*Journal of Survey Statistics and Methodology (2026) **14**, 209–237*

TOWARD A PRINCIPLED WORKFLOW FOR PREVALENCE MAPPING USING HOUSEHOLD SURVEY DATA

QIANYU DONG [†]
YUNHAN WU [†]
ZEHANG RICHARD LI *
JON WAKEFIELD *

Understanding the prevalence of key demographic and health indicators in small geographic areas and domains is of global interest, especially in low- and middle-income countries (LMICs), where vital registration data is sparse and household surveys are the primary source of information. Recent advances in computation and the increasing availability of spatially detailed datasets have led to much progress in sophisticated statistical modeling of prevalence. As a result, high-resolution prevalence maps for many indicators are routinely produced in the literature. However, statistical and practical guidance for producing prevalence maps in LMICs has been largely lacking. In particular, advice in choosing and evaluating models and interpreting results is needed, especially when data is limited. Software and analysis tools are also usually inaccessible to researchers in low-resource settings to conduct their own analysis or reproduce findings in the literature. In this paper, we propose a general workflow for prevalence mapping using household survey data. We consider all stages of the analysis pipeline, with particular emphasis on model choice and interpretation. We illustrate the proposed workflow using a case study mapping the proportion of pregnant women who had at least four antenatal care visits in

Qianyu Dong and Zehang Richard Li are with Department of Statistics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA. Yunhan Wu is with Department of Biostatistics, Hans Rosling Center for Population Health, University of Washington, Box 351617, Seattle, WA 98195, USA. Jon Wakefield is with Department of Statistics and Department of Biostatistics, Hans Rosling Center for Population Health, University of Washington, Room 331.1.C18, Box 351617, Seattle, WA 98195, USA
[†]These authors are co-first authors.
This work was supported by the National Institutes of Health [R01HD112421].
The authors do not have any conflicts of interest to report.
*Address correspondence to Zehang Richard Li, Department of Statistics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA; E-mail: lizehang@ucsc.edu and Jon Wakefield, Department of Statistics and Department of Biostatistics, Hans Rosling Center for Population Health, University of Washington, Room 331.1.C18, Box 351617, Seattle, WA 98195, USA; E-mail: jonno@uw.edu

https://doi.org/10.1093/jssam/smaf048 Advance access publication 6 February 2026

Kenya. The workflow is implemented using the R package *surveyPrev*, and all reproducible code is provided in the Supplementary Materials. It can be readily extended to a wide range of indicators.

KEY WORDS: Bayesian hierarchical models; Reproducibility; Small area estimation; Spatial models; Survey sampling.

Statement of Significance

We propose a principled workflow for estimating and mapping the prevalence of demographic and health indicators in small administrative areas using household survey data in low- and middle-income countries (LMICs). The workflow considers all stages of the data analysis pipeline and can be robustly adopted for a wide range of binary indicators. The aim of the workflow is to enable researchers and analysts in LMICs to carry out their own prevalence analysis according to their local needs and priorities. The workflow is illustrated through a case study mapping the proportion of pregnant women who had at least four antenatal care visits in Kenya, with reproducible code and R package in the Supplementary Materials.

1. INTRODUCTION

Accurate estimation and tracking of key demographic and health indicators is critical for assessing health progress and disparities. National statistical institutions and global organizations have long recognized the need to produce disaggregated estimates at fine resolutions. For example, the United Nations General Assembly's 2030 Agenda for Sustainable Development (United Nations 2015) established targets "to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts." The problem of estimating quantities of interest for small subpopulations based on limited data is known as small area estimation (SAE). SAE has a long history in the survey sampling literature. Traditional SAE approaches have been largely developed in the high-income country context and often utilize auxiliary administrative data to improve inference (Rao and Molina 2015). Recent applications of SAE approaches in LMICs include work on mapping poverty (Molina et al. 2019; Chi et al. 2022), child mortality (Mercer et al. 2015; Li et al. 2019; Wakefield et al. 2019; Wu et al. 2021), fertility (Saha et al. 2023), HIV prevalence (Wakefield et al. 2020), and vaccination coverage (Dong and Wakefield 2021). More recently, another line of research using geostatistical models has been adopted by major research organizations to map the prevalence of a number of important health indicators (see e.g. Diggle and Giorgi 2016, 2019;

Burstein et al. 2018; Osgood-Zimmerman et al. 2018; Utazi et al. 2018; Mayala et al. 2019). Geostatistical models theoretically enable prevalence estimation at a fine spatial resolution, for example, 100 m grids, and they leverage the growing availability of spatially referenced covariates such as population density (Tatem 2017) and night time light (Gaughan et al. 2019). A recent review and comparison of the two types of approaches is provided in Wakefield et al. (2025).

While rapid progress has been made in methodological developments of mapping prevalence in small areas, the existing literature often focuses on improving specific steps of the full SAE workflow. Guidance on conducting prevalence mapping from the beginning to the end has been largely lacking, especially for practitioners. Two recent papers, Tzavidis et al. (2018) and Giorgi et al. (2021), discuss the significance of having a coherent process across different stages of analysis when producing small area official statistics. However, scenarios where data is scarce remain challenging under both frameworks. Tzavidis et al. (2018) examined the production of official statistics from the perspective of national statistical institutes. Their framework follows a traditional SAE approach and focuses on continuous outcomes, for example, average income. The models for continuous outcomes are usually inappropriate for estimating the prevalence of categorical indicators, which is the focus of this paper. In addition, their framework emphasizes using covariates from census or administrative microdata in order to construct synthetic estimates at fine geographic scale. In LMICs, the issues with data availability are quite different from those faced in high-income countries. Auxiliary datasets from census are usually unavailable, or of poor quality and are often not updated in a timely fashion. Thus, it is important to exploit the spatial dependence in the response variable that cannot be explained by the covariates that are available to the analyst. This aspect is largely ignored in Tzavidis et al. (2018). On the other hand, Giorgi et al. (2021) provided a parallel set of guidance on developing prevalence models using geostatistical models, focusing on using high-quality spatially referenced covariates for point-level spatial prediction. In contrast with Tzavidis et al. (2018), some key considerations in the traditional SAE methods, for example, accounting for the sampling design of a survey, are mostly ignored. Notably, both Tzavidis et al. (2018) and Giorgi et al. (2021) were primarily written for statisticians capable of building and comparing sophisticated models. Translating the ideas to routine analytic procedures involving large numbers of indicators can be difficult for practitioners.

In this paper, we aim to fill this gap by developing a principled and accessible workflow for common types of prevalence mapping tasks in data-limited contexts with fully worked-out examples. Specifically, we focus on estimating the prevalence of binary indicators, that is, the proportion of individuals affected by certain conditions. We develop a 'default' analysis pipeline with a series of models for prevalence mapping that can be adopted robustly for a wide range of indicators, and all computation can be carried out using the

surveyPrev package (Dong et al. 2024) in R (R Core Team, 2024). Our proposed procedures acknowledge the sampling design explicitly and account for spatial dependence in the prevalence. The models are computationally scalable, making them feasible to implement in low-resource settings. Altogether, the workflow generates a robust set of prevalence estimates and insights that can form the basis for further extensions when bespoke analyses are needed.

The rest of the paper is organized as follows. Section 2 discusses the common types of household survey data in LMICs. Section 3 describes the steps in the workflow. Section 4 concludes with a discussion of the proposed workflow, active research areas, and open questions. We demonstrate the proposed workflow via a case study mapping the proportion of pregnant women who had at least four antenatal care visits using a Demographic and Health Survey (DHS) that was carried out in Kenya, in 2022.

2. HOUSEHOLD SURVEY DATA

In many LMICs, household surveys are routinely conducted to collect information on health and demographic variables. Prominent examples include the DHS (USAID 2019), the Malaria Indicators Survey (MIS) (Roll Back Malaria 2005), and the Multiple Indicator Cluster Surveys (MICS) (Khan and Hancioglu 2019). We consider the 2022 Kenya DHS (KNBS and ICF 2023) as our working example. The survey is based on a sampling frame derived from the 2019 Kenya Population and Housing Census, in which a total of 129,067 enumeration areas (EAs), or clusters, were formed. In the 2022 Kenya DHS, the country is first stratified into rural and urban strata within each of the 47 counties, which results in 92 strata (Nairobi City and Mombasa are purely urban). Within each stratum, clusters are sampled independently using probability proportional to size (PPS) random sampling, with households being the size variable. This resulted in 1,692 clusters being sampled across strata. Then 25 households were selected from each cluster and members of the households were interviewed to provide information on health and demographic variables. The design weight for each individual is calculated as the reciprocal of the sampling probability for the individual. The final weight also includes a non-response adjustment.

Household surveys in LMICs, including DHS, MICS, and MIS, often follow similar sampling designs. For ease of description, we refer to the principal administrative divisions used to define the stratification, for example, the 47 counties in Kenya, as the Admin-1 areas, and the next level of subdivisions as the Admin-2 areas. Note that the geographical stratification is survey-specific and may be Admin-1 or Admin-2, depending on the country. The number of administrative areas varies dramatically across countries. For example, there are 33 Admin-2 areas in Côte d'Ivoire and 774 in Nigeria. The sample size in most household surveys is designed to provide estimates of key indicators at

the national and Admin-1 level only. Data are typically sparse at the Admin-2 level.

The DHS program routinely calculates the national estimates of over 2,500 indicators for each survey (Croft et al. 2023). The majority of these indicators are binary. As a working example, we consider the prevalence of attending four or more antenatal care visits (ANC4+) among women who had a live birth or a stillbirth in the 5 years prior to the survey. Antenatal care is a crucial component in reducing maternal and perinatal mortality. The World Health Organization (WHO) recommended at least four ANC visits with a qualified healthcare provider during pregnancy (World Health Organization 2002) and a global ANC4+ coverage target of 90 percent by 2025, in order to meet the SDG goals on maternal mortality by 2030. Several studies have examined the ANC4+ coverage in Kenya using past DHS surveys, see for example Wairoto et al. (2020) and Macharia et al. (2022). The national estimates of ANC4+ coverage in Kenya have been steadily increasing, from 44 percent in 2009 to 66 percent in 2022 (KNBS and ICF 2023).

