

Article

Not peer-reviewed version

An Age-Structured Spatially Varying Coefficient Model for High Resolution Mapping of Vaccination Coverage

[Chigozie Utazi](#)*, [Somnath Chaudhuri](#), Oghenebrume Wariri, [Iyanuloluwa Olowe](#), Mohamed Megheib, Andrew Tatem

Posted Date: 17 July 2025

doi: 10.20944/preprints202507.1508.v1

Keywords: vaccination coverage; measles; spatially varying coefficient model; INLA-SPDE; demographic and health survey



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

An Age-Structured Spatially Varying Coefficient Model for High Resolution Mapping of Vaccination Coverage

C. Edson Utazi ^{1,2,3,*}, Somnath Chaudhuri ^{1,4}, Oghenebrume Wariri ^{5,6}, Iyanuloluwa D. Olowe ¹, Mohamed Megheib ^{1,7} and Andrew J. Tatem ¹

¹ WorldPop, School of Geography and Environmental Science, University of Southampton, SO17 1BJ, Southampton, United Kingdom

² Southampton Statistical Sciences Research Institute (S3RI), University of Southampton, SO17 1BJ, Southampton, United Kingdom

³ Nnamdi Azikiwe University, PMB 5025, Awka, Nigeria

⁴ CIBER of Epidemiology and Public Health (CIBERESP), Spain

⁵ Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK

⁶ Vaccines and Immunity Theme, Medical Research Council Unit The Gambia at London School of Hygiene and Tropical Medicine, Atlantic Boulevard, Fajara, The Gambia

⁷ Department of Mathematical Statistics, Faculty of Graduate Studies for Statistical Research, Cairo University, Egypt

* Correspondence: c.e.utazi@soton.ac.uk

Abstract

High-resolution maps of vaccination coverage are valuable for uncovering heterogeneities in coverage to inform vaccine delivery strategies. Coverage maps stratified by age can reveal additional heterogeneities in the timeliness of vaccination and critical immunity gaps among birth cohorts. Here, we propose a spatially varying coefficient model relying on a Bayesian approach for age-structured mapping of vaccination coverage using geolocated individual level household survey and geospatial covariate data. Our flexible modelling framework includes parameterizations capturing spatial (non-)stationarity in differences in coverage between age groups, as well as a modification to allow coverage mapping for single age points through the inclusion of a smoother over age. The proposed models are fitted using the INLA-SPDE approach implemented in the *inlabru* package in R. We choose between competing model parameterizations by examining their out-of-sample predictive performance via cross-validation and using Bayesian model choice criteria. The methodology is applied to age-structured mapping of measles vaccination coverage in Cote d'Ivoire using the 2021 Demographic and Health Survey. Our results reveal a significant delay in measles vaccination in the first year of life and substantial spatial differences in coverage by age, highlighting the need for targeted interventions to achieve equity and attain vaccine-derived immunity goals.

Keywords: vaccination coverage; measles; spatially varying coefficient model; INLA-SPDE; Demographic and Health Survey

Introduction

Vaccination coverage is an important metric for monitoring the performance of immunization programmes [1,2]. Accurate estimates of vaccination coverage are essential at global, national, and subnational levels to assess the relative effectiveness of vaccine delivery strategies such as routine immunization and campaigns [3,4], to guide and evaluate targeted interventions [5], and to track progress toward national and global immunization goals [1,6,7], among other applications. For

certain vaccines such as the measles-containing vaccine, estimates of coverage by age are particularly valuable since achieving coverage targets across all age groups and geographic areas is integral to disease elimination strategies [8]. Such estimates enable the monitoring of the timeliness of vaccination [9,10], identification of key immunity gaps across birth cohorts and the design of catch-up vaccination activities [11–13], thereby equipping national immunization programmes with the evidence needed to develop and implement effective approaches to achieve and sustain disease elimination.

Despite their importance, reliable estimates of vaccination coverage remain challenging to obtain, particularly at fine geographic scales. Data for estimating coverage are usually obtained from administrative sources and household surveys. The quality of data available from these sources can vary greatly due to inherent limitations such as incomplete reporting, recall bias and inaccuracies in population denominators, with survey estimates of coverage often taken to be the “gold standard” [2]. However, survey estimates are mostly produced for large administrative areas such as the national and first administrative levels due to the operational costs and logistical challenges of conducting more intensive sampling to produce estimates for more granular levels. These large area estimates mask epidemiologically and programmatically important heterogeneities in coverage. In line with increasing emphasis on more granular coverage estimates, as encapsulated in global policy frameworks such as the Immunization Agenda 2030 [1], there have been growing efforts to estimate coverage at more operationally relevant geographic scales to better inform vaccination programming [14–18]. Furthermore, survey-based estimates are typically reported for standard age groups, as determined by relevant vaccination schedules [7]; and when surveys are conducted infrequently, these estimates often fail to capture critical immunity gaps in other eligible age groups.

Few published studies such as Utazi et al [17] and Takahashi et al [15] have developed approaches for age-structured, high-resolution mapping of vaccination coverage using geolocated survey data. Utazi et al [17] employed a *birth cohort approach*, fitting independent geostatistical models to predict measles vaccination coverage in children aged 9–11 months, 12–23 months and 24–35 months. However, their methodology provides no mechanism to exploit shared patterns - such as the underlying spatial structures - across the age groups, which could potentially improve coverage estimation. Moreover, because separate models were fitted for each age group, their approach may perform poorly in contexts where overall or cluster-level sample sizes for a given age group are small (see, for example, the data section), since this can result in greater uncertainties in the modelled estimates for the age group [19]. Takahashi et al [15] adopted a *single age approach*, using a generalized additive model to map measles vaccination coverage by age in the African Great Lakes region. Their method incorporates a smooth function of age, enabling estimation of coverage at single-month age points but not for broader age groups. Additionally, the inclusion of a smooth function of survey cluster-level geocoordinates in their model primarily captures broad spatial trends and may be inadequate for modelling structured spatial dependence [20].

Spatially varying coefficient (SVC) models [21,22] are particularly useful in spatial regression contexts when relationships between an outcome and certain covariates exhibit spatial non-stationarity. These models allow the regression coefficients associated with these covariates to vary across space, yielding spatial surfaces/maps for these, and are also suitable for capturing differences in space between the levels of a given categorical covariate in relation to an outcome (see methods section). Gelfand et al [21] introduced a Bayesian framework for SVC models and these have been used in ecological settings in both discrete [23] and continuous [22] spatial domains to uncover nuanced spatial variation in ecological processes such as bird species abundance and forest biomass. In particular, Finley [22] compared different SVC model specifications, including geographically weighted regression [24], using both simulated and real-world datasets. The study accounted for directional (anisotropic) and nonstationary dependence structures, demonstrating the flexibility of these models to capture complex spatial patterns in the data.

In this study, we develop a novel and flexible methodology for age-structured mapping of vaccination coverage. The proposed approach is based on a spatially varying coefficient model that

captures spatial non-stationarity in the differences in coverage between age groups, with alternative model specifications characterizing different forms of age-related variation, and which also enable the estimation of coverage using the single age approach. Unlike previous work [17,19,25], the proposed models are fitted to individual level data, enabling more robust estimation of covariate-coverage relationships and potentially improving estimation in areas with few cluster-level observations. Model selection is based on Bayesian model choice criteria and using cross-validation to evaluate out-of-sample predictive performance. The methodology is applied to age-disaggregated mapping of the coverage of the first dose of the measles-containing vaccine (MCV1) in Cote d'Ivoire, using data from the 2021 Demographic and Health Survey [26]. Cote d'Ivoire was chosen as a case study due to being a priority country for an implementing/donor organization and has therefore been the focus of recent work [27].

Methods

Vaccination Coverage Data

Geolocated individual-level data on the coverage of MCV1 were obtained from the 2021 Cote d'Ivoire DHS (CDHS) [26]. The survey was designed to be representative at the national and first administrative (14 areas known as districts – supplementary Figure 1) levels, and for urban and rural areas. The sampling design involved a random, stratified two-stage selection, with urban and rural areas of each of the 14 districts constituting separate sampling strata, yielding a total of 28 strata. The first stage involved selecting the primary sampling units known as clusters (or enumeration areas) from a sampling frame, while the second stage involved selecting households within each sampled cluster, with 28 households sampled per cluster. In all, a total of 539 clusters were included in the survey [26].

Routine MCV1 is recommended for administration at 9 months of age in Cote d'Ivoire [28]. Hence, for each eligible child aged 9 to 35 months, we extracted information on their age, vaccination status (based on health records or caregiver recall) and the (latitude/longitude) geocoordinates of the cluster they were sampled from. Each individual level record was classified into one of three age groups - 9-11 months, 12-23 months or 24-35 months - representing different birth cohorts, and aggregated to the survey cluster level for validation purposes. The data extracted contained a total of 4,236 children, of whom 507, 1,920 and 1,809 belonged to the respective age groups. A summary of the extracted data is displayed in Figure 1 and supplementary Figure 2. This includes national survey weighted and unweighted estimates of MCV1 coverage for single age points and age groups, unweighted coverage estimates at the cluster level for age groups and corresponding distribution of cluster level sample sizes for each age group.

