SSM - Population Health 14 (2021) 100748

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: http://www.elsevier.com/locate/ssmph

Mapping changes in district level prevalence of childhood stunting in India 1998-2016: An application of small area estimation techniques

Swati Srivastava [a,*], Hukum Chandra [b], Shri Kant Singh [a], Ashish Kumar Upadhyay [a]

[a] International Institute for Population Sciences, Mumbai, India
[b] ICAR-Indian Agricultural Statistics Research Institute (IASRI), India

ARTICLE INFO

Keywords:
Small area estimation
Stunting
Generalized liner mixed model
Spatial analysis
National family health survey

ABSTRACT

The four rounds of National Family Health Survey (NFHS) conducted during 1992–93, 1998–99, 2005-06 and 2015-16 is main source to track the health and development related indicators including nutritional status of children at national and state level in India. Except NFHS-4, first three rounds of NFHS were unable to provides district-level estimates of childhood stunting due to the insufficient sample sizes. The small area estimation (SAE) techniques offer a viable solution to overcome the problem of small sample size. Therefore, this study uses SAE techniques to derive district level prevalence of childhood stunting corresponding to NFHS-2 (1998–99). Study further estimated GIS maps, univariate Local indicator of spatial autocorrelation (LISA) and Moran's I to understand the trend in district level childhood stunting between NFHS-2 and NFHS-4. Estimates obtained by SAE techniques suggest that prevalence of childhood stunting ranges from 20.7% (95% CI: 18.8–22.7) in South Goa district of Goa to 64.4% (95%CI: 63.1–65.7) in Dhaulpur district of Rajasthan during 1998–99. The diagnostic measures used to validate the reliability of estimates obtained by SAE techniques indicate that the model-based estimates are reliable and representative at district level. Results of geospatial analysis indicates substantial reduction in childhood stunting between 1998 and 2016. Out of 640 district,about 81 district experience reduction of more than 50%. At the same time 60 district experience less than 10% of reduction between 1998 and 2016. Spatial clustering of childhood stunting remains same over the study period except few additional cluster in Maharashtra, Andhra and Meghalaya in 2016. The district level estimates obtained from this study might be helpful in framing decentralized policies and implementation of vertical programs to enhance the efficacy of various nutrition interventions in priority districts of the country.

Introduction

Childhood stunting is a major public health issue and have both short and long term consequences on health. Short term effect includes morbidity, mortality and poor cognitive development during early childhood (Rice et al., 2000; Ijarotimi, 2013; Singh et al., 2017; Myatt et al., 2018). Recent estimates from India State-level Disease Burden Initiative suggest that more than 20% deaths and disease burden among Indian children under age five-years can be attributed to childhood malnutrition (Dandona et al., 2020; Swaminathan et al., 2019). Long term consequences are poor cognitive development, poor school performance, low productive and short adult stature (Victora et al., 2008; Adair et al., 2013; Karra & Fink, 2019; Paul & Singh, 2020). Reducing childhood undernutrition by 50% till 2015 from its level in 1990 was one of the essential objectives of Millennium Development Goal (MDG 2000). In post MDG plan, Sustainable Development Goal (SDG) was launched and aimed to eradicate childhood malnutrition by 2030. India has experienced a remarkable decline in prevalence of childhood stunting from 54% in 1992 to 38% in 2016 but fail to achieve the target set by MDG. However, improvements have been unevenly distributed across different states of India. At state level childhood stunting ranging from 20% in Kerala to 48% in Bihar, wasting from 7% in Mizoram to 29% in Jharkhand and underweight from 12% in Mizoram to 48% in Jharkhand (IIPS & ICF, 2017). Variations are expected at district level within the states due to difference in socioeconomic, demographic and ecological conditions which may affect child health outcome.

In India, the main source of data for child nutrition comes from National Family Health Survey (NFHS) (Demographic and Health

* Corresponding author.
E-mail addresses: sswati146@gmail.com (S. Srivastava), hchandra12@gmail.com (H. Chandra), sksingh31962@gmail.com (S.K. Singh), ashu100789@gmail.com (A.K. Upadhyay).

https://doi.org/10.1016/j.ssmph.2021.100748
Received 16 October 2020; Received in revised form 8 January 2021; Accepted 30 January 2021
Available online 3 February 2021

Survey in other countries). Till date four rounds of NFHS have been conducted during 1992–93, 1998–99, 2005-06 and 2015–16. Unfortunately, estimates on child health at smallest administrative unit (for example, district-level) were not available for child health indicators before fourth round of NFHS (2015–16). There is some evidence of subnational estimates of childhood malnutrition that uses data from single round of NFHS (Singh et al., 2011; Khan & Mohanty, 2018; Singh et al., 2019). However, there is lack of district level trends in childhood malnutrition over a long period of time. In the absence of reliable district level estimates for health indicators, it is difficult to track the progress on childhood stunting at lowest administrative units.

In absence of representative sample at a lower administrative level Small Area Estimation (SAE) techniques became the best option to estimate key developmental indicators at fine geographic level. SAE has been widely used in India and other countries to provide reliable estimates at local level in absence of representative sample. Therefore, firstly present study will provide district level childhood stunting for the years 1998-99 by using SAE techniques. Notably, prior studies from India have used this technique mainly to provide estimate on agricultural indicators. Sud et al. (2011) provided district-level estimates of crop yield for paddy in Uttar Pradesh. Another study by Chandra et al. (2011) also uses SAE to determine the district level estimates of indebtedness in Uttar Pradesh. Further, the SAE was used to derive estimates of poor households at the districts of Uttar Pradesh and Orissa (Chandra et al., 2018; Mohanty & Swain, 2018). Recently using data from National Crime Records Bureau, Vicente et al. (2018) estimated the district level rape incidence risk in Uttar Pradesh . Apart from this a few study used SAE to examine the vaccination coverage and infant mortality rate at district level in India (Pramanik et al., 2015). However, study from other countries have used SAE techniques for estimating under-five mortality, childhood undernutrition, institutional delivery, smoking and contraceptive prevalence rate (Dwyer-Lindgren et al.; Aliaga & Muhuri, 1994; Fujii, 2010; Johnson et al., 2010; Song et al., 2016; Chandra et al., 2018). So far, majority of the previous studies from India uses SAE methodology to provide the estimates in the field of agriculture and socioeconomic status, however there is lack of use of SAE techniques in public health in India (Elbers et al., 2003). To the best of our knowledge, this is the first study from India which intended to use SAE techniques to provide estimates for childhood stunting at district level. Second, the estimates obtained by SAE techniques for the year 1998-99 and direct estimates of childhood stunting obtained by NFHS-4 (2015–16) will be used to see the district level changes in prevalence of childhood stunting between 1998-99 and 2015–16. Distribution of childhood stunting must be varying across the geographical regions; thus, mapping of nutritional status and tracking the districts lagging in reduction of childhood stunting over the period may help to improve the program in terms of allocation of resources and policy decision which favors to child health.