3. THE WORKFLOW

Our prevalence mapping workflow starts by specifying a level of inference. Typically, household surveys powered to Admin-1 areas can robustly produce national and Admin-1 level estimates. Inference at finer resolutions relies more heavily on model assumptions, as we detail later. In this paper, we consider prevalence mapping at both Admin-1 and Admin-2. The same workflow follows for finer resolutions, though data sparsity will be more severe. Our proposed workflow consists of four stages summarized in figure 1:

Stage 1: Understand the sampling design and data availability
Stage 2: Conduct analyses with a range of models
Stage 3: Evaluate and compare estimates
Stage 4: Summarize, map, and visualize the prevalence estimates

The full workflow is carried out using the *surveyPrev* package (Dong et al. 2024). A reproducible report with all analysis scripts is included in the Supplementary Materials.

3.1 Stage 1: The Sampling Design and Data Availability

Understanding the survey data and how they are collected is critical in any modeling tasks, including prevalence mapping. Data availability has important implications for model choices. In our working example, we use the GPS coordinates of the clusters, which are available for most DHS surveys and some MICS surveys, and administrative boundaries from the GADM database (Global Administrative Areas 2022) to determine which area each cluster

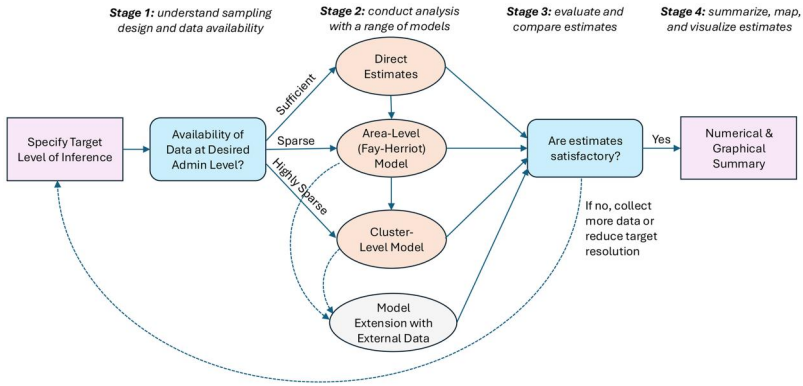

Figure 1. Workflow for prevalence mapping using household survey data.

belongs to. Mapping coordinates to areas may create incorrect assignments for clusters due to jittering or imprecise boundary maps. Comparing the area assignments with spatial information recorded in the survey, for example, Admin-1 labels, is recommended whenever such information exists. An example of the detailed steps in correcting potentially mis-assigned clusters is included in the Supplementary Materials. In practice, we expect the potential impact to be minor unless the geography is very small. Further discussion of this issue can be found in Altay et al. (2025). Figure 2 shows the location of sampled clusters across Admin-1 and Admin-2 areas, by urban and rural strata. The median number of clusters is 35 across the 47 Admin-1 areas and 5 across the 300 Admin-2 areas. There are 6 Admin-2 areas without a sampled cluster and 5 with only one sampled cluster. The sample size of women who had a birth or stillbirth within the 5 years preceding the survey has a median size of 218 across Admin-1 areas and 29 across Admin-2 areas. While it is difficult to provide a specific threshold for sample size, both the number of clusters and the individuals are reasonably large for Admin-1 areas but limited in most of the Admin-2 areas.

Another important consideration is the effect of within-area urban/rural stratification. Most of the household surveys in LMICs, including DHS and MICS, are stratified by urban/rural status within Admin-1 areas. The population proportions of urban/rural clusters are usually released in the survey report or are available from the census. Figure 3 shows the sampled proportion of urban clusters in each Admin-1 area compared to population proportions. We observe that urban clusters are over-sampled in most of the areas. The coverage of ANC4+ is higher in urban areas (the odds ratio from a weighted logistic regression 1.55). Therefore, models that do not use the weights and do not account for the stratification will overestimate the prevalence. The exact impact of the over-sampling of urban areas can only be determined by analyzing data under a specific analysis model, which we investigate further in

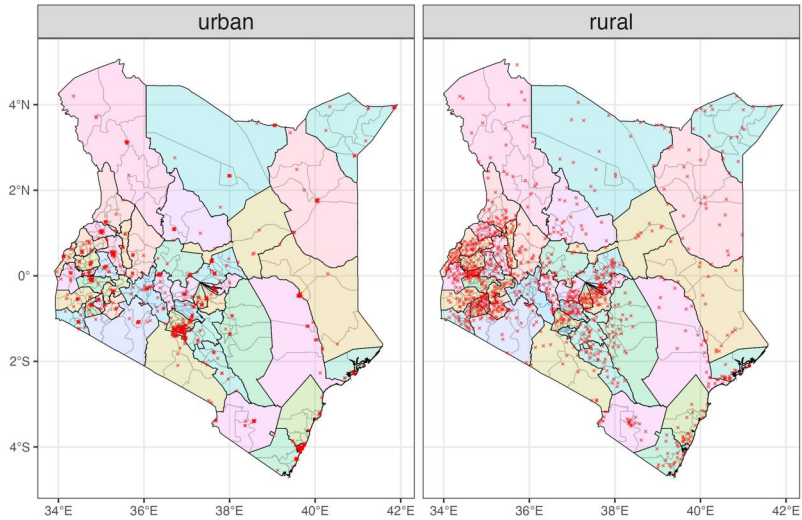

Figure 2. Map of the sampled clusters in the 2022 Kenya DHS by urban/rural status.

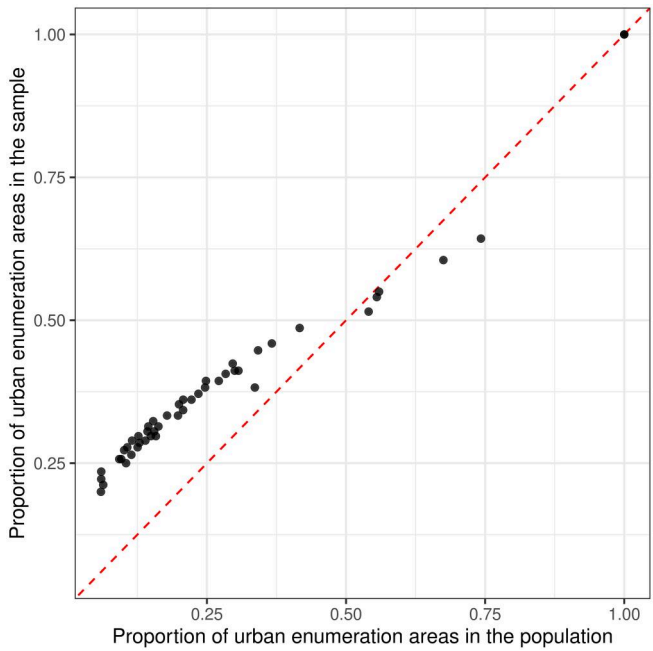

Figure 3. Comparison between the fraction of sampled clusters that are urban and the fraction of population urban clusters (enumeration areas) in each Admin-1 area.

Section 3.2.3. Details of the steps in this stage can be found in Sections 2.2 and 2.3 of the Supplementary Materials.

3.2 Stage 2: Models and Analysis

We propose three sets of models for routine analysis: direct estimation, area-level models, and cluster-level models. The overall approach we recommend is to fit all models, where possible, to assess sensitivity but to default to simpler models if they provide sufficiently precise estimates. In the rest of this section, we describe the three classes of models and discuss their advantages and limitations. We have omitted sensitivity analysis for some model specifications, such as prior choices, as the results in our example are not sensitive to these choices, but it is good practice to conduct more comprehensive sensitivity analysis in practice (Gelman et al. 2020; Wakefield et al. 2025).

To fix notation, consider a finite target population in a country with N individuals living in M areas (Admin-1 or Admin-2, depending on the target resolution). Let $y_j \in \{0, 1\}$ denote the outcome value for the j-th individual. We let p_i denote the area-level prevalence for the i-th area. We let U_i be the set of individuals in the target population in the i-th area with size $N_i = |U_i|$ and $S_i \in U_i$ denote the set of $n_i = |S_i|$ sampled individuals in the i-th area. Then the true prevalence in the i-th area is $p_i = (1/N_i) \sum_{j \in U_i} y_j$. For all sampled individuals, we observe the design weight w_j, which is the inverse probability of the j-th individual being included in the sample for $j \in S_1 \cup \ldots \cup S_M$.

3.2.1 Direct estimation.

Our starting point for prevalence analysis is direct estimation. The term 'direct' refers to estimating the prevalence of an area using only the response data from that area (Rao and Molina 2015). An example of a direct estimator is the design-based weighted estimator (Hájek 1971),

$$\widehat{p}_i^w = \frac{\sum_{j \in S_i} w_j y_j}{\sum_{j \in S_i} w_j}, \quad i = 1, \ldots, M.$$

The variance of $\widehat{p}_i$ can be easily estimated with standard software for survey data. The direct estimates and the associated uncertainty measure account for the weighting and clustering in the survey design, avoiding the bias due to informative sampling. When there are many clusters in the sample for a given area, confidence intervals constructed from the normality assumption will be accurate for the area. Direct estimation is based on minimal assumptions since there is no explicit model for the data. In practice, direct estimation often provides reasonable inference for Admin-1 areas, unless the outcome is very rare. The precision associated with the direct estimates, however, is usually comparatively small, since the estimates in each area only use data from that area

alone. This can be contrasted with the area-level and unit-level approaches, which borrow strength from the totality of data across all areas and thus leverage more information.

When data are sparse in an area, the uncertainty of the estimator can be too high to be usable. When there are a small number of clusters in an area, the point estimate of the prevalence may be 0 or 1, and in these cases, the usual formulas for calculating the variance equal zero. Other sparse data configurations may also produce unreasonably small estimates of uncertainty. In addition, even when point estimates and standard errors can be computed, the resultant uncertainty interval, based on the asymptotic normality assumption, can be inaccurate when the sample size is small.

