health indicators in India come from a number of surveys that vary
24 in periodicity, scope and detail. In the case of immunization, the most recent coverage
25 indicators are derived from the first round of Annual Health Survey (AHS-1, 2010-11), but
26 these were conducted only in 9 of 35 states and union territories. The most recent national
27 surveys of immunization coverage were conducted in 2009 (Coverage Evaluation Survey) by
28 UNICEF. Therefore, reliable immunization coverage data for the entire country since 2009 is
29 lacking. We used an established approach of small area estimation to predict coverage rates
30 of several vaccinations for the remaining 26 states (not covered by AHS-1) in 2011. In our
31 method, we considered a linear mixed model that combines data from five cross sectional
32 surveys representing five different time points. Our model encompasses sampling error of the
33 survey estimates, area specific random effects, autocorrelated area by time random effects
34 and hence, borrows strength across areas and time points both. Model-based estimates for
35 2011 are almost identical to the AHS-1 estimates for the nine states, suggesting that our
36 model provides reliable prediction of vaccination coverage as AHS-1 estimates are highly
37 precise because of their large sample size. Results indicate that coverage inequality between
38 rural and urban areas has been reduced significantly for most states in India. The National
39 Rural Health Mission has had both supply side and demand side effects on the immunization
40 program in rural India. In combination, these effects may have contributed to the reduction of
41 vaccination coverage gaps between urban and rural areas.

42 **1. Introduction**

43 Despite a long standing national program for universal immunization in India (UIP, since
44 1985), only 61% of India's birth cohort (consists of approximately 27 million infants, the
45 largest in the world) is fully immunized.¹ In the context of UIP, full immunization implies
46 conforming to the EPI vaccination schedule Bacillus Calmette-Guérin (BCG), diphtheria-
47 pertussis-tetanus (DPT), oral polio, and measles. It is estimated that one in three incompletely
48 vaccinated children in the world is in India. Vaccination coverage rates are much lower than
49 that of its South Asian neighboring countries. For example, based on the latest demographic
50 and health surveys, the coverage rate of three doses of DPT vaccine in India (72%) lags rates
51 in Sri Lanka (99%) and Bangladesh (93%).

52 According to the Coverage Evaluation Survey (CES 2009),¹ the reasons for low
53 immunization coverage pertain to issues on the demand and supply side. Lack of parental
54 knowledge about the whole vaccine schedule is responsible for high proportion of incomplete
55 immunization while fear of side effects is one of the major reasons for no immunization.^{1, 2}
56 Poor populations and those with lower levels of education are most vulnerable to impacts of
57 low levels of advocacy and communication. Limited cold chain infrastructure and capacity in
58 many states,³ vaccine stock outs,⁴ significant delay in placement of procurement orders and
59 irregular supply of vaccines⁵- all these supply side factors also have a direct impact on
60 vaccine availability affecting immunization coverage.

61 Reliable and current vaccination coverage rates are useful in monitoring progress, measuring
62 program performance, estimating the risk that children face for vaccine preventable diseases
63 (VPD) and the risk of VPD outbreaks that can jeopardize the disease prevention goals,
64 understanding infectious disease dynamics, among others.

65 In India, population level immunization coverage information is derived from the many
66 national level health surveys, although these vary in frequency, scope and detail. The most
67 recent estimates are available from the first round of Annual Health Survey (AHS 2010-11),
68 which are confined only to nine high focus states (Bihar, Jharkhand, Uttar Pradesh,
69 Uttarakhand, Madhya Pradesh, Chhattisgarh, Orissa, Rajasthan and Assam). Although these
70 high focus states are important in that their health indicators lag national averages, it is also
71 important to have reliable immunization coverage estimates for the remaining 26 states and
72 union territories (UTs). The latest nation-wide data that covers all 35 states and UTs are from
73 CES 2009. Since 2005, significant investments have been made in improving the
74 immunization program under the National Rural Health Mission (NRHM).⁶ However, in the
75 absence of precise and up-to-date data on immunization coverage, it is unclear if these
76 investments have improved immunization coverage.

77 We used a small area estimation⁷ approach to predict vaccination coverage rates for states
78 and UTs that are not covered by AHS-1 for 2011. To obtain small area estimates, we used a
79 linear mixed model that combines data from five cross sectional health surveys representing
80 five separate time periods. This type of model leads to considerable gain in efficiency in
81 estimating small area parameters,⁷⁻⁹ as compared to the estimates based on a single cross
82 sectional survey.

83 **2. Methods**

84 **2.1 Estimation Parameter and Geographical Scale**

85 We estimated coverage rates of the following vaccines: BCG (to prevent tuberculosis), DPT
86 (to prevent diphtheria, pertussis, tetanus- DPT3), OPV (to prevent poliomyelitis- Polio3), and
87 measles. The vaccination coverage rate is defined as percentage of children age 12-23 months
88 who received the full complement of vaccines, irrespective of their age at vaccination (at any

time before the survey) and source of information (vaccination card or mother's self-report). For most routine immunization coverage surveys WHO recommends using children aged 12–23 months if final primary immunization (measles, in the context of EPI) is at 9 months of age.¹⁰ In addition to these individual vaccines, we estimated the proportion of fully immunized children who received all EPI vaccines.

India's national level immunization coverage indicators mask considerable state level variations, as well as persistent rural-urban differences. Hence, we define the unit of analysis (the geographical level of estimation) as state (and UTs)^a by rural-urban, which results in $35 \times 2 = 70$ small areas or domains. We use the terms domain and small area interchangeably.

2.2 Data Sources

We fitted our model using data from five cross sectional health surveys: DLHS-2, NFHS-3, DLHS-3, CES-09, and AHS-1. In Table 1 we present specific details about the surveys. Direct survey estimates of vaccination coverage rates were obtained from the published reports of the surveys.^{1, 11-13} Sampling errors, as measured by standard errors (SEs), associated with the survey estimates and the corresponding design effects were available at the small area level for all vaccines from NFHS-3 report.¹² However, for the other four surveys, SEs were not readily available. We calculated SEs by multiplying the simple random sampling (SRS) SE by the NFHS-3 design effects to account for the inflation in SE due to cluster sampling. This is justified by the similarity in sampling design across the five surveys. Note that reported NFHS-3 design effects were different across small areas and vaccination coverage rates, hence, we account for the differential clustering effect across small areas and vaccination type. SRS sampling variances ($SE = \text{square root of sampling variance}$) were

^a A Union territory (UT) is a type of administrative division in India, similar to that of states. UTs include both urban and rural areas in them. Unlike states, which have their own elected governments, union territories are ruled directly by the Union Government (Central Government), hence the name union territory. UTs are much smaller in size compared to most states of India.

obtained using the formula $\hat{p}(1 - \hat{p})/n$; where $\hat{p}$ is the coverage rate and n is the number of 12-23 months children surveyed. Both $\hat{p}$ and n were available for all vaccines and small areas across surveys.