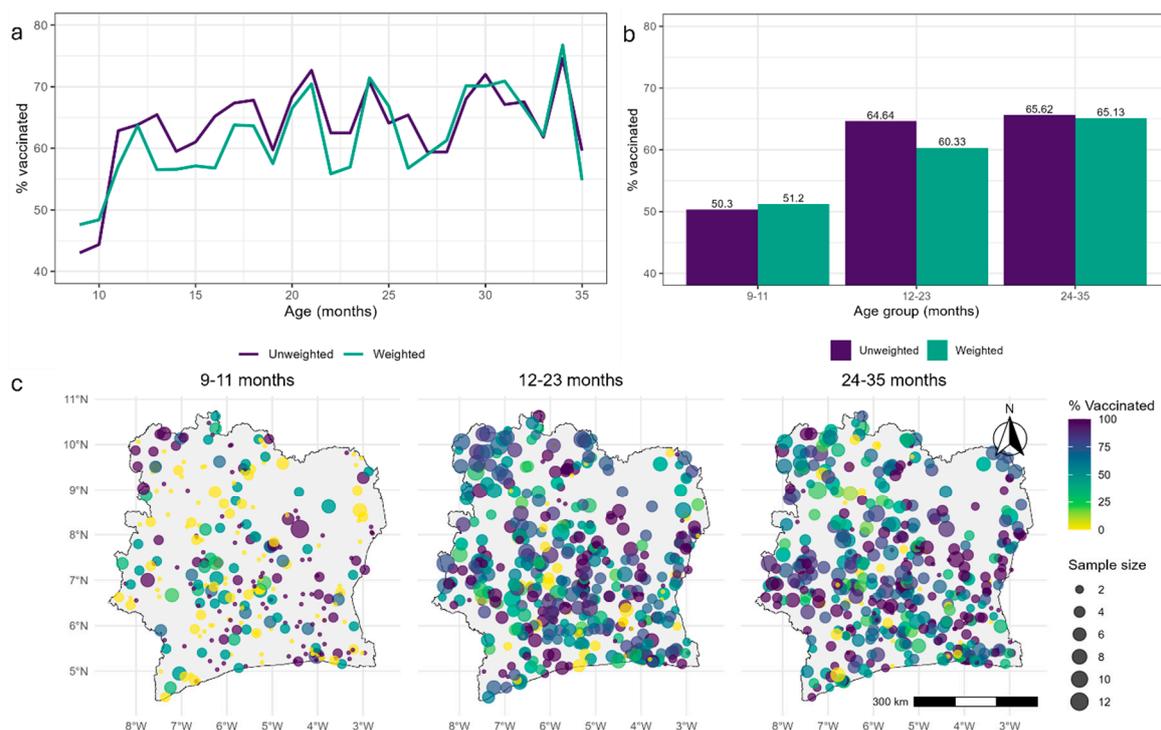


Figure 1. Empirical estimates of measles vaccination coverage for children aged 9 to 35 months at the national (a, b) and survey cluster (c) levels for (a) single age points and (b, c) age groups, according to the 2021 Cote d'Ivoire Demographic and Health Survey.

These plots generally reveal a sharp increase in coverage at lower ages, which then appears to stabilize and exhibit random fluctuations randomly in the later ages. The relatively lower coverage among children aged 9-11 months is evident both at the national and cluster levels. Cluster level sample sizes are much smaller for age 9-11 months, limiting the feasibility of a separate analysis for this age group. For the other age groups, cluster level sample sizes are also small (median = 3, interquartile range = 3). The distribution of these cluster-level sample sizes is particularly important for accurately estimating the underlying spatial structure in the data and for obtaining more robust point or grid level predictions of vaccination coverage across age groups.

Geospatial Covariates, Gridded Population and Boundary Data

Following previous studies [3,17,19,25], we assembled a wide range of geospatial covariate datasets for the analysis. These include geographic, socioeconomic, remoteness, climate, environmental, conflict and health-related variables, which have either direct or proximate associations with vaccination coverage. While most of the covariates were externally sourced, others such as ownership of a health card/document were obtained from the 2021 CDHS. These CDHS-derived covariates were first processed at the cluster level using the definitions provided in supplementary Table 1. Kriging interpolation was then used to create their interpolated surfaces at 1x1 km resolution as outlined in previous work [19,29]. Covariate selection was undertaken in a non-spatial framework, using a similar approach as outlined in Utazi et al [17]. This involved examining the relationships between the covariates and vaccination coverage, and applying log transformations to the covariates where appropriate; fitting single-covariate models and ranking the covariates based on their predictive ability (using predictive R^2 values); assessing multicollinearity and choosing between highly correlated covariates (with correlation > 0.8 or variance inflation factor > 4.0) based on their ranks; and finally, using stepwise regression (backward elimination based on the Akaike Information Criterion) to identify the best model or combination of covariates for modelling the outcome in a non-spatial framework with binomial regression models. A total of 13 covariates were selected for the analysis from an initial set of 38 covariates, including urban and rural areas, elevation,

vegetation index (veg_index, average number of wet days (wetdays), urban accessibility (urban_access), elevation, maximum temperature (max_temp), distance to conflict areas (dist_conf), walking travel time to the nearest health facility (walking_tt), malaria prevalence (mal_prev), maternal education (mat_educ), health card/document ownership (health_card), access to media (media) and household wealth (wealth) (see supplementary Table 1 for details). These selected covariates are displayed in supplementary Figures 3 and 4.

Gridded population data for children aged under 5 years in 2021 were obtained from WorldPop [30]. These were used as weights when aggregating the grid level estimates to the administrative level [19]. All boundary data sets used in the analysis were obtained from geoBoundaries [31].

Model Development

We begin by introducing the spatially varying coefficient model and its parameterizations for estimating vaccination coverage using the birth cohort approach, and then alternative specifications that enable estimation at single age points and for all eligible individuals.

Let $z_i(\mathbf{s}_j)$ denote the age (in months) of the i th eligible individual/child sampled from location \mathbf{s}_j , and $a_0, a_1, a_2, \dots, a_K$, the different $K + 1$ age groups (or birth cohorts), one of which includes $z_i(\mathbf{s}_j)$. For example, in our application, $K = 2$, and $a_0 = \{9, 10, 11 \text{ months}\}$; $a_1 = \{12, \dots, 23 \text{ months}\}$ and $a_2 = \{24, \dots, 35 \text{ months}\}$. Also, let $Y_i(\mathbf{s}_j)$ denote the corresponding binary outcome representing the individual's vaccination status (i.e., $Y_i(\mathbf{s}_j) = 1$ if vaccinated and 0 otherwise), and $p_i(\mathbf{s}_j)$ the (unobserved) true probability of being vaccinated. The spatially varying coefficient model can be expressed as

$$Y_i(\mathbf{s}_j) | p_i(\mathbf{s}_j) \sim \text{Binomial}(1, p_i(\mathbf{s}_j)), i = 1, \dots, n_j; j = 1, \dots, J,$$

$$\text{logit}(p_i(\mathbf{s}_j)) = \beta_0 + \sum_{k=1}^K (\beta_k + \beta_k(\mathbf{s}_j)) I(z_i(\mathbf{s}_j) \in a_k) + \mathbf{x}(\mathbf{s}_j)' \boldsymbol{\alpha} + \omega(\mathbf{s}_j) + \epsilon(\mathbf{s}_j), \quad (1)$$

where β_0 is the overall intercept, β_k and $\beta_k(\mathbf{s}_j)$ are fixed and spatially varying regression coefficients for age cohort a_k ($k > 0$), $I(\cdot)$ is an indicator function that returns a value of 1 if the specified condition is true and 0 otherwise, and $\mathbf{x}(\mathbf{s}_j)$ is a vector of geospatial covariates for location \mathbf{s}_j , with $\boldsymbol{\alpha}$ being the corresponding fixed regression coefficients. Thus, $\beta_k(\mathbf{s}_j)$ can be viewed as a random spatial adjustment to the average change in the log-odds of vaccination for the k th age cohort, β_k , at location \mathbf{s}_j . Note that the reference age group, a_0 , is omitted from equation (1) as this is absorbed into the intercept terms. In practice, the choice of reference age group may vary depending on the objective of the analysis. Also, the covariate vector, $\mathbf{x}(\mathbf{s}_j)$, includes an urban-rural variable which helps to account for the urban-rural stratification used in the survey design. Only cluster-level geospatial covariates are included in the model since the goal of inference is to obtain age-specific coverage estimates at the grid level, typically at 1x1 km resolution.

Furthermore, the terms $\omega(\mathbf{s}_j)$ and $\epsilon(\mathbf{s}_j)$ are modelled as Gaussian random effects, capturing residual spatial and non-spatial (or between-cluster) variation, respectively. Specifically, $\omega(\mathbf{s}_j)$ represents a random spatial adjustment to the overall intercept, β_0 , such that $\beta_0 + \omega(\mathbf{s}_j)$ defines a spatially varying intercept process (for the reference age group). We model $\epsilon(\mathbf{s}_j)$ as an independent and identically distributed (iid) process with mean 0 and variance σ_ϵ^2 . The spatial random effect $\omega(\mathbf{s}_j)$ is specified as a zero-mean Gaussian process with covariance function Σ_ω , that is, $\boldsymbol{\omega} = (\omega(\mathbf{s}_1), \dots, \omega(\mathbf{s}_j))' \sim N(\mathbf{0}, \Sigma_\omega)$, where Σ_ω is assumed to follow the Matérn covariance function [32] given by

$$\Sigma_\omega(\mathbf{s}_j, \mathbf{s}_{j'}) = \frac{\sigma_\omega^2}{2^{\nu-1} \Gamma(\nu)} (\kappa \|\mathbf{s}_j - \mathbf{s}_{j'}\|)^\nu K_\nu(\kappa \|\mathbf{s}_j - \mathbf{s}_{j'}\|).$$

Here, the notation $\|\cdot\|$ denotes the Euclidean distance, $\sigma_\omega^2 > 0$ is the marginal variance of $\omega(\mathbf{s}_j)$, $\kappa > 0$ is a scaling parameter related to the range $r(r = \frac{\sqrt{8\nu}}{\kappa})$ – the distance at which spatial correlation is close to 0.1 – and K_ν is the modified Bessel function of the second kind and order $\nu >$

0. Following Lindgren et al [33], it is common to fix the smoothness parameter ν to one to ensure identifiability.

For each of the spatially varying regression coefficients, $\beta_k(\mathbf{s}_j), k = 1, \dots, K$, we also assume zero-mean Gaussian processes with Matérn covariance functions denoted by Σ_{β_k} . We denote the corresponding range and marginal variance parameters using r_k and $\sigma_{\beta_k}^2$, respectively.