Data and methodology

Data

The present study is based on the analysis of secondary data collected under different rounds of Indian NFHS and Census of India. The NFHS is the repeated cross-sectional survey which was aimed to collect the detail information about the several health and nutritional indicators in India approximately with the gap of 5 years since 1992–93. The census of India is one of the important and largest data sources in India, which provides information on a variety of socio-economic, demographic, educational characteristics and migration status of people at disaggregate level. Details about used data set has been given in the appendix (see appendix). The study aims to derive the district level estimates of childhood stunting in all districts of India for the year 1998-99 using SAE technique by combining census and survey data.

Outcome variable

The outcome variable is childhood stunting, which is drawn from NFHS and binary at unit level. According to WHO criteria children have height-for-age z-score less than -2SD has been called as 'stunted'. The parameter of interest is to estimate the district level prevalence of childhood stunting under age 3 years for NFHS-2 (1998–99). Notably, NFHS-4 (2015–16) was designed to provide district level estimates for key maternal and child health indicators including childhood stunting, therefore, study did not used SAE technique to estimate prevalence of stunting for the NFHS-4 (IIPS and ICF 2017). The study could not include NFHS-1 (1992–93) because NFHS-1 did not collect children's height in five major states (Andhra Pradesh, Himachal Pradesh, Madhya Pradesh, Tamil Nadu and West Bengal) of India. NFHS-3 (2005–06) was also excluded from the analysis due to unavailability of district codes in the data. The study restricted our analysis for children under age 3 years as NFHS-2 (1998-99) were collected information on key indicators of child health and developemnt only for those who were born in three years prior to survey.

Exposure variables

Exposure variables for the study are taken from the population census of India, 2001. The exposure variables are those which is known for the entire populations and works as auxiliary information in SAE. The best-chosen auxiliary variables were household size, women's workforce participation, availability of separate kitchen for cooking, availability of the improved source of drinking water and availability of clean fuel for cooking. It may be possible that there are some other indicators which may affect the nutritional status of children, but the present study did not consider those indicators due to unavailability of the information in census data.

Methodology

First, direct estimates of childhood stunting for NFHS-2 (1998–99) were calculated by dividing the number of stunted children to number of sample children in each districts. Notably, number of sample children varies from 1 to 587 with an average sample of 57 children in each district (see Table S1 in the appendix). Likewise, the number of stunted children also varies from 0 to 253 with an average sample of 28 stunted children. Further, the direct estimates of childhood stunting vary from 0% to 100% at district level, which appears to be very unstable due to insufficient sample. Therefore, SAE technique was applied to produce precise estimates of prevalence of childhood stunting at district level for the year 1998–99. The fundamental tenet of SAE approach is to use statistical models to link the variable of interest with auxiliary information such as census and administrative data to produce model-based estimators for small areas. In other words, if the area-specific direct estimators do not provide adequate precision, then in making estimates for small area quantities it is necessary to employ model-based estimators that "borrow strength" from other areas. The small areas defined in this study are the districts. Small area models can be classified into two broad types namely area level or unit level model. Area-level modelling is typically used when unit-level data are unavailable, or, as is often the case, where model covariates or auxiliary variables are only available in aggregate form. We adopt the area level small area modelling because the auxiliary variables are available only at the district level. In particular, the present analysis considers the area level generalized linear mixed model (GLMM) with logit link function, also referred to as the logistic linear mixed model, which is generally fitted for binary variable (Chandra et al., 2011). In the fixed effect part of model, states were used dummy variable to account for unobserved state-specific heterogeneity. In general, in order to derive representative and precise estimates, sampling weight should be incorporated in SAE to account for complex sampling design. This study adopted Chandra et al. (2019) approach to incorporate the sampling design in SAE under an area level version of GLMM.

In small area modelling, certain diagnostics measures are applied to evaluate the validity and accuracy of estimates. In practice, two types of diagnostics measures are executed in SAE: the model diagnostics, and the diagnostics for the small area estimates (Chandra et al., 2011). The model diagnostics are examined to verify the assumptions of the underlying model, i.e., how well the model performs when it is fitted to data. In small area model, the random area specific effects are assumed to have a normal distribution with mean zero and fixed variance. If the model assumptions are satisfied, then the area or district level residuals are expected to be randomly distributed around zero. The second diagnostics are used to validate the reliability of model based district level estimates generated by SAE technique and are measured through a) bias diagnostic, b) percent coefficient of variation (CV) diagnostic and c) 95 per cent confidence interval (CI) diagnostic. The bias diagnostics are used to investigate if the model-based estimates of childhood stunting are less extreme than the direct survey-based estimates. The percent CVs were used to assess the improved precision of the model-based estimates compared to the direct survey based estimates. It is well known that the CVs show the sampling variability as a percentage of the estimate. Estimates with low CVs than larger CVs were considered as more reliable. However, there is no cutoff to define 'too large CVs' (Amoako Johnson et al., 2012; Chandra et al., 2011, 2018; Johnson et al., 2010). The 95 per cent CIs of the model-based estimates and direct estimates were compared to validate the robustness of our model-based estimates of childhood stunting (Johnson et al., 2010). The detail theoretical illustration of SAE techniques are given in next section.

Finally, we produced GIS maps, univariate local indicator of spatial autocorrelation (LISA) and Moran's I to understand the trend and pattern in district level childhood stunting between 1998 and 2016. LISA maps help to identifying local clusters and spatial outliers. LISA maps classify clusters into high-high, low-low, high-low, and low-high. Usually, high-high and low-low are known as spatial clusters. Whereas, high-low and low-high are known as spatial outliers (Anselin, 1995). On the other hand, Moran's I measures spatial autocorrelation i.e. to what degree the data points are similar or dissimilar to their spatial neighbors. While positive Moran's I indicates that points with similar attribute values are closely distributed in space, negative Moran's I indicates that closely associated points are more dissimilar. The value of Moran's I ranges between −1 and +1. A zero value indicates random spatial autocorrelation.

Notably, sampling framework adopted for NFHS-2 and NFHS-4 was Census of India 1991 and 2011 respectively. In NFHS-2 there was 438 districts whereas in NFHS-4 there was 640 districts in India. In order to see the changes in prevalence of childhood stunting between 1998-99 and 2015–16, study first calculated the district level prevalence of childhood stunting for 640 districts corresponding to year 1998–99. Three type of classification were made while adjusting increases in number of the districts over the years. 1) The unchanged districts pose no problem. 2) For the districts partitioned from one single district, estimates for the new-born districts have been assigned same as of their parent districts. 3) In cases where the districts have been carved from multiple districts, average values of the parent districts have been used for new districts. It was observed that 329 districts out of the 640 districts in 2011 were unaffected by boundary changes over 1991–2011. Two hundred eighty-six districts were clearly partitioned into multiple districts over the same period. Remaining 25 districts experienced more complex changes (see appendix Table S2).

Theoretical illustration of SAE

To start, let N_d and n_d be the population and samples sizes in the district $d(d=1,\ldots,D)$ respectively, where $D=438$ is the number of the district (or small area) in the population. In particular, in this study, we define N_d as the total number of children under age three years in d^{th} district recorded in the census and n_d as the number of children under age three years in d^{th} district recorded in the survey. The total number of units in the population is $N=\sum_{i=1}^{d} N_i$ with corresponding total sample size $n=\sum_{i=1}^{d} n_i$. We used two additional subscript 's' and 'r' to denote the quantities related to sample and non-sample part of the population such that y_{sd} and y_{rd} are the sampled and non-sample counts of the stunted children in district *d*. The response variable y_{sd} follows a binomial distribution with parameter n_d and π_d where π_d is the probability of being stunted in district *d*. Further, y_{sd} and y_{rd} are assumed to be independent binomial variables with common success probability π_d. This indicates that $y_{sd} \sim Bin(n_d,\pi_d)$ and $y_{rd} \sim Bin(N_d-n_d,\pi_d)$.

