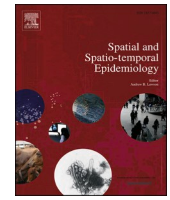
Spatial and Spatio-temporal Epidemiology 40 (2022) 100472

Contents lists available at ScienceDirect

Spatial and Spatio-temporal Epidemiology

journal homepage: www.elsevier.com/locate/sste



Effect of individual, household and regional socioeconomic factors and $PM_{2.5}$ on anaemia: A cross-sectional study of sub-Saharan African countries

Prince M. Amegbor [a,b,*], Sergio S. Borges [c], Alex Pysklywec [d], Clive E. Sabel [a,b]

[a] Big Data Centre for Environment and Health (BERTHA), Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[b] Department of Environmental Science, Aarhus University,Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[c] Department of Geography, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil
[d] LevelUp Planning Collaborative, Dawson City, YT, Canada

A R T I C L E I N F O

Keywords:
Anaemia
Sub-Saharan Africa
Demographic and health survey
Bayesian spatial analysis
Multilevel modelling
R-INLA

A B S T R A C T

There is limited knowledge on the effect of contextual and environmental factors on the risk of anaemia, as well as the spatial distribution of anaemia in the Sub-Saharan Africa region. In this study, we used multi-country data from the Demographic & Health survey (DHS) with 270,011 observations and $PM_{2.5}$ data from NASA, applied to the spatial risk pattern of anaemia in the SSA region. The prevalence of anaemia amongst women (41%) was almost twice that of men (22%). A Bayesian hierarchical model showed that individual household, neighbourhood and regional socioeconomic factors were significantly associated with the likelihood of being anaemic. 1 $\mu g/m^3$ increase in cumulative lifetime $PM_{2.5}$ exposure accounted for 1% ($\beta = 0.011$, CI = 0.008 – 0.015) increase in the likelihood of being anaemic. The results suggest the need for a multidimensional approach to tackle anaemia in the Sub-Saharan African region and identify high-risk areas for target intervention policies or programs.

1. Introduction

Anaemia is a condition where decreased numbers of red blood cells or low haemoglobin concentrations reduces the capacity for oxygen to be delivered throughout the body. The disease affects approximately 2.36 billion people around the globe (GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016). This non-communicable disease is increasingly burdensome in lower- and middle-income countries, including those in Sub-Saharan Africa (Chaparro and Suchdev, 2019; Stevens et al., 2013). In the Sub-Saharan Africa region, anaemia is most common in women and children under five years, affecting up to 57.1% of reproductive-aged women and about 60% of children in the region (Moschovis et al., 2018; Muriuki et al., 2020; WHO, 2015). However, there are significant variations amongst countries in the region. For instance, children aged 6–59 months in Rwanda had anaemia prevalence of 23.7% compared to 87.9% in Burkina Faso (Moschovis et al., 2018). Similarly, the prevalence of anaemia amongst women of reproductive age in Gabon is almost thrice (59.1%) that of Namibia (23.2%) and Ethiopia (23.4%) (FAO, IFAD, UNICEF, WFP AND WHO, 2017). Anaemia has negative effects on population health and can present significant individual- and societal-level economic costs; including increased rates of other diseases, reduced productivity and increased medical cost (Macharia et al., 2018; Teshale et al., 2020). Amongst adults, anaemia can result in mental and psychological disorders due to malfunction in neuromotor tasks (Chaparro and Suchdev, 2019; Hurtado et al., 1999; White, 2018).

The causes of anaemia are multifactorial, multidimensional and often interrelated (Roberts et al., 2020). Nutritional deficiencies are the most common cause of anaemia (FAO, IFAD, UNICEF, WFP AND WHO, WFP AND WHO, 2017; Teshale et al., 2020). Iron deficiency accounts for approximately 50% of cases worldwide (Elmardi et al., 2020) and 17%–51% of cases in some sub–Saharan Africa countries (Mwangi et al., 2017). Eating diets deficient in micronutrients, such as iron, folate, riboflavin, and vitamins A and B_{12}, is associated with the risk of anaemia (FAO, IFAD, UNICEF, WFP AND WHO, 2017). Pre-existing diseases, including hereditary blood disorders, malaria, tuberculosis, HIV, kidney disease, and renal disease, also contribute to the risk of anaemia (FAO, IFAD, UNICEF, WFP AND WHO, 2017). In addition to malnutrition and disease, studies show that individual, household, and community socioeconomic factors have significant effect on the risk of anaemia

; DHS, Demographic and Health Survey; INLA, Integrated Nested Laplace Approximations.

* Corresponding author.
E-mail address: pma@envs.au.dk (P.M. Amegbor).

https://doi.org/10.1016/j.sste.2021.100472
Received 21 July 2021; Received in revised form 28 October 2021; Accepted 29 November 2021
Available online 8 December 2021

amongst adults and children in the Sub-Saharan region (Adamu et al., 2017; Elmardi et al., 2020; Moschovis et al., 2018; Nankinga and Aguta, 2019; Roberts et al., 2020; Teshale et al., 2020). Evidence from previous research show that the risk of anaemia is high amongst urban dwellers, less educated, low-income households and households or communities with poor access to clean water and proper sanitation facilities (Adamu et al., 2017; Moschovis et al., 2018; Nankinga and Aguta, 2019). Regarding demographic factors, sex, age, marital status and race or ethnicity are associated with the risk of being anaemic (Adamu et al., 2017; Roberts et al., 2020; Teshale et al., 2020).

In addition, the findings of previous studies show that environmental factors, such as air quality, also has an effect on anaemia infection and prevalence (Accinelli and Leon-Abarca, 2017; Honda et al., 2017; Stanković et al., 2006). While the exact causal mechanism is unknown, researchers postulate that toxins from air pollutants can cause significant damage to red blood cells leading to anaemia via pathways, such as reduction in haemoglobin concentrations, erythrocytes and haematocrit (Stanković et al., 2006). Long term and short term increase in systemic inflammation and bone marrow stimulation due to exposure to $PM_{2.5}$ (fine particles with diameter of less than 2.5 μm) and other air pollutants have also been reported in existing studies (Honda et al., 2017). Studies report an association between poor air quality and anaemia in children (Accinelli and Leon-Abarca, 2017) and older adults (Honda et al., 2017). Socioeconomic development fuelled by industrialisation and transportation contribute to an increase in air pollutants and reduction in global air quality (Dechezleprêtre et al., 2020). In the Sub-Saharan African region, a growing population, increasing urbanisation and resource-intensive activities, coupled with low vehicle emission standards, cooking with solid fuels, and burning the household waste have all contributed to a downward trend in air quality in recent years (Abera et al., 2020).

There is currently little knowledge on the effect of contextual and environmental factors on the risk of anaemia, as well as, the spatial distribution of anaemia in the Sub-Saharan African region. Our study is the first to use pool data from the Demographic and Health Survey (DHS) and $PM_{2.5}$ data from the NASA Socioeconomic Data and Applications centre (SEDAC) to examine the effect of socioeconomic factors and air quality on anaemia in Sub-Saharan Africa. Specifically, this study sought to: 1) examine the effect of individual and contextual factors on risk of anaemia; 2) Map the distribution and risk of anaemia amongst regions in Sub-Saharan Africa – that is, identify high and low risk areas; and 3) discuss the public health and policy implications of the risk factors and risk distribution.

2. Data and methods

2.1. Data

The data for our study came from the DHS and SEDAC. The DHS program is a nationally representative survey of 95 countries, including 44 Sub-Saharan African countries. ICF International implements DHS in collaboration with Statistics Bureaus of representative countries with the aim of collecting and disseminating accurate nationally representative data on fertility, family planning, maternal and child health, gender, HIV/AIDS, malaria and nutrition. The program is mainly funded by the United States Agency for International Development (USAID) with financial and technical support from national governments other organisations, such as the United Nations Children's Fund (UNICEF), Bill and Melinda Gates Foundation (BMGF), the United Nations Population Fund (UNFPA), and the World Health Organization (WHO). In each country, women aged 15 to 49 years, who are members of selected households and others who spent the night before the survey in the selected households are eligible to be interviewed. In addition, men aged 15 to 54 years from half of the sampled households who are usual members and those who spent the night before the survey are interviewed. DHS surveys use a two-stage cluster sample design. The first stage entails the selection of random census enumeration areas (EAs) from a complete list of all EAs from the national population and housing census. EAs are the smallest geographic units defined for the census purpose. In the DHS surveys, the EAs are further clustered into geographic regions (administrative regions or zone) based the country's administrative regions and rural-urban residence. The second stage entails a random selection of households in each selected EA in each of the clusters. DHS data from twenty-six Sub-Saharan African countries from surveys conducted between 2006 and 2019 were used in this study. The data represent the most recent DHS survey for Sub-Saharan African countries with geographic information (sampled points or locations) and haemoglobin data.