Figure 4 shows the point estimates and 95 percent uncertainty intervals of different models we consider in the workflow. The point estimates are mostly comparable across all models at Admin-1. Figure 5 visually compares two commonly reported metrics of uncertainty: the coefficient of variation (CV), $\sqrt{var(\widehat{p}_i)}/\widehat{p}_i$, and the width of the 95 percent uncertainty intervals. Both metrics are typically larger for the direct estimates compared to the smoothing models. Figure 6 shows the same comparison for Admin-2 estimates. The larger uncertainty of the direct estimates in finer resolutions can be observed more clearly. The widest 95 percent uncertainty interval for Admin-1 direct estimates is 0.242, and more than 60 percent of Admin-2 areas have 95 percent uncertainty intervals that surpass this. Details of the implementation of direction estimation can be found in Section 3 of the Supplementary Materials.

3.2.2 Area-level models.

Area-level models, or Fay-Herriot models (Fay and Herriot 1979), are the most common SAE method and have seen widespread use in official statistics.

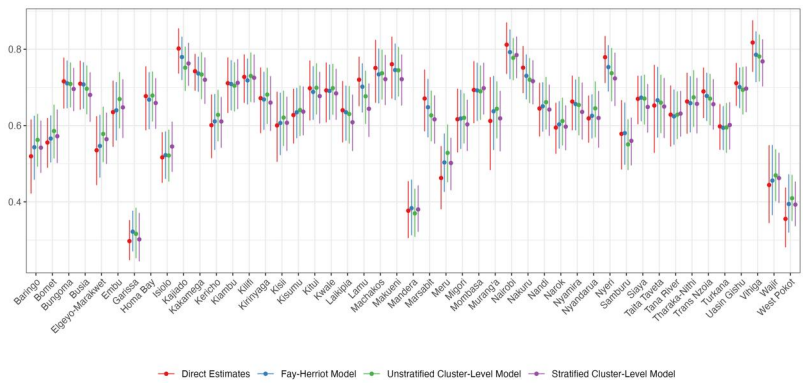

Figure 4. Point estimates and 95 percent uncertainty intervals for ANC4+ coverage across Admin-1 areas in Kenya, estimated using different models. Covariates are not included in all models.

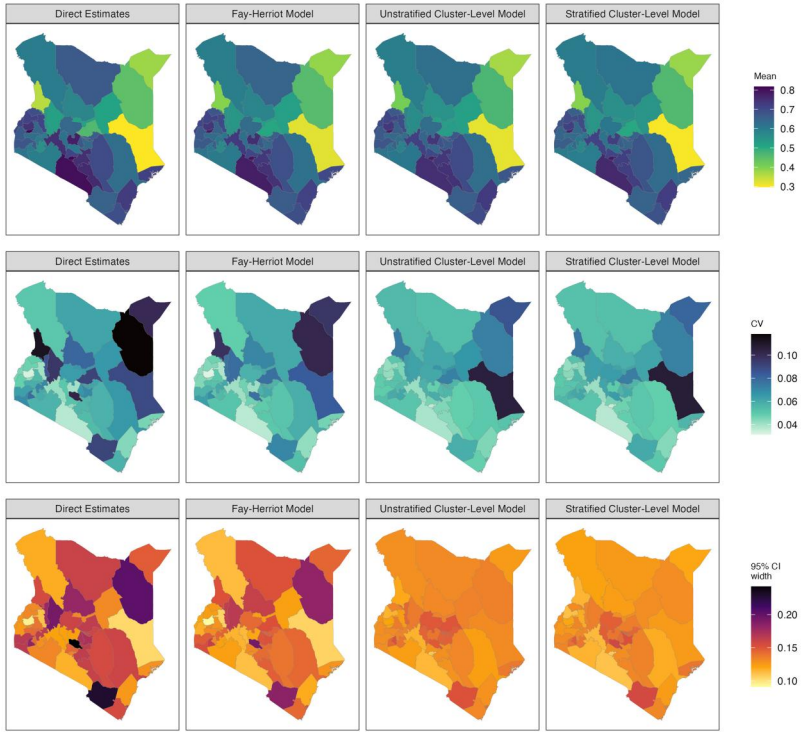

Figure 5. Maps of estimated prevalence (top row), coefficient of variation (middle row), and the width of the 95 percent uncertainty interval (bottom row) for ANC4+ coverage across Admin-1 areas in Kenya, estimated using different models. Covariates are not included in all models.

Under this approach, the direct estimates are linked together via a hierarchical model. For prevalence modeling, we first transform the direct estimates to the entire real line via $\widehat{\theta}_i^w = \text{logit}(\widehat{p}_i^w)$, where $\text{logit}(p) = \log(p/(1-p))$, and derive the associated asymptotic variance estimate of $\widehat{\theta}^w$ using the delta method, that is, $\widehat{V}_i = \widehat{var}(\widehat{p}_i^w)/[\widehat{p}_i^w(1-\widehat{p}_i^w)]^2$. The Fay-Herriot model assumes

$$\widehat{\theta}_i^w | \theta_i \sim N(\theta_i, \widehat{V}_i) \tag{1}$$

$$\theta_i = \alpha + \boldsymbol{x}_i^T \boldsymbol{\beta} + u_i, \quad i = 1, \ldots, M, \tag{2}$$

where α is the intercept, $\boldsymbol{x}_i$ are area-level covariates and $\boldsymbol{\beta}$ are the associated coefficients. Equation (1) is referred to as the *sampling model* and equation (2) the *linking model*. The latent prevalence at the i-th area is then $\text{expit}(\theta_i)$, where $\text{expit}(\theta) = \exp(\theta)/(1+\exp(\theta))$. The original Fay-Herriot model assumed that the random effects u_i are independent and identically distributed (IID) from a normal distribution. As discussed before, we usually expect residuals to

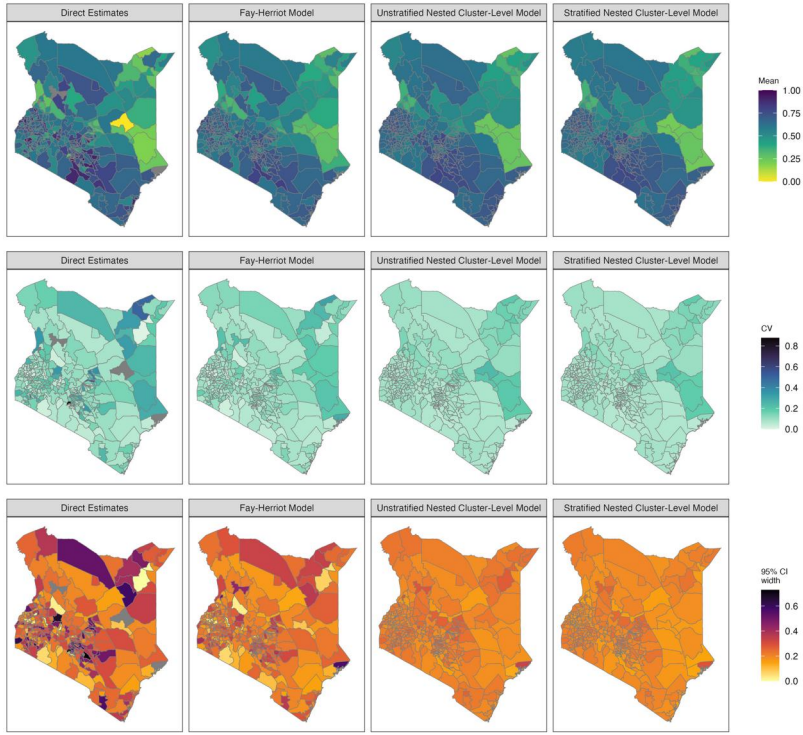

Figure 6. Maps of estimated prevalence (top row), coefficient of variation (middle row), and the width of the 95 percent uncertainty interval (bottom row) for ANC4+ coverage across Admin-2 areas in Kenya, estimated using different models. Covariates are not included in all models.

remain spatially correlated when covariates are limited. A wide range of spatial random effect models have been proposed and compared in the literature (see e.g. Chandra et al. 2019; Chung and Datta 2020). One approach that we recommend as a default choice is a model that decomposes the random effect into the sum of an unstructured IID normal random effect and a spatially structured intrinsic conditional autoregressive (ICAR) random effect, which is known as the BYM model (Besag et al. 1991). In our analysis, we adopt the reparameterization known as the BYM2 model (Riebler et al. 2016),

$$\boldsymbol{u} = \sigma(\sqrt{1-\phi}\mathbf{e}^{\text{unstruct}} + \sqrt{\phi}\boldsymbol{e}^{\text{struct}}),$$

where σ is the total standard deviation, ϕ is the proportion of the variance that is spatial, $\mathbf{e}^{\text{unstruct}}$ is IID standard normal random variable and $\boldsymbol{e}^{\text{struct}}$ follows a scaled ICAR prior so that the geometric mean of the marginal variances of e_i^{struct} is equal to 1. We adopt the penalised complexity (PC) priors for the

hyperparameters (Simpson et al. 2017). The PC priors provide robust shrinkage of model components to simpler base models and have been used successfully in various contexts (see e.g. Sørbye and Rue 2017; Fuglstad et al. 2020). Compared to the standard conjugate priors, the PC priors are specified by more intuitive probability statements. In our working example, for the total standard deviation, we specify $\text{Prob}(\sigma > 1) = 0.01$ and for the proportion of the variation that is spatial, we have $\text{Prob}(\phi > 0.5) = 2/3$. Posterior computation is performed with the integrated nested Laplace approximation (INLA) approach (Rue et al. 2009). INLA is extremely popular for carrying out Bayesian inference for dependent data in general and spatial data in particular (Blangiardo and Cameletti 2015; Krainski et al. 2018; Osgood-Zimmerman and Wakefield 2023) because it is accurate and very fast. Under the Bayesian framework, inference is conducted by evaluating the posterior distributions and credible intervals of the model parameters. In our context, the parameter of interest is the latent prevalence, $\text{expit}(\boldsymbol{\theta})$.

The rationale for the Fay-Herriot model is that the similarity of prevalences across the study areas may be leveraged to fine-tune the estimates in each area. Since data from all areas are used for estimation in each area, this is an example of an "indirect" estimate. The model accounts for the sampling design by using the design-based estimates, $\widehat{p}_i^w$, and their variances, $\widehat{var}(\widehat{p}_i^w)$ as the input. The latter goes to zero as the sampling fraction in an area goes to 1, and the weighted estimate equals the population prevalence in the area so that the estimates produced by the model are design consistent, which is very appealing. By collectively modeling all the data, uncertainty in each area is reduced on average. The model that links the areas together introduces additional modeling assumption, though this is typically not strong. We expect on average the mean squared errors of the estimates to be reduced compared to direct estimation, as the gain in precision will offset the bias due to shrinkage. Interpreting the uncertainty in each area based on hierarchical models, however, is more tricky. Across the map the coverage of interval estimates will be close to nominal, but they do not have the usual frequentist coverage when each area is viewed in isolation because of the shrinkage. This issue is discussed in detail in Burris and Hoff (2020).