NFHS-3 covered all 29 states but not the union territories. To obtain the SEs for 6 UTs, we used DLHS-3 survey that reports SEs for BCG and measles vaccination coverage.¹⁴ For UTs, design effects were calculated from BCG coverage rate sampling errors (except for Lakshadweep, for which measles was considered) and used for all other vaccinations across surveys (DLHS-2, 3 and CES-2009). For a particular UT, same design effect is used for rural and urban areas.

Survey estimates were available for all five time points for some, but not all, domains. For example, DLHS-3 does not cover the state Nagaland. Hence, for Nagaland-urban and Nagaland-rural domain we have data from 3 time points (DLHS-2, NFHS-3, and CES-2009). See Table 1 for further details on the survey coverage. For model fitting, it is not essential to have equal number of data points corresponding to all domains (see Section 2.5).

2.3 Small Area Model

We used the following area level model:

$$y_{it} = \mu + \alpha_t + \beta_s + \gamma_a + h'_{it}\lambda + v_i + u_{it} + e_{it}, \text{ where} \quad (1)$$

- y_{it} is the vaccination coverage estimate (or some transformed version of it to ensure normality) corresponding to the i th small area (defined as state by rural-urban) and t th time point; $t = 1, \dots, T_i$ (total number of time points corresponding to area i) and $i = 1, \dots, m$ (total number of areas: $35 \times 2 = 70$).
- μ is the common mean effect.
- α_t is the fixed effect due to the t th time point.

133 ▪ β_s is the fixed effect due to the s th geographical location, $s = 1, \dots, S$ (total number of
 134 unique locations). Location can be state or a broader region where the state belongs.
 135 ▪ γ_a is the fixed effect due to the type of residence, $a = \text{rural or urban}$.
 136 ▪ h'_{it} is the vector of area specific fixed covariates which may change with time.
 137 ▪ v_i is the area specific random effect which captures the variability not explained by
 138 the fixed effects; $v_i \sim \text{iid } N(0, \sigma_v^2)$.
 139 ▪ u_{it} is the area by time random effect to capture the spatio-temporal interactions. We
 140 assume that u_{it} 's are correlated across time for each i . First order autoregressive
 141 process is commonly used to incorporate such correlations:
 142 $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, $|\rho| < 1$; $\varepsilon_{it} \sim \text{iid } N(0, \sigma_\varepsilon^2)$ and independent of v_i . Note that when
 143 ρ takes on the value 1, u_{it} follows a random walk model. If true value of ρ is close to
 144 1, estimation of ρ could be problematic. It might fall outside the parameter space.^{8,9}
 145 Random walk model for the area by time interaction term is a safer choice from
 146 model fitting perspective.
 147 ▪ e_{it} is the sampling error of vaccination coverage estimate y_{it} ; where $e_{it} \sim \text{ind } N(0,$
 148 $\psi_{it})$, ψ_{it} is the (known) sampling variance of y_{it} . The sampling errors are
 149 independent of v_i and ε_{it} .
 150 True vaccination coverage (our parameter of interest) can be expressed as $\theta_{it} = \mu + \alpha_t +$
 151 $\beta_s + \gamma_a + h'_{it}\lambda + v_i + u_{it}$, $t = 1, \dots, T_i$; $i = 1, \dots, m$. These models, as described in (1), have
 152 been extensively used in econometrics.¹⁵⁻¹⁷ Anderson and Hsiao (1982) have formulated
 153 dynamic models with time invariant and time varying covariates and studied asymptotic
 154 properties of various estimators under different scenarios.¹⁶ Those models do not include the
 155 sampling error term as unit level observations are usually modeled. In the context of area-
 156 level small area modelling, it is important to incorporate the uncertainty of survey estimates.⁸

2.4 Choice of Covariates

Several covariates are known to be associated with vaccination coverage across vaccine types. Previous analysis based on three rounds of NFHS undertaken between 1992 and 2006 shows considerable variations in child immunization coverage across six geographical regions in India.¹⁸ Despite a decline in urban-rural and gender differences over time, children residing in rural areas and girls were found to have significantly lower coverage.¹⁸⁻²¹ Moreover, birth order of child, mother's education, caste, religion and wealth index of the household represent an influence on vaccination coverage.²² In addition to the above mentioned demand side predictors,²³⁻²⁶ supply side covariates such as access to health care facilities and other infrastructure also affect vaccination coverage.^{3, 24, 25}

Reliable area level estimates of some of these covariates were either unavailable or too resource intensive to obtain from all five time points. In our initial model, we included three potentially time varying predictors that were available across all time points. These small area level predictors are child sex ratio (to account for the boy-girl differential coverage), proportion of female illiteracy (proxy for mother's education), and proportion of tribal population (accounting for the effect of caste and hard to reach population). Based on ecological correlation (see Appendix A.1) and ordinary regression models, we identified female illiteracy as the only significant predictor and included that in the final small area model.

2.5 Model-based Estimates

For the ease of presentation, we used the notation T to denote the number of time points corresponding to a small area for which data are available. Ideally, the notation would be T_i for the i th small area and T_i 's vary across small areas. The best linear unbiased predictor

180 (BLUP) of θ_{iT} (vaccination coverage of the i th area in 2011), as follows from the general
 181 theory of linear mixed model (Rao 2003, Chapter 6),⁷ is given by

$$\hat{\theta}_{iT} = w_{iT}y_{iT} + (1 - w_{iT})x'_{iT}\tilde{\beta} + \sum_{t=1}^{T-1} w_{it}(y_{it} - x'_{it}\tilde{\beta}) \quad (2)$$

182 The estimator $\hat{\theta}_{iT}$ is also the best predictor (BP) of θ_{iT} , under the assumption of normality
 183 (which we have assumed). By “best” (both in BLUP and BP) we mean that the estimator has
 184 the smallest mean square error. It is interesting to note that the BLUP estimator of θ_{iT} , as
 185 given in equation (2), is a weighted sum of the direct survey estimator y_{iT} at the recent time
 186 point, the regression synthetic estimator $x'_{iT}\tilde{\beta}$ at the recent time point and the prediction
 187 errors $y_{it} - x'_{it}\tilde{\beta}$ for previous time points ($t = 1, \dots, T - 1$). The weights are governed by the
 188 relative precision of direct survey estimators and regression synthetic estimators, and also by
 189 the proximity of the survey estimates relative to the current time point. Further details on the
 190 derivation of $\hat{\theta}_{iT}$ and definition of different terms are available in Appendix A.2.