Let $\mathbf{x}_d$ be the *k*-vector x_d be the k vector of the covariates for the district *d*,. The model linking this success probability with the covariates is the logistic linear mixed model of the form

$$logit(\pi_d)=\ln\left\{\frac{\pi_d}{1-\pi_d}\right\}=\eta_d=\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d, .d=1,\ldots,110. \tag{1}$$

Here,
$\pi_d=\exp(\eta_d)\{1+\exp(\eta_d)\}^{-1}=\exp(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d)\{1+\exp(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d)\}^{-1}$ and $\boldsymbol{\beta}$ is the k vector of unknown fixed-effects parameters. $u_d\sim N(0,\phi)$ **is** the random effect that accounts for between district variability beyond that explained by the covariates included in the model. It is evident that model (1) relates the area (or district) level proportions (direct estimates) from the survey to area (district) level covariates. Often, this type of model is called as 'area-level' model in SAE terminology (Fay & Herriot, 1979; Rao, 2014, pp. 1–8). But the Fay and Herriot method was based on the area level linear mixed model, and their approach applies to a continuous variable. In contrast, here the model (1) is the special case of a GLMM with logit link function which is suitable for binary outcome variable (Breslow & Clayton, 1993). Saei and Chambers (2003) have described this model in the context of SAE and by definition the means of Y_{sd}, Y_{rd} given u_d under model (1) are-

$$E(y_{sd}\mid u_d)=n_d\left[\exp\left(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d\right)\left\{1+\exp\left(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d\right)\right\}^{-1}\right] \tag{2}$$

$$E(y_{rd}\mid u_d)=(N_d-n_d)\left[\exp\left(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d\right)\left\{1+\exp\left(\mathbf{x}_d^{'}\boldsymbol{\beta}+u_d\right)\right\}^{-1}\right] \tag{3}$$

Let T_d is the total number of stunted children in district *d*, then.
$T_d = y_{sd} + y_{rd}$ $(d, = 1, 2 \ldots .438)$.

The first term y_{sd} is the sample count known from the survey whereas the second term y_{rd} is the nonsample count that is unknown. Thus, an estimate $\hat{T_d}$ **of** the total number of stunted children in district *d*, which is obtained by replacing y_{rd} by its predicted value under model (1). That is-

$$\widehat{T}_d=y_{sd}+\widehat{y}_{rd}=y_{sd}+(N_d-n_d)\left[\exp\left(\mathbf{x}_d^{'}\widehat{\boldsymbol{\beta}}+\widehat{u}_d\right)\left\{1+\exp\left(\mathbf{x}_d^{'}\widehat{\boldsymbol{\beta}}+\widehat{u}_d\right)\right\}^{-1}\right] \tag{4}$$

Here $\hat{T_d}$ was estimated using only children within the census window to ensure consistency between N_d and n_d. The proportion (p_d) of stunted children in a district d is obtained as the total number of children within the district. Thus, an estimate of p_d is-

$$\widehat{p}_d=\frac{\widehat{T}_d}{N_d} \tag{5}$$

The estimator (5) defined under model (1) is widely used for the estimation of small area proportions, see for example, Chandra et al. (2018) and references therein, although it is not the most efficient predictor under that model. An alternative to (5) is the empirical best predictor (Jiang, 2003). But, this predictor does not have a closed form and can only be computed via numerical approximation. This is generally not straightforward, however, and so national statistical agencies favour computation of an approximation like the estimator (5). It is worth noting model (1) is based on unweighted sample count which assumes that sampling within areas is non-informative. As a result, equation (4) ignores the complex survey design. If the sampling design is informative and survey weighted counts are available, there are two main difficulties. First, the values for the weighted sample counts will not necessarily be the integers 0, 1, 2, 3 …. n_d; rather they will take a value from a finite set of unequally-spaced numbers (not necessarily integers) determined by the survey weights of the sample cases in area *d*. Second, the estimated sampling variance of the weighted sample counts, y_{sd} implied by the binomial distribution, i.e., $v(y_{sdw}) \approx n_d p_{iw}(1 - p_{iw})$ will be incorrect. Several authors suggested that one should use the 'effective sample size' instead of actual sample size in model while analysing area level estimates as a binomial proportion (Korn & Graubard, 1998). The use of "effective sample size" has been discussed by several authors including (Mercer et al., 2014) and (Liu et al., 2007) as a way of incorporating the survey weights. Mercer et al. (2014) observe that the pseudo likelihood approach and effective sample size approach lead to identical estimates of small area proportions. Using the effective sample size rather the actual sample size allows for the survey weights under complex sampling. Furthermore, the precision of an estimate from a complex sample can be higher than for a simple random sample, because of the better use of population data through a representative sample drawn using a suitable sampling design. Here we uses a subscript of (e) in all the quantities associated with the "effective sample size". We address the above two issues by defining an "effective sample size" $n_{d(e)}$, and an "effective sample count" $y_{sd(e)}$ such that-

$$y_{sd(e)} = n_{d(e)} p_{iw} \tag{6}$$

This leads to p_{iw} with its corresponding estimator of variance estimate $v(p_{iw})$. Then the model equation (1) applied effective sample count $y_{sd(e)}$ in district *d* follows the binomial distribution-

$$y_{sd(e)} \sim \mathrm{Bin}(n_{d(e)}, \pi_d) \tag{7}$$

The effective sample size. $n_{d(e)} = \frac{\widehat{P}_i(1-\widehat{P}_i)}{v^*(p_{iw})}$

Here, $\widehat{P}_i$ is the preliminary model-based prediction of the population proportion P_i under the generalized liner mixed model, and the estimates of variance $v^*(p_{iw})$ depends on $\widehat{P}_i$ through the generalized variance function (GVF) (Liu et al., 2007). Again, the value of $y_{sd(e)} = 0$ if $p_{iw} = 0$. But this does not cause problem since $\widehat{P}_i > 0$, implies $n_{d(e)} > 0$. Note that we use a generalized variance function (GVF) to generate estimates of the sampling variance even for areas that have an observed count of zero. Consequently, we do not exclude any area from model fitting. The empirical predictor of T_d is finally obtained by replacing (n_d, π_d) by $(n_{d(e)}, y_{sd(e)})$ in model equation (1), thus ensuring that sampling weights are used in the small area estimation process. The plug-in empirical predictor (EP) of T_d is then

$$\widehat{T}^{EP}_{d(e)} = y_{sd(e)} + (N_d - n_{d(e)})*\widehat{T}^{EP}_{d(e)} \tag{8}$$

Therefore, based on the effective sample size $n_{d(e)}$ and effective sample count $y_{sd(e)}$ the estimate of the proportion in district d is

$$\widehat{T}^{EP}_{d(e)} = \frac{T^{EP}_{d(e)}}{N_d} \tag{9}$$

Results

The result of this study will be explained in two steps. First, the district level estimates of childhood stunting obtained by SAE techniques. Second, we will explain the changes in prevalence of childhood stunting between the period 1998–2016.