Air pollution data, $PM_{2.5}$ was derived from multiple satellite aerosol optical depth (AOD) observations compiled by the Atmospheric Composition Analysis Group at Dalhousie University (Heft-Neal et al., 2018; Van Donkelaar et al., 2016, 2018). The data is accessible through the SEDAC platform as geotiff files. The dataset consists of continuous mean annual $PM_{2.5}$ concentrations from 1998 to 2016 at a resolution of $0.01^\circ \times 0.01^\circ$ - equivalent to 1.11 km^2 resolution near the equator (Van Donkelaar et al., 2016). The purpose of this data is to provide an annual global surface concentration of $PM_{2.5}$ for environmental research (Van Donkelaar et al., 2016, 2018). Information about the data and the modelling of $PM_{2.5}$ estimates are described in detail elsewhere (Van Donkelaar et al., 2016).

2.2. Linking DHS dataset and $PM_{2.5}$ data

The DHS Program provides geographic information on the location of selected EAs or the sample clusters where respondents were selected. The GPS point data has a matching key with the DHS survey datasets; coded as DHS-CLUST in the point data attribute table and v001 in the DHS datasets. The GPS location points of the EAs were displaced by 2 km for urban clusters (or EAs) and 5 km for rural clusters (or EAs) to ensure the complete anonymity of respondents. However, these displaced points were restricted to stay within the same districts and sub-districts thus retaining their administrative locational information for spatial analysis. Using ArcGIS' extract multi values to point tool, we extracted the mean annual $PM_{2.5}$ values for the sampled locations (EAs) of the 26 countries included in this study. A new dataset with the point information and $PM_{2.5}$ values overlain on a polygon data of administrative level 1 (or regions) of SSA. The results data from this spatial join contains information from both the point data, $PM_{2.5}$ values and the corresponding region information (name and ID) where the points are located. The attribute data from the point spatial dataset was matched with the DHS datasets using the matching key – DHS-CLUST in the point data attribute table and v001 in the DHS datasets. Given the matching key in both datasets were integer variables starting from 1, we modified them by adding the country abbreviation to the integer value. That is a new alphanumeric key was created that matched point data from each DHS country (in this study) to their respective women and men sample in the DHS dataset.

2.3. Measures

2.3.1. Outcome measure

The outcome variable of this study is anaemia. Anaemia is defined as a haemoglobin concentration below 12.0 g/dl and 11.0 g/dl for non-pregnant and pregnant women, respectively, age 15 to 49 years according to the WHO guideline (WHO, 2011, 2015). For men, anaemia was defined as a haemoglobin concentration below 13.0 g/dl. In both sexes, haemoglobin levels are adjusted for cigarette smoking and for the altitude in enumeration areas that are above 1000 m given the effect of smoking and altitude on haemoglobin levels. The outcome is a dichotomous variable that indicates whether a person is anaemic based on the criteria defined above. During the DHS surveys, blood specimens for anaemia and other tests (malaria and Vitamin A deficiency) are voluntarily collected from consenting eligible women and men voluntarily. The blood samples were obtained from a drop of blood taken from a finger prick and analysed with portable HemoCue® Hb analyser devices. Follow-up care was recommended for adults with haemoglobin levels below the following cut-off points: 8 g/dl for non-pregnant women, 7 g/dl for pregnant women and 8 g/dl for men.

2.3.2. Individual, household and regional variables

Three broad categories of variables were of interest as predictors of anaemia amongst women and men in SSA. The first group of predictors for anaemia was individual sociodemographic status. In this study, age, sex, the highest level of education, employment, health insurance coverage, access to print and electronic media were used as individual sociodemographic or socioeconomic predictors of anaemia. Age was a five-year interval categorical variable: 15–19, 20–24, 25–29, 30–34, 35–39, 40–44 and 45–49. Respondents' highest level of education in the DHS surveys were categorised as: no formal education, primary education, secondary or high school education, and higher education (post-secondary education). Respondents' employment status was classified by sector of employment namely: professional or clerical, sales and services, elementary jobs (defined as skilled and unskilled manual jobs) and agriculture. Unemployed respondents were classified as such not-employed – the reference group. In the DHS surveys, respondents were asked how often they a newspaper, listened to the radio or watched television. Our media exposure variable was a binary variable coded and labelled as "1 = adults (women and men) who did not access any of the forms of media at least once a week" and "0 = adults (women and men) who accessed any of the forms of media at least once a week".

The second category of predictors were household measures. These include the location of the household, wealth index and the type of cooking fuel used in the household. Household location was defined as urban or rural. The household wealth index is based on wealth quintiles defined by a composite score based on the number and kinds of consumer goods the household-owned, such as television, bicycle, car, source of drinking water, toilet facilities, and flooring materials. The quintiles were defined and labelled as "0 = poorest", "1 = poorer", "2 = middle", "3 = rich", and "4 = richest". The type of cooking fuel used in the household was classified as electricity, gas (LPG, natural gas and biogas), petroleum-based (Kerosene), biomass (wood, charcoal, shrubs, straws, grass, agricultural crop, animal dung, coal or lignite) and other or no cooking in household. In addition to these household socioeconomic characteristics, we computed the cumulative lifetime mean $PM_{2.5}$ exposure based on available $PM_{2.5}$ data years and the country-specific DHS survey year. That is, the cumulative $PM_{2.5}$ exposure variable was confined to the 1998 (earliest available data) and DHS survey year.

The final group of predictors were administrative level 1 or region socioeconomic covariates, namely: education, employment, household wealth, health insurance coverage and media exposure. The region education variable was defined as the proportion of women and men with secondary education or above. Likewise, the proportion of adults employed in the professional or clerical sector and the proportion of adults working in agriculture were used as region-level employment variables. Region-level wealth index variable was defined as the proportion of adults living in richer households, that is, those living in middle, rich and richest households. We also computed the proportion of adults with health insurance coverage and the proportion of adults who did not access any of the forms of media at least once a week as region-level health insurance and media exposure covariates, respectively.

2.4. Analysis

Our first analysis consists of descriptive statistics of the distribution of the study variables and sample by SSA countries included in the research. Next, we performed a cross-tabulation analysis to assess the distribution of the study variables by sex (women and men), as well as, determine whether there is a significant association between the study variables and sex using the Pearson chi-square test of independence. Cramer's V test was used to assess the strength of the association.

2.4.1. Bayesian multilevel modelling

The third analytical procedure was multilevel modelling. We performed multilevel modelling to determine the extent to which individual, household and regional socioeconomic factors, as well as, $PM_{2.5}$ exposure determine anaemia amongst adults of reproductive age (15–49 years) in SSA. Multilevel modelling acknowledges the inherent hierarchical structure of the data used in this study; that is, the presence of lower (individual) and higher (household, neighbourhood and regional) levels measures and the nesting of individuals within sampled or administrative regions. The DHS survey data has a hierarchical structure with individuals or respondents nested in different spatial units, such as households, districts and regions. Insight from existing studies shows that differences in outcomes or phenomenon of interest amongst the respondents are usually due to contextual or ecological effects (Subramanian et al., 2003; Subramanian 2010). Socioeconomic, environmental and physical characteristics of the areas where respondents are sampled, can have a direct or an indirect influence on the outcome of interest. This contextual effect goes beyond the compositional characteristics of individuals in these spatial units or clusters (Subramanian et al., 2003, 2010). Multilevel modelling enables us to capture both the dependence effect within each hierarchical or spatial unit as well as account for the heterogeneity of covariate effects – that is, varying slopes for the hierarchical structure of the data. That is, with multilevel modelling we can determine the variation of anaemia between spatial or regional administrative units that cannot be explained by the parameter factors (Subramanian et al., 2002).

The multilevel modelling was implemented using the integrated nested Laplace approximations (INLA) framework. INLA is a computationally less-intensive deterministic algorithm for Bayesian inference based on the latent Gaussian model (LGMs). LGMs are mainly structured Bayesian models with conditionally independent likelihood function, latent Gaussian field and prior distribution assigned to the hyper-parameters (Wang et al., 2018). LGMs cover a wide range of models, including generalised linear models, dynamic linear models, spatial and spatio-temporal models (Blangiardo and Cameletti, 2015). The multilevel analysis is a two-level model with individuals nested in sampled or administrative regions. The modelling process entailed fitting a null (empty) model that provides baseline estimations of the administrative region of residence specific effect on the likelihood of being anaemic. A full model with individual, household, $PM_{2.5}$ and regional covariates followed the null model. Given our anaemia was a binary variable, the multilevel model can be written as:

$$y_{ij} \sim \text{Binomial}\left(\pi_{ij}, n_{ij} = 1\right) \tag{1}$$

Where n_{ji} is the logit transformation of the probability π_{ji} a person i in administrative region j being anaemic; expressed as:

$$\eta_{ij} = logit\left(\pi_{ij}\right) = b_0 + u_{ij};\ i = 1, \ldots, n_j; j = 1, \ldots, 407 \tag{2}$$

$$\eta_{ij} = logit\left(\pi_{ij}\right) = \ b_0 + u_j + \beta_1 x_{ij};\ i = 1, \ldots, n_j; j = 1, \ldots, 407 \tag{3}$$

$$u_j \sim N(0, \tau_u); j = 1\ldots, 407 \tag{4}$$

Where n_j is the number of respondents in administrative region $_{j,}$ b_0 is the intercept, β is the linear effect for individual, household and region covariates (e.g. age, sex, education and employment), u_j is region (administrative level 1) random effect with a Gamma distribution prior – τ_u – parameters 1 and 0.00005.