191 For small areas for which no recent direct estimates are available from AHS-1, the estimator
 192 is $\hat{\theta}_{iT} = x'_{iT}\tilde{\beta} + \sum_{t=1}^{T-1} w_{it}(y_{it} - x'_{it}\tilde{\beta})$. The BLUP estimator of θ_{iT} involves unknown
 193 variance components $\sigma = (\sigma_v^2, \sigma_\varepsilon^2)$. We estimated σ using maximum likelihood method
 194 (Appendix A.3). Estimates of σ_v^2 and σ_ε^2 were inserted in equation (2) to obtain the empirical
 195 best linear unbiased predictor (EBLUP) of θ_{iT} .

196 **3. Results**

197 We present results after fitting the model to data from five cross sectional surveys. Besides
 198 the surveys, we also used 2001 and 2011 census data to obtain covariates (Appendix A.1).
 199 We analyzed data using the open source statistical software R.²⁷

3.1 Estimates of Hyperparameters β and σ

Table 2 shows estimates of regression coefficients and variance components. From the regression coefficient estimates, we identified several interesting trends regarding vaccination coverage in India. Relative to the southern states, in general, other region states are performing worse (look at the negative coefficients in Table 2; north-east, east and central states are the worst). For measles and full vaccination coverage, western states are slightly better off than that of south. Rural area coverage rates are lower for all types of vaccination, except for BCG where the regression coefficient estimate is almost zero.

In terms of time trend, coverage of 3 doses of polio vaccines has increased significantly during 2005-06 (NFHS-3, T2) as compared to the base period (DLHS-2, T1). For other vaccinations, such rapid rise has not been observed within such a small time span. Inclusion of Pulse Polio doses (part of supplementary immunization activities) in NFHS-3 perhaps explain the rapid increase of polio vaccination coverage. All surveys specifically exclude Pulse Polio doses while asking about polio vaccination, except NFHS-3 (2005-06). Compared to the base period (2002-04), coverage has increased for all types of vaccination at the current time (2011, T5). However, the rise is not steady. In general, relative to the base period, coverage rates increase until DLHS-3 (2007-08) and then decrease during CES 2009 and again increase during AHS-1 (2010-11).

3.2 EBLUP Estimates

Model Evaluation

Using the estimated hyperparameters, we obtained the EBLUP estimates [Eq. (2)] of vaccination coverage rates for all small areas in 2011 across vaccines. We evaluated our model by comparing EBLUP estimates with AHS-1 (2010-11) estimates for 9 high focus states. Full vaccination coverage EBLUP estimates were almost identical to the 2011 AHS

estimates both for rural and urban areas of high focus states. For an explanation of the similarity between AHS and EBLUP estimates, see Appendix B.

Comparing 2011 EBLUP estimates with 2009 CES estimates: Full vaccination coverage

We compared current EBLUP estimates with the CES 2009 estimates at the state by rural-urban level. In Fig. 1 and Fig. 2, we compare urban and rural estimates of full vaccination coverage, respectively. In the plots, high focus states for which AHS estimates are almost identical to EBLUP estimates (because of the reason mentioned in Appendix B) are labelled in green colour. For state by urban small areas (Fig. 1), full vaccination coverage rates increase for all areas relative to CES09 estimates, with the exception of Uttar Pradesh, Odisha, and Jharkhand. Similarly, for state by rural small areas, the exceptional states are Odisha and Assam. Maximum improvements are observed in urban areas of Assam, Bihar and Rajasthan and in rural areas of Madhya Pradesh, Bihar, Chhattisgarh and Rajasthan.

Trend in measles vaccination coverage

We illustrate the trend in measles vaccination coverage in urban and rural areas of six states representing six different regions - Andhra Pradesh (south), Gujarat (west), Haryana (north), Madhya Pradesh (central), Manipur (north-east), and West Bengal (east). In all six states, initial rural area coverage rates in 2002-04 were lower than that of urban areas. But with time, coverage inequality diminishes and for some states (Andhra Pradesh, Gujarat, and West Bengal) rural area coverage exceeds the urban area coverage. Similar trends with respect to the diminishing coverage gap were observed for other vaccinations.

We illustrate this further using a bar plot in Fig. 4 exhibiting percentage differences between urban and rural area measles coverage during 2007-08 and 2011. Coverage inequality between urban and rural areas has declined in 2011 as compared to 2007-08 (DLHS-3); blue bars are smaller than the red bars for most states. Maximum reduction in inequality is

observed in Meghalaya, Tripura, Gujarat, Madhya Pradesh, Jharkhand, Jammu & Kashmir, Haryana and Andhra Pradesh. Substantial reduction in coverage gap is also noted in union territories such as Chandigarh, Dadra & Nagar Haveli and Andaman & Nicobar Islands.

4. Discussion

Vaccination coverage estimates predicted for the year 2011 indicate that the urban-rural coverage inequality has decreased significantly in many states. UNICEF, WHO, and USAID through its grant for MCHIP (Maternal and Child Health Integrated Program) have provided technical support to the UIP in India in the areas of cold chain and vaccine management, communications, and program monitoring and supervision. These have led to improvements in the capacity of health workers, and the strengthening of immunization service delivery in parts of India where these organizations are active. However, the Ministry of Health's flagship program NRHM,²⁸ launched in 2005, has been the most significant structural change in India's health system in general, and for the UIP in particular, resulting in significant improvements in immunization service delivery across India. The NRHM has had both supply side and demand side effects on the immunization program in rural India.

On the supply side, the two primary NRHM interventions to improve immunization performance are the hiring of additional Auxiliary Nurse Midwives (ANMs) and the introduction of the Alternate Vaccine Delivery (AVD) system. CES 2009 identified that majority of Indian children receive vaccinations from ANMs in outreach sessions either in, or close to, their villages.¹ Anand et al.²⁹ concluded that the density of health workers, and in particular nurses, in developing countries affects the availability of vaccination services over time and space, thereby being a key determining factor in the likelihood of children being vaccinated. After exploring rural health statistics (<https://nrhm-mis.nic.in/SitePages/Home.aspx>) and the yearly population projections based on 2001

272 census,³⁰ we found an increasing trend in the density of ANMs since 2007 in most of Indian
273 states.

274 The primary means by which the NRHM aims to improve demand for immunization services
275 is through the social mobilization and awareness-raising efforts of a cadre of health workers -
276 Accredited Social Health Activists (ASHA) – and the provision for performance linked
277 incentives for vaccination. According to NRHM Report (2012),³¹ in most Indian states, the
278 required number of ASHA as per the 2001 population estimate are already in place. Several
279 assessment studies in rural India found that ASHA workers actively carry out their
280 immunization mobilization responsibilities and serve as effective links between the local
281 community and healthcare providers.^{4, 32-34}

282 The NRHM, primarily through the interventions mentioned above, may have contributed
283 significantly to the improvement in immunization coverage in rural areas. However, urban
284 areas have not received the required attention and resources during the same period, despite a
285 growing urban population and the related complexity of immunization challenges. To
286 achieve a high national full immunization coverage rate, it is critical that urban and peri-
287 urban areas are reached in a comprehensive manner to provide all services planned under the
288 integrated reproductive, maternal, newborn, child and adolescent health (RMNCH+A)
289 approach, and this could be possible under the proposed National Urban Health Mission
290 (NUHM) established alongside the NRHM.