The general model in equation (1) can be further parameterized to yield simpler, more parsimonious models, depending on the nature of the dependencies assumed in the data. For example, if the spatially varying coefficients and the geospatial covariates are deemed sufficient to capture all the spatial dependencies, or more specifically, if the geospatial covariates are assumed to fully account for the spatial variation in the outcome for the reference age group, then model (1) reduces to

$$\text{logit}(p_i(\mathbf{s}_j)) = \beta_0 + \sum_{k=1}^K (\beta_k + \beta_k(\mathbf{s}_j)) I(z_i(\mathbf{s}_j) \in a_k) + \mathbf{x}(\mathbf{s}_j)' \boldsymbol{\alpha} + \epsilon(\mathbf{s}_j), \quad (2)$$

which excludes the spatially correlated residual term, $\omega(\mathbf{s}_j)$. Other variants of model (2) are also possible. For instance, if prior evidence suggests that step changes captured by β_k are sufficient to characterise the differences between certain age groups and the reference age group, then the corresponding $\beta_k(\mathbf{s}_j)$ terms may also be excluded, while retaining $\omega(\mathbf{s}_j)$. Such specifications may be particularly advantageous in contexts involving a large number of age groups.

Furthermore, if the differences in the odds of vaccination between each age group and the reference age group are assumed not vary spatially and can be adequately characterised using fixed step changes, then the spatially varying coefficients can be removed entirely from model (1), yielding

$$\text{logit}(p_i(\mathbf{s}_j)) = \beta_0 + \sum_{k=1}^K \beta_k I(z_i(\mathbf{s}_j) \in a_k) + \mathbf{x}(\mathbf{s}_j)' \boldsymbol{\alpha} + \omega(\mathbf{s}_j) + \epsilon(\mathbf{s}_j). \quad (3)$$

In practice, models (2) and (3) provide alternative model specifications against which to compare model (1) to determine the most appropriate modelling approach for capturing spatial dependencies in the data.

Additionally, when the goal is to predict coverage for single ages (in months), model (3) can be modified to include a smooth function of age, $f(z_i(\mathbf{s}_j))$:

$$\text{logit}(p_i(\mathbf{s}_j)) = \beta_0 + f(z_i(\mathbf{s}_j)) + \mathbf{x}(\mathbf{s}_j)' \boldsymbol{\alpha} + \omega(\mathbf{s}_j) + \epsilon(\mathbf{s}_j), \quad (4)$$

where $f(\cdot)$ is modelled using a second-order random walk: $f(z_r | z_{r-1}, z_{r-2}) \sim N(2z_{r-1} - z_{r-2}, \sigma_z^2)$, which corresponds to the Bayesian equivalent of a cubic smoothing spline [34]. A sum-to-zero constraint was imposed on $f(\cdot)$ for identifiability, since the model includes an intercept term [34]. We note that, as in the birth cohort model, it is possible to model the effect of age in equation (4) using a spatially varying coefficient, such that $f(z_i(\mathbf{s}_j)) = \beta(\mathbf{s}_j) z_i(\mathbf{s}_j)$. However, such parameterization is not ideal for estimating the overall smoothed effect of coverage by age as intended in model (4). Hence, model (4) captures the non-linear effect of age on vaccination but assumes constant residual spatial dependence for all ages as in model (3). Similar to model (3), model (4) can also be used to obtain predictions for age groups if desirable (see discussion section).

Finally, to predict coverage for all eligible age groups combined, the age-specific terms in model (1) can simply be dropped, yielding

$$\text{logit}(p_i(\mathbf{s}_j)) = \beta_0 + \mathbf{x}(\mathbf{s}_j)' \boldsymbol{\alpha} + \omega(\mathbf{s}_j) + \epsilon(\mathbf{s}_j). \quad (5)$$

Model (5) aligns with approaches commonly used in previous work to map vaccination coverage [16,19,25], except that it is estimated using individual rather than cluster level data. Note that in equation (5), as in all the models considered here, individual level variation is already accounted for in the binomial likelihood, with the variance of the logistic distribution fixed at $\frac{\pi^2}{3} \cong 3.29$ [35].

For conciseness, we subsequently refer to the candidate birth cohort model specifications in equations (1) to (3) as MODsvc1, MODsvc2 and MODnosvc, respectively. Similarly, we refer to the single-age and all-age models in equations (4) and (5) as MODsmooth and MODall, respectively.

Bayesian Inference, Model Fitting and Prediction

The proposed models are implemented in a fully Bayesian framework. We place the following prior distributions on the parameters:

$$\begin{aligned} \boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_K)' &\sim N(\mathbf{0}, 10^3 \mathbf{I}); \quad \boldsymbol{\alpha} \sim N(\mathbf{0}, 10^3 \mathbf{I}); \quad p(r < \tilde{r}) = 0.01; \quad p(r_k < \tilde{r}) = 0.01; \\ p(\sigma_\epsilon > 3) &= 0.01; \quad p(\sigma_\omega > 3) = 0.01; \quad p(\sigma_{\beta_k} > 3) = 0.01; \quad \sigma_z^{-2} \sim \text{Gamma}(1, 0.01). \end{aligned} \quad (6)$$

These comprise non-informative priors on the fixed regression parameters and a weakly informative prior on the precision of the second-order random walk in MODsmooth. The priors on the range parameters and the remaining variance parameters are penalized complexity priors, introduced by Simpson et al [36] and used in Fuglstad et al [37] in a spatial modelling context, which penalize deviations from a base (simpler) model, e.g., a constant field. For the range parameters, \tilde{r} is chosen to be 5% of the extent of the study area in the longer direction (north-south direction for the Cote d'Ivoire application).

Let $\boldsymbol{\theta} = (\beta_0, \beta_1, \dots, \beta_K, \boldsymbol{\alpha}, r, r_1, \dots, r_K, \sigma_\epsilon^2, \sigma_\omega^2, \sigma_{\beta_k}^2)'$ denote all the fixed parameters of MODsvc1 in equation (1), $\boldsymbol{\beta}_k^s = (\beta_k(s_1), \dots, \beta_k(s_j))'$, $\boldsymbol{\epsilon} = (\epsilon(s_1), \dots, \epsilon(s_j))'$ and \mathbf{y} denote all the observed data. Its joint posterior distribution can be expressed as

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\omega}, \boldsymbol{\beta}_1^s, \dots, \boldsymbol{\beta}_K^s, \boldsymbol{\epsilon} | \mathbf{y}) &\propto \pi(\boldsymbol{\theta}) \times N(\boldsymbol{\omega} | \mathbf{0}, \boldsymbol{\Sigma}_\omega) \times N(\boldsymbol{\epsilon} | \mathbf{0}, \sigma_\epsilon^2 \mathbf{I}) \times N(\beta_0 | 0, 10^3) \\ &\times \prod_{k=1}^K \{N(\boldsymbol{\beta}_k^s | \mathbf{0}, \boldsymbol{\Sigma}_{\beta_k}) \times N(\beta_k | 0, 10^3)\} \\ &\times \prod_{j=1}^J \prod_{i=1}^{n_j} \text{Binomial}(y_i(s_j); p_i(s_j), z_i(s_j), y), \end{aligned} \quad (7)$$

where $\pi(\boldsymbol{\theta})$ is the joint prior distribution on $\boldsymbol{\theta}$. The posterior distributions of the other models can be straightforwardly derived from equation (7).

For parameter estimation, we use the integrated nested Laplace approximation-stochastic partial differential equation (INLA-SPDE) approach [33,38] implemented in the inlabru package in R [39]. The INLA approach performs approximate Bayesian inference for latent Gaussian models by using the Laplace method to approximate the marginal posterior distributions of all the unknown quantities in the model, including the parameters and latent variables. The SPDE approach facilitates the estimation of the spatial terms in the models (e.g., $\boldsymbol{\omega}$ and $\boldsymbol{\beta}_k^s$ in model (1)). The approach works by representing each of these terms as a Gaussian Markov random field to reduce the computational burden inherent in the estimation of their respective covariance matrices [33].

Of particular interest in the proposed models is the prediction of vaccination coverage for the age groups or single age points of interest at unsampled locations, e.g. 1×1 km grids, throughout the study area. In a Bayesian framework, the prediction of vaccination coverage for age group $k \in \{0, 1, \dots, K\}$ at a new location, $\tilde{\mathbf{s}}$, proceeds by evaluating the posterior predictive distribution of $p_k(\tilde{\mathbf{s}})$ given by

$$\begin{aligned} \pi(p_k(\tilde{\mathbf{s}}) | \mathbf{y}) &= \int \text{logit}^{-1}(\pi(\eta_k(\tilde{\mathbf{s}}) | \boldsymbol{\omega}, \boldsymbol{\epsilon}, \boldsymbol{\beta}_k^s, \boldsymbol{\theta}, \mathbf{y})) \\ &\times \pi(\boldsymbol{\omega}, \boldsymbol{\epsilon}, \boldsymbol{\beta}_k^s, \boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\omega} d\boldsymbol{\epsilon} d\boldsymbol{\beta}_k^s d\boldsymbol{\theta}, \end{aligned} \quad (8)$$

where $\eta_k(\tilde{\mathbf{s}})$ is the linear predictor for MODsvc1, such that

$$\eta_k(\tilde{\mathbf{s}}) = \begin{cases} \beta_0 + \mathbf{x}(\tilde{\mathbf{s}})' \boldsymbol{\alpha} + \omega(\tilde{\mathbf{s}}) + \epsilon(\tilde{\mathbf{s}}) & \text{if } k = 0 \\ \beta_0 + \beta_k + \beta_k(\tilde{\mathbf{s}}) + \mathbf{x}(\tilde{\mathbf{s}})' \boldsymbol{\alpha} + \omega(\tilde{\mathbf{s}}) + \epsilon(\tilde{\mathbf{s}}) & \text{if } k \geq 1 \end{cases}$$

In the `inlabru` package, equation (8) can be computed post model fitting by using either the `generate()` or `predict()` function. We obtained 1,000 posterior samples from equation (8) for each prediction grid location, which were then summarized to produce the grid level estimates and associated uncertainties. Administrative level estimates were also obtained as population-weighted averages over all the grid cells falling within each administrative unit [19], as noted previously, using these posterior samples. Individual level predictions (in- and out-of-sample prediction for validation purposes) and predictions using models (2), (3), (4) and (5) can also be obtained in a similar manner.