2.4.2. Disease mapping – bayesian approach

We followed the multilevel analysis with a disease-mapping model. The disease mapping was also implemented using the INLA framework. While standardised incidence rates (SIR) are often used in mapping the distribution of diseases, the pattern can be misleading and unreliable in areas with a small population or when the disease is a rare outcome Moraga (2019). Additionally, SIR disease maps do not also account for potential spatial dependency in the disease distribution. Spatial dependency is based on the assumption that neighbouring spatial units have similar characteristics; thus, neighbouring units can affect an outcome in a spatial unit (structured spatial effect) (Darmofal, 2015; Riebler et al., 2016). The concept of spatial dependency is based on Tobler's first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). Employing this concept in disease mapping entails borrowing information from neighbouring units in the modelling process, as well as, incorporating covariate information to smoothen or shrink extreme values resulting from a small sample size Moraga (2019). Nonetheless, a spatial unit's characteristics can influence an outcome independent of neighbouring units – that is, the independent region-specific noise or non-spatial heterogeneity (unstructured spatial effect) (Riebler et al., 2016). To account for both scenarios (spatial dependency and non-spatial heterogeneity), the anaemia disease mapping was based on the Besag-York-Mollie (BYM) specification. The BYM method affords the ability to account for both spatial dependency or structured spatial effect and non-spatial heterogeneity or unstructured spatial effect in the modelling process.

The outcome variable for the disease mapping is a count of anaemia cases in each administrative level 1 and the covariates are administrative level 1 socioeconomic and $PM_{2.5}$ measures. The socioeconomic covariates were the proportion of adults with secondary education or above, working in professional or clerical jobs, working in sales and services, elementary jobs (skilled and unskilled manual jobs), agriculture, with health insurance coverage, without access to the three forms of media, and use clean fuel in household cooking. Clean fuel was defined as using electricity, natural gas, biogas or LPG for cooking. For $PM_{2.5}$, we compute the mean cumulative exposure for adults in each region. Similar to the multilevel modelling, we fitted two disease mapping models: the first model without covariates and the second with covariates. The disease mapping models for anaemia can be expressed:

$$y_i \sim \text{Poisson}(E_i \rho_i)$$

$$\eta_i = \log(\rho_i) = b_0 + u_i + v_i \tag{5}$$

Where b_0 is the intercept quantifying the average anaemia rate in all 407 administrative regions. E_i is the number of expected cases of anaemia in each administrative region (acting as an offset); u_i is the spatially structured residual (spatial dependency) modelled using the intrinsic Conditional Auto-Regressive (iCAR) specification and v_i is the unstructured residual modelled using exchangeability amongst the 407 regions. The covariate model or ecological model can then be expressed as:

$$\eta_i = b_0 + \beta_1 x_i + u_i + v_i \tag{6}$$

Where x_i is a matrix of the region-level covariates mentioned above. The multilevel and spatial (disease mapping) models were implemented in the open-access R software (R Core Team, 2020) using the R-INLA package (Bakka et al., 2018; Lindgren and Rue, 2015; Lindgren et al., 2011; Martins et al., 2013; Riebler et al., 2017; Rue et al., 2009). Results of the relative risk of anaemia and region-specific effect for the multilevel and disease mapping models were also visualised in R software using the tmap and tmaptools packages (Tennekes, 2018).

3. Results

3.1. Descriptive and cross-tabulation

Table 1 shows the descriptive summary statistics of the variables used in this study. The results shows that 36.0% of women and men between the ages of 15–49 years had anaemia. Most respondents were

Table 1
Descriptive summary of study variables.

	Frequency / Mean	Percentage / Range	Weighted Percentage / Range
Outcome			
Anaemia			
No	172,914	64•0	64•3
Yes	97,097	36•0	35•7
Individual - Level factors			
Age			
15–19 years	59,415	22•0	21•8
20–24 years	48,969	18•1	18•1
25–29 years	44,823	16•6	16•8
30–34 years	37,935	14•0	14•3
35–39 years	32,443	12•0	12•1
40–44 years	25,360	9•4	9•3
45–49 years	21,066	7•8	7•7
Sex			
Female	198,723	73•6	73•5
Male	71,288	26•4	26•5
Education			
No education	68,285	25•3	25•3
Primary education	92,237	34•2	34•3
Secondary or High school education	95,590	35•4	35•1
Post-secondary education	13,899	5•1	5•3
Sector of employment			
Not employed	85,959	31•8	31•1
Professional or clerical	15,135	5•6	5•6
Sales & services	52,062	19•3	19•8
Elementary jobs - manual (skilled and unskilled)	29,278	10•8	10•9
Agriculture	87,577	32•4	32•7
Health Insurance			
No	255,219	94•5	94•5
Yes	14,792	5•5	5•5
Access none of the three media			
No	145,776	54•0	54•5
Yes	124,235	46•0	45•5
Household and Neighbourhood Factors			
Location of household			
Urban	98,751	36•6	36•1
Rural	171,260	63•4	63•9
Wealth index - Household			
Poorest	53,561	19•8	17•5
Poorer	49,179	18•2	18•6
Middle	50,592	18•7	19•4
Rich	53,959	20•0	21•2
Richest	62,720	23•2	23•2
Cooking Fuel - source			
Electricity	17,233	6•4	5•9
Gas	18,188	6•7	7•7
Petroleum based	3293	1•2	1•3
Biomass	229,838	85•1	84•6
Other/No cooking in household	1459	0•5	0•6
PM2.5 exposure - Neighbourhood-level	11•9 (6•0)[a]	0•7 – 41•9	12•2 (6•1)[a]
Administrative Level 1 Factors			
Percentage of educated	40•6 (22•8)[a]	0•0 - 92•5	40•8 (23•8)[a]
Percentage of employed in professional or clerical sector	5•6 (3•9)[a]	0•0 - 22•3	5•7 (4•1)[a]
Percentage of employed in agriculture sector	32•4 (23•6)[a]	0•0 - 94•7	32•5 (23•8)[a]
Percentage living in non-poor household	62•0 (20•8)[a]	0•0 - 100•0	64•2 (19•7)[a]
Percentage of adults with health insurance	5•5 (10•1)[a]	0•0 - 88•2	5•7 (9•7)[a]
Percentage of adults without access to media	46•0 (21•1)[a]	4•2 - 90•8	45•3 (21•5)[a]
Number of Observations - N	270,011		

[a] = mean; with standard deviation in parenthesis.

between the ages 15–29 years – 56.7%, female (73.6%), had primary or secondary education (69.6%), unemployed (31.8%) or worked in the agriculture sector (32.4%), had no health insurance coverage (94.5%), lived in rural households (63.4%) and used biomass for cooking (85.1%). The mean lifetime $PM_{2.5}$ exposure per respondent was 19.9 μg/m^3.

Given gender inequalities in health outcomes and socioeconomic status in the study region, we considered the distribution of anaemia, individual and household characteristics by sex,. Table 2 shows the results of the cross-tabulation analysis. The results of the Pearson chi-square and Cramer's V tests of independence show that all study variables were significant but weakly associated with sex. The result indicates that anaemia was more prevalent amongst women than men: 41.1% of women were anaemic compared to 21.8% of men. The result also reveals men had better socioeconomic status compared to women for those included in the sample. The mean lifetime $PM_{2.5}$ exposure was also relatively higher amongst women (12.0 μg/m^3) than men (11.7 μg/m^3).

3.2. Multilevel models and region effect on anaemia

Our first multilevel model was an empty-model where we examine the extent of the log-odds of the probability of being anaemic varies between regions in the Sub-Saharan Africa. We then fitted a full model with all predictor variables to compare with the regions varying effect in the previous (empty model). Table 3 shows the results of the two models. For easy interpretation of the results, we report the odds ratios (OR) here. The result of the empty model shows that the OR of the probability being anaemic is 0.547; that is, on average, respondents are 45.3% less likely to be anaemic. The result of the full model shows that on average respondents were 50.5% more likely to be anaemic – if they have all the characteristics of the reference categories. Older adults were less likely to be anaemic compared to teenagers. The model also shows males, educated respondents and respondents working in any the four sectors of employment were less likely to be anaemic compared to females and unemployed respondents, respectively. For instance, males, respondents with secondary and post-secondary levels of education were 62.2%, 19.4%, and 30%, respectively, less likely to be anaemic. Respondents

Table 2
Cross-tabulation analysis of the distribution of individual and household study variables by sex.