The model-based district level estimates of childhood stunting for 1998-99 are presented in Fig. 1 . Figure indicates prevalence of childhood stunting varies considerably across the districts of India. Prevalence of childhood stunting ranges lowest from 20.7% in South Goa followed by 22.2% in north-Goa districts to 64%% in Dholpur districts followed by 63.7% in Balia district. The general pattern of stunted children shows the relatively lower prevalence of childhood stunting in the southern and north-eastern part of country, however it was relatively higher in the norther part of the country. For example- Ernakulum is the district with the minimum proportion of stunted children (22%) in the south region; however, Dholpur is the district with the maximum proportion of stunted children (64%).

Results of model diagnostic are presented in Fig. 2(a) and (b). Fig. 2a shows that residuals are randomly distributed, and also the line of fit does not look significantly different from line y = 0. The q-q plots also confirm the normality assumptions of the data (Fig. 2(b)). Hence, the diagnostic procedures related to the model are fully satisfied with the data in the present study.

To examine the bias diagnostic, scatter plot of the model-based estimates of childhood stunting against the direct survey based estimates are presented in Fig. 3. Fig. 3 shows that the model-based estimates of childhood stunting are less extreme than the direct survey based estimates. For illustration, the study found that model-based estimates of childhood stunting are shrinking towards the mean (survey-based mean stunted children are 49.8). The study also highlighted that the districts having extreme direct estimates of childhood stunting were mainly due to small sample size.

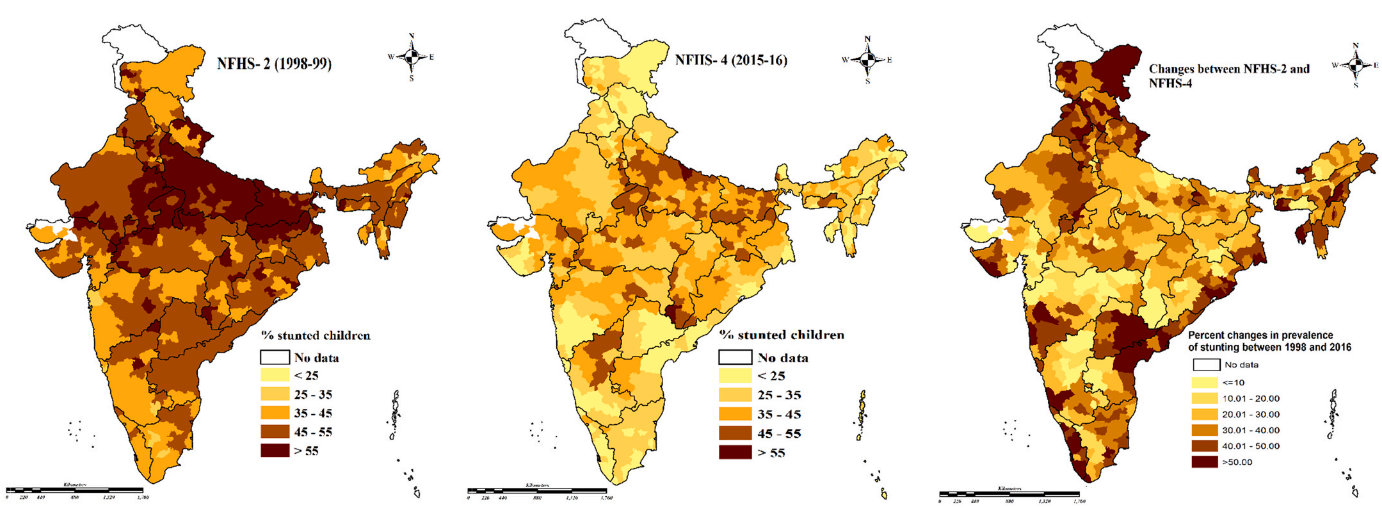

Fig. 1. District level prevalence of childhood stunting in India during 1998–99, 2015–16 and percent changes over the period of 1998–2016.

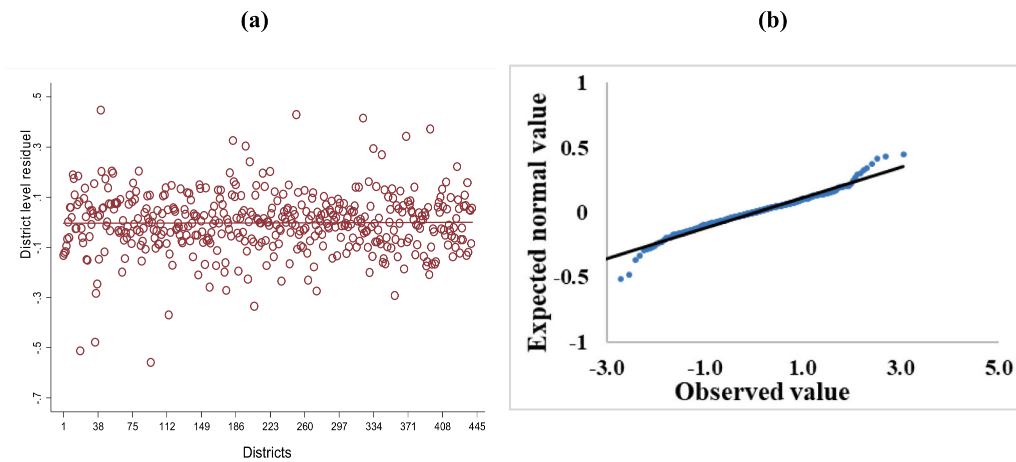

Fig. 2. **(a)** Model diagnostic plot showing the distribution of the district level residuals and **(b)** q-q plot for childhood stunting.

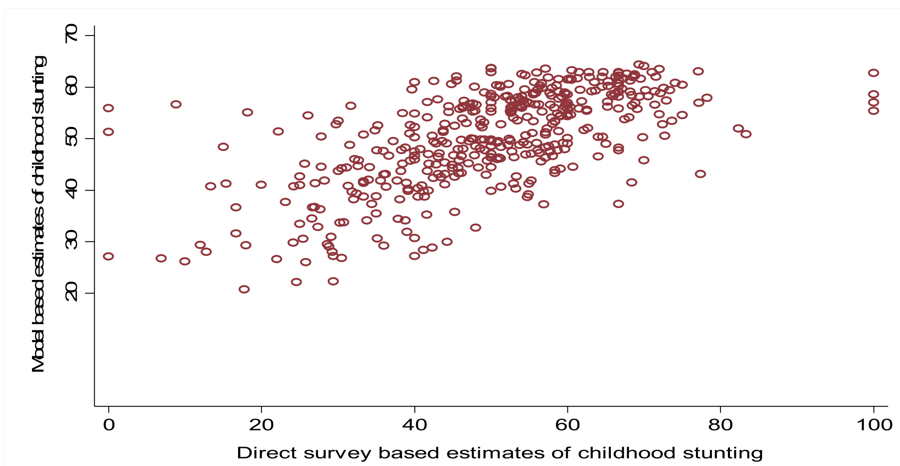

Fig. 3. Bias diagnostics for childhood stunting.