Model Choice and Validation

To compare the fits of the competing birth cohort models in equations (1) to (3) (and those of MODsmooth and MODall), we employed the Watanabe-Akaike Information Criterion (WAIC) [40,41] which uses the log of the predictive density of each data point to evaluate a model's out-of-sample predictive performance. The WAIC can be expressed as

$$WAIC = -2 \sum_{i=1}^n \log\{\pi(y_i|\mathbf{y})\} + 2p_{WAIC},$$

where $\pi(y_i|\mathbf{y}) = \int_{-\infty}^{\infty} \pi(y_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ is the posterior predictive density of data point y_i (out of n data points), which can be evaluated in practice as $\frac{1}{L} \sum_{l=1}^L \pi(y_i|\boldsymbol{\theta}^l)$ with $\boldsymbol{\theta}^l$ denoting the l -th posterior draw from a total of L posterior samples. The term p_{WAIC} represents the effective number of parameters, computed as $p_{WAIC} = \sum_{i=1}^n Var_{\boldsymbol{\theta}|\mathbf{y}}[\log(\pi(y_i|\boldsymbol{\theta}))]$. The WAIC balances model fit and complexity by penalizing overfitting through p_{WAIC} , which captures the sensitivity of the model's fit to individual data points. A lower WAIC indicates better predictive accuracy.

To further evaluate the out-of-sample predictive performance of the proposed models both at the cluster and individual levels, we conducted a k -fold cross-validation exercise, creating the cross-validation folds as random and spatially stratified splits of J cluster locations and setting $k = 10$. At the cluster level, we used the following metrics to evaluate predictive performance:

$$\text{Root mean square error, } RMSE = \sqrt{\frac{1}{m} \sum_{j=1}^m (\hat{p}(s_j) - p(s_j))^2};$$

$$\text{Mean absolute error, } MAE = \frac{1}{m} \sum_{j=1}^m |\hat{p}(s_j) - p(s_j)|;$$

$$\text{Average bias, } AVG_BIAS = \frac{1}{m} \sum_{j=1}^m (\hat{p}(s_j) - p(s_j));$$

Continuous ranked probability score,

$$CRPS = \frac{1}{m} \sum_{j=1}^m \left\{ \frac{1}{S} \sum_{s=1}^S |\hat{p}^{(s)}(s_j) - p(s_j)| - \frac{1}{2S^2} \sum_{s=1}^S \sum_{t=1}^S |\hat{p}^{(s)}(s_j) - \hat{p}^{(t)}(s_j)| \right\};$$

where m is the number of cluster-level observations in the k th cross-validation set, $p(s_j)$ denotes the observed coverage and $\hat{p}(s_j)$ the corresponding predicted values (i. e., the posterior means). With the CRPS [42], the entire S posterior samples are used to measure the discrepancies between the observed values and the predicted values for each data point rather than using a particular summary as in the other metrics. The smaller the values of these metrics, the better the predictive performance.

At the individual level, we evaluated out-of-sample predictive performance by computing the Brier score, $BS = \frac{1}{M} \sum_{j=1}^m \sum_{i=1}^{n_j} (\hat{p}_i(s_j) - y_i(s_j))^2$, where $M = \sum_{j=1}^m n_j$ is the number of individual level observations in m validation set locations and $y_i(s_j)$ denotes the binary outcomes [43]. A lower BS indicates better model performance. All the metrics were averaged over all k cross-validation folds in each case.

Results

Model Choice

Table 1 presents the WAIC values of the candidate birth cohort models (MODsvc1, MODsvc2 and MODnosvc), along with those of MODsmooth and MODall for comparison. Among the three birth cohort models, MODsvc1 had the best predictive performance. All models that account for age

clearly outperformed MODall, demonstrating better predictive accuracy at the individual level and substantial variation in coverage by age within the data. Notably, the better performance of MODsmooth relative to MODnosvc suggests that, in the absence of the spatially varying coefficients, the smooth age function in MODsmooth more effectively captured the gradients in age-specific coverage than the step changes in MODnosvc. Also, we observe that although MODsvc1 and MODsvc2 - both incorporating spatially varying coefficients - had the lowest WAIC, these also had the largest effective numbers of parameters relative to other models as expected given their greater flexibility.

Table 1. WAIC values of the proposed models.

Model	p_{WAIC}	WAIC
MODsvc1	210	5191
MODsvc2	219	5197
MODnosvc	186	5215
MODsmooth	187	5204
MODall	184	5262

The patterns observed in the WAIC results are corroborated by the BS values reported in supplementary Table 2. These results show that for individual level out-of-sample prediction under the random cross-validation scheme, MODsvc1 and MODsvc2 had the best predictive performance, although the differences between the models were modest. Under stratified cross-validation, the differences between the models were even smaller, with MODsvc1 still ranking among the best-performing models. Consistent with previous findings, for both cross-validation schemes, all models accounting for variation by age (except MODsvc2 under stratified cross-validation) outperformed MODall.

For out-of-sample prediction at the cluster level, Table 2 reports the validation metrics for MODsvc1, MODnosvc and MODsvc2, while supplementary Table 3 provides corresponding results for MODall for comparison. Due to sample size limitations, it was not feasible to evaluate cluster-level predictive performance for MODsmooth. Across the three models, out-of-sample predictive performance was broadly similar within each age group, though MODsvc1 performed slightly better overall, particularly for the 9–11 and 24–35 month age groups. For the 12–23 month age group, the differences between the models were marginal, with no model clearly outperforming the others. Importantly, all three models performed better for age groups 12–23 months and 24–35 months compared to 9–11 months (except when considering AVG_BIAS, which poorly differentiated between the models), likely reflecting the relatively larger cluster level sample sizes for both older age groups (Figure 1 and supplementary Figure 2). Indeed, for the aggregated 9–35 month age group, the validation metrics reported in supplementary Table 3 are consistently lower than those in Table 2, indicating improved predictive performance associated with larger cluster level sample sizes. As expected, predictive performance across all the models was generally poorer under the stratified cross-validation scheme, reflecting the increased difficulty of prediction in this setting.

Table 2. Results of cluster level k-fold cross-validation.

Model	Age group	RMSE	MAE	AVG_BIAS	CRPS
RANDOM					
MODsvc1	9 -11	0.439	0.403	0.006	0.332
MODsvc2	9 -11	0.440	0.404	0.006	0.335
MODnosvc	9 -11	0.440	0.403	0.008	0.334
MODsvc1	12 - 23	0.315	0.263	0.008	0.206
MODsvc2	12 - 23	0.315	0.262	0.009	0.203
MODnosvc	12 - 23	0.314	0.262	0.012	0.208
MODsvc1	24 - 35	0.298	0.249	-0.015	0.197

MODsvc2	24 - 35	0.299	0.249	-0.015	0.196
MODnosvc	24 - 35	0.300	0.251	-0.016	0.198
STRATIFIED					
MODsvc1	9 - 11	0.444	0.408	0.004	0.337
MODsvc2	9 - 11	0.445	0.408	0.004	0.341
MODnosvc	9 - 11	0.445	0.409	0.004	0.338
MODsvc1	12 - 23	0.318	0.266	0.003	0.206
MODsvc2	12 - 23	0.320	0.267	0.002	0.207
MODnosvc	12 - 23	0.319	0.265	0.009	0.209
MODsvc1	24 - 35	0.304	0.252	-0.017	0.196
MODsvc2	24 - 35	0.302	0.253	-0.017	0.198
MODnosvc	24 - 35	0.304	0.255	-0.018	0.200

Overall, our evaluation of the out-of-sample predictive performance of the competing birth cohort models, at both the individual and cluster levels, indicates that MODsvc1 offers the best performance. Accordingly, in the following sections, we discuss the results from MODsvc1 when using the birth cohort approach to estimate vaccination coverage.

Parameter Estimates

In Table 3, we present the estimates of the parameters of the best-fitting birth cohort model, MODsvc1. The covariates significantly associated with the odds of vaccination include: urban versus rural residence, distance to conflict areas, walking travel time to the nearest health facility, maternal education, health card/document ownership, access to media and age group. As expected, the odds of vaccination increased with greater distance to conflict areas, higher proportions of educated mothers, higher proportions of children who owned a health card and greater access to the media within households. Conversely, the odds of vaccination decreased as walking travel time to the nearest health facility increased.

Considering the estimated odds ratios, a unit increase in the proportion of children who owned a health card is associated with a sixfold increase in the odds of vaccination, for example. Notably, children residing in urban areas had 46% lower odds of being vaccinated compared to those in rural areas – a pattern that appears to be specific to the study country [26]. Furthermore, on average, children aged 12-23 months and 24-35 months had approximately twice the odds of being vaccinated compared to those aged 9-11 months, suggesting significant delays in MCV1 vaccination during the first year of life.