	Women Frequency	Percentage	Men Frequency	Percentage	2 (Cramer's V)	P-value
Outcome						
Anaemia						
No	117,156	59•0	55,758	78•2	8.5e+03 (−0.18)	0.000
Yes	81,567	41•1	15,530	21•8		
Individual - Level factors						
Age						
15–19 years	42,553	21•4	16,862	23•7		
20–24 years	36,538	18•4	12,431	17•4		
25–29 years	33,889	17•1	10,934	15•3		
30–34 years	28,297	14•2	9638	13•5	364.38 (0.04)	0.000
35–39 years	24,184	12•2	8259	11•6		
40–44 years	18,195	9•2	7165	10•1		
45–49 years	15,067	7•6	5999	8•4		
Education						
No education	55,462	27•9	12,823	18•0		
Primary education	69,448	35•0	22,789	32•0	4.7e+03 (0.13)	0.000
Secondary or High school education	65,316	32•9	30,274	42•5		
Post-secondary education	8497	4•3	5402	7•6		
Sector of employment						
Not employed	70,893	35•7	15,066	21•1		
Professional or clerical	9101	4•6	6034	8•5		
Sales & services	41,539	20•9	10,523	14•8	1.0e+04 (0.19)	0.000
Elementary jobs - manual (skilled and unskilled)	17,492	8•8	11,786	16•5		
Agriculture	59,698	30•0	27,879	39•1		
Health Insurance						
No	187,551	94•4	67,668	94•9	29.97 (−0.01)	0.000
Yes	11,172	5•6	3620	5•1		
Access none of the three media						
No	100,521	50•6	45,255	63•5	3.5e+03 (−0.11)	0.000
Yes	98,202	49•4	26,033	36•5		
Household and Neighbourhood Factors						
Location of household						
Urban	73,239	36•9	25,512	35•8	25.78 (0.01)	0.000
Rural	125,484	63•2	45,776	64•2		
Wealth index - Household						
Poorest	38,597	19•4	14,964	21•0		
Poorer	35,935	18•1	13,244	18•6		
Middle	37,376	18•8	13,216	18•5	118.57 (0.02)	0.000
Rich	39,975	20•1	13,984	19•6		
Richest	46,840	23•6	15,880	22•3		
Cooking Fuel - source						
Electricity	10,983	5•5	6250	8•8		
Gas	12,229	6•2	5959	8•4		
Petroleum based	2599	1•3	694	1•0	3.8e+03 (0.12)	0.000
Biomass	172,639	86•9	57,199	80•2		
Other/No cooking in household	273	0•1	1186	1•7		
PM2.5 exposure - Neighbourhood-level	12.0 (5.9)[a]	0.7 - 41.9	11.7 (6.2)[a]	0.7 - 41.9		
Number of Observations - N	**198,723**		**71,288**		**270,011**	

[a] = mean; with standard deviation in parenthesis.

Table 3
Results from the multilevel spatial mixed-effect models for determinants of anaemia amongst women and men in Sub-Saharan African countries.

	Intercept-only Model Mean (SD)	95% CI	OR	Full Model Mean (SD)	95% CI	OR
Intercept	**−0•603 (0•032)**	**−0•665 - −0•541**	**0•547**	**0•409 (0•069)**	**0•273 - 0•544**	**1•505**
Individual - Level factors						
Age						
15–19 years (ref)						
20–24 years				**−0•183 (0•014)**	**−0•209 - −0•156**	**0•833**
25–29 years				**−0•240 (0•014)**	**−0•268 - −0•212**	**0•787**
30–34 years				**−0•223 (0•015)**	**−0•252 - −0•193**	**0•800**
35–39 years				**−0•162 (0•016)**	**−0•193 - −0•131**	**0•850**
40–44 years				**−0•122 (0•017)**	**−0•156 - −0•088**	**0•885**
45–49 years				**−0•154 (0•018)**	**−0•191 - −0•118**	**0•857**
Sex						
Female (ref)						
Male				**−0•973 (0•015)**	**−1•004 - −0•943**	**0•378**
Education						
No education (ref)						
Primary education				**−0•144 (0•013)**	**−0•169 - −0•119**	**0•866**
Secondary or High school education				**−0•216 (0•015)**	**−0•245 - −0•187**	**0•806**
Post-secondary education				**−0•357 (0•027)**	**−0•411 - −0•304**	**0•700**
Sector of employment						
Not employed (ref)						
Professional or clerical				**−0•106 (0•023)**	**−0•152 - −0•060**	**0•899**
Sales & services				**−0•066 (0•013)**	**−0•092 - −0•040**	**0•936**
Elementary jobs - manual (skilled and unskilled)				**−0•114 (0•016)**	**−0•146 - −0•082**	**0•892**
Agriculture				**−0•044 (0•013)**	**−0•070 - −0•019**	**0•957**
Health Insurance						
No (ref)						
Yes				−0•029 (0•022)	−0•073 - 0•015	0•971
Access none of the three media						
No (ref)						
Yes				**0•049 (0•010)**	**0•029 - 0•068**	**1•050**
Household and Neighbourhood Factors						
Location of household						
Urban (ref)						
Rural				**−0•032 (0•012)**	**−0•056 - −0•007**	**0•969**
Wealth index - Household						
Poorest (ref)						
Poorer				**−0•077 (0•014)**	**−0•105 - −0•050**	**0•926**
Middle				**−0•128 (0•014)**	**−0•157 - −0•100**	**0•880**
Rich				**−0•201 (0•016)**	**−0•232 - −0•169**	**0•818**
Richest				**−0•327 (0•019)**	**−0•365 - −0•289**	**0•721**
Cooking Fuel - source						
Electricity (ref)						
Gas				**0•163 (0•029)**	**0•105 - 0•221**	**1•177**
Petroleum based				0•043 (0•046)	−0•047 - 0•133	1•044
Biomass				0•044 (0•025)	−0•006 - 0•093	1•045
Other/No cooking in household				−0•040 (0•070)	−0•178 - 0•096	0•961
$PM_{2.5}$ exposure - Neighbourhood-level				**0•011 (0•002)**	**0•008 - 0•015**	**1•011**
Administrative Level 1 Factors						
Percentage of educated				**−0•006 (0•001)**	**−0•007 - −0•004**	**0•994**
Percentage of employed in agriculture sector				0•000 (0•001)	−0•001 - 0•001	1•000
Percentage of employed in professional or clerical sector				**−0•033 (0•003)**	**−0•039 - −0•027**	**0•968**
Percentage living in non-poor household				0 •000 (0•000)	−0•001 - 0•001	1•000
Percentage of adults with health insurance				−0•001 (0•001)	−0•003 - 0•001	0•999
Percentage of adults without access to media				−0•003 (0•001)	−0•004 - 0•001	0•997
Random Effects - hyperparameter						
Precision for Administrative level 1	2•594 (0•19)5	2•215 - 2•981	13•38,319,747	3•198 (0•248)	2•718 - 3•693	24•484
Model Diagnostics						
DIC	333,638•94			324,661•12		
WAIC	333,636•99			324,659•72		
Marginal log-likelihood	−167,368•28			−163,068•44		

significant results boldened.

who did not have access to all three media (newspaper, radio and television) at least once a week were 5% more likely to be anaemic. With regards to household factors, respondents living in rural areas and those living in richer households were less likely to be anaemic compared to those in urban and poorest households. Respondents who lived in households that use natural gas for cooking were 17% more likely to be anaemic compared to those that use electricity for cooking in the household. One μg/m^3 increase in cumulative lifetime $PM_{2.5}$ exposure was associated with 1% increase in the likelihood of being anaemic.

For regional level factors, the model shows that the proportion of respondents with secondary education and above, and working in the professional sector were significantly inversely associated with the likelihood of being anaemic. A percentage increase in the proportion of respondents with secondary education and above, and employed in the professional or clerical sectors was associated with 0.6% and 3.2% reduction in the likelihood of being anaemic. The model diagnostics (DIC and WAIC) show that the full model is statistical preferable than the empty model. The full model had lower DIC and WAIC values of

324,661.12 and 324,659.72, respectively, compared to the empty model. Fig. 1 below displays the log-odds of regional effect on the likelihood of being anaemic for the empty and full models. The result for the empty model shows that people living in regions in Southern Africa, western Ethiopia, Democratic Republic of Congo (DRC), and Tanzania had lower probability of being anaemic. While those in West Africa generally had a higher probability of being anaemic. When we account for individual, household and region-level factors and $PM_{2.5}$, the probability of being anaemic increased for people living in some regions in Namibia, Southern Africa and the DRC. On the other hand, lower probability of being anaemic was observed from previously high-risk regions – in the null model – in Guinea and the Savanna regions of Ghana, Togo, and Benin.