291 There remain concerns about the method. Immunization coverage data based on maternal
292 recall can be inaccurate. Proportions of data from maternal recall vary by survey (hence by
293 time) and geography (Table 1). At the national level the percentage varies from 69% (DLHS-
294 2, 2002-04) to 48.5% (CES, 2009). This might have differential effects on the reliability of
295 vaccination coverage rates across surveys and small areas. We have not captured this

296 unknown effect in our model. We could not include many predictors of immunization
297 coverage, particularly the supply side ones, as they were either unavailable or too resource
298 intensive to obtain from all five time points. The logit transformation considered here makes
299 the data look approximately normally distributed. However, the normal Q-Q plots indicate
300 that the tail distribution does not match closely for some vaccination coverage estimates such
301 as BCG and Measles. This happens as some areas have extremely good coverage relative to
302 the rest of the areas. These outlying areas can be thought of coming from another normal
303 distribution with different parameters. Scale mixture of two normal distributions might be a
304 better alternative under such circumstances.³⁵ While combining different sources of
305 information, our model-based estimates incorporate sampling error of survey estimate.
306 However, reliability of survey estimates also depends on the magnitude of non-sampling
307 errors, such as measurement error, nonresponse error, which may vary across surveys. Our
308 model ignores non-sampling error components.

309 Effectiveness of health interventions can only be judged upon availability of updated
310 information on key indicators which, in turn, are crucial to better target these interventions. In
311 this paper, we demonstrate a statistical approach to combining previous time-point estimates
312 and determinants of outcome to predict indicators at the current time. The method could be
313 useful for other countries that lack precise yearly health indicators. Even if health indicators
314 are available for a particular year- from a single survey- it may not always be wise to rely on
315 one particular source given the limitations of data quality in these countries. Our model
316 combines information from various sources in an objective way to come up with reliable
317 estimates. The methodology strictly involved use of only publicly available data, among other
318 considerations, with the express intent of attempting to insure others could both access that
319 data while also replicating this model. After the application of the model, our findings reveal
320 significant reduction in coverage inequality between rural and urban areas. We discussed the

role of NRHM interventions as potential reasons behind the narrowing of coverage gap. Although change in coverage gap may be due to change in covariates (e.g., female literacy rate) and may not be due to change in program performance.

Acknowledgements: The authors would like to thank the reviewers for their useful comments which have led to substantial improvement of the manuscript. We acknowledge help from Moutushi Majumder and Melissa Schnure in extracting data from several survey reports and compiling census 2001 and 2011 data. We are thankful to Partha Lahiri (University of Maryland) and Jessica Metcalf (Princeton University) for useful comments and discussion. Part of R codes was updated based on codes written by N. Ganesh during a collaborative project with the first author at NORC at the University of Chicago. The work was funded by a grant from the Bill & Melinda Gates Foundation to support the Immunization Technical Support Unit at the Public Health Foundation of India, and a Health Grand Challenges grant from Princeton University. The funders had no role in the design of the study or interpretation of the results.

References

1. UNICEF India. Coverage Evaluation Survey (CES), 2009: All India report. 2010: Accessed on Oct 10, 2013 at http://www.unicef.org/india/1_CES_09_All_India_Report.pdf.
2. Vashishtha VM. Status of immunization and need for intensification of routine immunization in India. *Indian pediatrics* 2012; **49**(5): 357-61.
3. Datar A, Mukherji A, Sood N. Health infrastructure & immunization coverage in rural India. *Indian Journal of Medical Research* 2007; **125**(1): 31.
4. Prinja S, Gupta M, Singh A, Kumar R. Effectiveness of planning and management interventions for improving age-appropriate immunization in rural India. *Bulletin of the World Health Organization* 2010; **88**(2): 97-103.
5. Agarwal R. Routine Immunization: India's Achilles' heel! *Indian Pediatrics* 2008; **45**: 625:8.
6. Planning Commission. WG1 : Progress and Performance of National Rural Health Mission (NRHM) and suggestions, Report of the Working Group on National Rural Health Mission (NRHM) for the Twelfth Five Year Plan (2012-2017). New Delhi: Government of India, Planning Commission, 2011.
7. Rao JN. Small area estimation: Wiley; 2003.
8. Rao J, Yu M. Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics* 1994; **22**(4): 511-28.
9. Datta G, Lahiri P, Maiti T. Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference* 2002; **102**(1): 83-97.
10. Organization WH. Immunization coverage cluster survey: Reference manual. 2005; Available at http://whqlibdoc.who.int/hq/2005/who_ivb_04.23.pdf.

- 363 11. Office of the Registrar General & Census Commissioner. Annual Health Survey
364 (AHS-1), 2010-11: State Fact Sheets. New Delhi, India. 2012: Accessed on Oct 10, 2013 at
365 http://www.censusindia.gov.in/vital_statistics/AHSBulletins/Factsheets.html.
- 366 12. International Institute for Population Sciences. National Family Health Survey
367 (NFHS-3), 2005-06: State level reports, Mumbai: IIPS. 2008: Accessed on Oct 10, 2013 at
368 <http://www.rchiips.org/NFHS/report.shtml>.
- 369 13. International Institute for Population Sciences. District Level Household and Facility
370 Survey (DLHS-3), 2007-08: Key Indicators: States and Districts (comparison with DLHS-2,
371 2002-04), Mumbai: IIPS. 2010: Accessed on Oct 10, 2013 at
372 http://www.rchiips.org/pdf/DLHS-3_KI.pdf.
- 373 14. International Institute for Population Sciences. District Level Household and Facility
374 Survey (DLHS-3), 2007-08: India report, IIPS: Mumbai. 2010: Accessed on Oct 10, 2013 at
375 http://www.rchiips.org/pdf/INDIA_REPORT_DLHS-3.pdf.
- 376 15. Anderson TW, Hsiao C. Estimation of dynamic models with error components.
377 *Journal of the American statistical Association* 1981; **76**(375): 598-606.
- 378 16. Anderson TW, Hsiao C. Formulation and estimation of dynamic models using panel
379 data. *Journal of econometrics* 1982; **18**(1): 47-82.
- 380 17. Judge GG, Griffiths WE, Hill RC, Lütkepohl H, Lee T-C. The theory and practice of
381 econometrics: Wiley New York; 1980.
- 382 18. Singh PK. Trends in Child Immunization across Geographical Regions in India:
383 Focus on Urban-Rural and Gender Differentials. *PloS one* 2013; **8**(9): e73102.
- 384 19. Corsi D, Bassani D, Kumar R, et al. Gender inequity and age-appropriate
385 immunization coverage in India from 1992 to 2006. *BMC International Health and Human*
386 *Rights* 2009; **9**(Suppl 1): S3.

