

District-level geographic variation in National Health Insurance Scheme enrollment in Ghana: a three-level Bayesian hierarchical analysis of the 2022 Demographic and Health Survey

Welson Bentum

wbentum1@st.knust.edu.gh

Kwame Nkrumah University of Science and Technology

Research Article

Keywords: Bayesian hierarchical model, Multilevel analysis, Geographic variation, Intraclass correlation coefficient, Universal health coverage, Health insurance enrollment

Posted Date: June 10th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9953971/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

District-level geographic variation in National Health Insurance Scheme enrollment in Ghana: a three-level Bayesian hierarchical analysis of the 2022 Demographic and Health Survey

Welson Bentum^{1*}

^{1*}Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

Corresponding author(s). E-mail(s): wbentum1@st.knust.edu.gh;

Abstract

Ghana's National Health Insurance Scheme (NHIS) has achieved high aggregate enrollment nationally, yet substantial variation exists across administrative districts that aggregate figures conceal. Prior analyses have relied on flat logistic regression, which cannot quantify how much enrollment variation is geographic in origin or produce reliable district-level estimates for policy targeting. This study fit a three-level Bayesian hierarchical logistic regression to the 2022 Ghana Demographic and Health Survey, comprising **22,058** respondents across 618 enumeration area clusters, 237 sampled districts, and 16 administrative regions. The three-level model substantially outperforms the flat regression model ($\Delta\text{LOOIC} = \mathbf{844.4}$). Approximately 27.0 % of total enrollment variation among women ($\text{ICC} = \mathbf{0.270}$; 95 % CrI: 0.207–0.370) and 19.2 % among men ($\text{ICC} = \mathbf{0.192}$; 95 % CrI: 0.138–0.290) is attributable to shared district and regional context. Higher education confers over threefold enrollment odds for women (OR: 3.83) and over fivefold for men (OR: 5.43), with a monotonic wealth gradient and sex-differentiated employment, pregnancy, and religion effects also observed. Composition-adjusted posterior district enrollment probabilities range from 84.3 % to 99.2 % for women and 47.2 % to 91.4 % for men, with low-enrollment districts dispersed nationally. Geographic context is a first-order determinant of NHIS enrollment in Ghana that prior flat-regression analyses have been structurally unable to detect.

Keywords: Bayesian hierarchical model, Multilevel analysis, Geographic variation, Intraclass correlation coefficient, Universal health coverage, Health insurance enrollment.

1 Introduction

The goal that all people receive the quality health services they need to be effective, when and where they need them, without suffering financial hardship, popularly referred to as the Universal Health Coverage is enshrined in the United Nations Sustainable Development Goals (Goal 3.8) [1]. Ghana was among the first sub-Saharan African (SSA) countries to legislate a national health insurance framework, establishing the National Health Insurance Scheme (NHIS) in 2003 under the National Health Insurance Act (Act 650) [2], later replaced by the National Health Insurance Act, 2012 (Act 852) [3]. The NHIS provides a defined package of services covering approximately 95 % of commonly occurring disease conditions at accredited public and private facilities, financed primarily through a 2.5 % National Health Insurance Levy on goods and services and a 2.5 % deduction from formal sector Social Security and National Insurance Trust (SSNIT) contributions [3, 4].

Despite two decades of implementation, NHIS enrollment remains geographically uneven. The 2022 Ghana DHS records aggregate enrollment of 90 % among women aged 15–49 and 73 % among men aged 15–59 [5], yet regional figures from the 2021 Population and Housing Census reveal that female enrollment ranges from 56.2 % in the Oti Region to 88.9 % in the Upper East Region, and male enrollment from 48.1 % (Oti) to 83.8 % (Upper East), indicating substantial geographic heterogeneity that aggregate national figures conceal [6]. The NHIA administers the scheme through sixteen regional offices, one in each of Ghana’s political regions, supported by 166 district offices and 5 registration centers where members register and renew their membership [7]. Understanding which districts are underperforming and why is therefore directly relevant to the NHIA’s outreach strategy and Ghana’s Universal Health Coverage road map [8].

Prior studies have identified the socioeconomic determinants of NHIS enrollment, yet none has examined the geographic clustering or district-level variation that motivates this analysis. Early foundational studies established the core socioeconomic gradient: [9] assessed the determinants of enrolling in and dropping out from the NHIS by using a multinomial logit model on household survey data to evaluate respondent’s perceptions relating to schemes, providers and community health beliefs and attitudes, while [10] applied propensity score matching techniques to balance the relevant background characteristics and compare health indicators of recent mothers who are enrolled in the NHIS with those who are not. [11] used qualitative key in-depth interviews, observations and informal conversations across 15 communities in the Central and Eastern Regions (ERs) in Ghana to document that poverty, drug shortages, and dissatisfaction with service quality are among the principal barriers to enrollment and retention, and [12] applied logistic regression to five-district household data, identifying wealth status, age, health status, locality, perception about the quality of care at health facilities and perception of the NHIS, as the key factors that determine enrollment into the scheme. [4] examined enrollment trends using administrative time-series data, finding that persons below the age of 18 years and the informal sector workers had the highest renewal rates. The most recent analysis on the same 2022 GDHS used in this study, [13], applied multivariable logistic regression to examine education, residence, and socioeconomic predictors of enrollment, confirming the established gradient

but leaving geographic clustering entirely unmodeled. None of these studies estimate a VPC or ICC, produce district-level posterior enrollment estimates, or test variation across administrative levels. [14] is the sole multilevel study of NHIS enrollment using pooled national surveys, applying frequentist multilevel logistic regression to the 2014 DHS, Multiple Indicator Cluster Survey (MIC) conducted in 2011, and the sixth wave of the Ghana Living Standard Survey (GLSS) conducted in 2012–13 across three levels: geographical clusters, household, and individual units. Their analysis confirms the socioeconomic gradient and finds significant cluster-level variation, but uses data now a decade old, produces no posterior distributions, no district-level enrollment maps, and no formal decomposition of variation across administrative levels. A recent ecological study by [15] used spatial autocorrelation and machine learning on the 2021 Population and Housing Census to identify 42 high-uninsurance districts, predominantly in northern Ghana, with district-level illiteracy as the dominant predictor. That analysis operates at the aggregate district level and cannot separate individual-level socioeconomic disadvantage from contextual geographic effects, estimate posterior uncertainty at the district level, or formally decompose enrollment variation across administrative tiers. Hierarchical and spatial modeling frameworks have been applied in Ghana and sub-Saharan Africa to address analogous inferential problems, including Bayesian hierarchical spatiotemporal models for neonatal mortality [16], Bayesian space-time smoothing for small-area child mortality estimation [17], model-based geostatistical mapping of malaria risk [18], and Bayesian spatial modeling of HIV prevalence [19].

Flat regression has two critical limitations in this setting. First, because DHS data have an explicitly hierarchical structure, with individuals nested within enumeration area clusters, clusters within districts, and districts within regions, the independence assumption of flat logistic regression is highly questionable. Standard errors are underestimated, leading to overconfident inference on individual-level predictors [20]. Second, flat regression cannot produce reliable district-level enrollment estimates, as it offers no mechanism to borrow information across districts; estimates for small districts are therefore unstable. Under flat regression each district is estimated independently with no information borrowed across units, while under complete pooling all districts collapse to a single intercept. Partial pooling is the statistically optimal middle ground, where data-rich districts receive estimates close to their observed rate while data-sparse districts are shrunk toward the national mean in proportion to their within-district sample size, with optimality formalized by the James-Stein theorem [20, 21]. This shrinkage principle is the core inferential motivation for multilevel modeling [22].

To address this gap, this study presents the first Bayesian hierarchical analysis of NHIS enrollment in Ghana using individual-level survey data, jointly characterizing multi-level geographic variation and structural enrollment disadvantage with full posterior uncertainty quantification and systematic sensitivity analysis.

2 Materials and Methods

2.1 Data Source and Study Design

Data for this study come from the 2022 Ghana Demographic and Health Survey (Ghana DHS 2022), conducted from 17 October 2022 to 14 January 2023 by the Ghana Statistical Service, Ghana Health Service, and ICF [5]. The survey employs a two-stage stratified cluster sampling design: 618 enumeration areas (EAs) were selected with probability proportional to size from the 2021 Population and Housing Census frame, stratified into 16 strata defined by region and urban/rural residence, with approximately 30 households selected per EA. DHS Individual Recode (IR) and Male Recode (MR) files are available upon registration and approval at <https://dhsprogram.com>.

2.2 Spatial Architecture and Restructuring

To evaluate the National Health Insurance Scheme within Ghana’s contemporary decentralized governance framework, the geographic coordinates of all 618 primary sampling unit (PSU) enumeration area clusters from the 2022 Ghana Demographic and Health Survey (GDHS) were spatially linked to contemporary administrative district boundaries using R. Administrative boundaries were obtained from the Global Administrative Areas (GADM) database and represented the 261 officially recognized local government districts in Ghana.

The spatial linkage procedure was necessary because the Demographic and Health Survey (DHS) Program intentionally displaces cluster geographic coordinates to preserve respondent confidentiality. Consequently, a conventional point-in-polygon spatial intersection may fail to assign some displaced clusters to their correct administrative units, particularly when cluster centroids are located near district boundaries. Such displacement can result in clusters being positioned outside valid district polygons or within neighboring administrative jurisdictions, thereby introducing potential geographic misclassification.