The spatial variation in the differences in odds of vaccination between the age groups is illustrated by the maps of $\beta_1(\mathbf{s})$ and $\beta_2(\mathbf{s})$ in supplementary Figure 5. The estimated surface for $\beta_2(\mathbf{s})$ appears much smoother than that for $\beta_1(\mathbf{s})$, suggesting greater spatial variability in the differences (in the log-odds of vaccination) between age groups 9-11 months and 12-23 months than between 9-11 months and 24-35 months. This provides a strong justification for the use of spatially varying coefficients in our modelling approach, but it also indicates that to preserve parsimony in the model, the difference in the likelihood of vaccination between ages 9-11 months and 24-35 months could be modelled as a step change, as discussed previously in the methods section. We discuss the spatial patterns in the differences in coverage between the age groups in detail below.

Table 3. Table of parameter estimates. Estimates corresponding to significant regression coefficients are bolded.

Parameter	Mean	Odds ratio	Std. Dev.	2.5%	97.5%
$\hat{\beta}_0$	-3.835	0.022	5.557	-14.727	7.057
Urban	-0.611	0.543	0.143	-0.892	-0.33
Veg_index	-3.384	0.034	2.143	-7.585	0.816

Wetdays	-0.102	0.903	0.102	-0.303	0.099
Dist_conf	0.214	1.239	0.091	0.035	0.393
Elevation	-1×10^{-5}	1.000	1×10^{-5}	-3×10^{-5}	1×10^{-5}
Urban_access	-2×10^{-4}	1.000	5×10^{-4}	-0.001	0.001
Walking_tt	-0.002	0.998	0.001	-0.004	-0.001
Mal_prev	-0.558	0.572	0.834	-2.192	1.077
Max_temp	0.117	1.124	0.144	-0.166	0.4
Mat_educ	1.035	2.815	0.236	0.572	1.498
Health_card	1.824	6.197	0.302	1.231	2.417
Media	0.552	1.737	0.272	0.020	1.085
Wealth	-0.135	0.874	0.247	-0.619	0.349
$\hat{\beta}_{1(12-23 \text{ months})}$	0.716	2.046	0.124	0.473	0.959
$\hat{\beta}_{2(24-35 \text{ months})}$	0.827	2.286	0.305	0.228	1.425
\hat{r}	0.844	-	0.456	0.293	2.028
$\hat{\sigma}_{\omega}$	0.367	-	0.094	0.216	0.583
\hat{r}_{β_1}	0.702	-	0.344	0.283	1.595
$\hat{\sigma}_{\beta_1}$	0.399	-	0.112	0.214	0.649
\hat{r}_{β_2}	6.697	-	5.659	1.561	21.783
$\hat{\sigma}_{\beta_2}$	0.400	-	0.17	0.171	0.828
$\hat{\sigma}_{\epsilon}^{-2}$	4.056	-	1.228	2.151	6.94

The estimated spatial ranges for the spatially varying coefficients - $\beta_1(\mathbf{s})$ and $\beta_2(\mathbf{s})$ - are 78 km and 743 km, respectively, while that of the spatial random effect, ω , is 94 km. Again, these estimates suggest much greater spatial heterogeneities in the differences in coverage between age groups 9-11 months and 12-23 months than between 9-11 months and 24-35 months. Nevertheless, the corresponding marginal variance estimates for these terms are very close, and the iid residual term, ϵ , accounts for nearly twice as much variation in the model as ω .

Similar patterns are observed in the results of MODsmooth reported in supplementary Table 4 and supplementary Figure 6, both in terms of significant covariate effects and the effect of age on vaccination. In particular, the delay in the receipt of MCV1 is evident in the estimated nonlinear effect of age on vaccination, which shows a markedly lower probability of vaccination among children aged < 15 months relative to older ages. We observe a steep increase in the probability of vaccination from age 9 months to about 15 months, which then stabilizes at a high level with minor fluctuations beyond 23 months, suggesting immunity gaps in the 24-35 month birth cohort. The credible intervals for the smooth functions indicate greater uncertainty beyond age 12 months due to greater fluctuations in coverage at older ages (see Figure 1a).

Predicted 1x1 km Maps of Vaccination Coverage for Age Groups and Single age Points

Geographically, considerable heterogeneities in MCV1 coverage are observed across all age groups (Figure 2), with a pronounced effect of walking travel time to the nearest health facility (see supplementary Figure 4). The 1x1 km maps generally reveal lower coverage in the 9-11 month age group relative to other age groups, particularly concentrated in the western and northern districts of the country. The spatial distributions of coverage in these age groups show some interesting similarities, but there are also striking differences. Across all the age groups, coverage levels are consistently higher in and around Lacs, Vallée du Bandama and Yamoussoukro districts. Whilst overall coverage levels do not differ substantially between ages 12-23 months and 24-35 months, the

spatial patterns in coverage appear to be more similar between ages 9-11 months and 24-35 months than 12-23 months, with the latter showing more pronounced spatial gradients, as noted previously. The modelled estimates at the national level are 54.1%, 67.9% and 70.9% for the respective age groups. Compared to corresponding direct survey estimates of 51.2%, 60.3% and 65.1%, the model slightly overestimated coverage, especially in the 12-23 month age group. For the combined 9-35 month age group, similar spatial patterns are observed, although aggregating over this wider age range obscures differences in coverage between birth cohorts and is more suitable when interest is in estimating for broader age bands.

The widths of the 95% credible intervals (CIs) associated with the coverage maps indicate greater uncertainties in areas with sparse survey data (e.g., the Zanzan region; see Figure 1) and in areas of lower coverage, especially for age groups 12-23 months and 24-35 months. Uncertainties for the 9-11 month group show less spatial variation but are slightly higher overall, likely reflecting smaller sample sizes. The median widths of the 95% CIs are 50%, 48%, and 43% for the 9-11, 12-23, and 24-35 month groups, respectively, indicating considerable uncertainties in the grid-level estimates.

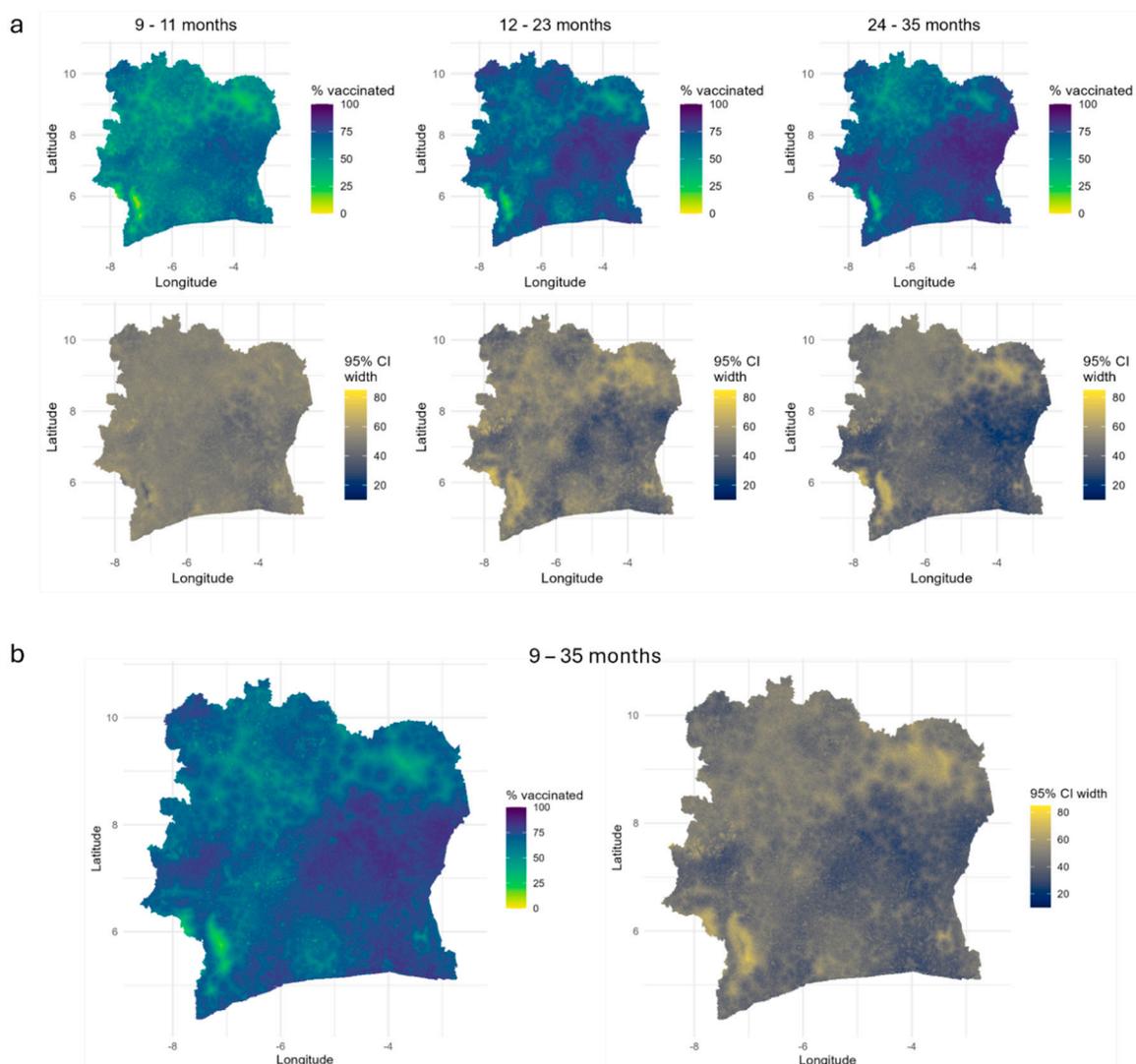


Figure 2. 1x1 km maps of MCV1 coverage and associated uncertainties for age groups 9-11 months, 12-23 months, 24-35 months and 9-35 months.