- 387 20. Prusty RK, Kumar A. Socioeconomic Dynamics of Gender Disparity in Childhood
388 Immunization in India, 1992–2006. *PloS one* 2014; **9**(8): e104598.
- 389 21. Singh A. Gender based within-household inequality in childhood immunization in
390 India: changes over time and across regions. *PloS one* 2012; **7**(4): e35045.
- 391 22. International Institute for Population Sciences. National Family Health Survey
392 (NFHS-3), 2005-06: India report, Mumbai: IIPS. 2007: Accessed on Oct 10, 2013 at
393 http://www.rchiips.org/NFHS/volume_1.shtml.
- 394 23. Patra N. Universal immunization programme in India: the determinants of childhood
395 immunization. 2006: Available at SSRN 881224.
- 396 24. Bhargava A, Guntupalli AM, Lokshin M, Howard LL. Modeling the Effects of
397 Immunizations Timing on Child Health Outcomes in India. *Health economics* 2013.
- 398 25. Mathew JL. Inequity in childhood immunization in India: a systematic review. *Indian*
399 *pediatrics* 2012; **49**(3): 203-23.
- 400 26. Vikram K, Vanneman R, Desai S. Linkages between maternal education and
401 childhood immunization in India. *Social Science & Medicine* 2012; **75**(2): 331-9.
- 402 27. R Core Team. R: A language and Environment for Statistical Computing. *R*
403 *Foundation for Statistical Computing, Vienna, Austria* 2012; **ISBN 3-900051-07-0**, URL
404 <http://www.R-project.org/>.
- 405 28. NRHM. National Rural Health Mission 2005-2012, Mission document *Ministry of*
406 *Health and Family welfare, Government of India* 2005: Accessed on Oct 10, 2013 at
407 http://www.nird.org.in/brgf/doc/Rural%20HealthMission_Document.pdf.
- 408 29. Anand S, Bärnighausen T. Health workers and vaccination coverage in developing
409 countries: an econometric analysis. *The Lancet* 2007; **369**(9569): 1277-85.

- 410 30. Office of the Registrar General of India. Population Projection for India and States
411 2001–2026. *Office of the Registrar General and Census Commissioner of India, New Delhi*,
412 2006 2006.
- 413 31. NRHM. Update on the ASHA programme. *NRHM report January 2012* 2012:
414 Accessed on Oct 10, 2013 at
415 [http://nhsrindia.org/pdf_files/resources_thematic/Community_Participation/NHSRC_Contri](http://nhsrindia.org/pdf_files/resources_thematic/Community_Participation/NHSRC_Contribution/Update%20on%20the%20ASHA%20Progra_477.pdf)
416 [bution/Update%20on%20the%20ASHA%20Progra_477.pdf](http://nhsrindia.org/pdf_files/resources_thematic/Community_Participation/NHSRC_Contribution/Update%20on%20the%20ASHA%20Progra_477.pdf).
- 417 32. BRD Medical College Gorakhpur. A Study of Interface of ASHA with the
418 Community and the Service Providers in Eastern Uttar Pradesh. 2009: Accessed on 10 Oct,
419 2013 at <http://www.nihfw.org/pdf/RAHI-II%20Reports/GORAKHPUR.pdf>.
- 420 33. Bhatnagar R, Singh K, Bir T, Datta U, Raj S, Nandan D. An assessment of
421 performance based incentive system for ASHA Sahyogini in Udaipur, Rajasthan. *Indian*
422 *journal of public health* 2008; **53**(3): 166-70.
- 423 34. Pala S, Kumar D, Jeyashree K, Singh A. Preliminary evaluation of the ASHA scheme
424 in Naraingarh block, Haryana. *The National medical journal of India* 2011; **24**(5): 315.
- 425 35. Gershunskaya J, Lahiri P. Robust Small Area Estimation Using a Mixture Model.
426 Proceedings of the Survey Research Methods Section. Alexandria, VA: American Statistical
427 Association; 2010.
- 428 36. Ybarra LM, Lohr SL. Small area estimation when auxiliary information is measured
429 with error. *Biometrika* 2008; **95**(4): 919-31.
- 430 37. Harville DA. Bayesian inference for variance components using only error contrasts.
431 *Biometrika* 1974; **61**(2): 383-5.
- 432 38. Harville DA. Maximum likelihood approaches to variance component estimation and
433 to related problems. *Journal of the American Statistical Association* 1977; **72**(358): 320-38.

434 39. Jiang J. REML estimation: Asymptotic behavior and related topics. *The Annals of*
435 *Statistics* 1996; **24**(1): 255-86.

436 40. Patterson HD, Thompson R. Recovery of inter-block information when block sizes
437 are unequal. *Biometrika* 1971; **58**(3): 545-54.

438 41. Wolter KM. Introduction to variance estimation: Springer; 2007.

439 42. Ganesh N, Smith PJ, Copeland KR, Wolter KM. Modeling County-Level Vaccination
440 Coverage Rates Across Time. Proceedings of the Survey Research Methods Section.
441 Alexandria, VA: American Statistical Association; 2009.

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

Tables and Figures

Table 1 Summary of different cross sectional health surveys used to predict the vaccination coverage rates at the current time

Name of the Survey and Year	Sampling Design	Responsible Office	Nodal Agency	Coverage of the Survey (representativeness)	Eligibility criteria for child selection**	Vaccination card seen (%) [†]
District Level Household Survey (DLHS-2) 2002-04	Multistage stratified systematic sampling design	MoHFW [§]	IIPS*	All 35 states & UTs 593 districts Currently married women age 15-44 and their husbands	Last and last but one living children born to selected women in the 3 years preceding the survey	India- 31 Urban- 39.8 Rural- 27.7 Rajasthan- 16 Goa- 69.8
National Family Health Survey (NFHS-3) 2005-06	Multistage stratified PPS sampling design	MoHFW	IIPS	All 29 states (but no UTs) Ever-married women (age 15-49), men (15-54) and never married women (15-49)	All living children born to selected ever-married women in 5 years preceding the survey	India- 37.5 Urban- 46.2 Rural- 34.5 UP- 20.3 Kerala- 75.3
District Level Household and Facility Survey (DLHS-3) 2007-08	Multistage stratified systematic PPS sampling design	MoHFW	IIPS	All states (except Nagaland) & UTs 601 districts Ever-married women (15-49), unmarried women (15-24)	Youngest living child born to selected ever-married women since 01.01.2004	India- 42.7 Urban- 47.7 Rural- 40.8 Rajasthan- 24 WB- 81.5
Coverage Evaluation Survey (CES-2009) 2009-10	Multi-stage stratified design	UNICEF	ORG Centre for Social Research, New Delhi	All state and UTs (eligible households: at least one child of age 12-23 months or at least one woman who delivered a baby in the previous 12 months)	Children aged between 12-23 months	India- 51.5 Urban- 58.0 Rural- 48.9 Rajasthan- 24 Sikkim- 85.2
Annual Health Survey (AHS-1) 2010-11	Single-stage stratified SRS, except in larger villages (two stage stratified)	MoHFW	Office of the Registrar General, India	9 high focus states 284 districts 48% of the total population Ever-married women (15-49)	Last and last but one living children born to selected ever-married women in the 3 years preceding the survey	Not Applicable UP- 68.2 Odisha- 94

[§]Ministry of Health and Family Welfare (MoHFW), Government of India

* International Institute for Population Sciences (IIPS), Mumbai

**Except for CES 2009, all other surveys collect immunization related information from more children (as specified in the last column of Table 1) than actually used to calculate the coverage rate. To estimate coverage rate, all surveys only used 12-23 month age group children (relative to survey date). This minimizes the time overlap between surveys as far as vaccination coverage rate is concerned.