Table S3 provides district-wise direct and the model-based estimates of childhood stunting, along with per cent CV and 95 confidence intervals. Also, the distribution of per cent CV of the direct and the model-based estimates of childhood stunting have been plotted in Fig. 4 which shows that estimated CVs for model-based estimates having more reliability than the direct survey-based estimates CVs. For stunting, the CVs for direct survey-based estimates are ranging from 0.5% to 3.7%, but the CVs for the model-based estimates range between 0.1% and 2.7%. Likewise to earlier study, the improvement in per cent CV is bigger for the districts with small sample size compared to districts with a larger sample size (Das et al., 2019). For a few districts (Delhi, Greater Bombay, Aizwal, Imphal, Udaipur, Madras, Kohima, Cuttack, Udhampur, Jaipur, Kangra Puri and Alwar) with higher sample size and more stunted children, the difference in per cent CV is between 0.62 and 1.04%. For some districts, it was not possible to compute CV and standard error for direct survey-based estimates because of no sample counts. However, the advantage of SAE technique worked here very well and helped to predict estimates even in such areas where no sample information was available. Fig. 5 also confirms that for districts where sample size or sample count are zero, was not possible to compute 95%

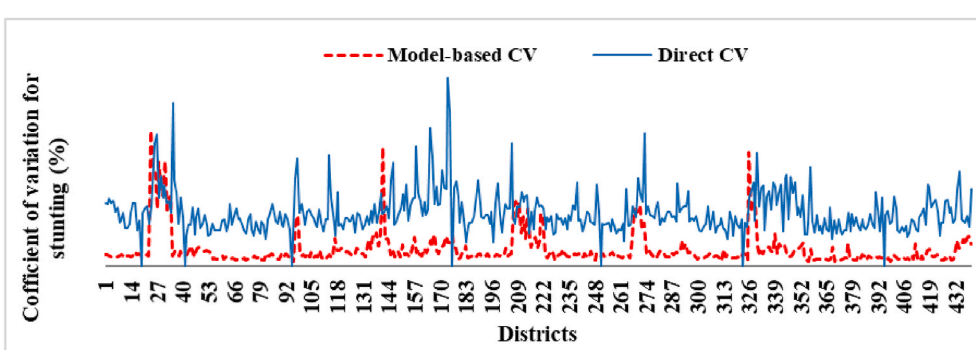

Fig. 4. District wise coefficient of variation for childhood stunting in India, 1998-99.

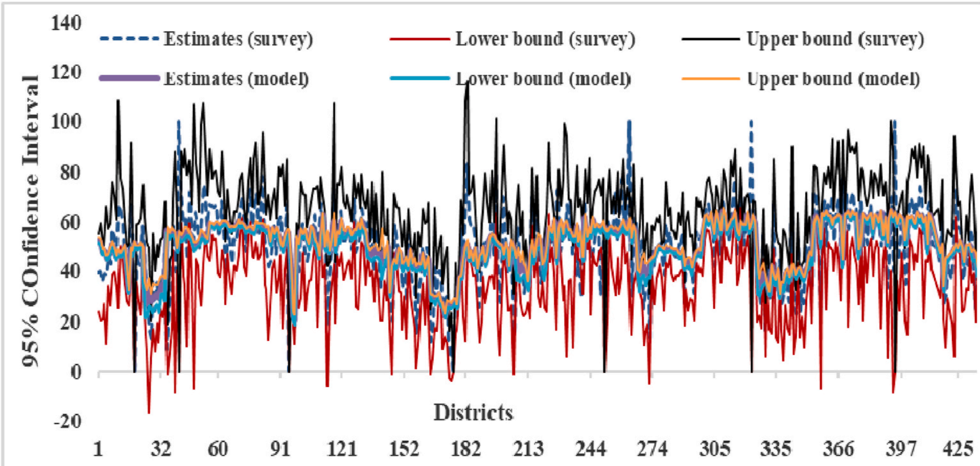

Fig. 5. District wise 95% CI for model-based estimates and direct survey-based estimates for stunting in India, 1998–99.

CI. However, it is evident from the respective figure that the model-based estimates are having tight confidence interval than direct survey-based estimates.

We also calculated the percent point difference between direct survey based estimates and model based estimates of childhood stunting (Table 1). About 160 district shows a difference of 5% point between direct survey based estimates and model based estimates. Difference of 10 or more percent were observed for 142 districts. For the remaining 126 districts difference between survey and model based estimates lies between 5% and 10%.

Spatial distribution and changes in childhood stunting in India during 1998–2016

Trends in spatial profile of childhood stunting has been shown through Fig. 1, which shows that childhood stunting has been declined from 50 percent in 1998/99 to 36 percent in 2015/16 with considerable amount of variation. During 1998–2016 subsequent changes were occurred in the spatial profile of childhood stunting with some intra-state disparities. For example-in 1998–99, childhood stunting varied from 25 percent in Papum Pare (lowest) district of Arunachal Pradesh to 72 percent in Etawah (highest) district of Uttar Pradesh. Whereas, in 2015–16, childhood stunting varied from 12 percent in Tawang (3rd lowest districts) district of Arunachal Pradesh to 62 percent in Bahraich (highest) district of Uttar Pradesh. There were altogether 285 districts where the prevalence of childhood stunting was higher than the national average (50%) in 1998-99 however; there were about 272 districts, where the prevalence of childhood stunting was higher than the national average (36%) in 2015–16.

Percent changes in prevalence of childhood stunting between 1998 and 2016 are presented in Fig. 1 . Result shows that out of 640 district, 81 districts experience more than 50% reduction. These districts are coming from the state of Andhra Pradesh (9 districts), Arunachal Pradesh (1 districts), Bihar (1 districts), Daman & Diu (1 districts), Gujarat (2 districts), Haryana (2 districts), Himachal Pradesh (5 districts), Jammu & Kashmir (10 districts), Karnataka (3 districts), Kerala (8 districts), Maharashtra (4 districts), Meghalaya (2 districts), Orissa (6 districts), Puducherry (1 districts), Punjab (8 districts), Rajasthan (1 districts), Manipur (3 districts), Nagaland (6 districts), Tamil Nadu (2 districts), Tripura (2 districts), Uttarakhand (3 districts) and West Bengal (1 districts). About 108 and 110 districts experience reduction 40–50% and 30–40% respectively. The highest number of districts (181 districts) were experienced the reduction of 20–30 percent in the prevalence of childhood stunting. Surprisingly, about 90 districts shows a reduction of 10–20 percent and 60 districts shows a decline of less than 10%.

Table 1
Percent difference in prevalence of childhood stunting between direct and survey based estimates in India 1998–99.