To address this challenge, a fully automated two-stage assignment procedure was implemented using the `sf` spatial analysis framework in R. In the first stage, a deterministic point-in-polygon intersection was used to assign each cluster coordinate to the district polygon within which it was geographically located. Let p_k denote the geographic coordinate of cluster k , and let M_j represent the polygon corresponding to district j , where $j = 1, \dots, J$ and $J = 261$. Cluster assignments were initially determined according to:

$$A_k = j \quad \text{if} \quad p_k \in M_j. \quad (1)$$

Clusters that could not be matched through direct spatial intersection were subsequently processed using an automated nearest-feature reassignment procedure. For each unmatched cluster, the shortest spatial distance between the cluster coordinate and all district polygons was evaluated, and the cluster was assigned to the nearest valid administrative unit according to

$$A_k = \arg \min_{j \in \{1, \dots, J\}} d(p_k, M_j), \quad (2)$$

where $d(p_k, M_j)$ denotes the minimum spatial distance between cluster coordinate p_k and district polygon M_j . This nearest-feature fallback mechanism ensured that all displaced clusters received valid district assignments while avoiding the loss of observations through selective exclusion.

The spatial intersection procedure was implemented using the `st_join()` function with an intersection predicate, whereas nearest-neighbor reassignment of unmatched clusters was performed using the `st_nearest_feature()` function within the `sf` package [23].

Following district assignment, a programmatic administrative crosswalk was applied to harmonize district-level classifications with Ghana’s post-2019 regional reorganization. Specifically, districts affected by the creation of the Oti, Bono, Bono East, Ahafo, Savannah, North East, and Western North Regions were systematically reclassified to align with the regional identifiers recorded in the GDHS. In the small number of instances where residual discrepancies persisted because of boundary revisions or geometric inconsistencies between data sources, the regional identifier contained within the DHS survey records was treated as authoritative and used to resolve the conflict.

Of Ghana’s 261 administrative districts, only districts containing at least one sampled DHS cluster contributed district-level posterior estimates to the analysis. Districts without sampled clusters were retained within the administrative boundary layer for cartographic completeness and were displayed separately in the resulting maps. The final spatial processing framework achieved complete district assignment for all sampled enumeration area clusters, thereby ensuring that every survey observation could be linked to a valid administrative healthcare delivery context for subsequent multilevel and spatial analyses.

2.3 Analytical Sample and Theoretical Framework

The analytical sample comprises 15,014 women aged 15–49 (n_w) and 7,044 men aged 15–59 (n_m) who completed individual questionnaires. The outcome y_{ijk} is binary NHIS enrollment status (1 = currently enrolled, 0 = otherwise) for individual i in district j and region k , measured by the DHS self-report indicator (`v481` for women; `mv481` for men). The self-report indicator is preferred over the card-seen indicator (`s1119/sm822`) because many enrolled respondents do not carry their card at interview, and card-seen coding introduces attrition that renders the subsample unrepresentative. Observations with missing or “don’t know” responses are excluded. A card-seen sensitivity analysis is reported in the Results section. Women and men are analyzed separately following [13, 14]. To analyze the determinants of National Health Insurance Scheme (NHIS) enrollment, variables were selected and structured according to the Andersen Behavioral Model [24]. This theoretical framework categorizes individual and structural predictors into three distinct dimensions:

- **Predisposing Factors:** Demographic traits that inherently influence an individual’s propensity to seek insurance coverage. In this analysis, these include age cohort, marital status, and religious affiliation.

- **Enabling Factors:** Personal, household, or economic resources that facilitate or inhibit an individual’s structural capacity to enroll. These include formal educational attainment, household wealth quintile, employment status, and urban-rural residential status.
- **Need Factors (Household and Individual):** Structural or biological vulnerabilities that dictate the immediate perceived or evaluated necessity of health insurance coverage. These indicators include total household size, female household headship, current pregnancy status, and parity.

2.4 Hierarchical Model Structure and Specification

To evaluate determinants of NHIS enrollment while accounting for the clustered structure of the survey data, this study adopts a three-level Bayesian hierarchical logistic regression model. The levels are:

- Level 1 (individual i): male or female interviewed
- Level 2 (district j): individuals nested within $j = 1, \dots, J_k$ districts
- Level 3 (region k): districts further nested within $k = 1, \dots, 16$ regions.

The model includes random effects at the district and region levels; no EA-level intercept is specified. EAs are densely nested within districts, so the district random effect $u_{j(k)}$ absorbs the localized spatial clustering operating at that scale. The district also corresponds to the primary operational tier for health insurance governance and public infrastructure financing in Ghana.

The first level of the model specifies the conditional data likelihood. Assuming individual enrollment decisions are independent conditional on the cluster-level random effects, each response is modeled as:

$$y_{ijk} \mid \pi_{ijk} \sim \text{Bernoulli}(\pi_{ijk}). \quad (3)$$

The conditional enrollment probability $\pi_{ijk} = P(y_{ijk} = 1 \mid \mathbf{X}_{ijk}, u_{j(k)}, v_k)$ is linked to the linear predictor via the logit function:

$$\text{logit}(\pi_{ijk}) = \ln\left(\frac{\pi_{ijk}}{1 - \pi_{ijk}}\right) = \mathbf{X}_{ijk}^\top \boldsymbol{\beta} + u_{j(k)} + v_k, \quad (4)$$

where

- $\mathbf{X}_{ijk}$ represents the vector of individual-level demographic covariates
- $\boldsymbol{\beta}$ is the vector of fixed-effect parameters
- $u_{j(k)} \sim \mathcal{N}(0, \sigma_u^2)$ represents the district-level random intercept nested within region k
- $v_k \sim \mathcal{N}(0, \sigma_v^2)$ represents the regional-level random intercept.

Building on Equations (3) and (4), four model specifications are compared.

Model 1 (M_1): Flat Logistic Regression

This is the baseline model, which assumes spatial homogeneity across Ghana, omitting all random intercepts:

$$\text{logit}(\pi_i) = \mathbf{X}_i^\top \boldsymbol{\beta} \quad (5)$$

where $\mathbf{X}_i$ is the vector of individual and household covariates.

Model 2 (M_2): Two-Level Cluster Model

This specification introduces localized community-level clustering using the survey enumeration areas (EAs) as a random intercept:

$$\text{logit}(\pi_{ij}) = \mathbf{X}_{ij}^\top \boldsymbol{\beta} + c_j \quad (6)$$

where $c_j \sim \mathcal{N}(0, \sigma_c^2)$ represents the unobserved random effect of survey cluster j .

Model 3 (M_3): Three-Level Hierarchical Model (Main Specification)

The main model accounts for administrative hierarchy, nesting individuals within districts (j) and regions (k). This specification is stratified separately by gender.

$$\text{logit}(\pi_{ijk}) = \mathbf{X}_{ijk}^\top \boldsymbol{\beta} + u_{j(k)} + v_k \quad (7)$$

To test for broader structural gender differences, a pooled version of (M_3) which includes a sex indicator (S_i) and its cross-product interactions with all individual covariates is considered:

$$\text{logit}(\pi_{ijk}) = \mathbf{X}_{ijk}^\top \boldsymbol{\beta} + \gamma S_i + (S_i \cdot \mathbf{X}_{ijk})^\top \boldsymbol{\delta} + u_{j(k)} + v_k \quad (8)$$

This study also performs a sensitivity analysis in which the district random effect is replaced by a survey cluster intercept $c_{j(k)}$, removing any dependence on the GPS-to-district spatial join and providing a direct check against boundary misclassifications:

$$\text{logit}(\pi_{ijk}) = \mathbf{X}_{ijk}^\top \boldsymbol{\beta} + c_{j(k)} + v_k \quad (9)$$

2.5 Weakly Informative Priors

To estimate the model parameters, weakly informative priors were chosen assigned to all parameters following foundational recommendations by [25]. This approach provides sufficient regularization to ensure computational stability and prevent overfitting within a highly stratified three-level model, while remaining intentionally broad so that the final posterior estimates are predominantly driven by the observed data.

For the fixed effects ($\boldsymbol{\beta}$) and the global intercept, independent normal priors were specified:

$$\boldsymbol{\beta} \sim \text{Normal}(0, 2.5^2) \quad (10)$$

Since the model operates on the log-odds scale via the logit link function, a prior standard deviation of 2.5 represents a wide, conservative range, fixing the probability mass within a realistic envelope, as a log-odds change of ± 2.5 corresponds to a twelvefold increase or decrease in the odds of NHIS enrollment ($e^{2.5} \approx 12.2$). This effectively prevents the model from exploring mathematically unstable, extreme values (or encountering complete separation) while placing minimal constraint on the direction and magnitude of covariate effects

For the group-level random effects representing geographic variation, specifically, the standard deviations of the district-level intercepts (σ_u) and region-level intercepts (σ_v), a half-Student-t prior was applied:

$$\sigma_u, \sigma_v \sim \text{Student-}t(3, 0, 1)\mathbb{I}(0, \infty) \quad (11)$$

As demonstrated by [26], traditional conditionally conjugate inverse-gamma priors can introduce undesirable biases and sensitivity when estimating variance components in hierarchical models. Consequently, a half-Student-t distribution (truncated at zero, $\mathbb{I}(0, \infty)$) with 3 degrees of freedom and a scale parameter of 1 was specified. This reflects a conservative prior expectation that geographic variation across districts and regions is modest. However, the choice of a Student-t distribution ensures heavy tails, hence allowing the model to capture large, unexpected geographical disparities across the 16 regions if such evidence is robustly supported by the survey data, without forcibly shrinking those real-world differences toward zero.