3.3. Mapping distribution of anaemia

Table 4 shows the posterior estimates of relative risk of anaemia. The results indicate that an increase in the proportion of adults employed in the professional and elementary sectors and an increase in regional $PM_{2.5}$ exposure reduces the relative risk of anaemia. Contrary, a proportional increase in adults working in sales and services and without access to media increases the relative risk of anaemia.

Fig. 2 shows the results of the relative risk of anaemia from the Bayesian spatial models. The upper most maps shows the predicted relative risk of anaemia from the two modelling process. The map on the top left corner shows the relative risk of anaemia from the empty model (without covariates) while the map on the top right corner shows the relative risk from the full model. The result of both models show similar patterns of relative risk of anaemia with minute differences. The results of the relative risk models, similar to the rate maps, show that regions in West Africa and coastal East Africa have a higher relative risk of anaemia; albeit greater heterogeneity in the relative risk models compared to the rate maps. The maps below show the region specific relative risk of anaemia for the empty model (bottom left) and full model (bottom right). The map on the bottom left shows that region specific relative risks of anaemia compared to the whole study region, while the bottom right map shows the residual relative risk of each administrative region after considering region-covariates. Again, the empty model shows the region-specific relative risk of anaemia was higher amongst regions in West Africa followed by those in coastal East Africa. However, the high-risk pattern in these regions diminishes in the full model (accounting for region covariates). The model diagnostics values for the full model were lower (DIC=3494.08, WAIC=3408.85) than the empty model; that is, the full model is statistically preferable compared to the empty model

4. Discussion

Using multilevel mixed-effect models, the study assessed and visualised the residual spatial effects or region-specific effects on the risk of anaemia amongst adults in Sub-Saharan Africa. The study also presents disease maps that provide knowledge on the spatial distribution of

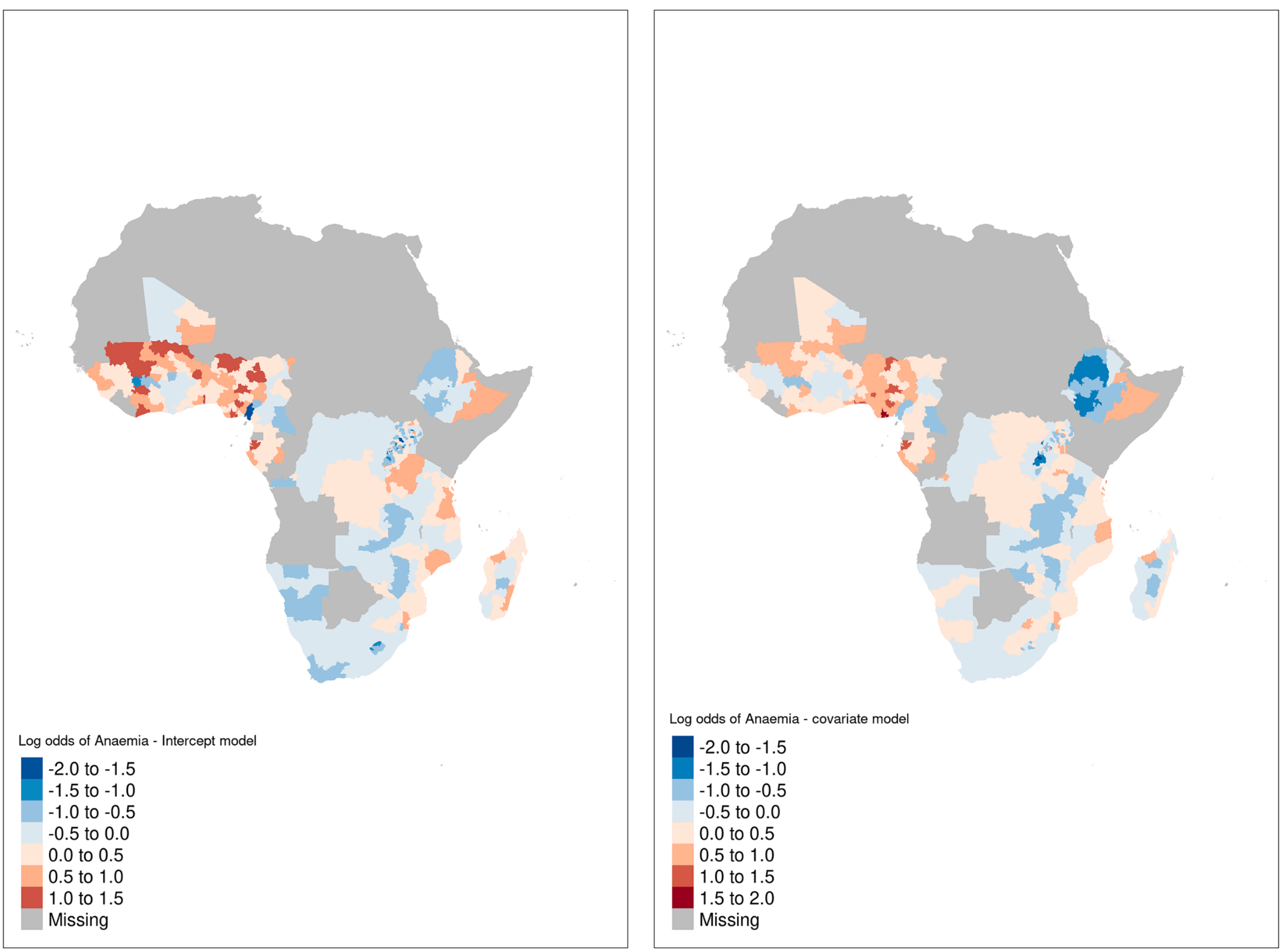


Fig. 1. Map of the posterior mean (log-odds) of region-level random effect for regions in 26 Sub-Saharan African countries.

Table 4
Posterior estimates of relative risk of anaemia amongst women and men in Sub-Saharan African countries.

	Intercept-only Model		Full Model	
	Mean (SD)	95% CI	Mean (SD)	95% CI
Intercept	−0•064 (0•022)	−0•107 - −0•022	−0•103 (0•152)	−0•402 - 0•194
Covariates				
Education			−0•001 (0•002)	−0•004 - 0•002
Professional employment			**−0•023 (0•006)**	**−0•034 - −0•011**
Sales and Services			**0•012 (0•002)**	**0•008 - 0•016**
Manual labour - skilled and unskilled			**−0•006 (0•002)**	**−0•01 - −0•001**
Agriculture			−0•001 (0•001)	−0•003 - 0•002
Health Insurance Coverage			0•000 (0•002)	−0•004 - 0•003
Access none of the three media			**0•004 (0•001)**	**0•001 - 0•006**
Household wealth			0•001 (0•001)	−0•001 - 0•003
Household Clean Fuel usage			−0•001 (0•001)	−0•004 - 0•002
$PM_{2.5}$ exposure			**−0•008 (0•004)**	**−0•015 - −0•001**
Random Effect - Hyperparameters				
Precision for regions - iid component	5•960 (0•553)	4•840 - 6•990	8•590 (0•028)	7•920 - 9•150
Precision for regions - spatial component	89•920 (94•913)	19•790 - 334•120	198•760 (9•99e+07)	69•920 - 664•630
Model Diagnostics				
Deviance Information Criterion - DIC	3507•24		3494•08	
Watanabe-Akaike information criterion - WAIC	3417•75		3408•85	
Marginal log-Likelihood	−2027•88		−2048•87	

significant results boldened.

the risk of anaemia in Sub-Saharan Africa after accounting for socio-economic and air quality factors. The results of the multiple multilevel model show a gendered and socioeconomic pattern in the risk of anaemia amongst adults living in Sub-Saharan Africa. Environmental factors, such as $PM_{2.5}$, percentage of person with secondary education or higher and persons working in professional or clerical jobs, were significantly associated with the anaemia (Table 3). The disease-mapping model also shows that employment, media access and $PM_{2.5}$ were significant predictors of the risk of anaemia (Table 4).