A key disadvantage of Fay-Herriot models is that they are based on direct estimates. When data are extremely sparse, direct estimates or their variance estimates are unavailable or unstable, which renders the Fay-Herriot model unusable or unreliable. To alleviate variance instability, the sampling variances may be modeled using generalized variance functions, which leverage the mean-variance relationship between θ_i and V_i, as well as covariates (Otto and Bell 1995; Mohadjer et al. 2012; Franco and Bell 2013; Liu et al. 2014; Gao and Wakefield 2023). When fitting the Admin-2 level Fay-Herriot model, the design-based variance formula does not produce an estimate, or produces an estimate that is zero or close to zero for 12 out of 300 areas. For these Admin-2 areas, we supplement the data with a "phantom" cluster that has prevalence

equal to the prevalence in the Admin-1 area within which the Admin-2 area is contained, and with the sum of the weights equal to the average sum across all observed clusters in the Admin-1 area. For further details, see Wakefield et al. (2025).

Figures 4 to 6 present the results of the Fay-Herriot model where no covariates are included We observe that the Fay-Herriot model estimates are, in general, shrunk from the direct estimates, though to a lesser extent compared to the cluster-level models that we describe next. Details of the implementation of area-level models can be found in Section 4 of the Supplementary Materials.

3.2.3 Cluster-level models.

SAE models dealing with individual outcomes directly are known as unit-level models. In our context, modeling individual responses (the "units") using a Bernoulli likelihood is equivalent to a cluster-level binomial model. More specifically, for a survey consisting of C clusters, we let $c[j]$ be the cluster index of the j-th individual. When there is no ambiguity of resolution, we use $i[c]$ to denote the index of the area that the c-th cluster resides in. For nested models where both Admin-1 and Admin-2 level components are included, we use $a[c]$ and $a[i]$ to denote the corresponding Admin-1 index for the c-th cluster or the i-th Admin-2 area. Let Y_c and n_c denote the number of positive outcomes and the number of sampled individuals from the c-th cluster, respectively, that is, $Y_c = \sum_{j:c[j]=c} y_j$ and $n_c = \sum_{j:c[j]=c} 1$. We consider a cluster-level model with a beta-binomial likelihood,

$$Y_c | p_c \sim \text{BetaBinomial}(n_c, p_c, d).$$

In this formulation, p_c is the latent prevalence for the c-th cluster, and d is the overdispersion parameter that captures the additional within-cluster variation. This class of beta-binomial models has been used in modeling a variety of indicators in the recent literature. A detailed review of different binomial models was carried out in Dong and Wakefield (2021).

To link the cluster-level prevalence to area-level prevalence, a simple unstratified model is

$$\text{logit}(p_c) = \alpha + \boldsymbol{x}_c^T \boldsymbol{\beta} + u_{i[c]}.$$

The intercept and spatial random effects are modeled in the same way as in the Fay-Herriot model, and $\boldsymbol{x}_c$ are cluster-level covariates. The covariates may also be at the area level, which simplifies the aggregation step discussed in Section 3.2.4. When modeling at Admin-2 level, it is often also reasonable to include a separate intercept for each Admin-1 level to reduce the shrinkage across Admin-1 areas. This leads to a nested unstratified model,

$$\text{logit}(p_c) = \alpha_{a[c]} + \boldsymbol{x}_c^T\boldsymbol{\beta} + u_{i[c]}.$$

The stratification over Admin-1 areas is either implicitly accounted for by the inclusion of random effects, $u_{i[c]}$ or explicitly by $\alpha_{a[c]}$. We use the term "unstratified" to refer to the fact that the model does not account for any within-area stratification, as discussed in Section 3.1. To account for the additional stratification by urbanicity, we consider the following three stratified models:

$$\text{(non-nested)}\quad \text{logit}(p_c) = \alpha + \gamma \boldsymbol{1}_{c\in\text{rural}} + \boldsymbol{x}_c^T\boldsymbol{\beta} + u_{i[c]},$$

$$\text{(nested)}\quad \text{logit}(p_c) = \alpha_{a[c]} + \gamma \boldsymbol{1}_{c\in\text{rural}} + \boldsymbol{x}_c^T\boldsymbol{\beta} + u_{i[c]},$$

$$\text{(nested with interaction)}\quad \text{logit}(p_c) = \alpha_{a[c]} + \gamma_{a[c]} \boldsymbol{1}_{c\in\text{rural}} + \boldsymbol{x}_c^T\boldsymbol{\beta} + u_{i[c]}.$$

All three models account for the within-area stratification by including terms corresponding to the urban/rural status. The first two models assume the urban/rural effect, γ, is constant across all areas. The third model further assumes that the urban/rural associations vary over Admin-1 areas. Since the number of Admin-1 areas is typically small, we treat γ_a in the third model as fixed effects. Selecting which model to use is context-specific. In our example, a likelihood ratio test (Lumley and Scott 2014) for the interaction terms is included in the Supplementary Materials, together with further model diagnostics of all three models in terms of the WAIC scores. Both procedures favor the nested model over the other two. There is not enough evidence for spatially varying urban/rural associations. Therefore, we focus on the nested model in the rest of the paper.

It is worth noting that most of the geostatistical models in the literature ignore the stratification variable. Giorgi et al. (2021) argued that the sampling design could be implicitly accounted for by including adequate covariates. In practice, however, it is difficult to assess whether a specific set of covariates is sufficient to adjust for the sampling design. Thus, we view the inclusion of the stratification variable as a necessary step, as long as the prevalences are associated with the urban/rural designation and there is evidence of over-sampling of urban or rural clusters.

The cluster-level models can deal with very sparse data, as they do not rely on the asymptotic properties of the direct estimates. The existence of data idiosyncrasies, such as zero response counts, can be easily accommodated by modeling the binary outcome instead of weighted estimators, since they are simply captured by a discrete probability model. If the parametric likelihood model is a reasonable approximation to the true underlying data generating mechanism, this approach is an efficient use of the data. Similar to the Fay-Herriot model, the cluster-level models also introduce bias, due to shrinkage, but reduce uncertainty. Hence, the mean squared errors of the estimates will be reduced on average. On the other hand, a major limitation of the

cluster-level models is the assumption of an individual-level sampling model, which cannot be easily checked. Dong and Wakefield (2021) provided a more in-depth discussion of different sampling models for binary outcomes, and the beta-binomial model we consider here is among the best-performing candidates in their empirical comparison. Also, cluster-level models are more computationally intensive to fit, though this is typically not an issue with a scalable implementation using INLA. Last, and not least, additional post-processing steps are necessary to aggregate cluster-level predictions into area-level estimates, and external information is often needed, which we discuss in the next two subsections. Details of the implementation of these cluster-level models can be found in Sections 5.1 and 5.2 of the Supplementary Materials.

3.2.4 Cluster-level aggregation with area-level covariates.
The inclusion of the urban/rural effect and cluster-level covariates comes at a cost when producing area-level estimates. We first consider the case with area-level covariates x_i. For the unstratified model, the prevalence of the i-th area is simply $\theta_i = \text{expit}(\alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i)$. For the stratified model, however, an additional aggregation step is needed. As an example, for the nested model,

$$\theta_i = (1 - q_i) \times \text{expit}(\alpha_{a[i]} + \gamma + \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i) + q_i \times \text{expit}(\alpha_{a[i]} + \boldsymbol{x}_i^T\boldsymbol{\beta} + u_i),$$

where q_i is the proportion of the population in area i that is urban. The population parameter q_i usually cannot be reliably estimated from the survey at fine resolutions. The interpretation of q_i is also subtle. The urban/rural designation in household surveys is usually based on a previous census classification. The actual status of a cluster can change over time. Thus, q_i represents the fraction of the target population who, at the time of the survey, reside in areas that were classified as urban in the last census.

Two challenges exist in obtaining these fractions. First, the complete list of clusters and their urban/rural designation is never released to the public. The proportion of urban population is sometimes reported, but usually at the Admin-1 or national level only. Second, the target population can vary significantly across indicators, and the total population of a country may not be a good proxy for it. For example, when estimating ANC4+ coverage, the target population consists of women who had a recent birth or stillbirth, while for estimating the neonatal mortality rate, the target population consists of all births within the relevant time period. Therefore, these aggregation fractions need to be carefully specified and estimated from external data sources.

Reconstructing q_i forms a line of research on its own. One recent proposal using classification models to reconstruct urban/rural partitions can be found in Wu and Wakefield (2024). Here we adopt a simple thresholding approach that does not require sophisticated model tuning or high-quality external data. The approach relies on the simplifying assumption that the urban locations

have higher population density than the rural locations. The process consists of the following steps:

(1) Identify the year when the sampling frame was defined (usually the last census). Denote this year as T_0, and the year the survey took place as T. Obtain the total population raster for the country at year T_0 from sources such as WorldPop (Tatem 2017).
(2) For each Admin-1 area, use the population raster to identify a threshold such that the proportion of the population living in pixels with higher population density matches the reported fraction of urban population in this area at T_0. These reported fractions can usually be found in the survey report. The pixels with population above the corresponding threshold form the spatial partition of urban areas.
(3) Identify the age-sex-specific sub-population with the closest match to the target population at year T. For example, for ANC4+, we use the female population of age 15-49 as the proxy target population from the 2022 gridded population estimates for Kenya (Gadiaga et al. 2023).
(4) Use the pixel-level population map of this sub-population at year T and the urban partition to compute the fraction of the target population residing in the area designated as urban at year T.