2.6 Variance Partition Coefficients

Geographic variation in NHIS enrollment is quantified using Variance Partition Coefficients (VPCs) and the intraclass correlation coefficient (ICC), estimated from the three-level model (M_3) separately for women and men. All variance components are computed on the latent scale following [27–29], with the individual-level residual fixed at $\sigma_e^2 = \pi^2/3 \approx 3.289$. The total latent variance is partitioned as:

$$\sigma_{\text{total}}^2 = \sigma_v^2 + \sigma_u^2 + \frac{\pi^2}{3} \quad (12)$$

where σ_v^2 and σ_u^2 are the between-region and between-district posterior variance components. District- and region-level VPCs are defined as:

$$\text{VPC}_{\text{Dist}} = \frac{\sigma_u^2}{\sigma_{\text{total}}^2}, \quad \text{VPC}_{\text{Reg}} = \frac{\sigma_v^2}{\sigma_{\text{total}}^2} \quad (13)$$

The ICC for two individuals residing in the same district is:

$$\text{ICC}_{\text{same-dist}} = \frac{\sigma_u^2 + \sigma_v^2}{\sigma_{\text{total}}^2} \quad (14)$$

All derived quantities (VPC_{Dist} , VPC_{Reg} , and $\text{ICC}_{\text{same-dist}}$) are computed at each post-warmup draw by evaluating the above expressions across the full joint posterior, yielding a posterior distribution for each quantity. Posterior medians and 95% credible intervals are reported for all variance components, VPCs, and the ICC.

2.7 Controlled Posterior Spatial Predictions

Evaluating raw enrollment proportions across districts can introduce severe compositional bias if certain areas contain higher concentrations of wealthy, highly educated,

or urban individuals. Hence, district-level enrollment probabilities are computed for a fixed reference profile $\mathbf{X}_{\text{ref}}$ to remove compositional differences across districts: a married, employed, Christian, urban woman or man with secondary education and middle-quintile wealth (Q3), not currently pregnant. For this reference vector, the conditional posterior enrollment probability for a specific district j nested within region k is calculated by passing the integrated linear predictor through an inverse-link logistic transformation:

$$\hat{\pi}_{jk}^{\text{ref}} = \mathbb{E} [\text{expit} (\mathbf{X}_{\text{ref}}^{\top} \boldsymbol{\beta} + u_{j(k)} + v_k) \mid \mathbf{Y}] \quad (15)$$

This expectation is numerically integrated via Monte Carlo averaging over the full array of 8,000 post-warm-up draws from the empirical posterior distribution. Predictions are computed separately for women and men and mapped to GADM Level-2 district polygons via the `district_id` key.

2.8 Survey Weights

Survey design weights (`v005/1e6`) are applied to all baseline descriptive statistics to guarantee that population mean estimates accurately reflect national demographic trends. They are omitted from the hierarchical model likelihoods, as the random intercepts for districts and regions directly account for the unequal selection probabilities induced by the stratified cluster design. As a robustness check, a survey-weighted model is estimated using weights scaled to a sample mean of one:

$$w_{\text{scaled}} = \frac{w_i}{\bar{w}} \quad (16)$$

and incorporated into the model likelihood for direct comparison with the unweighted results.

2.9 Posterior Computation

All analyses were conducted in R version 4.4.2. Models were fitted using `brms` version 2.23.0 [30], which interfaces with Stan version 2.32.7 [31] via the `rstan` package. Four parallel chains are run for 4,000 iterations each, with 2,000 warmup iterations, yielding 8,000 post-warmup draws. `adapt_delta` is set to 0.99 and `max_treedepth` to 12. Convergence is assessed by the Gelman-Rubin statistic ($\hat{R} \leq 1.01$) and bulk and tail effective sample sizes ($\text{ESS} > 400$). Model fit is compared across M_1 , M_2 , and M_3 using Leave-One-Out cross-validation (LOO-CV) and the Watanabe-Akaike Information Criterion (WAIC).

2.10 Sensitivity and Robustness Analyses

Five sensitivity analyses are conducted, each targeting a distinct source of uncertainty. To assess sensitivity to prior specification, M_3 (for women) is re-estimated under three alternative priors on the group-level standard deviations: Half-Normal ($\text{Normal}(0, 2.5)$, $\text{lb} = 0$), Half-Cauchy ($\text{Cauchy}(0, 1)$, $\text{lb} = 0$), and Exponential(1), with stability of VPC_{Dist} examined across specifications. To evaluate sensitivity to outcome definition, the self-reported enrollment indicator (`v481/mv481`) is replaced

with the card-verified indicator (`s1119/sm822`). Potential bias from missing covariate data is addressed by multiple imputation by chained equations [32] ($m = 10$, predictive mean matching), with posterior draws combined using Rubin’s rules [33], applied when missingness in any covariate exceeds 1 %. To test whether covariate effects differ between women and men, a pooled model with sex-by-covariate interaction terms is estimated on the combined sample. Finally, the unweighted specification is compared against a model in which the scaled design weights w_{scaled} are incorporated directly into the likelihood.

3 Results

3.1 Sample Description

Table 1 presents the sample composition by region, with unweighted and weighted counts reported separately. The analytical sample comprises 15,014 women and 7,044 men across 16 regions. Unweighted sample sizes are largest in Northern (women: $n = 1,169$; men: $n = 531$) and Ashanti (women: $n = 1,131$; men: $n = 491$), reflecting the DHS oversampling of smaller regions, and smallest in Western North (women: $n = 792$) and Western (women: $n = 797$).

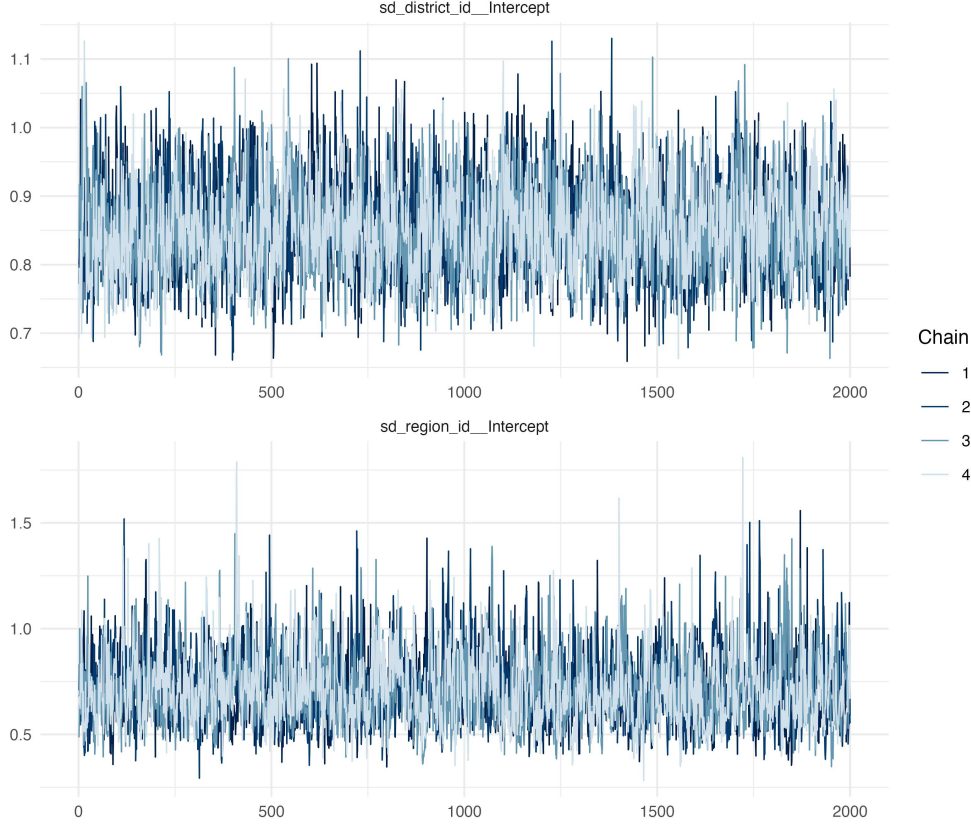
Table 1 Analytical sample composition by region, Ghana DHS 2022.

Region	n_w	n_m	n_w (weighted)	n_m (weighted)
Western	797	382	959	458
Central	979	442	1,703	766
Greater Accra	969	493	2,327	1,210
Volta	837	335	713	278
Eastern	854	389	1,220	557
Ashanti	1,131	491	2,928	1,315
Western North	792	402	411	209
Ahafo	849	422	317	158
Bono	835	351	567	241
Bono East	974	500	676	343
Oti	921	467	403	211
Northern	1,169	531	1,149	526
Savannah	999	548	319	176
North East	963	416	290	128
Upper East	987	464	640	301
Upper West	958	411	398	168
Total	15,014	7,044	15,014	7,044

3.2 MCMC Convergence

All parameters across both the women’s and men’s models achieved $\hat{R} \leq 1.01$, and bulk and tail effective sample sizes exceeded 400 for all fixed effects and 1,000 for the variance parameters σ_u and σ_v . No divergent transitions were recorded after warm-up in any of the three models. Trace plots for the variance standard deviation parameters from M_3 (Figs 1 and 2) confirm well-mixed chains with no evidence of drift.