At single age points (Figure 3), lower coverage at 9 months is also evident, with very similar spatial patterns observed from ages 12 to 35 months. (Note that MODsmooth does not capture spatial gradients in the differences between age groups.) The modelled national coverage estimates of 47.6%,

65%, 68.3%, 70%, 70.3% and 68.7% at 9, 12, 15, 18, 24 and 35 months, respectively, further confirm that coverage patterns stabilize and become more similar from age 15 months and above. The corresponding uncertainties, reported in supplementary Figure 7, show similar patterns and ranges as those in Figure 2.

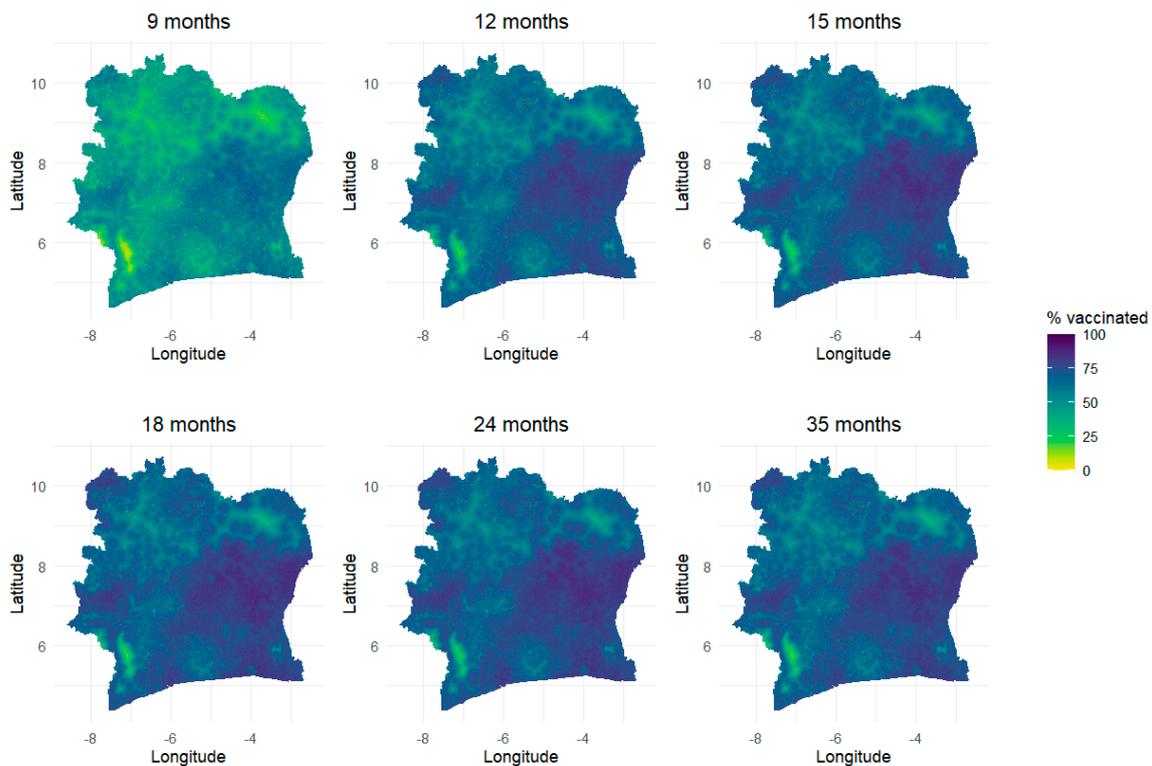


Figure 3. 1x1 km maps of MCV1 coverage for single age points at 9, 12, 15, 18, 24 and 35 months. Corresponding uncertainty estimates, obtained as the widths of the 95% credible intervals, are shown in the supplementary file.

Estimation of Vaccination Coverage at the Administrative Level

At the departmental level, differences in both coverage and spatial prioritization for the 9–11, 12–23, and 24–35 month age groups are shown in Figure 4 and Supplementary Figure 8. Consistent with the national-level estimates reported earlier, coverage remains suboptimal at the departmental level, with no department reaching the target of 95% coverage in any age group. In particular, for the 9–11 month age group, coverage is estimated to be below 60% in about 70% of departments. For the 24–35 month age group, clusters of departments with lower coverage (<60%) are observed in the neighboring Savanes (M'Bengué, Tengrela, Boundiali, and Kouto) and Denguélé (Kaniasso, Madinani, and Séguélon) districts, despite overall higher coverage at older ages - potentially contributing to the dip in coverage after 23 months noted in supplementary Figure 6.

Among 9–11 month olds, the highest coverage levels (>67%) are estimated in M'Bahiakro, Daoukro, and Prikro (all in the Lacs district), while the lowest (<44%) are in Dianra, Daloa, and Séguélon departments. For 12–23 months, coverage is highest (>81%) in M'Bahiakro, Tanda, and Transua (Lacs and Zanzan districts), and lowest (<47%) in Daloa, Dianra, and Guitry. In the 24–35 month group, coverage peaks (>85%) in Tanda, Koun-Fao, M'Bahiakro, and Transua (Zanzan and Lacs districts), and is lowest in Séguélon, Kouto, and Dianra. Although there are similarities in the locations of higher and lower coverage across age groups, differences also emerge, as reflected in the departmental rankings shown in Figure 4b. For example, even within Lacs district - where coverage is consistently higher across all age groups - the ranks of departments differ when comparing the 12–23 month group with the 9–11 and 24–35 month groups. The rankings of the 9–11 and 24–35 month groups are more similar, corroborating earlier findings about the resemblance in their spatial coverage patterns. The median absolute differences in departmental ranks between age groups are:

9 (IQR=12) for 9–11 vs. 12–23 months, 8 (IQR=11) for 9–11 vs. 24–35 months, and 14 (IQR=18) for 12–23 vs. 24–35 months. Furthermore, the rankings appear to vary most in the Zanzan, Denguélé, and Montagnes districts (based on the larger IQRs of ranks across all age groups within these districts).

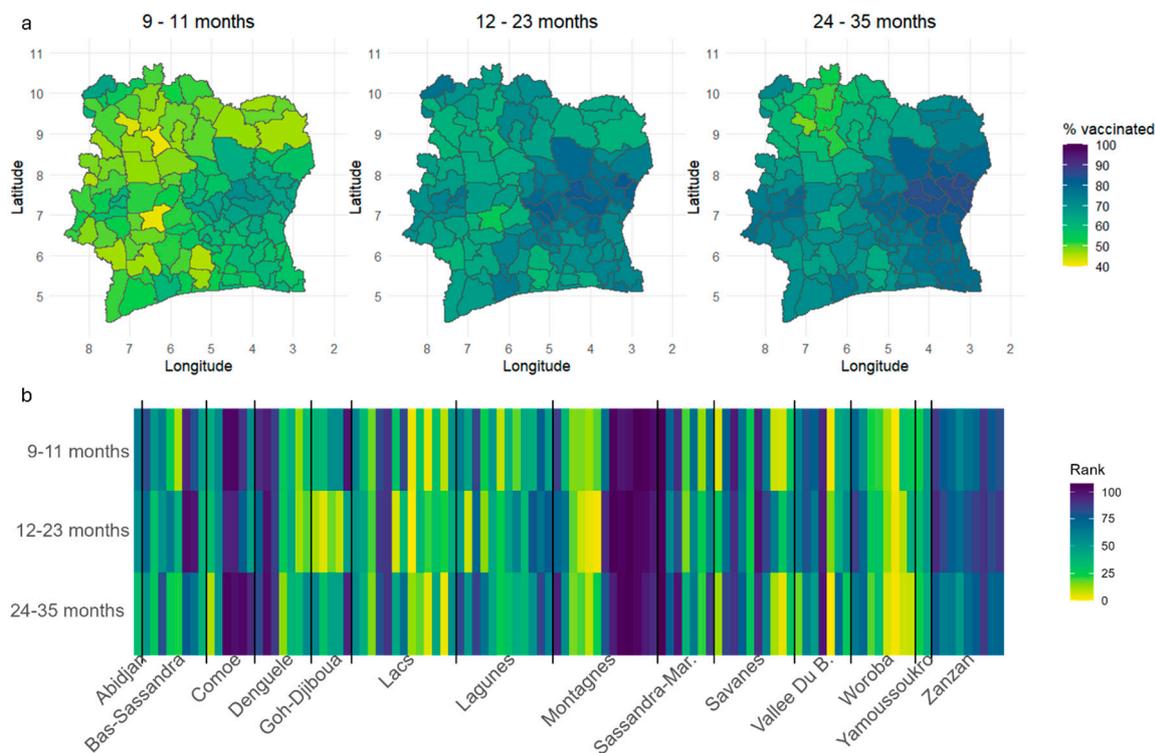


Figure 4: (a) Estimates of MCV1 coverage and (b) corresponding ranks at the departmental level for each age group. The vertical lines differentiate between departments in each of the 14 districts.

The uncertainties associated with the departmental-level estimates (supplementary Figure 9) are considerably smaller than those of the grid-level estimates, with median 95% CI widths of 23%, 22%, and 18% for the respective age groups.

We also estimated coverage at the district level for the 12–23 month age group to enable comparison with direct survey estimates at this level (supplementary Figure 10). Overall, there is good agreement between the modelled and direct survey estimates; however, the relative overestimation of coverage observed earlier is most pronounced in the Abidjan, Comoé, and Lagunes districts.