Difference between direct estimates and model based estimates (%)	Number of districts
<-10	71
−9.99 to −5.00	65
−4.99 to −0.01	83
0.00 to 5.00	87
5.01–10.00	61
>10.00	71

The univariate LISA cluster results for childhood stunting has been presented in Fig. 6. The trends in univariate LISA map depicts high-high clusters mainly comprised in districts of Jharkhand, Uttar Pradesh, Madhya Pradesh, Gujarat, Bihar, Rajasthan during 1998-99 and 2015–16. However, we found some additional high-high clusters in the districts of Maharashtra, Karnataka, Andhra and Meghalaya in 2015–16. Low-low clusters of childhood stunting coming from districts of the southern and northeastern part of the country along with Jammu & Kashmir, Himachal Pradesh and Punjab. Few of them were also coming from the state of West Bengal, Uttarakhand, Maharashtra, Orissa, Tripura and Gujarat. Also, the cold-spots for childhood stunting have been changed between 1998 and 2016. The study further computed the Moran-I statistics, which shows the magnitude of geospatial clustering of childhood stunting. The value of Moran's-I was 0.59 in both the rounds of NFHS. Such a high positive value of Moran's I indicates that districts with similar prevalence of childhood stunting are closely distributed in space.

The univariate LISA cluster maps according to magnitude of reduction in childhood stunting between 1998 and 2016 are presented in Fig. 7. Finding suggest that districts with higher reduction in stunting are clustered in Andhra Pradesh, Kerala, Punjab, Himachal Pradesh, Nagaland and Tripura. High-high cluster was also noted in a few districts of Maharashtra, Rajasthan, Manipur, Karnataka, West Bengal, Odisha, Tamil Nadu, and Mizoram. Majority of the districts with low-low clustering was belongs to the Maharashtra, Karnataka, Gujarat and Chhattisgarh.

Discussion

This study uses SAE method defined under an area-level GLMM with

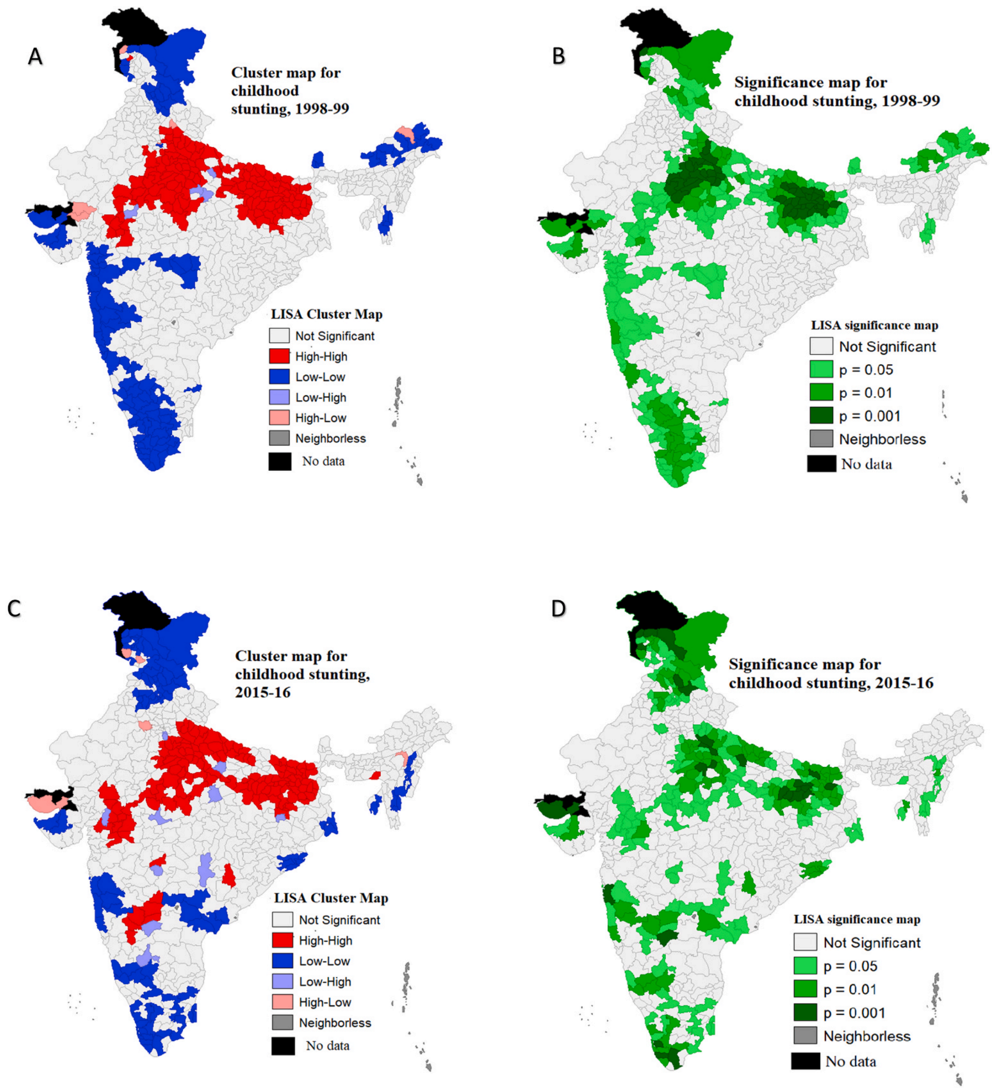

Fig. 6. Univariate LISA (Cluster and Significance) maps depicting spatial clustering and spatial outliers of childhood stunting in India, 1998–16. **A.** Univariate LISA Cluster map of childhood stunting across 640 districts of India 1998-99 **B.** Univariate LISA Significance map of childhood stunting across 640 districts of India 1998-99 **C.** Univariate LISA Cluster map of childhood stunting across 640 districts of India 2015-16 **D.** Univariate LISA Significance map of childhood stunting across 640 districts of India 2015-16.

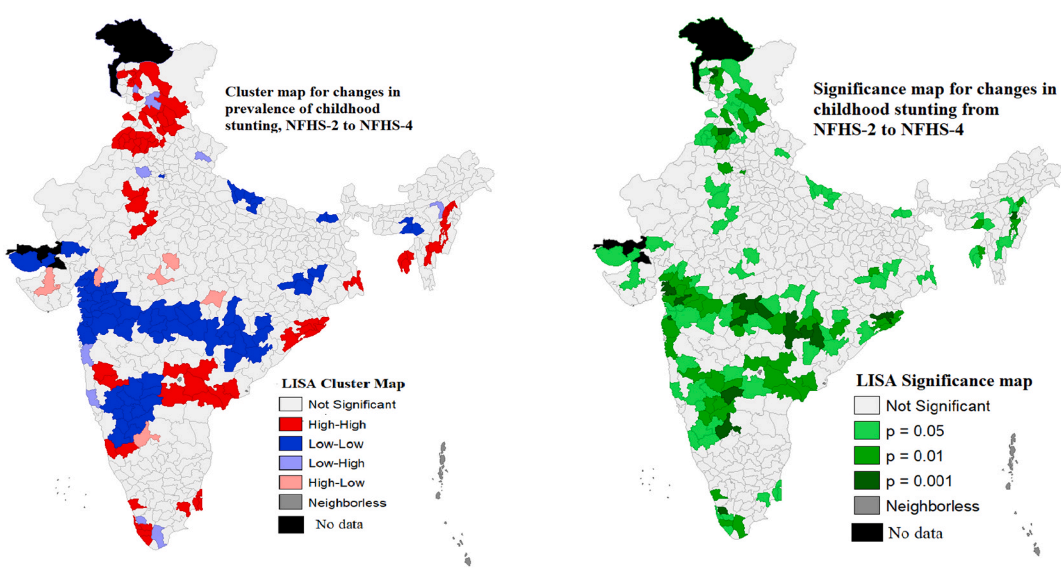

Fig. 7. Results of univariate LISA map for percent changes in prevalence of childhood stunting in India 1998–2016.

logit link function for estimating the district-wise prevalence of childhood stunting in India. In particular, this analysis combined data from second round of NFHS conducted during 1998-99 and the 2001 population census data to develop the reliable district level estimates of childhood stunting in India during 1998–99. The study further applied several diagnostic tools to examine the reliability and validity of the model-based estimates generated by using SAE technique. The model-based estimates generated in this analysis were also used to observe the level of changes in spatial profile of childhood stunting in 640 districts of India from 1998 to 2016.