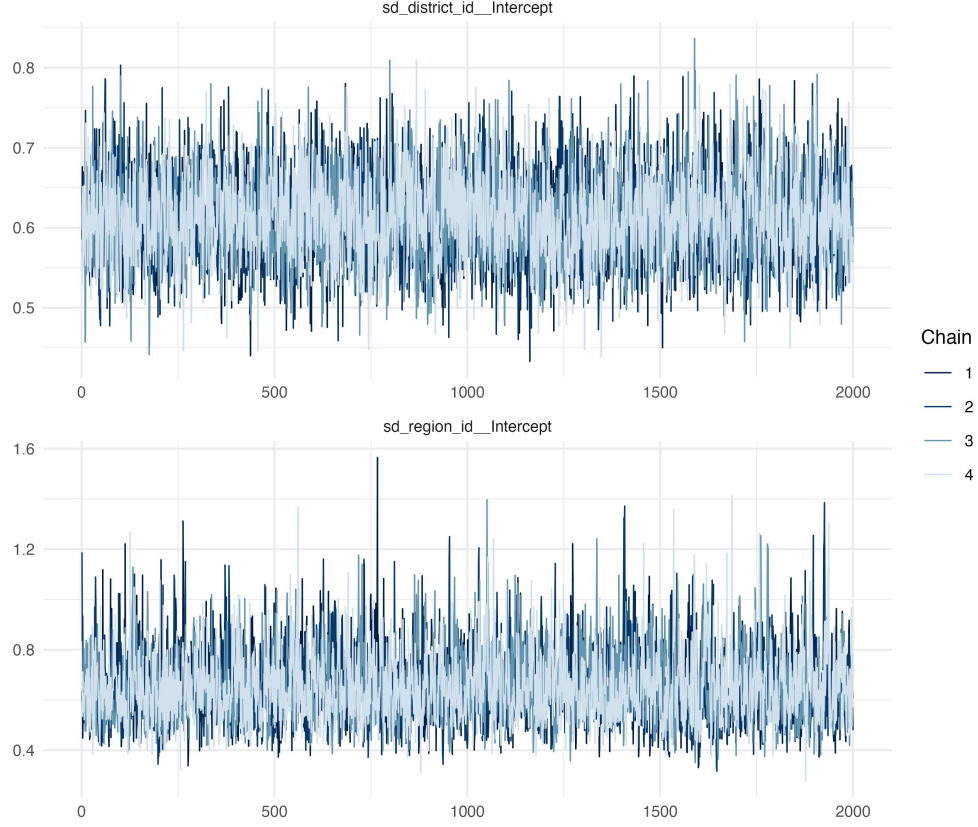
Fig. 1 MCMC trace plots for variance standard deviation parameters (σ_u , σ_v) from the three-level women’s model.



3.3 Model Comparison

Table 2 reports LOO-CV and WAIC for M_1 , M_2 , and M_3 . For both sexes, M_3 is preferred over M_1 : $\Delta\text{LOOIC} = 844.4$ for women and 547.8 for men, confirming that the hierarchical structure captures genuine clustering absent from the flat model. The comparison between M_2 and M_3 differs by sex: for men, M_3 improves substantially on M_2 ($\Delta\text{LOOIC} = 62.0$), while for women the difference is small ($\Delta\text{LOOIC} = 5.4$) relative to the individual-model SE of 149.7, indicating approximately equivalent out-of-sample

Fig. 2 MCMC trace plots for variance standard deviation parameters (σ_u , σ_v) from the three-level men’s model.



predictive accuracy between the two-level and three-level specifications for women. M_3 is nonetheless retained as the primary model on substantive grounds: the district and region levels correspond to the administrative structure of NHIS governance, and their separate variance components are of direct policy interest.

3.4 Variance Decomposition

Table 3 reports posterior variance components and VPC/ICC estimates from M_3 . For women, the district standard deviation is $\hat{\sigma}_u = 0.850$ (95 % CrI: 0.735–0.992) and the region standard deviation is $\hat{\sigma}_v = 0.690$ (95 % CrI: 0.452–1.084). $VPC_{\text{Dist}} = 0.159$ (95 % CrI: 0.120–0.209), indicating that 15.9 % of total enrollment variation on the latent scale arises from between-district differences net of regional heterogeneity, and $VPC_{\text{Reg}} = 0.106$ (95 % CrI: 0.049–0.227), indicating that a further 10.6 % is attributable to broader differences across regions. The combined ICC is 0.270 (95 % CrI: 0.207–0.370), meaning that 27.0 % of total enrollment variation among women is attributable to shared geographic context.

Table 2 Model comparison via LOO-CV and WAIC, stratified by sex.

Model	LOOIC	SE	WAIC	Δ LOO	Δ WAIC
<i>Women</i> ($n_w = 15,014$)					
M1: Flat	8,678.3	158.0	8,678.3	844.4	845.4
M2: Two-level (EA)	7,839.3	149.7	7,833.7	5.4	0.8
M3: Three-level	7,833.9	149.1	7,832.9	0.0	0.0
<i>Men</i> ($n_m = 7,044$)					
M1: Flat	7,379.9	86.1	7,379.8	547.8	548.5
M2: Two-level (EA)	6,894.0	90.0	6,889.4	62.0	58.1
M3: Three-level	6,832.0	90.6	6,831.3	0.0	0.0

For men, $\hat{\sigma}_u = 0.610$ (95 % CrI: 0.511–0.725) and $\hat{\sigma}_v = 0.633$ (95 % CrI: 0.430–0.975). $VPC_{\text{Dist}} = 0.091$ (95 % CrI: 0.065–0.125), indicating that 9.1 % of total enrollment variation on the latent scale arises from between-district differences net of regional heterogeneity, and $VPC_{\text{Reg}} = 0.099$ (95 % CrI: 0.048–0.206), indicating that a further 9.9 % is attributable to broader differences across regions. The combined ICC is 0.192 (95 % CrI: 0.138–0.290), meaning that 19.2 % of total enrollment variation among men is attributable to shared geographic context, a smaller share than for women.

Table 3 Posterior variance components and Variance Partition Coefficients from the three-level model (M_3), stratified by sex. Values are posterior medians with 95 % credible intervals.

Parameter	Women	Men
σ_u (district SD)	0.850 (0.735, 0.992)	0.610 (0.511, 0.725)
σ_v (region SD)	0.690 (0.452, 1.084)	0.633 (0.430, 0.975)
VPC_{Dist}	0.159 (0.120, 0.209)	0.091 (0.065, 0.125)
VPC_{Reg}	0.106 (0.049, 0.227)	0.099 (0.048, 0.206)
$ICC_{\text{same-dist}}$	0.270 (0.207, 0.370)	0.192 (0.138, 0.290)

3.5 Fixed Effects: Individual-Level Predictors

Table 4 reports posterior odds ratios and 95 % credible intervals for all fixed effects. A strong positive wealth gradient is present for both sexes: among women, odds ratios rise from 1.39 (Q2) to 3.04 (Q5) relative to Q1; among men from 1.29 (Q2) to 2.64 (Q5). These associations are attenuated relative to what flat regression would produce, as the hierarchical structure separates individual wealth effects from district-level compositional differences. The education gradient is steeper for men than women at the highest level. Higher education yields $OR = 3.83$ (2.60–5.75) for women and OR

= 5.43 (3.96–7.48) for men, suggesting education is a particularly strong enabler of enrollment among men.

Age patterns differ markedly by sex. Women’s enrollment is stable at ages 25–34 (OR = 1.00) before declining at ages 35–44 (OR = 0.74) and ages 45+ (OR = 0.53), while men’s decline is steeper and begins earlier (OR = 0.46 at 25–34, 0.47 at 35–44). Currently married women have substantially higher enrollment odds (OR = 1.90; 95 % CrI: 1.56–2.32), consistent with shared insurance decisions within households; the effect for men is smaller and its credible interval includes one (OR = 1.16; 0.95–1.39). Among women, current pregnancy is strongly associated with enrollment (OR = 2.56; 1.88–3.57), consistent with the NHIS premium exemption for pregnant women, and parity is positively associated (OR per SD = 1.35; 1.21–1.51). A negative employment effect is observed for men (OR = 0.66; 0.53–0.82), likely reflecting coverage through SSNIT-linked or employer schemes, while the effect is nearly null for women (OR = 0.93; 0.80–1.09). Urban residence is associated with higher enrollment among men (OR = 1.29; 1.10–1.52), with a weaker and less certain association among women (OR = 1.18; 1.00–1.40). Among men, traditionalist or no religious affiliation is associated with lower enrollment odds (OR = 0.79; 0.68–0.92), while the Muslim effect is borderline (OR = 0.83; 0.68–1.01); neither religion coefficient has a credible interval excluding one for women. Household size and female-headed household status are near-null for both sexes.

3.6 District-Level Random Effects

Fig 3 shows the caterpillar plot of posterior district random effects for women. Of the 237 sampled districts, 26 show lower-than-expected enrollment net of individual-level covariates (red colored) and 10 have credible intervals entirely above zero, signifying higher-than-expected enrollment (blue colored). The remaining 201 are not reliably distinguishable from the national mean (gray colored). The spread is considerably wider for women than for men, consistent with the higher VPC_{Dist} .

For men, the spread of random effects is narrower (Fig 4), with 9 districts having credible intervals entirely below zero, 3 entirely above zero, and 225 indistinguishable from the national mean, consistent with the lower VPC_{Dist} reported for men.

3.7 Districts with Lowest Composition-Adjusted Enrollment

Tables 5 and 6 list the 30 districts with the lowest posterior mean enrollment probabilities for the reference profile, separately for women and men. Because these estimates hold individual-level characteristics constant at the reference profile, they reflect structural geographic disadvantage rather than population composition.

For women, all 30 districts remain above 84.0%, consistent with the compressed overall distribution. Greater Accra and Northern each contribute five districts to the list, followed by Central and Ashanti with four each, suggesting that low relative enrollment for women is concentrated in urban southern districts and selected northern districts rather than dispersed nationally. The lowest district is Pru East in the Bono East Region at 84.3% (95 % CrI: 71.7–92.6%).

Table 4 Fixed-effect posterior odds ratios (ORs) and 95 % credible intervals (CrIs) from the three-level model (M_3), stratified by sex. Reference: age 15–24; never married; Christian; no education; wealth Q1; not employed; rural; mean household size; male-headed household; not pregnant; mean parity.