[†] For the remaining children, vaccination coverage status was obtained from mother's recall. This column provides percentage of children (12-23) with a vaccination card seen for India and separately by urban and rural. It also includes minimum and maximum percentages across states.

Table 2 Regression coefficients and variance component estimates (along with standard errors in smaller font) for different vaccination coverage model

Regression coefficients	BCG		DPT3		Polio3		Measles		Full	
Intercept	4.98	0.74	3.14	0.52	2.78	0.57	3.14	0.60	2.26	0.50
<i>Region</i> (reference South)										
North-East	-2.08	0.48	-2.24	0.35	-2.35	0.37	-1.82	0.38	-2.30	0.36
East	-1.44	0.68	-1.42	0.5	-1.35	0.53	-1.37	0.56	-1.31	0.52
Central	-1.31	0.76	-1.48	0.56	-1.28	0.59	-1.15	0.62	-1.22	0.58
North	-1.04	0.49	-0.85	0.37	-0.94	0.38	-0.87	0.41	-0.81	0.38
West	-0.40	0.53	0	0.38	-0.47	0.41	0.14	0.42	0.09	0.39
Islands	-0.55	0.79	-0.79	0.53	-2.35	0.64	-0.19	0.58	-0.99	0.54
rural	-0.004	0.38	-0.30	0.27	-0.23	0.29	-0.29	0.31	-0.38	0.27
<i>Time</i> (reference T1: DLHS2)										
T2 (NFHS-3)	-0.16	0.16	-0.08	0.11	0.68	0.13	0.003	0.14	0.05	0.09
T3 (DLHS-3)	0.72	0.20	0.27	0.13	0.48	0.15	0.70	0.16	0.41	0.11
T4 (CES09)	-0.26	0.24	0.19	0.15	0.31	0.19	0.30	0.19	0.40	0.13
T5 (AHS-1)	0.73	0.39	0.52	0.26	0.69	0.31	0.82	0.32	0.64	0.22
UT or not	1.44	0.49	-	-	1.09	0.39	-	-	-	-
High-focus state or not	-0.82	0.52	-0.58	0.39	-0.57	0.41	-0.71	0.43	-0.58	0.40
Female Illiteracy	-0.26	0.14	-0.13	0.10	-0.11	0.11	-0.16	0.11	-0.11	0.09
$\hat{\sigma}_v^2$	0	-	0.18	0.13	0	-	0	-	0.39	0.14
$\hat{\sigma}_\varepsilon^2$	1.29	0.13	0.53	0.06	0.78	0.08	0.88	0.09	0.37	0.04

Fig. 1 Comparison between CES 2009 full immunization coverage estimates and predicted estimates at the current time point (2011) for small areas defined by state x

urban: arranged by increasing CES 2009 urban area estimates. High focus states are labeled in green colour.

Fig. 2 Comparison between CES 2009 full immunization coverage estimates and predicted estimates at the current time point (2011) for small areas defined by state x rural: arranged by increasing CES 2009 rural area estimates. High focus states are labeled in green colour.

Fig. 3 Trends in measles vaccination coverage in urban and rural areas of six states representing six different regions of India

Fig. 4 Percentage difference [(urban-rural)×100] in measles vaccination coverage between urban and rural areas (arranged by increasing difference as in DLHS-3). Bars falling below the zero line suggest better rural area coverage for the corresponding states.

498 **Appendices (Supplementary Materials)**

499 **A. Technical Details on Methodology**

500 **A.1 Selection of Predictors**

501 In our model, to account for the location effect, we included region rather than state (35
502 levels) as a predictor. Inclusion of state led to an overfitted model and consequently, the
503 estimate of the small area variance component ($\hat{\sigma}_v^2$) turned out to be zero for all five models
504 (BCG, DPT3, Polio3, Measles, Full). The area specific random effect term was included to
505 explain the variability in vaccination coverage rates across areas not explained by the fixed
506 covariates. Moreover, for some states there are only six data points to estimate the state
507 effect. To avoid the overfitting problem, we grouped 35 states and UTs in 7 regions: South
508 (Andhra Pradesh, Karnataka, Kerala, Tamil Nadu, Puducherry), North-East (Sikkim,
509 Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Meghalaya, Assam), East (Bihar,
510 West Bengal, Jharkhand, Odisha), Central (Rajasthan, Uttar Pradesh, Chhattisgarh, Madhya
511 Pradesh), North (Jammu & Kashmir, Himachal Pradesh, Punjab, Chandigarh, Uttarakhand,
512 Haryana, Delhi), West (Gujarat, Maharashtra, Goa), and Islands (Andaman & Nicobar and
513 Lakshadweep). Even after considering region, instead of state, $\hat{\sigma}_v^2$ is zero for BCG, Polio3
514 and Measles model.

515 Past survey estimates suggest that states within a region are similar in terms of vaccination
516 coverage rates. Hence we think region can explain the geographical variation and this leads to
517 a more parsimonious model. Within a region, to differentiate the effect of progressive and
518 high-focus states we considered an indicator variable in our model. For example, in Eastern
519 region, historically, West Bengal had better health indicators as compared to Bihar,
520 Jharkhand and Odisha. Union territories seem to have better coverage rates compared to most
521 states in India. That effect was also included in the model, wherever it was declared as
522 significant (for BCG and Polio3) by the ordinary regression model.