Given the complexity of constructing the aggregation weights, in practice, we recommend first fitting both the unstratified and stratified models without aggregation. The urban- and rural-specific estimates from the stratified model allow one to gauge how sensitive the aggregated results are regarding the unknown urban fraction. If the log odds ratio associated with urban/rural, γ, is estimated to be very close to 0, the aggregated results will be very similar to the unstratified model regardless of the aggregation weights. In our analysis, γ is estimated to be -0.453 (95 percent credible interval: $[-0.564, -0.343]$), which indicates lower coverage of ANC4+ in rural areas on average. Given the over-sampling of urban clusters discussed in Section 3.1, and the significant association between ANC4+ and urbanicity of clusters, the unstratified models are likely to be biased and produce higher estimates. This is confirmed in figure 7. Therefore, the stratified model is preferred.

Figure 6 shows the estimates produced by the nested cluster-level models without covariates. Compared to the Fay-Herriot model, Admin-2 level estimates show greater shrinkage towards the Admin-1 average. This is expected when data are sparse and do not contain sufficient information for the model to deviate from being 'flat' within each Admin-1 area. Without the Admin-1 fixed effect, the estimates are more likely to be shrunk towards the overall mean, which produces a map that is too flat. Details of the implementation of the area-level covariate models can be found in Section 5.3.1 of the Supplementary Materials.

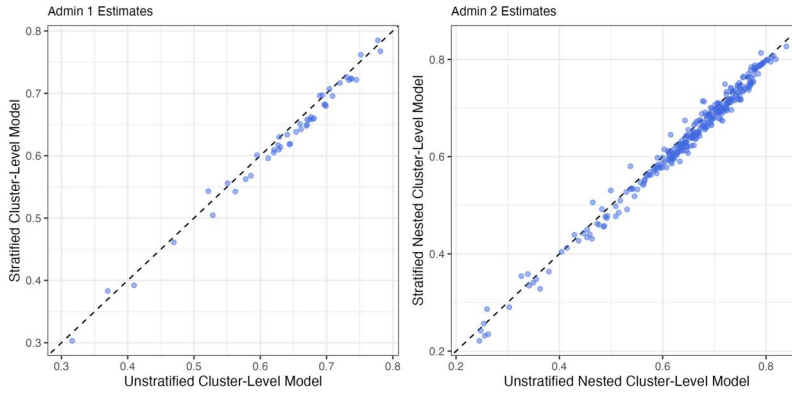

Figure 7. Comparing the estimated prevalence using the unstratified and stratified models at the Admin-1 level (left) and Admin-2 level (right). Covariates are not included in the models.

3.2.5 Cluster-level aggregation with cluster-level covariates.
When cluster-level models are used, the aggregation step becomes more complicated, as predictions must be made for both the sampled and unsampled clusters. Take the nested stratified model, for example,

$$\theta_i = \sum_{c:i[c]=i} q'_c \times \text{expit}(\alpha_{a[c]} + \gamma \mathbf{1}_{c\in\text{rural}} + \boldsymbol{x}_c^T\boldsymbol{\beta} + u_i),$$

where q'_c is the proportion of population in the c-th cluster within i-th area. Notice that the summation is over all clusters in the population, rather than the sampled clusters. In practice, the population frame is almost always unknown to researchers, thus the summation over clusters are usually approximated by a summation over a fixed set of grids where population and covariate values are known. Compared to the models with only area-level covariates in Section 3.2.4, stronger modeling assumptions are needed here as the aggregation step requires classifying each pixel as urban or rural, instead of just estimating the urban population proportion. We follow the same thresholding procedure in Section 3.2.4 to assign urban/rural status to each pixel. Details of the implementation of the cluster-level covariate models can be found in Section 5.3.1 of the Supplementary Materials.

3.2.6 Inclusion of covariates.
In general, covariates strongly related to the indicator of interest can usually improve prevalence estimates. Cluster-level covariates are often more informative than area-level covariates. However, with non-linear models, aggregation requires them to be known for all clusters in the population. In practice, this requirement is rarely attained and instead, we require the covariates to be

available on a grid, which approximates the cluster-level availability but leads to a further approximation step when aggregation from cluster to area is carried out. We also focus on improving prevalence estimates rather than interpreting the association between covariates and the outcome. Therefore, while some covariates may not be causally related to the outcome of interest, we may still choose to include them in the model if they are correlated with the outcome. Given the common scarcity of available covariates, we omit the discussion of variable selection in this paper. Often, this step will be carried out in a more informal manner and will be strongly based on the availability of well-measured covariates. We have a preference for avoiding covariates that are the subject of excessive modeling before they are used. Readers who wish to examine more formal procedures are referred to procedures in the literature, for example, Hoeting et al. (2006), Molina et al. (2019) and Michal et al. (2023).

As an illustrative example, we consider four covariates in our analysis: the total population (Tatem 2017), night time lights (Román et al. 2018), vegetation index (Didan et al. 2015), and travel time to the nearest healthcare facility (Weiss et al. 2020). Both nighttime lights and vegetation index are averaged over the period of 2018–2022. We consider both the pixel-level covariates and area-level covariates by aggregating them into Admin-1 or Admin-2 level averages, weighted by the population at each pixel. The covariates are standardized to have mean 0 and unit variance. Figure 8 shows the estimated regression coefficients for Admin-2 level cluster-level models using either area- or pixel-level covariates. The signs of the associations between each covariate and the outcome are consistent among models. Comparing the two cluster-level models, the effects of covariates are slightly attenuated after accounting for the urban/rural designation of the clusters. Figure 9 shows that the final Admin-2 level estimates do not change much with the inclusion of covariates, which has also been found in other work (see e.g. Wakefield et al. (2019)). However, we would rather explain between-area differences in terms of covariates when possible. We emphasize that the four predictors in this analysis are chosen primarily for illustration, and more future work, both methodological and through practical examples, is required with respect to covariate choice and modeling.

3.3 Stage 3: Model Evaluation and Estimate Comparison

Evaluating and comparing estimates from different models is an important step in producing small area statistics. It is often helpful to first rule out models that are inappropriate to use. While not exhaustive, we find the following steps often provide a useful path towards this goal.

(1) Are there sufficient data? Areas without samples cannot produce a direct estimate. Thus, if too many areas have no data or only sparse data, direct

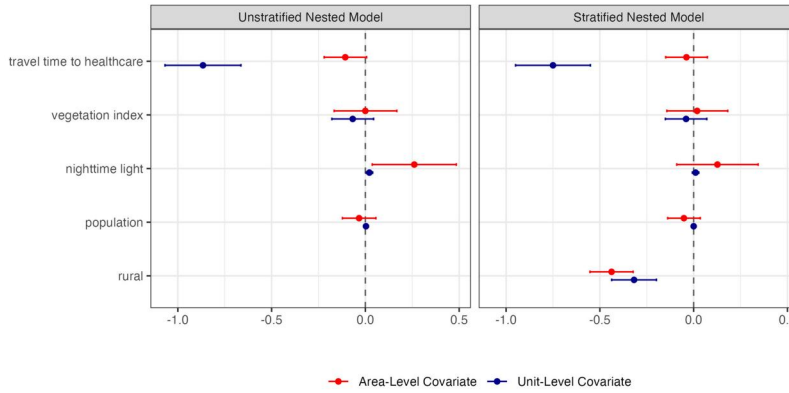

Figure 8. Comparing the estimated mean and 95 percent credible intervals of the coefficients of the fixed effects in the three models at both Admin-1 and Admin-2 levels.

estimates will contain many missing values. Consequently, estimates of Fay-Herriot models are also less trustworthy, as they rely heavily on random effects and covariates to impute the prevalence of those areas. This is more likely to occur when modeling at fine resolutions, and in this scenario, cluster-level models may be the only viable approach. It may also be reasonable to consider adjusting the target resolution of the estimates if data are highly sparse, since one must accept that below a certain level of data availability, no model can resuscitate the analysis.

(2) Are estimates consistent with other models when aggregated to the national and the Admin-1 level? When data are abundant, different methods should lead to similar estimates. For example, if there exist systematic differences between the national estimates, the smoothing model needs to be examined closely. As examples, the difference may be due to incorrect model implementation or specification, inappropriate covariates or external information (e.g. in aggregation of a unit-level model), or informative sampling not being properly accounted for. This diagnostic, however, is not sufficient to conclude a model is reliable. Models with systematic bias may still provide consistent aggregated estimates if the area-level biases cancel out.

(3) Are the estimates over-smoothed? Over-smoothing usually happens when observed data is sparse or when the model is misspecified. Severe over-smoothing can usually be identified by comparing the model-based estimates against another model or direct estimates through scatter plots. Figure 10 compares the point estimates of three smoothing models against the direct estimates. Usually, we expect the shrinkage of smoothed estimates to lead to the point cloud showing a tilted (attenuated towards the

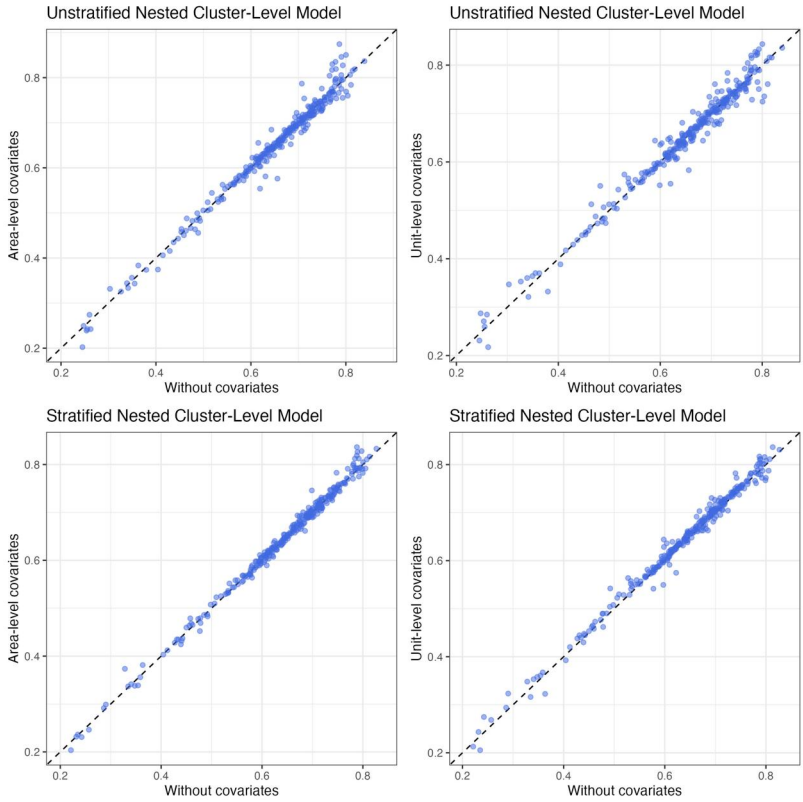

Figure 9. Comparing Admin-2 level model with and without area-level covariates (left) and unit-level covariates (right) in terms of the estimated prevalence for the nested models that do not account for urban/rural stratification (top row) and nested and stratified model (bottom row).