The findings of this study show that prevalence of childhood stunting obtained from model-based estimates are more precise as compared to the direct-survey based estimates. Likewise to other studies, present study also went through some testing tools to verify the reliability and validity of model-based estimates using the methods described in (Johnson et al., 2010), which shows that the distribution of the district level residuals and normality curves verified the assumptions of the underlying model. Further, the bias diagnostic (CV and 95% CI) also confirmed the validity and reliability of the model-based estimates, as model-based estimates had less extreme values than direct survey-based estimates, The CV, which shows the sample variability as a percentage of estimates of model-based estimates was found smaller than the CV of direct survey-based estimates. Furthermore, the confidence interval of model-based estimates was found closure than the confidence interval of direct survey-based estimates. Hence, our study follows all the criteria of the diagnostic test.

The present study highlighted that the estimates of childhood stunting obatined by direct survey based methods have more extreme values than estimates obtained by SAE techniques. Hence, study clearly shows the advantage of SAE technique to cope up the small sample size problem in producing the estimates or reliable confidence intervals. Findings of the study are consistent with the earlier studies conducted in Bangladesh, which also combined the survey data with census data and found that the estimates at district level were sufficiently accurate (Das et al., 2019). In addition, we have also applied SAE techniques to estimate prevalence of underweight and wasting for NFHS-2 and examine the changes in these indicators between 1998 and 2016 (appendix Figure S3 & S4).

This study further attempted some spatial tools to see the changing pattern of childhood stunting in districts of India over the period of 1998-2016. Importance of using spatial analysis is to generate policy at the fine geographic level i.e., district level, to explain intrastate disparity more clearly than individual-level factors. This study not only measures the spatial clustering of childhood stunting but also looks for the changes in the spatial profile of childhood stunting in India for the last two decades. Earlier to this, few of the studies were attempted to understand the association of spatial factors and child health outcomes in India (Singh et al., 2011; Gupta et al., 2016). Our findings suggest that though there has been a decline in stunted children, the proportion of decline is far away to achieve SDG goals. Not only the rate of decrease was slow in the country, but their considerable disparities were observed in the rate of decline across the districts of India. Finding of this study are consistent with previous study conducted in India (Hemalatha et al., 2020). The analysis further suggested that India did not experience the same level of reduction in childhood stunting across the country, some of the districts experienced a low level of decline however some of the districts experienced a fast decline in stunted children. The prime concern of this study is to highlight the districts which are still grappling with the high burden of childhood stunting.

A key contribution of the present study is to identify the changes in hotspots (i.e., districts with high burden of childhood stunting also surrounded by high burden of childhood stunting) over the period of 1998–2016. The study is also interested to highlight some cold spots (i.e., districts with low burden of childhood stunting also surrounded by low burden of childhood stunting) and outliers (i.e., district with high burden of stunted children surrounded by district with low burden of stunted children and vice versa). The finding of this study is consistent with the other research conducted in India (Singh et al., 2011).

Limitation of the study must be stated. First, anthropometric data such as height and weight collected in different round of NFHSs are subject to measurement error. Second, estimates obtained from NFHS may underestimated due to the fact that survey do not collect information on anthropometric outcomes who had died due to malnutrition. However recent evidence from India suggest that at aggregate level child anthropometric outcome do not change substantially even after taking HAZ-score of dead children into account (Upadhyay & Singh, 2020). Third, spatial covariates used to estimate childhood stunting for 1998-99 do not include complete set of confounders due to lack of information on other indicators at district level.

In 2018, the Government of India launch National Nutrition Mission (NNM), also known as POSHAN Abhiyaan has emphasized targeting efforts at the district as well as sub-district levels to accelerate improvement in childhood malnutrition (Maiti, 2016; Paul et al., 2018). Our study indicates substantial district level variations in prevalence of childhood stunting and the rate of decline, which can use to understand the extent of efforts required to achieve the target set under NNM. If childhood stunting decline with similar rates in future, the target set by SDGs cannot be achieved. Therefore, the district level variation observed in childhood stunting and rate of decline calls the urgent need for appropriate planning and policy interventions to combat the associated problems with childhood malnutrition in India.

Ethical statement

The analysis presented in the paper is based on National Family Health Survey 1998-99 and 2015–16 which is a publically available dataset with no identifiable information on the survey participants. All the ethical concerns, including informed consent, are strictly followed in the survey. Given these, no ethical approval or informed consent was required for the current study.

Author statement

Swati Srivastava (SS) conceived the idea, SS, Ashish Kumar Upadhyay (AKU) and Hukum Chandra (HC) design the experiment, SS and AKU analyses data, SS, AKU and HC drafted the manuscript, Shri Kant Singh (SKS) edited the manuscript.

Financial disclosure statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2021.100748.

References

Adair, L. S., Fall, C. H., Osmond, C., Stein, A. D., Martorell, R., Ramirez-Zea, M., et al. (2013). Associations of linear growth and relative weight gain during early life with adult health and human capital in countries of low and middle income: Findings from five birth cohort studies. *The Lancet, 382*(9891), 525–534.

Aliaga, A., & Muhuri, P. K. (1994). *Methods of estimating contraceptive prevalence rates for small areas: Applications in the Dominican Republic and Kenya Methods of estimating contraceptive prevalence rates for small areas: Applications in the Dominican Republic and Kenya* (p. 19).

Amoako Johnson, F., Padmadas, S. S., Chandra, H., Matthews, Z., & Madise, N. J. (2012). Estimating unmet need for contraception by district within Ghana: An application of small-area estimation techniques. *Population Studies, 66*(2), 105–122.

Anselin, L. (1995). Local indicators of spatial association—LISA. *Geographical Analysis, 27*(2), 93–115.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association, 88*(421), 9–25.

Chandra, H., Aditya, K., & Sud, U. (2018). Localised estimates and spatial mapping of poverty incidence in the state of Bihar in India—an application of small area estimation techniques. *PloS One, 13*(6), Article e0198502.

Chandra, H., Chambers, R., & Salvati, N. (2019). Small area estimation of survey weighted counts under aggregated level spatial model. *Survey Methodology, 45*(1), 31–59.

Chandra, H., Salvati, N., & Sud, U. (2011). Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India: An application of small-area estimation technique. *Journal of Applied Statistics, 38*(11), 2413–2432.