Discussion

Our study proposed and evaluated a spatially varying coefficient model for age-structured mapping of vaccination coverage, along with alternative specifications offering the flexibility to map coverage for both age groups and single age points. The methodology is also applicable to mapping other health and development indicators where understanding age-related differences is of interest. The proposed models can be readily implemented using the INLA-SPDE approach, although predictions at a high resolution (e.g., 1×1 km) and over large study areas may require high-memory computing environments.

Among the three competing birth cohort models, our validation results based on the 2021 CDHS data provided sufficient support for the spatially varying coefficient model, MODsvc1. We also found that explicitly modelling age improved predictive performance at the individual level. As expected, however, out-of-sample prediction across multiple age categories is constrained by smaller sample sizes compared to predictions for all ages combined. The impact of cluster-level sample sizes on out-of-sample prediction has been explored in detail elsewhere [19]. For larger age groups, more parsimonious parameterizations of MODsvc1 - incorporating fixed step changes between certain age

groups - can be informed by the data, guided by model selection criteria, or refined through inspection of the predicted surfaces from the fully parameterized model and subsequent adjustments in later analyses.

The estimated covariate effects in our analysis highlight the need for interventions aiming to improve coverage in urban areas, conflict areas, areas with limited access to health facilities/vaccination services, as well as those targeted towards improved maternal/caregiver education, improved vaccination card retention (also an indication of improved access to vaccination services) and improved vaccine information communication in Cote d'Ivoire. These findings are corroborated by previous studies [44,45] which analysed similar indicators of vaccination coverage in Cote d'Ivoire. Our study also demonstrated a profound effect of age on vaccination. In the birth cohort model MODsvc1, we estimated that children aged 12–23 months and 24–35 months had approximately twice the odds of being vaccinated compared to those aged 9–11 months. We also observed substantial spatial heterogeneity in the differences in coverage between the 9–11 and 12–23 month age groups. In the single-age model MODsmooth, we found markedly lower vaccination probabilities among children younger than 15 months, as well as slight dips in coverage within the 24–35 month cohort. Together, these findings provide strong evidence of significant delays in MCV1 vaccination during the first year of life. If these patterns persist, interventions targeting eligible children during the first year of life and up to 15 months (e.g., improved maternal/caregiver education and improved access to vaccination services) are likely to be most impactful.

Our predicted maps clearly showed that MCV1 coverage is lowest at younger ages (around the first year of life) in the northern and western districts, and further corroborated the relatively higher and more uniform coverage at older ages. These spatial patterns in coverage are also consistent with Ahoussou et al's [46] study, which analysed measles case-based surveillance data in Cote d'Ivoire and found relatively higher odds of detection of positive measles-specific immunoglobulin M (IgM) in health regions in the west of the country. At the grid, departmental, and district levels, estimated coverage remained well below the 95% target across the country. Specifically, in the 9–11 month age group, coverage was estimated to be below 60% in about 70% of departments. In the 24–35 month age group, despite generally higher coverage, areas of lower coverage (<60%) were concentrated in the neighboring Savanes (M'Bengué, Tengrela, Boundiali, and Kouto) and Denguélé (Kaniasso, Madinani, and Séguélon) districts. Across all age groups, we identified notable differences in departmental rankings, highlighting distinct priority areas for each age group, although some consistent patterns also emerged. For example, the departments of Dianra (Woroba district), Daloa (Sassandra–Marahoué), and Séguélon (Denguélé) were consistently among the three lowest-coverage areas in at least two age groups, whereas Guitry and Kouto were among the lowest-coverage areas only for the 12–23 and 24–35 month groups, respectively. Departmental rankings were most variable in the Zanzan, Denguélé, and Montagnes districts, indicating greater heterogeneity in age-specific coverage within these regions. The uncertainties in these modelled estimates were higher at the grid level (with median 95% CI widths ranging between 43% and 50%) than at the departmental level (with median 95% CI widths ranging between 18% and 23%), suggesting that departmental-level estimates are more actionable for programmatic purposes. Nevertheless, producing estimates at the grid level first provides the flexibility to aggregate to operationally relevant spatial scales where precision improves, while retaining the spatial granularity necessary for targeted interventions.

Our results have several implications for measles control and elimination efforts in Cote d'Ivoire. WHO and UNICEF estimates indicate that national MCV1 coverage remained suboptimal at 70% in 2023 [47]. In 2024 alone, Cote d'Ivoire reported 1,872 measles cases to WHO - a 37.5% increase from the previous year [48]. Addressing existing immunity gaps is therefore critical for progress towards measles control and elimination, and our study provides evidence to help guide intervention strategies. First, our modelled estimates for the 9–11 month age group can support routine immunization interventions aimed at improving coverage in the first year of life, or in children up to 15 months of age, by identifying and prioritizing the areas most at risk. Second, these estimates can inform the planning of catch-up vaccination activities, particularly for children currently aged 5–9

years, which encompasses two of the birth cohorts analyzed in our study (12–23 months and 24–35 months at the time of the survey). For spatial prioritization, the rankings of subnational areas, as demonstrated in Figure 4, could be enhanced by using estimates of the number of zero-dose children, which may better capture areas of highest disease risk. These zero-dose estimates can also feed into microplanning for vaccination activities, while field data collected during such activities can in turn be used to validate and refine the modelled estimates.

Our findings could be further strengthened through triangulation with measles case-based surveillance data or serosurveillance data that reflect the national immunity profile, helping to build a more robust evidence base for targeted interventions [25,46,49]. On the technical side, the spatially varying coefficient model developed here can be extended to age-dependent analyses of these aggregate programmatic data at the district or provincial level [23]. Such extensions will typically involve modelling the spatially-varying regression coefficients and residual spatial dependence in the model using conditional autoregressive priors [50] and considering alternative probability distributions for the outcome. Additionally, it is possible to adapt MODsmooth to produce estimates over a specified age range. This will typically involve computing $\frac{1}{z_2-z_1} \int_{z_1}^{z_2} f(z) dz$ by defining a fine grid within the interval $[z_1, z_2]$, computing the posterior mean of $f(z)$ at each grid point and then averaging over these points. In R-INLA, the integration points over $[z_1, z_2]$ can be included in $f(\cdot)$ during model fitting and then the marginal posterior distributions of $f(\cdot)$ can be obtained post model-fitting and averaged to compute the integral. However, as noted earlier, this approach does not account for spatial non-stationarity when estimating differences in coverage between the age groups. Moreover, estimating age effects at single age points using a spatially varying coefficient model is not well suited for capturing the overall smoothed effect of age (see Supplementary Figure 11). Indeed, in our 2021 CDHS application, such a specification underperformed (WAIC = 5232) relative to MODsmooth in equation (4). In future work, we plan to explore spatiotemporal extensions of the proposed models to allow age- and time-specific estimation of coverage. We anticipate that this analysis will yield additional programmatic insights into how the timeliness of vaccination and differences in coverage between the age groups have changed over space and time within and across countries.

Our analysis has some limitations. In the CDHS application, small cluster-level sample sizes - particularly in the 9–11 month age group - resulted in greater uncertainty in the predictions. We also observed that the fitted model slightly overestimated coverage in some southeastern areas of the country, potentially due to limited sample sizes, poor spatial coverage of clusters in these areas, or oversmoothing by the model. Furthermore, our analysis relied on information from both caregiver recall and vaccination records, which could introduce recall bias. The age range considered in our application was also constrained by data availability – DHS surveys only collect information on vaccination coverage for children aged 0–35 months at the time of the survey. As noted earlier, analyses covering broader age ranges are likely to yield greater insights for program planning and decision-making.

In summary, our study demonstrates the value of the proposed spatially varying coefficient model and alternative specifications for producing high-resolution maps of vaccination coverage by age and identifying priority areas for interventions. By highlighting substantial geographic and age-related heterogeneities in MCV1 coverage in our application, our findings underscore the need for targeted strategies to close immunity gaps, particularly in younger children and underserved areas. The proposed methodological framework is flexible, scalable, and applicable to other health indicators, offering a useful tool to support evidence-based decision-making for immunization programs and beyond. Future extensions to the spatiotemporal setting will further enhance its utility.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Data and code availability: The data used in this study are available from the cited sources. The authors do not have the permission to redistribute these data. All R scripts used in the analyses are available on GitHub: https://github.com/EdsonUtazi/SVC_mod_paper_code.

Author contributions: C. Edson Utazi: Conceptualization, Data Curation, Methodology, Formal analysis, Software, Supervision, Visualization, Funding acquisition, Writing – original draft, Writing – review and editing. Somnath Chaudhuri: Data Curation, Methodology, Formal analysis, Software, Visualization, Writing – review and editing. Oghenebrume Wariri: Methodology, Writing – review and editing. Iyanuloluwa D. Olowe: Data Curation, Visualization, Writing – review and editing. Mohamed Megheib: Formal analysis, Writing – review and editing. Andrew J. Tatem: Supervision, Funding acquisition, Writing – review and editing.

Funding: This study was supported by funding from UNICEF [grant number: 43387656] and Gavi, the Vaccine Alliance.

Acknowledgements: This work represents an output of the VaxPop project led by CEU. The authors would like to thank colleagues from the WorldPop program at the University of Southampton for their support during the study.

Conflict of interest: The authors declare no competing interests.