national prevalence) pattern, with more shrinkage in models of higher resolution. Another useful check of over-smoothing is by comparing the variation in estimates against a more reliable model at a coarser level. Consider a set of prevalence estimates, θ_i, for Admin-2 areas, we can calculate the posterior distribution of $v = \frac{1}{M-1}\sum_{i=1}^{M}(\theta_i - \bar{\theta})^2$. We then compare this with the same Admin-2 summary statistic computed from an Admin-1 model, where the prevalences in all Admin-2 areas nested within the same Admin-1 area are assumed to be constant. If the Admin-2 model has significantly smaller variation, over-smoothing is likely present. In our example, if we populate Admin-2 prevalence with the Admin-1 Fay-Herriot model, the estimated v is 0.0122 (95% credible interval: (95% credible interval: [0.0096, 0.0147]). For the stratified nested model at Admin-2, v is estimated to be 0.0161 (95% credible interval: [0.0136,

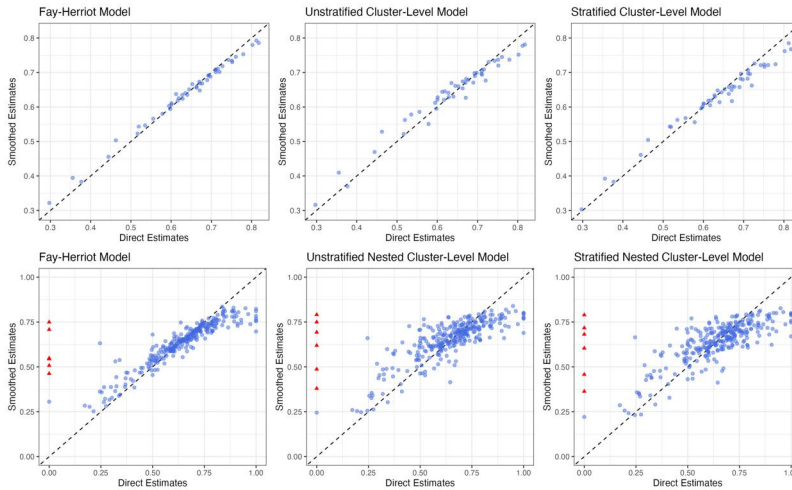

Figure 10. Comparing point estimates of the three smoothed models against the direct estimates. Top row: Admin-1 estimates. Bottom row: Admin-2 estimates. The red triangles correspond to areas where direct estimates cannot be computed.

0.0187]), indicating the Admin-2 model indeed produces more variation than Admin-1 alternatives.

(4) Are the uncertainties too large to be useful? Government agencies often require certain precision when reporting estimates. For example, Statistics Canada has guidelines for area-level estimates, specified with respect to the coefficient of variation (CV) (Cloutier and Langlet 2014). Areas with CV less than 16.7% can be used without restriction. Other metrics, for example, width of uncertainty intervals, may also be used, and the choice of threshold for any metric is context-specific. For Admin-1 modeling in our example, direct estimates are probably good enough with reasonable uncertainty intervals and CVs, as can be seen in figure 5. Further modeling may not provide much gain in practice. For Admin-2 modeling, as can be observed in figure 6, smoothed estimates are usually necessary to obtain the desired precision.

(5) Are model parameters reasonably estimated? It is a good practice to visualize and check the model components of a smoothing model to confirm that the model is free from numerical or computational issues. Figure 11 shows the estimated Admin-1 intercepts and Admin-2 random effects for the nested stratified cluster-level model. The Admin-2 random effects have much smaller magnitude than the Admin-1 fixed effects, as expected due to data sparsity.

(6) Which final model to report? When multiple models pass these previous checks, it is often of interest to pick a final model to report. General

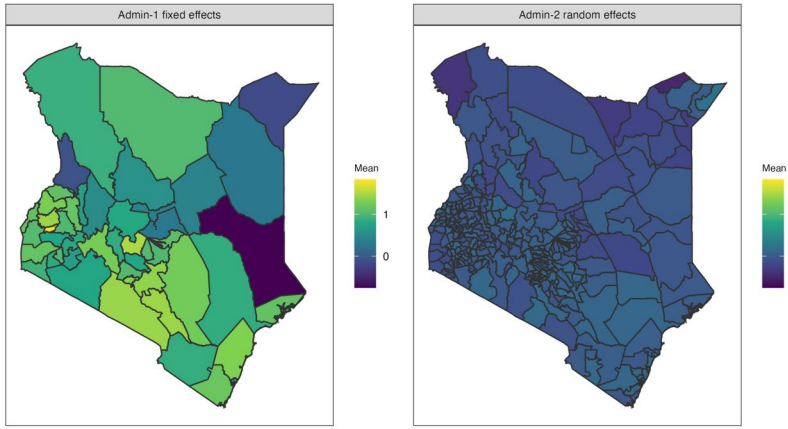

Figure 11. Posterior mean of the Admin-1 fixed effects and Admin-2 random effects in the nested stratified model without covariates.

guidelines for model comparison metrics are far from well-established and are an open area of ongoing research in the SAE literature. The gold-standard of model comparison is to compare the estimates with external 'truth', usually the census (Merfeld et al. 2022, 2024), but for many indicators in LMICs, such information is rarely available. Therefore, model performance is commonly assessed with simulations that are either model-based, that is, samples are generated from a pre-specified true model, or design-based, that is, re-sampled from a fixed population (see e.g. Torabi and Rao 2008; Molina and Rao 2010). Different methods can be used to estimate the prevalence on the replicate samples and be compared against the truth (Tzavidis et al. 2018). However, population data are also rarely available to researchers outside of the national statistical agencies. Another common practice is leaving some areas out of the model fitting process and evaluating the out-of-sample predictive performance on those areas (Pfeffermann 2013). Merfeld et al. (2024) argues that cross validation is usually the second-best option after census validation, though the best practice to implement cross validation is unclear in the literature.

3.4 Stage 4: Summarization and Visualization

In addition to numerical summaries, it is essential to develop visualization pipelines that enable practitioners to easily interpret the large number of model outputs. The interval plot in figure 4 and choropleth map in figures 5 and 6 are the most natural tools to examine the estimates and their uncertainties. The ridge plot in figure 12 provides another useful and more detailed view of the posterior marginal distributions of the estimates. A common theme behind

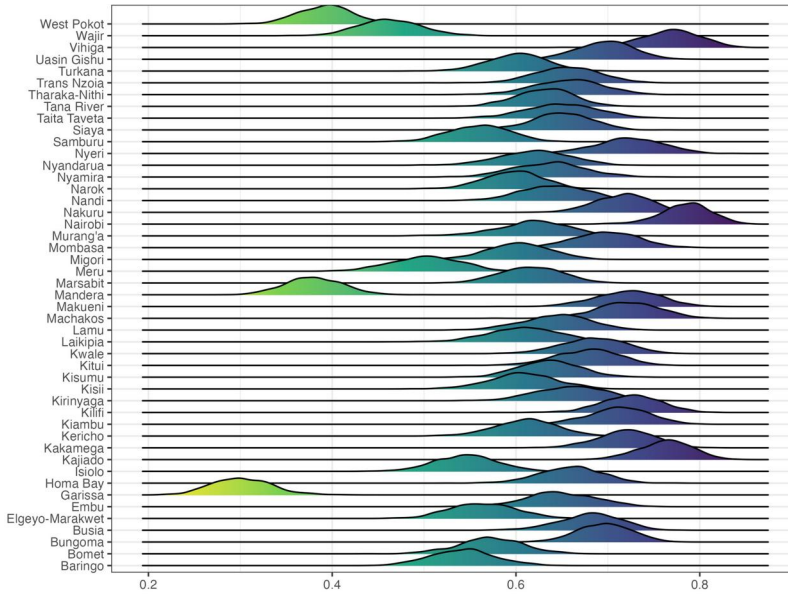

Figure 12. Ridge plot of the posterior distributions of the Admin-1 prevalence, based on the stratified cluster-level model.

these graphical summaries is the high uncertainty of the prevalence estimates. This is unavoidable, given limited data, but it has important implications for the interpretation of the results. For example, figure 12 shows a large number of counties with similar distributions of ANC4+ coverage, between 50 percent and 70 percent. Naively ranking these areas by their point estimates while ignoring such uncertainty could produce misleading results. As the number of areas in consideration becomes larger, all of these plots can be difficult to examine visualy. Web-based interactive visualizations can be very useful.

In terms of model comparison, scatter plots such as figures 9 and 10 are useful in identifying potential over-smoothing. Figure 13 further provides a more concise summary of estimates at different levels by combining point estimates at both Admin-1 and Admin-2 levels. The plot provides a useful visual check of whether Admin-2 level models lead to estimates that are reasonably consistent with the estimates based on Admin-1 level models. The shrinkage of Admin-2 estimates, however, should not be interpreted as a lack of variation within any given Admin-1 areas, but rather it indicates the lack of information in the data that allows us to capture such variation reliably. Similar cautions should be taken in interpreting the 'flat' map of point estimates in figure 6.