Predictor	Women OR (95 % CrI)	Men OR (95 % CrI)
Intercept	4.52 (2.86–7.11)	2.60 (1.68–3.99)
Age 25–34	1.00 (0.82–1.22)	0.46 (0.37–0.57)
Age 35–44	0.74 (0.57–0.94)	0.47 (0.37–0.60)
Age 45–49/59	0.53 (0.39–0.72)	0.82 (0.64–1.06)
Currently married	1.90 (1.56–2.32)	1.16 (0.95–1.39)
Formerly married	1.21 (0.95–1.57)	0.80 (0.58–1.09)
Muslim	1.06 (0.87–1.31)	0.83 (0.68–1.01)
Traditionalist/none	0.95 (0.80–1.12)	0.79 (0.68–0.92)
Education: Primary	1.02 (0.84–1.23)	1.13 (0.91–1.39)
Education: Secondary	1.99 (1.65–2.40)	2.05 (1.70–2.47)
Education: Higher	3.83 (2.60–5.75)	5.43 (3.96–7.48)
Wealth Q2	1.39 (1.16–1.66)	1.29 (1.07–1.55)
Wealth Q3	1.67 (1.34–2.09)	1.85 (1.47–2.33)
Wealth Q4	2.36 (1.81–3.08)	1.98 (1.53–2.57)
Wealth Q5 (richest)	3.04 (2.23–4.16)	2.64 (1.94–3.58)
Employed	0.93 (0.80–1.09)	0.66 (0.53–0.82)
Urban	1.18 (1.00–1.40)	1.29 (1.10–1.52)
HH size (per SD)	0.99 (0.92–1.05)	1.05 (0.98–1.13)
Female-headed HH	1.01 (0.86–1.17)	1.17 (0.96–1.43)
Currently pregnant	2.56 (1.88–3.57)	N/A
Parity (per SD)	1.35 (1.21–1.51)	N/A

For men, enrollment falls substantially lower. The lowest district is Sunyani West in the Bono Region at 47.2 % (95 % CrI: 28.7–66.9 %), and the bottom 30 draws from 14 different regions spanning the entire country, from Upper West and Upper East in the north to Volta, Eastern, and Western in the south. Credible intervals for men are considerably wider than for women, with several districts spanning more than 40 percentage points, reflecting greater posterior uncertainty in men’s district-level estimates. Nine districts appear among the bottom 30 for both sexes: Ketu South (Volta), Mion (Northern), Gomoa East and Gomoa West (Central), Banda (Bono), Ayawaso North and Ablekuma West (Greater Accra), Adansi Akrofuom (Ashanti), and Awutu Senya West (Central).

3.8 Posterior Enrollment Maps

Figs 5 and 6 show posterior mean enrollment probabilities $\hat{\pi}_{jk}^{\text{ref}}$ for the reference profile across all 237 sampled districts. For women, enrollment ranges from 84.3 % to 99.2 % with a mean of 95.7 %, a compressed distribution spanning 14.9 percentage points. For men, the range is markedly wider, from 47.2 % to 91.4 % with a mean of 79.9 %,

Fig. 3 Posterior district random effects $\hat{u}_{j(k)}$ for women, sorted by posterior mean. Error bars are 95 % credible intervals.

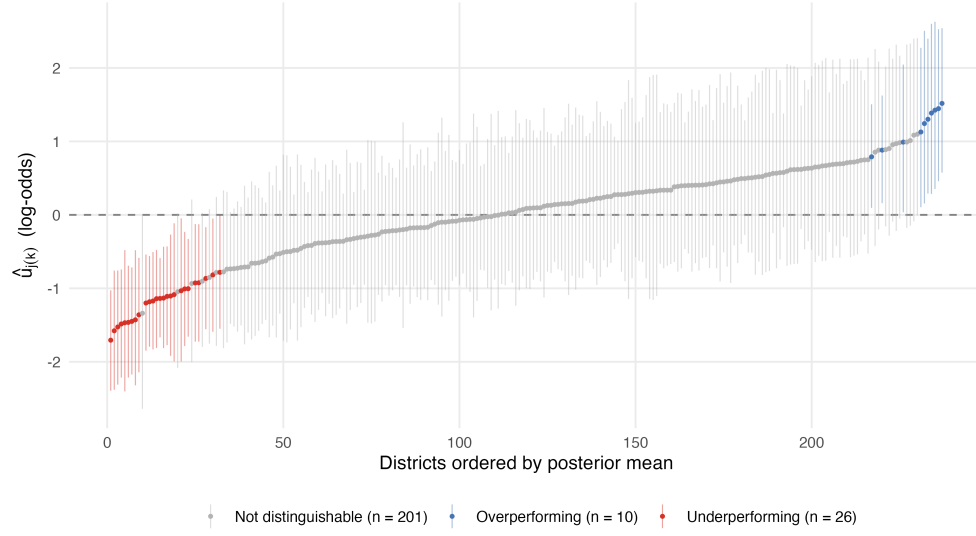
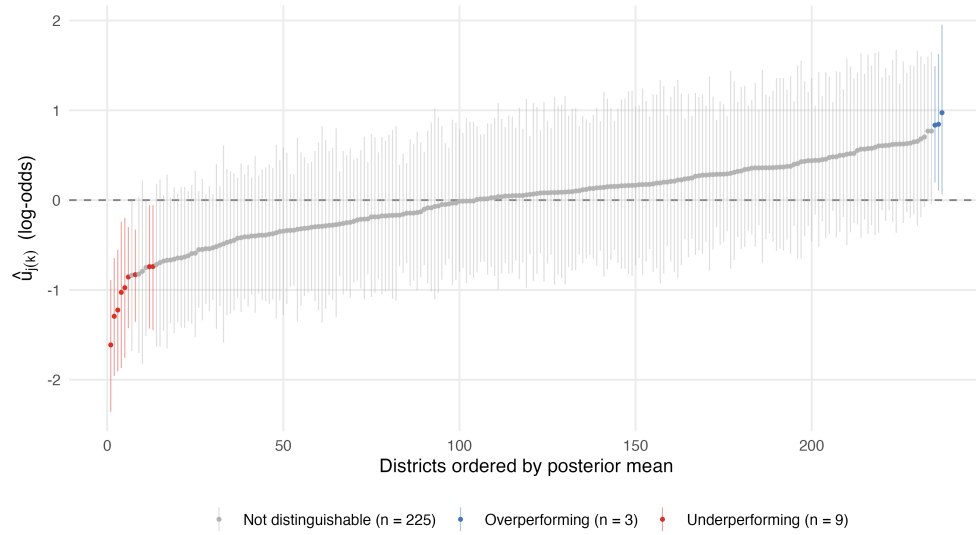


Fig. 4 Caterpillar plot of posterior district random effects $\hat{u}_{j(k)}$ for men, sorted by posterior mean. Error bars are 95 % credible intervals.



spanning 44.2 percentage points, roughly three times the women's range. The mean enrollment gap between the two groups is 15.9 percentage points. The 24 districts with no sampled clusters are shown in gray in both maps.

Table 5 Thirty districts with the lowest posterior mean NHIS enrollment probabilities for the reference woman. Reference: married, wealth Q3, secondary education, employed, urban, Christian, not pregnant.

District	Region	Enrollment (%)	95 % CrI
Pru East	Bono East	84.3	71.7–92.6
Ablekuma West	Greater Accra	85.7	72.5–94.1
East Gonja	Savannah	86.4	73.9–94.2
Adansi Akrofuom	Ashanti	86.7	71.9–95.5
Awutu Senya West	Central	86.9	75.1–94.4
Atwima-Kwanwoma	Ashanti	87.1	75.7–94.5
Atwima-Nwabiagya North	Ashanti	87.2	73.8–95.4
Banda	Bono	87.3	68.0–97.1
Ayawaso North	Greater Accra	87.3	76.3–94.4
Ledzokuku	Greater Accra	88.2	76.8–95.1
Nanumba North	Northern	89.9	81.4–95.4
Ablekuma North	Greater Accra	90.1	82.1–95.3
Ketu South	Volta	90.1	82.1–95.5
North East Gonja	Savannah	90.4	82.1–95.6
Cape Coast	Central	90.4	82.0–95.7
Mion	Northern	90.5	80.5–96.3
Adansi Asokwa	Ashanti	90.5	79.3–96.8
Gomoa East	Central	90.5	83.6–95.2
Upper West Akim	Eastern	90.6	78.4–97.3
Gomoa West	Central	90.7	82.5–95.8
West Akim	Eastern	90.8	79.4–97.1
Mamprugu Moagduri	North East	91.4	82.8–96.4
Pru West	Bono East	91.4	79.7–97.7
Prestea-Huni Valley	Western	91.5	84.0–96.2
Zabzugu	Northern	91.9	83.5–96.9
Tatale Sanguli	Northern	91.9	83.5–96.9
South Dayi	Volta	91.9	82.2–97.2
South Tongu	Volta	92.3	83.2–97.4
Gushegu	Northern	92.5	85.6–96.7
Krowor	Greater Accra	92.6	82.6–97.9

3.9 Sensitivity Analyses

Covariate missingness did not exceed 1% in any variable for either sex; multiple imputation was therefore not required. Table 7 reports VPC_{Dist} across all sensitivity specifications for women. Prior choice has no material effect on conclusions: posterior means under Half-Normal, Half-Cauchy, and Exponential priors are 0.161, 0.162, and 0.161 respectively, with credible intervals virtually identical to the main model. Replacing district random effects with EA-cluster random effects yields a modestly higher $VPC = 0.184$ (95 % CrI: 0.145–0.226), consistent with EA-level clustering capturing homogeneity both within and across districts. Switching to the card-verified outcome (**s1119**) produces a substantially wider and lower posterior ($VPC = 0.080$; 95 % CrI: 0.000–0.359), reflecting the reduced sample size under card-seen coding and confirming that self-reported enrollment is the more stable outcome for geographic variance

Table 6 Thirty districts with the lowest posterior mean NHIS enrollment probabilities for the reference man. Reference: married, wealth Q3, secondary education, employed, urban, Christian.