523 Table 3 shows sources of three time varying predictors across different time points. For
524 CES09, covariates were unavailable and hence substituted by DLHS-3 estimates. Given the
525 small time gap between these two surveys, use of same values for the covariates is not
526 unreasonable. When the covariates are derived from census data, they are believed to be free
527 of sampling errors. However, that does not necessarily apply to the covariates estimated from
528 surveys. For sufficiently large sample size at the domain level, the sampling errors associated
529 with the covariates were usually small. In this paper, we ignored the sampling errors of
530 covariates. For small area estimation when covariates are measured with error, the readers are
531 encouraged to follow the approach outlined in Ybarra and Lohr.³⁶

532 **Table 3 Sources of predictors (child sex ratio, proportion of female illiteracy, and**
533 **proportion of tribal population) corresponding to different time points**

Time points	Source
DLHS-2 (2002-04)	Census 2001
NFHS-3 (2005-06)	NFHS-3

DLHS-3 (2007-08)	DLHS-3
CES-2009 (2009)	DLHS-3
AHS-1 (2010-11)	Census 2011

We studied the influence of these covariates on vaccination coverage rates for BCG, DPT3, Polio3, Measles, and Full immunization. After calculating the ecological correlation and fitting ordinary regression models we identified covariates that are significant predictors of vaccination coverage and then included those in the small area model. Table 4 shows that the correlation between female illiteracy (a proxy of mother's education in the area) and vaccination coverage is the highest for all types of vaccination, except for Polio3 for which Scheduled Tribe proportion has higher ecological correlation. The signs corresponding to all correlation seem to make sense intuitively; however, it's very weak for child sex ratio. Ordinary regression models identified female illiteracy as the only significant predictor, even for Polio3, after adjusting for location, area of residence, and time effect.

Table 4 Ecological correlation between vaccination coverage and covariates across vaccines: calculations are based on 280 data points for which vaccination coverage rates are available (ideally 70 small areas and 5 time points would lead to 350 observations, but the remaining ones are unavailable)

Covariates	BCG	DPT3	Polio3	Measles	Full Immunization
Child sex ratio	-0.11	-0.14	-0.16	-0.13	-0.15
Proportion of female illiteracy	-0.34	-0.43	-0.31	-0.40	-0.39
Proportion of Scheduled Tribe	-0.25	-0.32	-0.38	-0.27	-0.34

A.2 Estimating θ_{it}

We rewrite the linear mixed model (1) in the form of hierarchical models as follows:

- **Level 1:** $y_{it}|\theta_{it} \sim \text{ind } N(\theta_{it}, \psi_{it}); t = 1, \dots, T_i; i = 1, \dots, m.$
- **Level 2:** $\theta_{it}|u_{it} \sim \text{ind } N(x'_{it}\delta + u_{it}, \sigma_v^2);$ where $x'_{it}\delta = \mu + \alpha_t + \beta_s + \gamma_a + h'_{it}\lambda.$
- **Level 3:** $u_{it} = u_{it-1} + \varepsilon_{it}; \varepsilon_{it} \sim \text{iid } N(0, \sigma_\varepsilon^2)$

In level 1, θ_{it} is the true vaccination coverage for area i and time t ; $t = 1, \dots, T_i; i = 1, \dots, m.$ In level 2, $x'_{it}\delta$ is the combined fixed effect term which includes all covariates (p). For area i , if u_{it} 's were uncorrelated across time points, it would have been simpler to express the above models as a linear mixed normal model. In that case, unconditional variance of y_{it} would be the sum of variances of all random components and hence would have a block diagonal structure. However, because of the random walk model assumption, u_{it} 's are not independent. To specify the distribution of u_{it} , we derived the first two moments. After successive substitution, we expressed u_{it} as follows:

$$u_{it} = u_{i0} + \sum_{j=1}^t \varepsilon_{ij} \quad (3)$$

Following Datta et al.,⁹ we assumed that the initial condition u_{i0} is a fixed unknown quantity (say, u) as opposed to considering it to be random. No distribution was assumed for u_{i0} to avoid identifiability problem. For more on the identifiability issue, under a variety of assumptions, see Anderson and Hsiao.¹⁶ Now, from equation (3), we can write

$$\begin{aligned} E(u_{it}) &= u \\ V(u_{it}) &= t \cdot \sigma_\varepsilon^2 \\ \text{Cov}(u_{it}, u_{it'}) &= \begin{cases} t' \cdot \sigma_\varepsilon^2; & t' \leq t \\ t \cdot \sigma_\varepsilon^2; & t' > t \end{cases} \end{aligned} \quad (4)$$

Once the distributions of all random components were known up to second order moments, the estimates of θ_{iT} (vaccination coverage of the i th area in 2011) followed from the general theory of linear mixed model; see Rao (2003, Chapter 6).⁷

In equation (2) of Section 2.5, different quantities are defined as follows (using common notation T to denote the number of time points corresponding to a small area): The weights $w'_i = (w_{i1}, \dots, w_{iT}) = (\sigma_v^2 1_T + \sigma_\varepsilon^2 g_T)' \Sigma_i^{-1}$, where g_T is the T th column of $\Gamma\Gamma'$, Γ is a $T \times T$ lower triangular matrix with all diagonal and non-zero off-diagonal elements unity, 1_T is the $T \times 1$ vector of 1's and $\Sigma_i = \sigma_v^2 1_T 1_T' + \sigma_\varepsilon^2 \Gamma\Gamma' + \Psi_i$. The quantity Ψ_i is a $T \times T$ diagonal matrix with diagonal elements being (known) sampling variances ψ_{it} , $t = 1, \dots, T$. Unlike Datta et al.,⁹ our Ψ_i matrix takes a simpler (diagonal) form as we considered independent surveys across time. Datta et al. used data from the current population survey (CPS), which is a panel survey and hence they included all possible covariances between direct estimates from consecutive years. We express all fixed effects in the form of a $mT \times p$ matrix X , which can be explicitly written as $X = \text{Col}_{1 \leq i \leq m} X_i$, $X_i = \text{Col}_{1 \leq t \leq T} x'_{it}$; where x'_{it} is the vector of p covariates corresponding to the i th area and t th time point. To denote the regression coefficient we use a new notation β (a vector of order $p \times 1$) since the initial condition for the random walk model (u) is confounded with the common mean term of model (1). In other words, the intercept term includes the initial state of vaccination coverage, which is assumed to be same for all areas. $\tilde{\beta}$ is the weighted least square estimator (same as the maximum likelihood estimator) of β . Assuming that the variance components $\sigma = (\sigma_v^2, \sigma_\varepsilon^2)$ are known,

$$\tilde{\beta} = \tilde{\beta}(\sigma) = \left(\sum_{i=1}^m X_i' \Sigma_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^m X_i' \Sigma_i^{-1} y_i \right), \quad (5)$$

where $y'_i = (y_{i1}, \dots, y_{iT})$.