Lastly, the posterior distribution of the prevalence are usually the input to compute other quantities of interest. For example, exceedance probabilities, that is, the probability that the prevalence of a given area exceeds certain

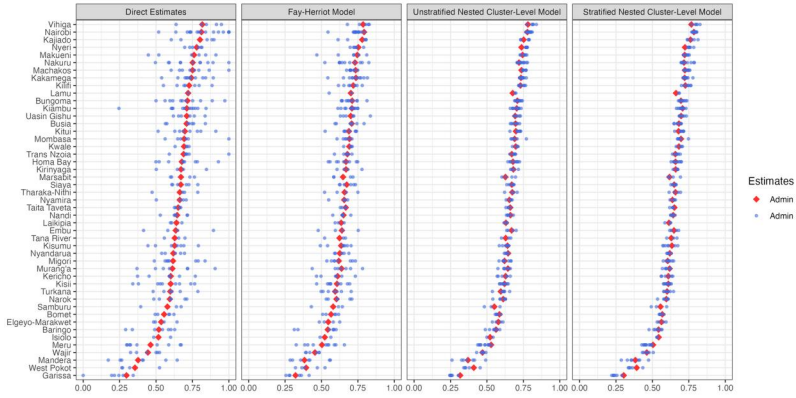

Figure 13. Point estimates of prevalence from different models. Blue points correspond to estimates from Admin-2 level models, and red points correspond to estimates from Admin-1 level models. The Admin-2 level estimates are grouped by their Admin-1 level membership. For the smoothing models, the Admin-1 estimates are produced by fitting a separate model and are not directly aggregated from Admin-2 level results.

thresholds, are often used to assess progress towards specific targets. Figure 14 shows the posterior probability of ANC4+ coverage exceeding the 70 percent goal (Macharia et al. 2022), for each Admin-1 and Admin-2 area under the stratified nested cluster-level model. The striking band of red in the north and east is largely due to the great rurality of these regions of Kenya (see Figure 2), and the general lack of health facilities (where women would travel for ANC visits) and increased travel times, which are an obvious deterrent.

4. DISCUSSION

In this paper, we propose a general workflow for mapping the prevalence of a binary outcome using household survey data in LMICs. The workflow is illustrated through an example estimating the subnational ANC4+ coverage in Kenya. The workflow includes steps to assess data availability and how they are collected, fitting a series of prevalence mapping models, model evaluation and comparison, and visualization and summarization. Throughout the discussion, we focus on the comparison of strengths and limitations across models, as well as potential adaptations to other contexts with varying data availability.

There are tasks in each of the steps where clear-cut recommendations cannot be provided. Careful analysis and reasoning remain indispensable, and one must never forget the context of the data in question, with local knowledge

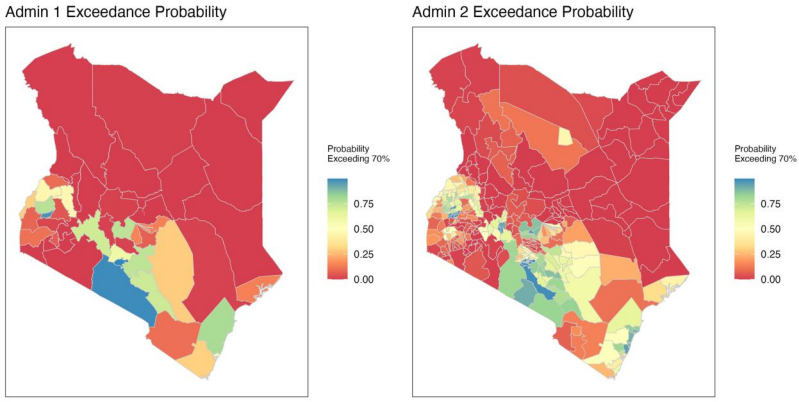

Figure 14. Probability of prevalence exceeding 70 percent in each Admin-1 area (left) and Admin-2 area (right) using the nested stratified cluster-level model.

being invaluable. In addition, the survey considered in this paper includes a reasonable number of samples. For smaller surveys, mapping the prevalence at fine geographic resolution can be more challenging and require careful model development, checking, and evaluation. We emphasize that this workflow should be treated as a baseline analysis pipeline to assist analysts in further adaptations and extensions.

There are many open questions and challenges in prevalence mapping. On the methodological side, more work needs to be done to connect the SAE and geostatistical approaches. The stratified cluster-level model in our workflow is one example of utilizing flexible spatial models at the unit level while accounting for the sampling design. Model evaluation and comparison frameworks are ever more critical with these flexible models. The workflow discussed in this paper focuses on one binary indicator from a single household survey. The same principles extend to modeling more complex indicators, joint modeling of multiple related indicators, and modeling multiple surveys, for example, mapping child mortality in space and time (Mercer et al. 2015). More complex models are needed to borrow information across the additional dimensions effectively, and further modeling considerations, such as seasonality of the covariates and outcomes, may be needed, particularly when modeling at finer spatial and temporal scales. Finally, a major goal of this paper on workflow is to enable the prevalence mapping models and tools to be accessible to researchers and analysts in LMICs so that they can analyze their own data to answer their specific questions of priority, rather than having to rely on estimates that are often difficult to reproduce. Understanding the limitations of the models and the uncertainties of the estimates is critical in designing better and more targeted data collection schemes in LMICs.

SUPPLEMENTARY MATERIALS

Supplementary materials are available online at academic.oup.com/jssam.

DATA AVAILABILITY

The DHS dataset used in this paper can be obtained using the script provided in the Supplementary Material after registering with the DHS program for data access. All results can be reproduced following the script in the Supplementary Materials.

REFERENCES

Altay, U., Paige, J., Riebler, A., and Fuglstad, G.-A. (2025), "Impact of Jittering on Raster-and Distance-Based Geostatistical Analyses of DHS Data," *Statistical Modelling*, 25, 55–74.

Besag, J., York, J., and Molli, A. (1991), "Bayesian Image Restoration with Two Applications in Spatial Statistics," *Annals of the Institute of Statistics and Mathematics*, 43, 1–20.

Blangiardo, M., and Cameletti, M. (2015), *Spatial and Spatio-Temporal Bayesian Models with R-INLA*, United Kingdom: John Wiley and Sons.

Burris, K. C., and Hoff, P. D. (2020), "Exact Adaptive Confidence Intervals for Small Areas," *Journal of Survey Statistics and Methodology*, 8, 206–230.

Burstein, R., Wang, H., Reiner, R. C. Jr., and Hay, S. I. (2018), "Development and Validation of a New Method for Indirect Estimation of Neonatal, Infant, and Child Mortality Trends Using Summary Birth Histories," *PLoS Medicine*, 15, e1002687.

Chandra, H., Chambers, R., and Salvati, N. (2019), "Small Area Estimation of Survey Weighted Counts under Aggregated Level Spatial Model," *Survey Methodology*, 45, 31–59.

Chi, G., Fang, H., Chatterjee, S., and Blumenstock, J. E. (2022), "Microestimates of Wealth for All Low-and Middle-Income Countries," *Proceedings of the National Academy of Sciences*, 119, e2113658119.

Chung, H. C., and Datta, G. S. (2020), *Bayesian Hierarchical Spatial Models for Small Area Estimation*. Technical report, Washington, DC: Center for Statistical Research & Methodology, U.S. Census Bureau.

Cloutier, E. C., and Langlet, É. (2014), *Aboriginal Peoples Survey, 2012: Concepts and Methods Guide, Aboriginal Peoples Survey, 2012: Concepts and Methods Guide*, Canada: Statistics Canada=Statistique Canada.

Croft, T. N., Allen, C. K., and Zachary, B. W. (2023), *Guide to DHS Statistics*, Rockville, Maryland, USA: ICF.

Didan, K., Munoz, A. B., Solano, R., Huete, A, et al. (2015), "MODIS Vegetation Index User's Guide (MOD13 Series)," *University of Arizona: Vegetation Index and Phenology Lab*, 35, 2–33.

Diggle, P., and Giorgi, E. (2016), "Model-Based Geostatistics for Prevalence Mapping in Low-Resource Settings," *Journal of the American Statistical Association*, 111, 1096–1120.

Diggle, P. J., and Giorgi, E. (2019), *Model-Based Geostatistics for Global Public Health: Methods and Applications*, New York: Chapman and Hall/CRC.

Dong, Q., Li, Z. R., Wu, Y., Boskovic, A., and Wakefield, J. (2024), *surveyPrev: Mapping the Prevalence of Binary Indicators using Survey Data in Small Areas* R package version 1.1.0.

Dong, T., and Wakefield, J. (2021), "Modeling and Presentation of Health and Demographic Indicators in a Low- and Middle-Income Countries Context," *Vaccine*, 39, 2584–2594.

Fay, R., and Herriot, R. (1979), "Estimates of Income for Small Places: An Application of James–Stein Procedure to Census Data," *Journal of the American Statistical Association*, 74, 269–277.

Franco, C., and Bell, W. R. (2013), Applying bivariate binomial/logit normal models to small area estimation. In *Proceedings of the American Statistical Association*, *Survey Research Section*, pp. 690–702.

Fuglstad, G.-A., Hem, I. G., Knight, A., Rue, H., and Riebler, A. (2020), "Intuitive priors for variance parameters," *Bayesian Analysis,* 15, 1109–1137. https://dx.doi.org/10.1214/19-BA1185.

Gadiaga, A. N., Abbott, T. J., Chamberlain, H., Lazar, A. N., Darin, E., and Tatem, A. J. (2023), Census disaggregated gridded population estimates for Kenya (2022), version 2.0.

Gao, P. A., and Wakefield, J. (2023), "A Spatial Variance-Smoothing Area Level Model for Small Area Estimation of Demographic Rates," *International Statistical Review*, 91, 493–510.

Gaughan, A. E., Oda, T., Sorichetta, A., Stevens, F. R., Bondarenko, M., Bun, R., Krauser, L., Yetman, G., and Nghiem, S. V. (2019), "Evaluating Nighttime Lights and Population Distribution as Proxies for Mapping Anthropogenic CO2 Emission in Vietnam, Cambodia and Laos," *Environmental Research Communications*, 1, 1.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C.., and Modrák, M. (2020), Bayesian workflow. *arXiv preprint arXiv: 2011.01808*, preprint: not peer reviewed.

Giorgi, E., Fronterrè, C., Macharia, P. M., Alegana, V. A., Snow, R. W., and Diggle, P. J. (2021), "Model Building and Assessment of the Impact of Covariates for Disease Prevalence Mapping in Low-Resource Settings: To Explain and to Predict," *Journal of The Royal Society Interface*, 18, 20210104.

Global Administrative Areas (2022), GADM database of global administrative areas, version 4.1.

Hájek, J. (1971), Discussion of, "An Essay on the Logical Foundations of Survey Sampling, Part I" by In D, Basu, V. Godambe and D. Sprott (Eds.), *Foundations of Statistical Inference*, Toronto: Holt, Rinehart and Winston.