As a nutritional disease, anaemia is associated with poor dietary habits (FAO, IFAD, UNICEF, WFP AND WHO, 2017; Teshale et al., 2020). Poor dietary habits amongst any given population are often due to socioeconomic challenges and inadequate nutritional knowledge (Moschovis et al., 2018; Teshale et al., 2020). The number of people living in poor households in Sub-Saharan African region is significant higher compared to other regions of the world. Due to individual and household poverty, the majority of population in this region have difficulties accessing proper nutrition; thus, contributing to the high burden of anaemia in Sub-Saharan Africa. The findings of the study show that the prevalence of anaemia is approximately 36% amongst women of reproductive age and men aged 15–49 years. Although this rate is lower than a recent reported estimate of 39.2% in 2016 (FAO, IFAD, UNICEF, WFP AND WHO, 2017), the reported prevalence in this report was for women of reproductive age only. Knowledge from existing studies shows men have lower prevalence rate of anaemia, compared to women (Adamu et al., 2017). Thus, the inclusion of men in the estimate may account for this lower rate. Indeed, this becomes clear in the cross-tabulation analysis (Table 2) where the prevalence of anaemia amongst women is almost twice (41.05%) that of men (21.78%) and also higher than the reported rate for women of reproductive age. This gendered pattern is also revealed in the multiple multilevel analysis, where men were 62% less likely to be anaemic compared to women. While biological explanations such as menstrual periods, pregnancy, and lactation, can make women vulnerable to the risk of anaemia; we argue womens' risk of anaemia mainly stem from their socioeconomic vulnerabilities. Gender inequalities in employment, education, and economic resources, as well as, some bad sociocultural practices (restriction of women and girls from eating certain nutrient rich diets) means women in Sub-Saharan Africa may not able to meet their dietary needs (Arzoaquoi et al., 2015). As evident in Table 2, women have lower socioeconomic status (education, employment and access to information) compared to their male counterparts.

Consistent with knowledge from existing studies (Adamu et al., 2017; Elmardi et al., 2020; Moschovis et al., 2018; Nankinga and Aguta, 2019; Teshale et al., 2020), the findings of the multilevel analysis also reveals that other individual and household characteristics are significantly associated with anaemia. Regarding individual socioeconomic factors, being older than 19 years, educated, employed, and having access to print and electronic media reduces the likelihood of being anaemic amongst women and men in Sub-Saharan Africa. Being older, having a higher education, and employment are closely associated with accumulation of wealth hence adults who are educated and employed adults are more likely to have the financial means and resources to meet their dietary requirements. Closely linked with education, adults with access to print and electronic media are more likely to have adequate knowledge or information on essential nutrients and foods items that contain these nutrients (Elmardi et al., 2020; Teshale et al., 2020).

At the household and neighbourhood level, the findings indicate place of residence, household wealth, household cooking fuel and $PM_{2.5}$ have significant association with the risk of anaemia. The association between rural/urban residence and risk of anaemia is inconclusive. While some studies reported high risk of anaemia amongst rural populations (Nankinga and Aguta, 2019); others have found higher risk amongst urban populations (Adamu et al., 2017) or observed no statistical difference between the two (Elmardi et al., 2020). In our study, the findings reveal the risk of being anaemic is lower amongst rural adult populations. We argue that the burden of anaemia in rural areas is mainly as a result of nutritional challenges induced by socioeconomic difficulties (Nankinga and Aguta, 2019). However, access to land for subsistence farming in rural areas means residents have improved access to cheaper, diverse and nutritious food sources, reducing their risk of anaemia (Adamu et al., 2017). On the other hand, the high risk of anaemia amongst urban populations may be attributable to more expensive food and anaemia-inducing chronic diseases, such as diabetes and cancer or environmental risk exposures, which are common amongst urban populations (Adamu et al., 2017). We also hypothesise that there may be other potential confounders which influence the association between rural/urban residence and the risk of being anaemic. Consistent with knowledge from existing studies (Adamu et al., 2017; Chaparro and Suchdev, 2019; Moschovis et al., 2018; Nankinga and Aguta, 2019; Teshale et al., 2020), the finding also reveals adults in richer households have lower likelihood of being anaemic. Individuals in richer households have the economic means to meet their nutritional and health needs, hence their lower risk of anaemia compared to those from poorer households. Exposure to indoor and outdoor pollution increases the risk of anaemia amongst any given population. Existing studies shows that outdoor pollutants, such as $PM_{2.5}$, and indoor pollution from the use of biofuel, natural gas, liquid-petroleum-gas and other petroleum based fuels have significant influence on the risk of anaemia and cardiovascular diseases (Accinelli and Leon-Abarca, 2017; Honda et al., 2017; Stanković et al., 2006). Air pollutants may biological interfere with red blood cells production through the synthesis of haem, the formation of the red blood cells and reduce the life expectancy of red

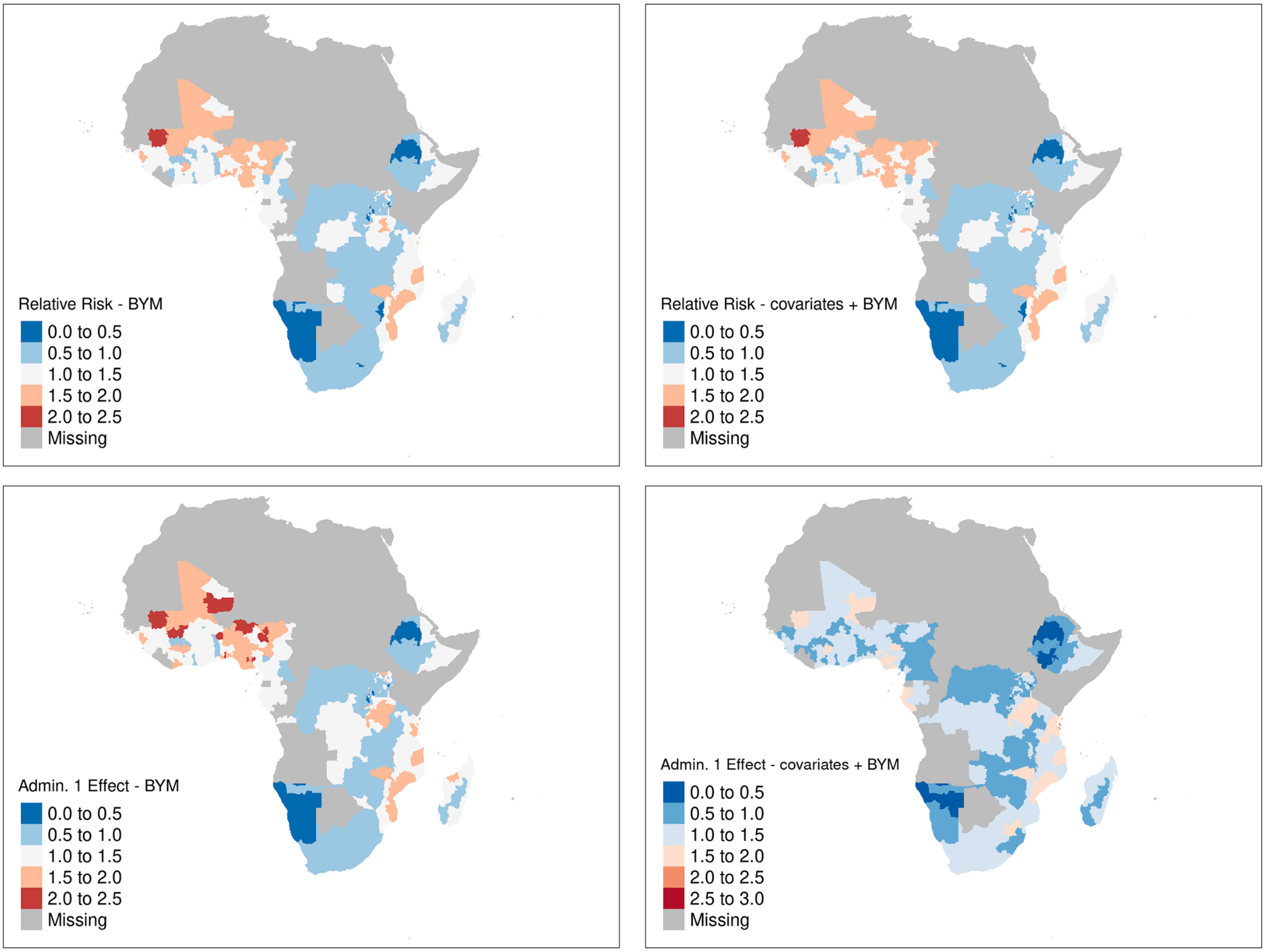


Fig. 2. Map of the distribution of relative risk of anaemia from the empty model (top-left), relative risk of anaemia from full model (top-right), region-specific relative risk of anaemia null model (bottom-left) and region-specific relative risk of anaemia full model (bottom-right).

blood cells (Stanković et al., 2006).