District	Region	Enrollment (%)	95 % CrI
Sunyani West	Bono	47.2	28.7–66.9
Ketu South	Volta	54.8	36.7–71.8
Wa East	Upper West	56.5	37.4–73.6
Mion	Northern	61.0	39.0–79.2
Binduri	Upper East	62.2	41.5–80.1
Nkwanta North	Oti	65.1	49.5–78.4
Akatsi South	Volta	65.1	43.3–82.9
Gomoa West	Central	65.4	43.4–83.0
Gomoa East	Central	65.6	50.8–78.7
Abuakwa North	Eastern	65.8	39.7–85.6
Wassa Amenfi East	Western	67.1	47.4–83.1
Tarkwa Nsuaem	Western	67.4	49.3–82.3
North Gonja	Savannah	67.4	48.7–82.2
Amansie South	Ashanti	67.6	44.9–85.1
Upper Manya	Eastern	68.0	44.8–86.0
Banda	Bono	68.5	44.1–86.6
Upper Denkyira East	Central	68.6	49.0–84.0
Ayawaso North	Greater Accra	68.9	47.7–85.2
Wassa Amenfi Central	Western	68.9	50.9–83.5
Asene-Manso-Akroso	Eastern	69.3	49.8–84.9
Asutifi South	Ahafo	69.4	51.2–83.9
Ablekuma West	Greater Accra	69.5	49.3–85.1
Sissala East	Upper West	69.9	51.7–84.6
Adansi Akrofuom	Ashanti	70.1	47.0–87.1
Sefwi-Akontombra	Western North	70.3	50.9–85.0
Awutu Senya West	Central	71.2	52.1–86.1
Sene West	Bono East	71.3	54.7–84.7
Aowin	Western North	71.5	54.2–85.1
Yendi	Northern	71.6	55.5–84.2
Wassa East	Western	71.6	49.6–87.9

estimation. The pooled sex model yields $VPC = 0.093$ (95 % CrI: 0.070–0.120), lower than the sex-stratified estimates as pooling averages across the two models. Incorporating scaled DHS survey weights in the likelihood yields $VPC = 0.183$ (95 % CrI: 0.139–0.234), consistent with the main model and confirming that the unweighted specification is not materially affected by the complex survey design.

4 Discussion

4.1 Principal findings

Geographic context is a first-order determinant of NHIS enrollment in Ghana, operating at a scale that individual and household characteristics alone cannot explain. Approximately 27.0 % of total enrollment variation among women and 19.2 % among

Fig. 5 Posterior mean NHIS enrollment probabilities for the reference woman (married, wealth Q3, secondary education, employed, urban, Christian, not pregnant) across 237 sampled Ghanaian districts. Enrollment probabilities increase from purple through orange to yellow. Unsampled districts are shown in gray.

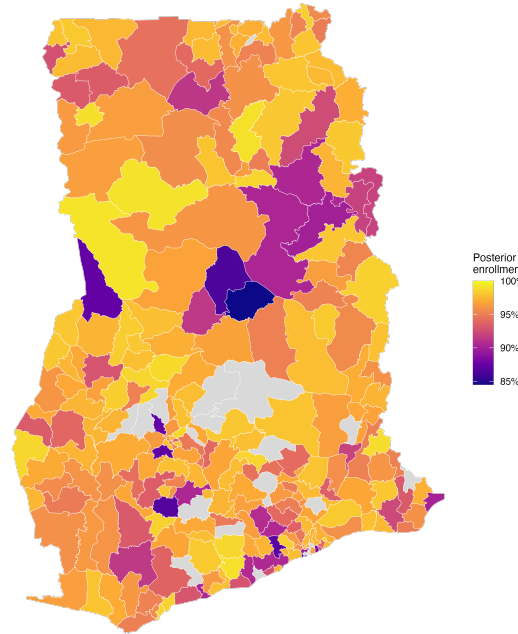
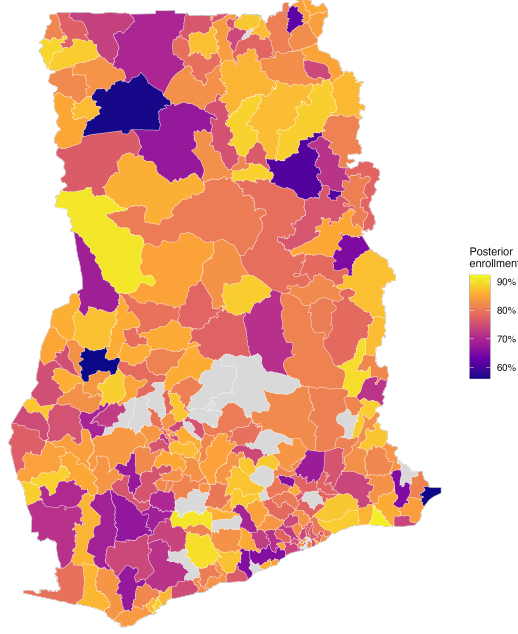


Table 7 Sensitivity analysis: posterior VPC_{Dist} (posterior mean) under alternative specifications for women.

Specification	Mean	95 % CrI
Main model (self-reported enrollment)	0.161	0.120–0.209
Cluster as Level-2 (no spatial join)	0.184	0.145–0.226
Card-seen (verified) outcome	0.080	0.000–0.359
Pooled sex model	0.093	0.070–0.120
Survey-weighted likelihood	0.183	0.139–0.234
Prior: Half-Normal(0,1)	0.161	0.119–0.210
Prior: Half-Cauchy(0,1)	0.162	0.119–0.210
Prior: Exponential(1)	0.161	0.119–0.209

men is attributable to the shared district and regional environment in which people live, with district-level differences accounting for 15.9 and 9.1 percentage points respectively. Composition-adjusted posterior district enrollment probabilities, computed by holding all individual-level characteristics constant at the reference profile, span 14.9

Fig. 6 Posterior mean NHIS enrollment probabilities for the reference man (married, wealth Q3, secondary education, employed, urban, Christian) across 237 sampled Ghanaian districts. Enrollment probabilities increase from purple through orange to yellow. Unsampled districts are shown in gray.



percentage points for women and 44.2 percentage points for men, with men’s enrollment more than twice as sensitive to geographic context as women’s. Education and wealth are the strongest individual-level enablers for both sexes, while current pregnancy, employment, age, and religion show sex-specific patterns consistent with the structural differences in how women and men interact with the NHIS. Nine districts appear among the lowest-performing 30 for both sexes simultaneously, identifying specific geographic units where structural barriers to enrollment operate independently of population composition. These findings also highlight that flat logistic regression, the predominant analytic approach in the NHIS enrollment literature [9–13], does not fully accommodate the nested structure of DHS data. It cannot capture within-group correlation, tends to produce underestimated standard errors, and offers no mechanism for reliable district-level enrollment estimation. As a result, more than a quarter of enrollment variation among women has remained analytically invisible in prior work.

4.2 Geographic variation and district-level heterogeneity

Posterior district enrollment probabilities for a standardized reference individual (married, middle wealth, secondary education, employed, urban, Christian) ranged from 84.3 % to 99.2 % among women and from 47.2 % to 91.4 % among men across the 237 sampled districts. The markedly wider range for men (44.2 percentage points

versus 14.9 for women) indicates that men’s enrollment decisions are more sensitive to geographic context than women’s. Prior regional analyses [14] using the 2014 DHS found that the Upper West, Volta, and Brong Ahafo regions had the highest proportions of enrolled individuals, with Brong Ahafo used as the regression reference in that study because it was where the NHIS was first piloted. The present composition-adjusted district-level estimates extend this picture and reveal two important details. First, the 2019 administrative reorganization divided the former Brong Ahafo Region into three successor regions (Bono, Bono East, and Ahafo), and districts within Bono and Bono East include some of the lowest-performing units in the current analysis, underscoring the added diagnostic value of district-level posterior estimates over regional aggregates. Second, although Upper West was among the highest-enrolling regions in the earlier survey, several of its districts contribute to the lower tail of the composition-adjusted men’s distribution in the current analysis. This apparent reversal does not necessarily reflect a decline in raw enrollment. Rather, the composition-adjusted estimates remove the influence of individual wealth and education, so a region can exhibit high aggregate enrollment while specific districts within it remain structurally disadvantaged once those individual-level differences are held constant. The bottom 30 districts for men span 14 different regions, from Upper West and Upper East in the north to Volta, Eastern, and Western in the south, confirming that low enrollment among men is a nationally dispersed rather than regionally concentrated phenomenon. This finding contrasts with the spatial autocorrelation analysis of [15], which identified 42 high-uninsurance hotspot clusters predominantly in northern savanna districts using 2021 Census aggregate data. The divergence is methodological in nature. Composition-adjusted estimates hold individual-level characteristics constant at the reference profile, so districts that appear to underperform in aggregate analyses may do so partly because of who lives there rather than where they live. Once those individual-level differences are accounted for, low-enrollment districts emerge across all regions rather than clustering in the north, suggesting that geographic targeting based on raw aggregate rates alone may misidentify the districts most in need of structural intervention. Nine districts appear among the bottom 30 for both sexes, including Mion (Northern Region), Ketu South (Volta Region), Gomoa East and Gomoa West (Central Region), Banda (Bono Region), Ayawaso North and Ablekuma West (Greater Accra Region), and Adansi Akrofuom (Ashanti Region), suggesting that geographic disadvantage in these districts operates independently of sex and warrants prioritized attention from the NHIA.

The wider geographic dispersion of low-enrollment districts for men is consistent with several structural mechanisms. Irregular income streams and poverty constrain premium payment continuity, a barrier documented across enrollment studies [11, 12], and these constraints may fall more heavily on men as primary household earners in districts where formal employment is limited. Women’s enrollment may additionally be stabilized by the maternal health pathway, which provides repeated contact with the scheme independent of income. Variation in NHIA district office capacity and the density of accredited facilities across Ghana’s administrative geography may also reduce the perceived benefit of enrollment in under-served districts, a concern reflected in national health policy [8]. The fixed-effects results further show that men identifying

as traditionalist or reporting no religion have credibly lower enrollment odds than Christian men (OR = 0.79; 0.68–0.92), adding a cultural dimension to the structural geographic gradient.