A.3 Estimation of the Variance Components

The BLUP estimator of θ_{iT} , as given in equation (2), involves unknown variance components $\sigma = (\sigma_v^2, \sigma_\varepsilon^2)$ that needs to be estimated. We considered likelihood based methods that maximize the likelihood function of σ to estimate it. Maximum likelihood (ML) and residual maximum likelihood (REML) are two widely used variance component estimation methods. In classical statistical inference, the REML method is preferred to ML as it takes into account

the loss of degrees of freedom resulting from the estimation of the model's fixed effects and hence reduces the downward bias of ML.³⁷⁻⁴⁰ The REML estimates of σ_v^2 and σ_ε^2 can be obtained by maximizing the residual log-likelihood function given below:

$$\ln RL = \text{const} - \frac{1}{2} \sum_{i=1}^m \left[\ln |X_i' \Sigma_i^{-1} X_i| + \ln |\Sigma_i| + (y_i - X_i \tilde{\beta})' \Sigma_i^{-1} (y_i - X_i \tilde{\beta}) \right] \quad (6)$$

For our model, the log-ML function is same as (6), except that the first term inside the summation ($\ln |X_i' \Sigma_i^{-1} X_i|$) is absent. To estimate the variance components we considered maximum likelihood method. The additional term in the REML function resulted in negative infinite value of the likelihood, hence REML was avoided.

A.4 Transformation of small area direct estimates

Fig. 5 An Illustrative plot using measles vaccination coverage estimates from five time points. Direct survey estimates in the original scale are negatively skewed and hence deviate from normal distribution. However, logit transformation of the direct estimates leads to approximate normality.

Prior to fitting the model, direct survey estimates (z_{it}) of vaccination coverage proportions were transformed to logit (log-odds) scale: $y_{it} = \log[z_{it}/(1 - z_{it})]$. The rationale behind this transformation is two-fold: first to ensure normality, at least approximately (see Fig. 5) and secondly to restrict the EBLUP estimates within the parameter space $[0, 1]$. For known sampling variance of z_{it} , we derived the variance of the transformed non-linear estimator y_{it} using Taylor series linearization technique (delta method, Wolter, Chapter 6).⁴¹ The variance of y_{it} is given by

$$\psi_{it} = \frac{1}{z_{it}^2(1 - z_{it})^2} [\text{SE}(z_{it})]^2 \quad (7)$$

Standard errors of y_{it} are governed by the direct estimates, which appear in the denominator of the expression (7). So for small areas having higher vaccination coverage rate, ψ_{it} 's are going to be high, irrespective of the sample size in that area. This is not desirable as such direct estimates would receive an extremely small weight in the small area estimate (hence suffer from overshrinkage problem). For these small areas, SEs were defined differently; following the usual formula of variance for proportion.⁴²⁴²⁴² For each vaccination coverage, small areas having proportion above the 75th percentile (q_3), we defined $\psi_{it} = [q_3(1 - q_3)/n_{it}] \times \text{deff}$, where n_{it} is the number of children surveyed from the i th area and t th time point and deff is the median design effect across areas. Note that q_3 and deff vary by vaccination type.

B. Model Evaluation

To explain the similarity between AHS and EBLUP estimates, we recall the expression of weights (see Appendix A.2) involved in the EBLUP equation (2). The weights are given by $w'_i = (w_{i1}, \dots, w_{iT}) = (\sigma_v^2 1_T + \sigma_\varepsilon^2 g_T)' \Sigma_i^{-1}$, where g_T is the T th column of $\Gamma\Gamma'$, Γ is a $T \times T$ lower triangular matrix with all diagonal and non-zero off-diagonal elements unity. In essence, $g_T = (1, 2, 3, 4, 5)$. This clearly illustrates that closer we move towards the recent time more weights are assigned to the direct survey estimates, with AHS receiving the highest weight. Also note that the weights are governed by the precision of direct survey estimates through Ψ_i (part of Σ_i). But AHS estimates are extremely precise, hence, Σ_i^{-1} does not contribute much towards reducing the corresponding weights. Standard errors of AHS estimates are much lower than that of any other survey estimates. Consequently, the distributions of weights are extremely skewed towards AHS (see the right panel of Table 5).

Table 5 Comparison between full vaccination coverage EBLUP estimates and AHS-1 survey estimates for the high focus states: Model Evaluation

State	Rural		Urban		Distribution of weights across time points for urban area EBLUP estimates					
	EBLUP	AHS	EBLUP	AHS	DLHS2	NFHS3	DLHS3	CES09	AHS	Total
Assam	0.5821	0.582	0.6366	0.637	0.0	0.00000	0.00014	0.00374	0.99611	1
Bihar	0.6430	0.643	0.6606	0.661	0.0	0.00001	0.00033	0.00692	0.99274	1
Chhattisgarh	0.7300	0.730	0.7870	0.787	0.0	0.00000	0.00000	0.00018	0.99982	1
Jharkhand	0.6110	0.611	0.7142	0.714	0.0	0.00001	0.00026	0.00369	0.99604	1
MP	0.5110	0.511	0.6491	0.649	0.0	0.00002	0.00042	0.00406	0.99550	1
Odisha	0.5431	0.543	0.5953	0.595	0.0	0.00001	0.00014	0.00323	0.99662	1
Rajasthan	0.6900	0.690	0.7730	0.773	0.0	0.00000	0.00001	0.00015	0.99984	1
Uttarakhand	0.7380	0.738	0.7990	0.799	0.0	0.00000	0.00001	0.00014	0.99985	1
UP	0.4470	0.447	0.4813	0.481	0.0	0.00000	0.00015	0.00222	0.99763	1

Distribution of weights for some non-high focus states are given in Table 6. For non-high focus states, we do not have AHS estimates available; hence, it does not receive any weight. The maximum weight is given to the CES09 direct estimates which are the most recent ones

for non-high focus states. Because of small sample size, NFHS-3 does not produce rural area estimates in Delhi. The state Nagaland is not covered by DLHS-3. This explains the zero weight at the corresponding table cells. Weights corresponding to CES09 vary across states depending on the precision of CES09 estimates. For example, CES09 estimates receive weights 0.97, 0.95 and 0.84, respectively, for Andhra Pradesh, Arunachal Pradesh and Delhi. Differential weights for these three states can be explained by the differential standard errors of CES09 estimates which are 0.005, 0.030 and 0.039, respectively. This illustrates the intuitive nature of our EBLUP estimator that follows from model (1). Table 5 and Table 6 show results for full vaccination coverage estimates; similar pattern holds true for other vaccination coverage rates.

Table 6 Distribution of weights (as defined in the EBLUP expression) across time points for some non-high focus states: Estimating rural area full vaccination coverage rates

State	DLHS-2	NFHS-3	DLHS-3	CES09	AHS-1	Total
Andhra Pradesh	0.00002	0.00025	0.02424	0.97549	0	1
Arunachal Pradesh	0.00034	0.00202	0.04416	0.95347	0	1
Delhi	0.01124	0	0.14159	0.84583	0	1
Gujarat	0.00003	0.00031	0.03698	0.96268	0	1
Haryana	0.00019	0.00119	0.08825	0.91037	0	1
Karnataka	0.00000	0.00002	0.00135	0.99863	0	1
Maharashtra	0.00002	0.00024	0.03389	0.96585	0	1
Mizoram	0.00055	0.00200	0.06205	0.93539	0	1
Nagaland	0.00215	0.02234	0	0.97544	0	1