It is widely acknowledged that contextual socioeconomic economic factors effect health outcomes. Regional or area socioeconomic factors may amplify the effect of individual and household deprivation on health (Macintyre et al., 2008). The findings of our study are consistent with this observation. Higher proportion of educated and adults employed in the professional and clerical sectors in the region of residence reduced an individual's likelihood of being anaemic. A similar effect was observed in relation to the rate of anaemic adults, where rate of employment in low-skill jobs and professional and clerical jobs reduced the relative risk of anaemia at the region-level (Table 4). Interestingly, a percentage increase in the proportion of adults working in sales and services increases the region-level relative risk of anaemia. The majority of adults in Sub-Saharan Africa working in the sales and services sector are mainly in the subsistence level jobs that are characterised by street trading, cottage enterprises and other labour intensive low skill jobs Aryeetey (2015). People working in this sector are usually poor and less educated; thus, a possible explanation for the high risk of anaemia in regions with a high percentage of the adult population in this sector of the economy. The number of adults with access to media also reduces the regional level relative risk of anaemia given with greater access to media and information may be more likely have adequate knowledge on nutritional requirements and also better health behaviours (Elmardi et al., 2020).

Remarkably, 1 $\mu g/m^3$ increase in regional mean exposure of $PM_{2.5}$ reduces the relative risk of anaemia by 1% (RR = exp(−0.008)). The mechanism whereby air pollution appears be related to anaemia needs further consideration. It would be a mistake to infer a simple causal relationship between the two variables. We argue that regional level $PM_{2.5}$ levels may be indicative of wider socioeconomic development in Sub-Saharan Africa, hence the lower likelihood of anaemia in higher $PM_{2.5}$ areas. Indeed, evidence from existing studies and reports show that there is a positive correlation between air pollution and GDP growth (Dechezleprêtre et al., 2020). It is argued higher socioeconomic development in Sub-Saharan Africa often implies more industrial, transportation and energy consumption activities that lead to pollutant emissions (Dechezleprêtre et al., 2020) but that these same higher socioeconomic areas contain individuals with better access to nutritious food, which prevents anaemia.

The results from the disease mapping shows that the burden of anaemia is higher amongst adult populations in West Africa and the coastal regions of East Africa. Both disease prevalence and relative risk maps reveal Mali, northern Benin, and northern and central Nigeria as high-risk regions in West Africa; while Mozambique is a high-risk region in East Africa. This finding supports evidence from existing reports that shows that the burden of anaemia is high in the West Africa region (FAO, IFAD, UNICEF, WFP AND WHO, 2017). The prevalence of anaemia amongst women of reproductive age (15–49 years) in Western Africa was 49.3 in 2016, compared to 43.5 in Central Africa, 31.2 in Eastern Africa, 32.1 in Northern Africa and 26.0 in Southern Africa

(FAO, IFAD, UNICEF, WFP AND WHO, 2017). The reasons for this spatial distribution pattern are multifactorial; major amongst these factors being the role of malaria. Countries in the high-risk areas in the disease distribution maps are known to have a rate of Plasmodium falciparum infection – the malaria-causing vector (Stevens et al., 2013). Existing studies show that Plasmodium falciparum induce anaemia during the blood stages of infection and malaria is known to be a major cause of anaemia in tropical regions (White, 2018). Other contributing factors could be sickle cell traits that are high in western and eastern Africa regions (Chaparro and Suchdev, 2019; Macharia et al., 2018). Nonetheless, the finding also reveals that when we consider regional socioeconomic and $PM_{2.5}$ the burden of anaemia associated with region of residence in these high-risk areas becomes similar to other parts of Sub-Saharan Africa. That is, regional socioeconomic factors and air quality have a potential levelling effect on the region-specific risk burden of anaemia.

Our study has some limitations. First, the spatial scale used in this study is very large (average 36,391.93 km^2 per areal unit) hence does not account for potential intra-regional heterogeneity in the risk of anaemia amongst women and men in Sub-Saharan Africa. There are variations in socioeconomic indicators within the regions used in the study that could have a significant influence on the prevalence and risk distribution of anaemia within the regions of study. Nonetheless, the stratified sampling approach used in the DHS surveys means geographic or spatial estimates from the data are only accurate at regional and rural/urban levels. This informed our decision to use regions as the unit of analysis for the spatial models. Second, only one air pollutant was used in this study to access the effect of air quality on the risk of anaemia. Further studies can explore the effect of other air pollutants, such as carbon monoxide, lead, nitrogen oxides, ozone and sulphur dioxide, on the risk and spatial distribution of anaemia amongst the adult populations in Sub-Saharan Africa. Reliable, continent-wide data on these other air pollutants are scarce hence the decision to use publicly available $PM_{2.5}$ data from NASA's SEDAC and Dalhousie University's Atmospheric Composition Analysis Group. The third limitation of this study pertains to the difference in DHS survey years for the Sub-Saharan African countries included in this study. The discontinuation of the survey is some countries and the implementation of the survey in different years in different countries mean the datasets do not have similar years of implementation. These discrepancies can have an effect of obtaining a spatial distribution pattern that does not reflects current trend in anaemia risk or show distribution patterns for similar year(s). Lastly, the study is based on secondary data from cross-sectional surveys; hence, we cannot infer causality in the effect of individual, household and other contextual factors on the risk of anaemia. These limitations notwithstanding, our study offers novel insights on the multidimensional risk factors of anaemia and the spatial distribution of the risk of the disease across the Sub-Saharan African region.

Notwithstanding the limitations, the findings of this study show that improving individual, household and local socioeconomic status can profoundly reduce the risk of anaemia amongst adult populations in Sub-Saharan Africa. Similarly, the findings reveal that at the individual level improving air quality by minimising local $PM_{2.5}$ concentrations can also help reduce the risk of anaemia amongst women and men in Sub-Saharan African countries. The multidimensional nature of these risk factors indicate that comprehensive intervention policies that address socioeconomic inequalities and improves the quality of environment, are needed to tackle the burden of anaemia in Sub-Saharan Africa. Public health and social programs that improve access to media, education, employment, household wealth, clean cooking fuel and minimises $PM_{2.5}$ exposure are needed to address the pathways to the risk of anaemia. Such programs will equip the population in Sub-Saharan Africa with the knowledge and resources, such as access to improved or balance diet; thus, minimising their risk of anaemia. Knowledge from the disease mapping and risk distribution patterns can also guide target intervention measures in identified high-risk areas – West Africa and the coastal regions of East Africa – in the context of resource scarcity in Sub-Saharan Africa.

5. Authors' contribution statement

PMA, SSB, AP and CES designed the study concept. PMA curated data, developed statistical methods, analysed data with input from SSB, AP and CES. PMA prepared results. All authors contributed to writing the discussion. All authors contributed to writing the different drafts of the manuscript, editing, and revising the final version. SSB and AP wrote the first draft of the report. PMA and CES contributed to subsequent drafts. All authors had access to all data used in the study. SSB and AP accessed and verified the underlying study data. All authors had final responsibility for the decision to submit for publication.

Declaration of Competing Interest

The authors declare no conflict of interest in carrying out this research. There was no funding for this research.

Acknowledgement

Prince M. Amegbor and Clive E. Sabel were supported by Big Data Centre for Environment and Health (BERTHA) funded by the Novo Nordisk Foundation Challenge Programme (grant NNF17OC0027864). We are also grateful to the NASA Socioeconomic Data and Applications centre (SEDAC), the USAID DHS programme, and Uganda Bureau of Statistics for the data used in this study.

References

Abera, A., Friberg, J., Isaxon, C., Jerrett, M., Malmqvist, E., Sjöström, C., Vargas, A.M., 2020. Air quality in Africa: public health implications. Ann. Rev. Public Health 42, 193–210. https://doi.org/10.1146/annurev-publhealth-100119-113802 https://doi.org/.

Accinelli, R.A., Leon-Abarca, J.A., 2017. Solid fuel use is associated with anemia in children. Environ. Res. 158 (June), 431–435. https://doi.org/10.1016/j.envres.2017.06.032 https://doi.org/.

Adamu, A.L., Crampin, A., Kayuni, N., Amberbir, A., Koole, O., Phiri, A., Fine, P., 2017. Prevalence and risk factors for anemia severity and type in Malawian men and women: urban and rural differences. Popul Health Metr 15 (1), 1–15. https://doi.org/10.1186/s12963-017-0128-2 https://doi.org/.

Aryeetey, E., 2015. The informal economy, economic growth, and poverty in Sub-Saharan Africa. In: McKay, A., Thorbecke, E. (Eds.), Economic Growth and Poverty Reduction in Sub-Saharan Africa: Current and Emerging Issues, pp. 583–605. https://doi.org/10.1093/acprof:oso/9780198728450.003.0006 https://doi.org/.

Arzoaquoi, S.K., Essuman, E.E., Gbagbo, F.Y., Tenkorang, E.Y., Soyiri, I., Laar, A.K., 2015. Motivations for food prohibitions during pregnancy and their enforcement mechanisms in a rural Ghanaian district. J. Ethnobiol. Ethnomed. 11 (1), 1–9. https://doi.org/10.1186/s13002-015-0044-0 https://doi.org/.

Bakka, H., Fuglstad, G., Riebler, A., Bolin, D., Krainski, E., Simpson, D., and Lindgren, F. (2018). *Spatial modelling with R-INLA : a review*.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R-INLA. John Wiley and Sons.