The partial-pooling property of the Bayesian hierarchical framework is particularly consequential for interpreting low-enrollment districts. Under flat regression, small-sample districts receive unstable, high-variance enrollment estimates, while under complete pooling their geographic distinctiveness is erased [20, 21]. Partial pooling produces estimates that are statistically optimal in the James-Stein sense [21], where data-rich districts are pulled minimally toward the national mean and data-sparse districts are pulled strongly, with the degree of shrinkage calibrated by the within-district posterior variance [22]. Districts with caterpillar-plot credible intervals entirely below zero represent genuine geographic barriers to enrollment that NHIA raw-count performance metrics cannot detect, as those metrics confound structural disadvantage with administrative under-performance.

4.3 Individual-level determinants

The individual-level fixed-effects results are broadly consistent with the socioeconomic gradient documented across two decades of NHIS enrollment research. The finding that wealthier and more educated individuals are more likely to enroll confirms patterns reported by [9], who found clear inequity in enrollment across consumption quintiles, and by [14], who found education and wealth to be the most consistent predictors across three national surveys. The hierarchically-adjusted estimates reported here are, however, likely more precise than those from prior flat-regression studies. By separating individual-level enabling resources from district-level compositional effects, the hierarchical model removes the upward bias that arises when educated and wealthy households are geographically clustered into higher-enrolling districts, suggesting that prior analyses may have overstated the independent contribution of education and wealth to enrollment. Within this gradient, the magnitude of the education effect is striking. Higher education confers over threefold enrollment odds for women (OR = 3.83; 95 % CrI: 2.60–5.75) and over fivefold for men (OR = 5.43; 3.96–7.48), the largest single predictor in the model for both sexes. The steeper male education gradient extends the argument made by [13] that enrollment barriers operate at the intersection of education and residential context. Men who lack education may be less likely to navigate NHIS registration processes or access employer-linked entry points, whereas women with low education may still encounter the scheme through maternal health services, consistent with the high enrollment documented among pregnant and recently delivered women in administrative data [4]. The wealth gradient is monotonic and substantial for both sexes, with ORs rising from 1.39 (Q2) to 3.04 (Q5) for women and from 1.29 to 2.64 for men relative to the poorest quintile, a pattern that is consistent with [12]’s finding that wealth status was among the strongest determinants of enrollment among women in five Ghanaian districts, and with [9]’s call for targeted premium subsidies toward the poorest quintiles. What is more instructive is that this gradient survives composition-adjustment and remains credibly non-zero even at the upper wealth quintiles, suggesting that the problem is not simply one of absolute poverty but of perceived value relative to competing expenditures at every

wealth level. Subsidy targeting at lower quintiles therefore remains relevant even after two decades of implementation, not as a novel recommendation but as an enduring structural gap that the scheme has not yet closed.

The strong association between current pregnancy and enrollment ($OR = 2.56$; 95 % CrI: 1.88–3.57) reflects a rational response to the NHIS premium exemption tied to the Free Maternal Health Care Policy [4]. That parity is also independently associated with enrollment (OR per SD = 1.35; 1.21–1.51), beyond the effect of current pregnancy, suggests a durable pathway: women who have passed through the maternal health system accumulate familiarity with the scheme and remain connected to health services through child-related visits. This durable maternal pathway may partly explain why women’s enrollment is generally higher and geographically less dispersed than men’s, a finding that has received little attention in the prior literature.

The contrasting employment effects for women and men illuminate an important structural asymmetry in how the NHIS intersects with formal sector work. For men, formal employment is associated with lower enrollment odds ($OR = 0.66$; 0.53–0.82), most plausibly because employed men with SSNIT contributions may access employer-sponsored or SSNIT-affiliated health coverage, reducing the marginal incentive to register separately with the NHIS [2]. For women, the effect is essentially null ($OR = 0.93$; 0.80–1.09), suggesting that women’s employment status and NHIS membership are driven by different mechanisms, with the maternal health pathway providing an enrollment route that is independent of occupational structure. This asymmetry has direct implications for policy. Outreach strategies designed around formal employment linkages will reach men who lack SSNIT coverage but will miss the enrollment dynamics that govern women’s participation. The non-monotonic age gradient for men is notable: enrollment odds are substantially below the 15–24 reference level at ages 25–34 ($OR = 0.46$) and 35–44 ($OR = 0.47$) but partially recover at ages 45 and above ($OR = 0.82$), consistent with a life cycle dynamic in which young adult men de-prioritize NHIS premiums while facing competing financial obligations and re-enroll as chronic morbidity risk increases. Women’s enrollment declines more monotonically with age (OR s of 1.00, 0.74, and 0.53 for 25–34, 35–44, and 45+ respectively), possibly reflecting the sustaining role of child-bearing years and maternal health pathways among younger women.

4.4 Comparison with existing literature

Ghanem [15] confirmed significant spatial autocorrelation in district-level uninsurance (Global Moran’s $I = 0.422$) using 2021 Census aggregate data, establishing that geographic dependence in NHIS coverage is a real structural feature of the Ghanaian health insurance landscape. The present analysis extends this by modeling that dependence at the individual level within a Bayesian hierarchical framework, with composition-adjusted posterior estimates and full uncertainty quantification. The present analysis also extends and substantiates the contribution of [14], the sole prior multilevel analysis of NHIS enrollment, which applied frequentist multilevel logistic regression to pooled 2014 DHS, MICS 2011, and GLSS 2012–13 data across three levels: enumeration area, household, and individual. That analysis confirmed cluster-level clustering and the socioeconomic gradient but uses data now over a decade old,

predates the 2019 administrative reorganization creating six new regions and 38 new districts, did not estimate posterior credible intervals on variance components, produced no district-level enrollment maps, and could not formally decompose geographic variation across the three administrative levels that now govern NHIA operations. The district-specific VPC of 15.9% for women demonstrates that the administrative district tier captures a non-trivial share of enrollment variation not reducible to EA-level community context; this is the tier at which NHIA offices operate, district health directorates plan outreach, and Ghana’s Universal Health Coverage strategy allocates resources [8]. The regional VPC of 10.6% for women further confirms macro-level institutional variation that persists after accounting for district-level heterogeneity.

The combined ICC of 0.270 for women situates NHIS enrollment within the broader pattern of geographic clustering documented for health outcomes in Ghana and sub-Saharan Africa. Bayesian hierarchical and spatial frameworks applied to analogous binary outcomes in the region have consistently identified significant geographic dependence, including for malaria risk [18], child mortality [17], neonatal mortality [16], and HIV prevalence [19]. That health insurance enrollment exhibits geographic clustering at a scale consistent with these health status outcomes suggests that the structural and contextual forces shaping disease risk and survival also shape access to health financing, reinforcing the case for geographic targeting in supply-side investment.

4.5 Policy implications

The finding that posterior district enrollment probabilities fall as low as 47.2% for men, even for a reference individual with secondary education, middle wealth, and urban residence, has direct implications for Ghana’s Universal Health Coverage road map [8] and the NHIA’s district-level strategy. Premium subsidy allocation should be guided by composition-adjusted posterior enrollment probabilities rather than raw district enrollment counts, which conflate structural disadvantage with administrative under-performance and may divert subsidies toward districts whose low enrollment reflects population poverty rather than operational failure. The shrinkage-adjusted posterior maps produced here provide a ready operational tool for this reallocation. Outreach and registration infrastructure investment should prioritize districts whose caterpillar-plot credible intervals lie entirely below the national mean, as these districts under-perform relative to their own population composition and represent potentially correctable inefficiencies. Critically, these districts are not exclusively northern; nine of them appear in the bottom 30 for both sexes simultaneously, suggesting structural barriers that cut across sociodemographic lines. The large male education gradient further suggests that NHIS communications targeting low-education male populations could yield substantial enrollment gains nationally, rather than in any single geographic zone. Finally, the strong pregnancy-exemption effect confirms the efficacy of event-triggered demand-side subsidies as enrollment catalysts, indicating that extending analogous incentives to other high-need groups could shift enrollment curves without requiring structural reform.

4.6 Limitations

Several limitations warrant acknowledgment. First, the cross-sectional survey design precludes causal inference; the observed geographic associations reflect a combination of structural, economic, historical, and cultural factors that cannot be disentangled within a single survey cross-section. Second, 24 of Ghana’s 261 districts have no sampled DHS clusters, preventing district-level posterior estimates for approximately 9.2 % of the country’s districts, and these unsampled districts are not randomly distributed. Third, although the self-reported enrollment indicator is the appropriate primary outcome for stable geographic variance estimation, confirmed by the extremely wide and near-zero VPC under the card-seen sensitivity analysis ($VPC = 0.080$; 95 % CrI: 0.000–0.359), self-report may introduce differential recall bias across education and wealth strata. Fourth, the exclusion of the enumeration area level as a modeling tier means that the district random effect absorbs some within-district EA clustering, potentially leading to modest upward bias in district-level variance estimates. Finally, the absence of district-level supply-side covariates such as health facility density, NHIA office staffing, and road infrastructure prevents formal cross-level interaction analysis that would quantify how much of the district random effect is mediated by observable structural factors.