Hoeting, J. A., Davis, R. A., Merton, A. A., and Thompson, S. E. (2006), "Model Selection for Geostatistical Models," *Ecological Applications*, 16, 87–98.

Khan, S., and Hancioglu, A. (2019), "Multiple Indicator Cluster Surveys: Delivering Robust Data on Children and Women across the Globe," *Studies in Family Planning*, 50, 279–286.

KNBS, I. C. F. (2023), Kenya Demographic and Health Survey 2022: volume 1.

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2018), *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R AND INLA*, New York: Chapman and Hall/CRC.

Li, Z. R., Hsiao, Y., Godwin, J., Martin, B. D., Wakefield, J., and Clark, S. J., with support from the United Nations Inter-agency Group for Child Mortality Estimation and its technical advisory group (2019), "Changes in the Spatial Distribution of the under Five Mortality Rate: Small-Area Analysis of 122 DHS Surveys in 262 Subregions of 35 Countries in Africa," *PLoS One*, 14, e0210645.

Liu, B., Lahiri, P., and Kalton, G. (2014), "Hierarchical Bayes Modeling of Survey-Weighted Small Area Proportions," *Survey Methodology*, 40, 1–13.

Lumley, T., and Scott, A. (2014), "Tests for Regression Models Fitted to Survey Data," *Australian & New Zealand Journal of Statistics*, 56, 1–14.

Macharia, P. M., Joseph, N. K., Nalwadda, G. K., Mwilike, B., Banke-Thomas, A., Benova, L., and Johnson, O. (2022), "Spatial Variation and Inequities in Antenatal Care Coverage in Kenya, Uganda and Mainland Tanzania Using Model-Based Geostatistics: A Socioeconomic and Geographical Accessibility Lens," *BMC Pregnancy and Childbirth*, 22, 908.

Mayala, B., Dontamsetti, T., Fish, T. D.., and Croft, T. N. (2019), *Interpolation of DHS Survey Data at Subnational Administrative Level 2*. DHS Spatial Analysis Reports No. 17. Rockville, Maryland, USA.

Mercer, L., Wakefield, J., Pantazis, A., Lutambi, A., Mosanja, H., and Clark, S. (2015), "Small Area Estimation of Childhood Mortality in the Absence of Vital Registration," *Annals of Applied Statistics*, 9, 1889–1905.

Merfeld, J., Chen, H., Lahiri, P., and Newhouse, D. (2024), Small area estimation with geospatial data: A primer.

Merfeld, J. D., Newhouse, D. L., Weber, M., and Lahiri, P. (2022), Combining survey and geospatial data can significantly improve gender-disaggregated estimates of labor market outcomes.

Michal, V., Wakefield, J., Schmidt, A. M., Cavanaugh, A., Robinson, B., and Baumgartner, J. (2023), Small area estimation with random forests and the LASSO. *arXiv preprint arXiv : 2308.15180*, preprint: not peer reviewed.

Mohadjer, L., Rao, J. N. K., Liu, B., Krenzke, T., and de Kerckhove, W. V. (2012), "Hierarchical Bayes Small Area Estimates of Adult Literacy Using Unmatched Sampling and Linking Models," *Journal of the Indian Society of Agricultural Statistics*, 55–63.

Molina, I., Rao, J., and Guadarrama, M. (2019), "Small Area Estimation Methods for Poverty Mapping: A Selective Review," *Statistics and Applications*, 17, 11–22.

Molina, I., and Rao, J. N. (2010), "Small Area Estimation of Poverty Indicators," *Canadian Journal of Statistics*, 38, 369–385.

Osgood-Zimmerman, A., Millear, A. I., Stubbs, R. W., Shields, C., Pickering, B. V., Earl, L., Graetz, N., Kinyoki, D. K., Ray, S. E., Bhatt, S., Browne, A., Burstein, R., Cameron, E., Casey, D., Deshpande, A., Fullman, N., Gething, P., Gibson, H., Henry, N., Herrero, M., Krause, L., Letourneau, I., Levine, A., Liu, P., Longbottom, J., Mayala, B., Mosser, J., Noor, A., Pigott, D., Piwoz, E., Rao, P., Rawat, R., Reiner, R., Smith, D., Weiss, D., Wiens, K., Mokdad, A., Lim, S., Murray, C., Kassebaum, N., and Hay, S. (2018), "Mapping Child Growth Failure in Africa between 2000 and 2015," *Nature*, 555, 41–47.

Osgood-Zimmerman, A., and Wakefield, J. (2023), "A Statistical Review of Template Model Builder: A Flexible Tool for Spatial Modelling," *International Statistical Review*, 91, 318–342.

Otto, M. C., and Bell, W. R. (1995), Sampling error modelling of poverty and income statistics for states. In *American Statistical Association, Proceedings of the Section on Government Statistics*, pp. 160–165.

Pfeffermann, D. (2013), "New Important Developments in Small Area Estimation," *Statistical Science*, 28, 40–68.

R Core Team (2024), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.

Rao, J., and Molina, I. (2015), *Small Area Estimation*, (2nd ed.). New York: John Wiley.

Riebler, A., Sørbye, S., Simpson, D., and Rue, H. (2016), "An Intuitive Bayesian Spatial Model for Disease Mapping That Accounts for Scaling," *Statistical Methods in Medical Research*, 25, 1145–1165.

Roll Back Malaria (2005), *Malaria Indicator Survey: Basic Documentation for Survey Design and Implementation*, Calverton MD: RBM Monitoring and Evaluation Research Group, WHO, UNICEF, Measure Evaluation, US Centers for Disease Control and Prevention.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., Pandey, B., Seto, K. C., Hall, D., Oda, T., Wolfe, R. E., Lin, G., Golpayegani, N., Devadiga, S., Davidson, C., Sarkar, S., Praderas, C., Schmaltz, J., Boller, R., Stevens, J., Ramos González, O. M., Padilla, E., Alonso, J., Detrés, Y., Armstrong, R., Miranda, I., Conte, Y., Marrero, N., MacManus, K., Esch, T., and Masuoka, E. J. (2018), "Nasa's Black Marble Nighttime Lights Product Suite," *Remote Sensing of Environment*, 210, 113–143.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 71, 319–392.

Saha, U. R., Das, S., Baffour, B., and Chandra, H. (2023), "Small Area Estimation of Age-Specific and Total Fertility Rates in Bangladesh," *Spatial Demography*, 11, 2.

Simpson, D., Rue, H., Riebler, A., Martins, T., and Sørbye, S. (2017), "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors (with Discussion)," *Statistical Science*, 32, 1–28.

Sørbye, S. H., and Rue, H. (2017), "Penalised Complexity Priors for Stationary Autoregressive Processes," *Journal of Time Series Analysis*, 38, 923–935.

Tatem, A. J. (2017), WorldPop, open data for spatial demography. *Scientific data 4*.

Torabi, M., and Rao, J. (2008), "Small Area Estimation under a Two-Level Model," *Survey Methodology*, 34, 11.

Tzavidis, N., Zhang, L.-C., Luna, A., Schmid, T., and Rojas-Perilla, N. (2018), "From Start to Finish: A Framework for the Production of Small Area Official Statistics," *Journal of the Royal Statistical Society: Series A*, 181, 927–979.

United Nations (2015), *Sustainable Development Goals*. https://sdgs.un.org/2030agenda.
USAID (2019), *Demographic and Health Surveys*. http://www.dhsprogram.com: United States Agency for International Development.
Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Takahashi, S., Metcalf, C. J. E., Lessler, J., and Tatem, A. J. (2018), "High Resolution Age-Structured Mapping of Childhood Vaccination Coverage in Low and Middle Income Countries," *Vaccine*, 36, 1583–1591.
Wairoto, K. G., Joseph, N. K., Macharia, P. M., and Okiro, E. A. (2020), "Determinants of Subnational Disparities in Antenatal Care Utilisation: A Spatial Analysis of Demographic and Health Survey Data in Kenya," *BMC Health Services Research*, 20, 665.
Wakefield, J., Fuglstad, G.-A., Riebler, A., Godwin, J., Wilson, K., and Clark, S. (2019), "Estimating under Five Mortality in Space and Time in a Developing World Context," *Statistical Methods in Medical Research*, 28, 2614–2634.
Wakefield, J., Gao, P. A., Fuglstad, G.-A., and Li, Z. R. (2025), The two cultures for prevalence mapping: small area estimation and model-based geostatistics. *Statistical Science*
Wakefield, J., Jitong, J., and Wu, Y. (2025), Variance adjustment in the Fay-Herriot model using a pseudo prior. *Manuscript under Preparation.*
Wakefield, J., Okonek, T., and Pedersen, J. (2020), "Small Area Estimation for Disease Prevalence Mapping," *International Statistical Review*, 88, 398–418.
Weiss, D. J., Nelson, A., Vargas-Ruiz, C. A., Gligorić, K., Bavadekar, S., Gabrilovich, E., Bertozzi-Villa, A., Rozier, J., Gibson, H. S., Shekel, T., Kamath, C., Lieber, A., Schulman, K., Shao, Y., Qarkaxhija, V., Nandi, A. K., Keddie, S. H., Rumisha, S., Amratia, P., Arambepola, R., Chestnutt, E. G., Millar, J. J., Symons, T. L., Cameron, E., Battle, K. E., Bhatt, S., and Gething, P. W. (2020), "Global Maps of Travel Time to Healthcare Facilities," *Nature Medicine*, 26, 1835–1838.
World Health Organization (2002), *WHO Antenatal Care Randomized Trial: Manual for the Implementation of the New Model*. Technical report, Geneva: World Health Organization.
Wu, Y., Li, Z. R., Mayala, B., Wang, H., Gao, P., Paige, J., Fuglstad, G.-A., Moe, C., Godwin, J., Donohue, R., Janocha, B., Croft, T., and Wakefield, J. (2021), *Spatial Modeling for Subnational Administrative Level 2 Small-Area Estimation*. DHS Spatial Analysis Reports No. 21. Rockville, Maryland, USA.
Wu, Y., and Wakefield, J. (2024), "Modelling Urban/Rural Fractions in Low-and Middle-Income Countries," *Journal of the Royal Statistical Society Series A: Statistics in Society*, 187, qnae003–830.

Journal of Survey Statistics and Methodology, 2026, 14, 209–237
https://doi.org/10.1093/jssam/smaf048
Article