Dandona, R., Kumar, G. A., Henry, N. J., Joshua, V., Ramji, S., Gupta, S. S., et al. (2020). Subnational mapping of under-5 and neonatal mortality trends in India: The global burden of disease study 2000–17. *The Lancet, 395*(10237), 1640–1658.

Das, S., Chandra, H., & Saha, U. R. (2019). District level estimates and mapping of prevalence of diarrhoea among under-five children in Bangladesh by combining survey and census data. *PloS One, 14*(2), Article e0211062. https://doi.org/10.1371/journal.pone.0211062

Dwyer-Lindgren, L., Kakungu, F., Hangoma, P., Ng, M., Wang, H., Flaxman, A., et al. (1980-2010). *Applying small area models to estimate mortality from birth history data: Under-5 mortality in Zambian districts*.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica, 71*(1), 355–364.

Fay, R. E., & 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*(366a), 269–277.

Fujii, T. (2010). Micro-level estimation of child undernutrition indicators in Cambodia. *The World Bank Economic Review, 24*(3), 520–553.

Gupta, A. K., Ladusingh, L., & Borkotoky, K. (2016). Spatial clustering and risk factors of infant mortality: District-level assessment of high-focus states in India. *Genus, 72*(1), 2.

Hemalatha, R., Pandey, A., Kinyoki, D., Ramji, S., Lodha, R., Kumar, G. A., et al. (2020). *Mapping of variations in child stunting, wasting and underweight within the states of India: The global burden of disease study 2000–2017*. *EClinicalMedicine*.

Iips, and Icf. (2017). *National family health survey (NFHS-4), 2015–16*. India: IIPS Mumbai.

Ijarotimi, O. S. (2013). Determinants of childhood malnutrition and consequences in developing countries. *Current Nutrition Reports, 2*(3), 129–133. https://doi.org/10.1007/s13668-013-0051-5

Jiang, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of Statistical Planning and Inference, 111*, 117–127.

Johnson, F. A., Chandra, H., Brown, J. J., & Padmadas, S. S. (2010). *District-level estimates of institutional births in Ghana: Application of small area estimation technique using census and DHS data*.

Karra, M., & Fink, G. (2019). Long run height and education implications of early life growth faltering: A synthetic panel analysis of 425 birth cohorts in 21 low- and middle-income countries. *BMC Public Health, 19*(1), 876. https://doi.org/10.1186/s12889-019-7203-5

Khan, J., & Mohanty, S. K. (2018). Spatial heterogeneity and correlates of child malnutrition in districts of India. *BMC Public Health, 18*(1), 1027.

Korn, E. L., & Graubard, B. I. (1998). Confidence intervals for proportions with small expected number of positive counts estimated from survey data. *Survey Methodology, 24*, 193–201.

Liu, B., Lahiri, P., & Kalton, G. (2007). Hierarchical Bayes modeling of survey-weighted small area proportions. In *Paper presented at the proceedings of the American statistical association, survey research section*.

Maiti, K. (2016). *Ministry of women and child development*. Government of India. Retrieved from January.

Mercer, L., Wakefield, J., Chen, C., & Lumley, T. (2014). A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial Statistics, 8*, 69–85.

Mohanty, B., & Swain, A. (2018). District level poverty estimation for rural Odisha (India) using different estimation techniques. *Model Assisted Statistics and Applications, 13*(1), 5–17.

Myatt, M., Khara, T., Schoenbuchner, S., Pietzsch, S., Dolan, C., Lelijveld, N., et al. (2018). Children who are both wasted and stunted are also underweight and have a high risk of death: A descriptive epidemiology of multiple anthropometric deficits using data from 51 countries. *Archives of Public Health, 76*(1), 28. https://doi.org/10.1186/s13690-018-0277-1

Paul, R., & Singh, A. (2020). Does early childhood adversities affect physical, cognitive and language development in indian children? Evidence from a panel study. *SSM - Population Health, 12*, 100693. https://doi.org/10.1016/j.ssmph.2020.100693

Paul, V. K., Singh, A., & Palit, S. (2018). POSHAN Abhiyaan: Making nutrition a janandolan. *Proceedings of the Indian National Science Academy, 84*(4), 835–841.

Pramanik, S., Muthusamy, N., Gera, R., & Laxminarayan, R. (2015). Vaccination coverage in India: A small area estimation approach. *Vaccine, 33*(14), 1731–1738.

Rao, J. (2014). *Small-area estimation*. Wiley StatsRef: Statistics Reference Online.

Rice, A. L., Sacco, L., Hyder, A., & Black, R. E. (2000). Malnutrition as an underlying cause of childhood deaths associated with infectious diseases in developing countries. *Bulletin of the World Health Organization, 78*, 1207–1221.

Saei, A., & Chambers, R. (2003). *Small area estimation under linear and generalized linear mixed models with time and area effects*.

Singh, A., Pathak, P. K., Chauhan, R. K., & Pan, W. (2011). Infant and child mortality in India in the last two decades: A geospatial analysis. *PloS One, 6*(11), Article e26856.

Singh, S., Srivastava, S., & Upadhyay, A. K. (2019). Socio-economic inequality in malnutrition among children in India: An analysis of 640 districts from national family health survey (2015–16). *International Journal for Equity in Health, 18*(1), 203.

Singh, A., Upadhyay, A. K., Singh, A., & Kumar, K. (2017). The association between unintended births and poor child development in India: Evidence from a longitudinal study. *Studies in Family Planning, 48*(1), 55–71. https://doi.org/10.1111/sifp.12017

Song, L., Mercer, L., Wakefield, J., Laurent, A., & Solet, D. (2016). *Using small-area estimation to calculate the prevalence of smoking by subcounty geographic areas in King County, Washington, Behavioral Risk Factor Surveillance System, 2009–2013*.

Sud, U., Chandra, H., & Srivastava, A. (2011). Crop yield Estimation at district Level by combining Improvement of crop statistics Scheme Data and census data. In *Paper presented at the wye city group on statistics on Rural development and agricultural household income, 4th Meeting, Rio de Janeiro, Brazil*.

Swaminathan, S., Hemalatha, R., Pandey, A., Kassebaum, N. J., Laxmaiah, A., Longvah, T., et al. (2019). The burden of child and maternal malnutrition and trends in its indicators in the states of India: The global burden of disease study 1990–2017. *The Lancet Child & Adolescent Health, 3*(12), 855–870.

Upadhyay, A. K., & Singh, A. (2020). *Are Indian girls really better nourished than Indian boys? Evidence from Indian national family health survey 1992–2016. Clinical Epidemiology and global health*.

Vicente, G., Goicoa Mangado, T., Puranik, A., & Ugarte Martínez, M. D. (2018). Small area estimation of gender-based violence: Rape incidence risks in Uttar Pradesh, India. *Statistics and Applications, 16*(No. 1), 71–90, 2018 (New Series).

Victora, C. G., Adair, L., Fall, C., Hallal, P. C., Martorell, R., Richter, L., et al. (2008). Maternal and child undernutrition: Consequences for adult health and human capital. *Lancet (London, England), 371*(9609), 340–357. https://doi.org/10.1016/S0140-6736(07)61692-4