5 Conclusion

Across 237 sampled districts of Ghana, where a person lives determines their likelihood of NHIS enrollment independently of who they are, with approximately 27.0 % of total enrollment variation among women and 19.2 % among men reflecting shared district and regional context after full adjustment for individual and household characteristics, and district-level differences alone accounting for 15.9 percentage points of that share for women and 9.1 for men. This geographic signal has been structurally invisible to every prior analysis of NHIS enrollment precisely because flat logistic regression cannot detect within-group correlation or reliably estimate district-level effects, yet its consequences are substantial, as composition-adjusted enrollment probabilities span 14.9 percentage points for women and 44.2 percentage points for men, and the nine districts that appear among the lowest-performing thirty for both sexes simultaneously, drawn from Greater Accra, Ashanti, Central, Bono, Volta, and Northern regions, represent genuine structural failures not reducible to the poverty or low education of their populations. Education and wealth remain the strongest individual-level predictors, with higher education conferring over threefold enrollment odds for women and over five-fold for men, yet these associations are attenuated in the hierarchical model relative to prior flat-regression estimates, because separating individual-level resources from district-level compositional effects reveals that some of what past analyses attributed to education and wealth was partly a geographic effect. For the NHIA, the practical implication is direct in that premium subsidies and outreach directed at districts that underperform relative to their own population composition, identifiable in the posterior caterpillar plots, will generate greater returns than those directed at districts whose low raw enrollment simply reflects population poverty, and the composition-adjusted maps produced here give the Authority a ready operational basis for making

that distinction. Building on these findings will require incorporating district-level supply-side data to move the analysis from describing where enrollment is low to explaining why, and linking these estimates to repeated DHS rounds or NHIA administrative records to reveal whether the geographic disparities documented here are narrowing or widening over time.

6 Acknowledgments

The author thanks the DHS Program for providing access to the 2022 Ghana Demographic and Health Survey data, and colleagues and lecturers in the Department of Statistics and Actuarial Science at the Kwame Nkrumah University of Science and Technology for their support and guidance throughout this work.

7 Competing interests

The author has declared that no competing interests exist.

8 Ethics statement

This study is based on secondary data from the 2022 Ghana Demographic and Health Survey, which are available to registered and approved users through the DHS Program. The DHS Program obtains ethical approval from the ICF Institutional Review Board and the Ghana Health Service Ethics Review Committee before each survey round. No new human subjects data were collected for this analysis, and individual records are fully anonymized prior to release. No additional ethical approval was required.

9 Data availability

The 2022 Ghana Demographic and Health Survey Individual Recode, Male Recode, and GPS cluster files are available upon registration, application, and approval through the DHS Program at <https://dhsprogram.com>. The GADM Version 4.1 Level-2 district polygon shapefile for Ghana is freely available at <https://gadm.org>. DHS GPS cluster coordinates are subject to the DHS Program's confidentiality restrictions and cannot be redistributed. The complete R analysis code used in this study, including all model specifications, spatial processing steps, and figure generation scripts, is available from the corresponding author upon reasonable request.

10 Funding

The author received no funding for this work.

References

- [1] United Nations: Transforming our world: the 2030 agenda for sustainable development. Resolution adopted by the General Assembly **A/RES/70/1** (2015)

- [2] National Health Insurance Authority: National health insurance act (act 650). Technical report, Parliament of Ghana, Accra, Ghana (2003)
- [3] National Health Insurance Authority: National health insurance act (act 852). Technical report, Parliament of Ghana, Accra, Ghana (2012). p. 1–55
- [4] Nsiah-Boateng, E., Aikins, M.: Trends and characteristics of enrolment in the National Health Insurance Scheme in Ghana: a quantitative analysis of longitudinal data. *Global Health Research and Policy* **3**, 32 (2018) <https://doi.org/10.1186/s41256-018-0087-6>
- [5] Ghana Statistical Service, ICF: Ghana Demographic and Health Survey 2022. GSS and ICF, Accra, Ghana and Rockville, Maryland, USA (2024)
- [6] Ghana Statistical Service: Ghana 2021 Population and Housing Census General Report, Volume 3C: Background Characteristics. Ghana Statistical Service, Accra, Ghana (2021). <https://census2021.statsghana.gov.gh>
- [7] National Health Insurance Scheme: Districts. <https://www.nhis.gov.gh/districts>. [Internet; accessed 25-May-2026] (2026)
- [8] Ministry of Health, Ghana: National health policy: Ensuring healthy lives for all ghanaians. Technical report, Ministry of Health, Accra, Ghana (2020)
- [9] Jehu-Appiah, C., Aryeetey, G., Spaan, E., Hoop, T., Agyepong, I., Baltussen, R.: Equity aspects of the National Health Insurance Scheme in Ghana: who is enrolling, who is not and why? *Social Science & Medicine* **72**(2), 157–165 (2011) <https://doi.org/10.1016/j.socscimed.2010.10.025>
- [10] Mensah, J., Oppong, J.R., Schmidt, C.M.: Ghana’s national health insurance scheme in the context of the health mdgs: An empirical evaluation using propensity score matching. *Health Economics* **19**, 95–106 (2010) <https://doi.org/10.1002/hec.1633>
- [11] Kotoh, A.M., Aryeetey, G.C., Geest, S.: Factors that influence enrolment and retention in ghana’ national health insurance scheme. *International Journal of Health Policy and Management* **7**(5), 443–454 (2018) <https://doi.org/10.15171/ijhpm.2017.117>
- [12] Kusi, A., Fenny, A.P., Arhinful, D.K., Asante, F.A., Parmar, D.: Determinants of enrolment in the NHIS for women in Ghana — a cross sectional study. *International Journal of Social Economics* **45**(9), 1318–1334 (2018) <https://doi.org/10.1108/IJSE-10-2016-0291>
- [13] Antabe, R., Anfaara, F.W., Sano, Y., Amoak, D.: Ghana’s national health insurance enrollment: Does the intersection of educational and residential status matter? *PLOS ONE* **20**(2), 0318202 (2025) <https://doi.org/10.1371/journal>

- [14] Salari, P., Akweongo, P., Aikins, M., Tediosi, F.: Determinants of health insurance enrolment in Ghana: evidence from three national household surveys. *Health Policy and Planning* **34**(8), 582–594 (2019) <https://doi.org/10.1093/heapol/czz079>
- [15] Ghanem, V.G.: Spatial and machine learning analysis of district-level health insurance inequities in Ghana. *Cureus* **18**(1), 101984 (2026) <https://doi.org/10.7759/cureus.101984>
- [16] Takramah, W.K., Dwomoh, D., Aheto, J.M.K.: Spatio-temporal variations in neonatal mortality rates in Ghana: an application of hierarchical Bayesian methods. *PLOS Global Public Health* **2**(9), 0000649 (2022) <https://doi.org/10.1371/journal.pgph.0000649>
- [17] Li, Z., Hsiao, Y., Godwin, J., Martin, B.D., Wakefield, J., Clark, S.J., GBD 2015 Eastern Mediterranean Region Child and Adolescent Health Collaborators: 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**(1), 0210645 (2019) <https://doi.org/10.1371/journal.pone.0210645>
- [18] Yankson, R., Anto, E.A., Chipeta, M.G.: Geostatistical analysis and mapping of malaria risk in children under 5 using point-referenced prevalence data in Ghana. *Malaria Journal* **18**, 67 (2019) <https://doi.org/10.1186/s12936-019-2709-y>
- [19] Allotey, P., Harel, O.: Bayesian spatial modeling of incomplete data with application to HIV prevalence in Ghana. *Sankhya B: The Indian Journal of Statistics* **85**(2), 307–329 (2023) <https://doi.org/10.1007/s13571-023-00308-6>
- [20] Gelman, A., Hill, J.: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge (2007)
- [21] Efron, B., Morris, C.N.: Stein’s paradox in statistics. *Scientific American* **236**(5), 119–127 (1977) <https://doi.org/10.1038/scientificamerican0577-119>
- [22] McElreath, R.: *Statistical Rethinking: A Bayesian Course with Examples in R And Stan*, 2nd edn. Chapman and Hall/CRC, Boca Raton, FL (2020)
- [23] E, P.: Simple features for R: standardized support for spatial vector data. *R J* **10**(1), 439–446 (2018) <https://doi.org/10.32614/RJ-2018-009>
- [24] Andersen, R.: *A Behavioral Model of Families’ Use of Health Services*. Research Series No. 25. Center for Health Administration Studies, University of Chicago, Chicago, IL (1968)
- [25] Gelman, A., Jakulin, A., Pittau, M.G., Su, Y.-S.: A weakly informative default

- prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2**(4), 1360–1383 (2008) <https://doi.org/10.1214/08-AOAS191>
- [26] Gelman, A.: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis* **1**(3), 515–534 (2006) <https://doi.org/10.1214/06-BA117A>
 - [27] Goldstein, H., Browne, W., Rasbash, J.: Partitioning variation in multilevel models. *Understanding Statistics* **1**(4), 223–231 (2002) https://doi.org/10.1207/S15328031US0104_02. See also: Goldstein H, Browne W, Rasbash J (2002). *Stat. Methods Med. Res.* 11(1):49–58
 - [28] Snijders, T.A.B., Bosker, R.J.: *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling*, 2nd edn. SAGE Publications, London (2012)
 - [29] Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L., Larsen, K.: A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology & Community Health* **60**(4), 290–297 (2006) <https://doi.org/10.1136/jech.2004.029454>
 - [30] Bürkner, P.-C.: *brms*: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**(1), 1–28 (2017) <https://doi.org/10.18637/jss.v080.i01>
 - [31] Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan: A probabilistic programming language. *Journal of Statistical Software* **76**(1), 1–32 (2017) <https://doi.org/10.18637/jss.v076.i01>
 - [32] Buuren, S., Groothuis-Oudshoorn, K.: *mice*: Multivariate imputation by chained equations in R. *Journal of Statistical Software* **45**(3), 1–67 (2011) <https://doi.org/10.18637/jss.v045.i03>
 - [33] Rubin, D.B.: *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, New York (1987)