Chaparro, C.M., Suchdev, P.S., 2019. Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. Ann. N. Y. Acad. Sci. 1450 (1), 15–31. https://doi.org/10.1111/nyas.14092 https://doi.org/.

Darmofal, D., 2015. Spatial analysis for the social sciences. Spatial Analysis for the Social Sciences. https://doi.org/10.1017/cbo9781139051293 https://doi.org/.

Dechezleprêtre, A., Rivers, N., and Stadler, B. (2020). *The economic cost of air pollution: evidence from Europe*(No. 1584).

Elmardi, K.A., Adam, I., Malik, E.M., Abdelrahim, T.A., Elhag, M.S., Ibrahim, A.A., Kremers, S., 2020. Prevalence and determinants of anaemia in women of reproductive age in Sudan: analysis of a cross-sectional household survey. BMC Public Health 20 (1), 1–12. https://doi.org/10.1186/s12889-020-09252-w https://doi.org/.

FAO, IFAD, UNICEF, W. and W. (2017). *The State of Food Security and Nutrition in the World 2017: Building resilience For Peace and Food Security*. Rome.

GBD 2015 Disease and Injury Incidence and Prevalence Collaborators, 2016. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet 388, 1545–1602. https://doi.org/10.1016/S0140-6736(16)31678-6 https://doi.org/.

Heft-Neal, S., Burney, J., Bendavid, E., Burke, M., 2018. Robust relationship between air quality and infant mortality in Africa. Nature 559 (7713), 254–258. https://doi.org/10.1038/s41586-018-0263-3 https://doi.org/.

Honda, T., Pun, V.C., Manjourides, J., Suh, H., 2017. Anemia prevalence and hemoglobin levels are associated with long-term exposure to air pollution in an older population. Environ. Int. 101, 125–132. https://doi.org/10.1016/j.envint.2017.01.017 https://doi.org/.

Hurtado, E.K., Claussen, A.H., Scott, K.G., 1999. Early childhood anemia and mild or moderate mental retardation. American J. Clin. Nutr. 69 (1), 115–119. https://doi.org/10.1093/ajcn/69.1.115 https://doi.org/.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R - INLA. J. Stat. Softw. (19), 63. Retrieved from. http://www.jstatsoft.org/v63/i19/.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. Royal Stat. Soc. Series B 73 (4), 423–498.

Macharia, A.W., Mochamah, G., Uyoga, S., Ndila, C.M., Nyutu, G., Makale, J., Williams, T.N., 2018. The clinical epidemiology of sickle cell anemia in Africa. Am. J. Hematol. 93 (3), 363–370. https://doi.org/10.1002/ajh.24986 https://doi.org/.

Macintyre, S., Macdonald, L., Ellaway, A., 2008. Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. Soc. Sci. Med. 67 (6), 900–914. https://doi.org/10.1016/j.socscimed.2008.05.029 https://doi.org/.

Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2013. Bayesian computing with INLA : new features. Comput. Stat. Data Analysis 67, 68–83. https://doi.org/10.1016/j.csda.2013.04.014 https://doi.org/.

Moraga, P., 2019. Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny. Chapman and Hall/CRC.

Moschovis, P.P., Wiens, M.O., Arlington, L., Antsygina, O., Hayden, D., Dzik, W., Hibberd, P.L., 2018. Individual, maternal and household risk factors for anaemia among young children in Sub-Saharan Africa: a cross-sectional study. BMJ Open 8 (5), 1–14. https://doi.org/10.1136/bmjopen-2017-019654 https://doi.org/.

Muriuki, J.M., Mentzer, A.J., Webb, E.L., Morovat, A., Kimita, W., Ndungu, F.M., Atkinson, S.H., 2020. Estimating the burden of iron deficiency among African children. BMC Med 18 (1), 1–14. https://doi.org/10.1186/s12916-020-1502-7 https://doi.org/.

Mwangi, M.N., Phiri, K.S., Abkari, A., Gbané, M., Bourdet-Sicard, R., Braesco, V.A., Prentice, A.M., 2017. Iron for Africa-report of an expert workshop. Nutrients 9 (6), 1–15. https://doi.org/10.3390/nu9060576 https://doi.org/.

Nankinga, O., Aguta, D., 2019. Determinants of Anemia among women in Uganda: further analysis of the Uganda demographic and health surveys. BMC Public Health 19 (1), 1–9. https://doi.org/10.1186/s12889-019-8114-1 https://doi.org/.

R Core Team. (2020). *R: a language and environment for statistical computing*. Retrieved from https://www.r-project.org/.

Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., Lindgren, F.K., 2017. Bayesian computing with INLA : a Review. Ann. Rev. Stat. Appl. 4, 395–421. Retrieved from. http://arxiv.org/abs/1604.00860.

Riebler, A., Sørbye, S.H., Simpson, D., Rue, H., Lawson, A.B., Lee, D., MacNab, Y., 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. Stat. Methods Med. Res. 25 (4), 1145–1165. https://doi.org/10.1177/0962280216660421 https://doi.org/.

Roberts, D.J., Matthews, G., Snow, R.W., Zewotir, T., Sartorius, B., 2020. Investigating the spatial variation and risk factors of childhood anaemia in four Sub-Saharan African countries. BMC Public Health 20 (1), 1–10. https://doi.org/10.1186/s12889-020-8189-8 https://doi.org/.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. Royal Stat. Soc. Series B 71 (2), 319–392. https://doi.org/10.1111/j.1467-9868.2008.00700.x https://doi.org/.

Stanković, A., Nikić, D., Nikolić, M., 2006. Relationship between exposure to air pollution and occurrence of anemia in pregnancy. Med. Biol. 13 (1), 54–57. Retrieved from. http://facta.junis.ni.ac.rs/mab/mab200601/mab200601-11.pdf.

Stevens, G.A., Finucane, M.M., De-Regil, L.M., Paciorek, C.J., Flaxman, S.R., Branca, F., Ezzati, M., 2013. Global, regional, and national trends in haemoglobin concentration and prevalence of total and severe anaemia in children and pregnant and non-pregnant women for 1995-2011: a systematic analysis of population-representative data. The Lancet Global Health 1 (1), 16–25. https://doi.org/10.1016/S2214-109X(13)70001-9 https://doi.org/.

Subramanian, S.V., Jones, K., Duncan, C., 2003. Multilevel methods for public health research. In: Kawachi, I., Berkman, L.F. (Eds.), Neighborhoods and Health, pp. 583–605. https://doi.org/10.1093/acprof https://doi.org/.

Subramanian, S.V., Kim, D.J., Kawachi, I., 2002. Social trust and self-rated health in US communities: a multilevel analysis. J. Urban Health 79 (4 SUPPL. 1), S21–S34. https://doi.org/10.1093/jurban/79.suppl_1.s21 https://doi.org/.

Subramanian, S.V., 2010. Multilevel modeling. In: Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, pp. 507–525. https://doi.org/10.1007/978-3-642-03647-7 https://doi.org/.

Tennekes, M., 2018. tmap : thematic maps in R. J Stat Softw 84 (6), 1–39. https://doi.org/10.18637/jss.v084.i06 https://doi.org/.

Teshale, A.B., Tesema, G.A., Worku, M.G., Yeshaw, Y., Tessema, Z.T., 2020. Anemia and its associated factors among women of reproductive age in eastern Africa: a multilevel mixed-effects generalized linear model. PLoS ONE 15 (9 September), 1–16. https://doi.org/10.1371/journal.pone.0238957 https://doi.org/.

Tobler, W.R., 1970. A computer movie simulation urban growth in Detroit region. Econ Geogr 46, 234–240.

Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C.,… Winker, D.M. (2016). Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors. https://doi.org/ 10.1021/acs.est.5b05833.

Van Donkelaar, A., Martin, R.V., Brauer, M., Hsu, N.C., Kahn, R.A., Levy, R.C.,… Winker, D.M. (2018). *Global annual PM2.5 grids from MODIS, MISR and SeaWiFS aerosol optical depth (AOD) with GWR, 1998-2016*. Retrieved from https://doi.org/ 10.7927/H4ZK5DQS.

Wang, X., Yue, Y.R., Faraway, J.J., 2018. Bayesian Regression Modeling With INLA. CRC Press - Taylor and Francis Group, New York.

White, N.J., 2018. Anaemia and malaria 11 medical and health sciences 1108 medical microbiology 11 medical and health sciences 1103 clinical sciences. Malar. J. 17 (1), 1–17. https://doi.org/10.1186/s12936-018-2509-9 https://doi.org/.

WHO, 2011. Haemoglobin Concentrations For the Diagnosis of Anaemia and Assessment of Severity. World Health Organization, Geneva, Switzerland https://doi.org/2011.

WHO. (2015). *The Global Prevalence of Anaemia in 2011*. Geneva, Switzerland.