References

1. World Health Organization. Immunization Agenda 2030: A global strategy to leave no one behind; 2020. Available from: https://www.who.int/immunization/immunization_agenda_2030/en/. [Accessed on 25/06/2020].
2. Cutts FT, Claquin P, Danovaro-Holliday MC, Rhoda DA. Monitoring vaccination coverage: Defining the role of surveys. *Vaccine*. 2016;34(35):4103–9.
3. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. Mapping vaccination coverage to explore the effects of delivery mechanisms and inform vaccination strategies. *Nature Communications*. 2019;10(1):1633.
4. Auzenberg M, Fu H, Abbas K, Procter SR, Cutts FT, Jit M. Health effects of routine measles vaccination and supplementary immunisation activities in 14 high-burden countries: a Dynamic Measles Immunization Calculation Engine (DynaMICE) modelling study. *The Lancet Global Health*. 2023;11(8):e1194–e204.
5. Nelson KN, Wallace AS, Sodha SV, Daniels D, Dietz V. Assessing strategies for increasing urban routine immunization coverage of childhood vaccines in low and middle-income countries: A systematic review of peer-reviewed literature. *Vaccine*. 2016;34(46):5495–503.
6. Gavi The Vaccine Alliance. Gavi Strategy 5.0, 2021-2025; 2020. Available from: <https://www.gavi.org/our-alliance/strategy/phase-5-2021-2025>. [Accessed on 25 June 2021].
7. Haeuser E, Byrne S, Nguyen J, Raggi C, McLaughlin SA, Bisignano C, et al. Global, regional, and national trends in routine childhood vaccination coverage from 1980 to 2023 with forecasts to 2030: a systematic analysis for the Global Burden of Disease Study 2023. *The Lancet*. 2025.
8. Cutts FT, Ferrari MJ, Krause LK, Tatem AJ, Mosser JF. Vaccination strategies for measles control and elimination: time to strengthen local initiatives. *BMC Medicine*. 2021;19(1):2.
9. Wariri O, Utazi CE, Okomo U, Metcalf CJE, Sogur M, Fofana S, et al. Mapping the timeliness of routine childhood vaccination in The Gambia: A spatial modelling study. *Vaccine*. 2023;41(39):5696–705.
10. Manaksha M, Danya Arif S, Vijay Kumar D, Mubarak Taighoon S, Sundus I, Hamidreza S, et al. Coverage, timeliness of measles immunisation and its predictors in Pakistan: an analysis of 6.2 million children enrolled in the Provincial Electronic Immunisation Registry. *BMJ Global Health*. 2025;10(3):e016717.
11. World Health Organization, United Nations Children’s Fund (UNICEF), Gavi the Vaccine Alliance. The Big Catch-Up: An Essential Immunization Recovery Plan for 2023 and Beyond; 2023. Available from: <https://www.who.int/publications/i/item/9789240075511>. [Accessed on 12 July 2025].

12. O'Brien KL, Lemango E. The big catch-up in immunisation coverage after the COVID-19 pandemic: progress and challenges to achieving equitable recovery. *The Lancet*. 2023;402(10401):510–2.
13. Nambiar D, Hosseinpoor AR, Bergen N, Danovaro-Holliday MC, Wallace A, Johnson HL. Inequality in Immunization: Holding on to Equity as We 'Catch Up'. *Vaccines*. 2023;11(5):913.
14. Sbarra AN, Rolfe S, Nguyen JQ, Earl L, Galles NC, Marks A, et al. Mapping routine measles vaccination in low- and middle-income countries. *Nature*. 2021;589(7842):415–9.
15. Takahashi S, Metcalf CJE, Ferrari MJ, Tatem AJ, Lessler J. The geography of measles vaccination in the African Great Lakes region. *Nature Communications*. 2017;8(1):15585.
16. Utazi CE, Nilsen K, Pannell O, Dotse-Gborgbortsi W, Tatem AJ. District-level estimation of vaccination coverage: Discrete vs continuous spatial models. *Stat Med*. 2021;40(9):2197–211.
17. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Takahashi S, Metcalf CJE, et al. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine*. 2018;36(12):1583–91.
18. Utazi CE, Thorley J, Alegana VA, Ferrari MJ, Nilsen K, Takahashi S, et al. A spatial regression model for the disaggregation of areal unit based data to high-resolution grids with application to vaccination coverage mapping. *Statistical Methods in Medical Research*. 2018;28(10-11):3226–41.
19. Utazi CE, Aheto JMK, Chan HMT, Tatem AJ, Sahu SK. Conditional probability and ratio-based approaches for mapping the coverage of multi-dose vaccines. *Statistics in Medicine*. 2022;41(29):5662 – 78.
20. Utazi C, Yankey O, Chaudhuri S, Olowe I, Danovaro-Holliday C, Lazar A, et al. Geostatistical and Machine Learning Approaches for High-Resolution Mapping of Vaccination Coverage. *Preprints: Preprints*; 2024.
21. Gelfand AE, Kim H-J, Sirmans CF, Banerjee S. Spatial Modeling With Spatially Varying Coefficient Processes. *Journal of the American Statistical Association*. 2003;98(462):387–96.
22. Finley AO. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*. 2011;2(2):143–54.
23. Meehan TD, Michel NL, Rue H. Spatial modeling of Audubon Christmas Bird Counts reveals fine-scale patterns and drivers of relative abundance trends. *Ecosphere*. 2019;10(4):e02707.
24. Fotheringham AS, Brunsdon C, Charlton M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: Wiley; 2003.
25. Utazi CE, Aheto JMK, Wigley A, Tejedor-Garavito N, Bonnie A, Nnanatu CC, et al. Mapping the distribution of zero-dose children to assess the performance of vaccine delivery strategies and their relationships with measles incidence in Nigeria. *Vaccine*. 2023;41(1):170–81.
26. Institut National de la Statistique, ICF. Côte d'Ivoire Enquête Démographique et de Santé 2021 Rapport final. Rockville, Maryland, USA et la Côte d'Ivoire: INS et ICF; 2023.
27. UNICEF. Reaching the unreached with life-saving vaccines through data science and geospatial technologies; 2025. Available from: <https://data.unicef.org/resources/reaching-the-unreached-with-life-saving-vaccines-through-data-science-and-geospatial-technologies/>. [Accessed on 12 July 2025].
28. World Health Organisation. Vaccination schedule for Measles (Cote d'Ivoire); 2025. 2025; Available from: https://immunizationdata.who.int/global/wiise-detail-page/vaccination-schedule-for-measles?ISO_3_CODE=CIV&TARGETPOP_GENERAL=. [Accessed on 12 July 2025].
29. Cressie N. *Statistics for Spatial Data*: Wiley; 2015.
30. Tatem AJ. WorldPop, open data for spatial demography. *Scientific Data*. 2017;4(1):170004.
31. Runfola D, Anderson A, Baier H, Crittenden M, Dowker E, Fuhrig S, et al. geoBoundaries: A global database of political administrative boundaries. *PLoS One*. 2020;15(4):e0231866.
32. Matérn B. *Spatial Variation*. 2nd ed. Berlin, Germany: Springer-Verlag; 1960.
33. Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J Roy Stat Soc Series B (Stat Methodol)*. 2011;73(4):423–98.
34. Wang X, Yue YR, Faraway JJ. *Bayesian regression modeling with INLA*: Chapman and Hall/CRC; 2018.
35. Browne WJ, Subramanian SV, Jones K, Goldstein H. Variance partitioning in multilevel logistic models that exhibit overdispersion. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2005;168(3):599–613.

36. Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*. 2017;32(1):1–28.
37. Fuglstad G-A, Simpson D, Lindgren F, Rue H. Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*. 2019;114(525):445–52.
38. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009;71(2):319–92.
39. Lindgren F, Bachl F, Illian J, Suen MH, Rue H, Seaton AE. inlabru: software for fitting latent Gaussian models with non-linear predictors. arXiv preprint arXiv:240700791. 2024.
40. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*. 2010;11(12):3571–94.
41. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*. 2014;24(6):997–1016.
42. Gneiting T, Raftery AE. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*. 2007;102(477):359–78.
43. Brier GW. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*. 1950;78(1):1–3.
44. Ogundele OA, Ogunwemimo HS, Fehintola FO, Ogundele T, Olorunsola A, Bello OE, et al. Predictors of incomplete childhood vaccination in four West African countries: a population based cross-sectional study. *Sci Rep*. 2025;15(1):17119.
45. Douba A, Aka Lepri Bernadin N, Attoh-Toure H, Bangaman Akani C, Yao GHA, Konan Ng, et al. An analysis of risk factors for incomplete immunization for children in Côte d'Ivoire: Examination of 1998-1999 and 2011-2012 Demographic and Health Survey. *HEALTH SCIENCES AND DISEASE*. 2016;17(1).
46. Ahoussou EMK, Ekra KD, Dananche C, Aka LBN, Rabilloud M, Vanhems P. Factors associated with detection of measles-specific immunoglobulin M (IgM-positive) in Côte d'Ivoire from 2015 to 2019: A nested case-control study from a measles national surveillance programme. *Journal of Public Health and Epidemiology*. 2023;15(2):97–105.
47. World Health Organization. WHO/UNICEF Estimates of National Immunization Coverage (WUENIC) 2024 revision; 2024. Available from: <https://www.who.int/teams/immunization-vaccines-and-biologicals/immunization-analysis-and-insights/global-monitoring/immunization-coverage/who-unicef-estimates-of-national-immunization-coverage>. [Accessed on 05 July 2025].
48. World Health Organisation. Côte d'Ivoire Reported cases of vaccine-preventable diseases (VPDs); 2025. Available from: <https://immunizationdata.who.int/dashboard/regions/african-region/CIV>. [Accessed on 12 July 2025].
49. Carcelen AC, Winter AK, Moss WJ, Chilumba I, Mutale I, Chongwe G, et al. Leveraging a national biorepository in Zambia to assess measles and rubella immunity gaps across age and space. *Scientific Reports*. 2022;12(1):10217.
50. Lee D. A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spatiotemporal Epidemiol*. 2011;2(2):79–89.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.