BMJ Open

High-resolution mapping of essential maternal and child health service coverage in Nigeria: a machine learning approach

Yoshito Kawakatsu [1], Jonathan F Mosser [2], Christopher Adolph,[3] Peter Baffoe,[4] Fatima Cheshi,[4] Hirotsugu Aiga,[5,6] D A Watkins [7], Kenneth H Sherr[1]

To cite: Kawakatsu Y, Mosser JF, Adolph C, *et al*. High-resolution mapping of essential maternal and child health service coverage in Nigeria: a machine learning approach. *BMJ Open* 2024;**14**:e080135. doi:10.1136/bmjopen-2023-080135

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-080135).

Received 21 September 2023
Accepted 12 May 2024

Check for updates

For numbered affiliations see end of article.

Correspondence to
Yoshito Kawakatsu;
y.kawakatsu.0829@gmail.com

ABSTRACT

Background National-level coverage estimates of maternal and child health (MCH) services mask district-level and community-level geographical inequities. The purpose of this study is to estimate grid-level coverage of essential MCH services in Nigeria using machine learning techniques.

Methods Essential MCH services in this study included antenatal care, facility-based delivery, childhood vaccinations and treatments of childhood illnesses. We estimated generalised additive models (GAMs) and gradient boosting regressions (GB) for each essential MCH service using data from five national representative cross-sectional surveys in Nigeria from 2003 to 2018 and geospatial socioeconomic, environmental and physical characteristics. Using the best-performed model for each service, we map predicted coverage at 1 km^2 and 5 km^2 spatial resolutions in urban and rural areas, respectively.

Results GAMs consistently outperformed GB models across a range of essential MCH services, demonstrating low systematic prediction errors. High-resolution maps revealed stark geographic disparities in MCH service coverage, especially between rural and urban areas and among different states and service types. Temporal trends indicated an overall increase in MCH service coverage from 2003 to 2018, although with variations by service type and location. Priority areas with lower coverage of both maternal and vaccination services were identified, mostly located in the northern parts of Nigeria.

Conclusion High-resolution spatial estimates can guide geographic prioritisation and help develop better strategies for implementation plans, allowing limited resources to be targeted to areas with lower coverage of essential MCH services.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This study used representative data from five national household surveys conducted in Nigeria from 2003 to 2018, including over 150 000 households.
- ⇒ Health service coverage of 10 essential maternal and child health services were estimated under the best-performing model with a wide range of geospatial data.
- ⇒ We compared gradient boosting models and generalised additive models for predicting health service coverage by assessing four performance criteria: weighted root-mean-squared error, mean bias error, mean squared error and nominal coverage of 95% prediction intervals.
- ⇒ Our modelling approach can be applicable to other countries with similar household survey datasets and health services which are associated with geospatial data.
- ⇒ However, our approach is less applicable for health services like treatments of childhood illnesses, which are more associated with individual and contextual factors.

BACKGROUND

National and subnational coverage of essential maternal and child health (MCH) services in low/middle-income countries (LMICs) are typically estimated using nationally representative cross-sectional surveys, including Demographic and Health Surveys (DHS) and Multiple Indicator Cluster Surveys (MICS). Because these surveys are generally designed to produce estimates at the national and subnational levels (eg, administrative level 1) exclusively, these estimates can mask district-level and community-level inequities in service coverage. As a result, it can be difficult for policymakers to identify underserved geographic locations and prioritise interventions that match each community's context and needs.

Although estimates of essential MCH service coverage are rarely available at the lowest administrative level in LMICs, several earlier studies reported service coverage of diphtheria-tetanus-pertussis (DTP) vaccination, measles vaccinations and contraceptive use at subdistrict or community levels in LMICs.[1–4] Prevalence rates of HIV[5] and

Plasmodium falciparum[6] were also estimated at the subnational level in some LMICs. Finally, the DHS Programme Map Surfaces provides grid-level estimates of coverage for multiple health service estimated using a Bayesian geostatistical approach with a standardised set of covariates across countries.[7 8] However, these studies focus only on a single survey or a single indicator, whereas studies combining data from multiple surveys to investigate multiple indicators over time are lacking. The DHS Programme Map Surfaces reports recommend the use of additional geospatial covariates and integrated data from multiple surveys to further improve their methodologies and monitor progress towards universal health coverage (UHC) at subnational level.[7]

Nigeria has the highest population and GDP in Africa. However, a high burden of maternal and child mortality in Nigeria significantly contributes to global maternal and child deaths.[9] Previous surveys and studies identified noticeable disparities in health outcomes and disease burden across regions in Nigeria.[10 11] Also, its diverse sociocultural and economic characteristics influences health-seeking behaviours and service utilisation differently across states and lower administrative areas,[2 12–14] necessitating a granular analysis.

There is a critical need for district-level and community-level data on coverage of essential MCH services, so that countries may effectively target interventions, including new health facilities and outreach services, to communities with low coverage of multiple MCH services. Therefore, this study is aimed at estimating grid-level coverage of essential MCH services (ie, antenatal care (ANC), facility-based deliveries, childhood vaccinations and treatments of childhood illnesses) in Nigeria using machine learning methods. We further compare the prediction precision of two machine learning techniques, generalised additive models (GAMs) and gradient boosting (GB). We then use these methods to create high-resolution maps to visualise inequities in MCH service coverage across specific locations in Nigeria.

METHODS

Survey data in Nigeria

This study uses data from five nationally representative cross-sectional surveys administered in Nigeria between 2003 and 2018 that include geolocation data of primary sampling units (PSUs): DHS 2003, DHS 2008, DHS 2013, DHS 2018 and MICS 2016–2017. DHS data were extracted from IPUMS DHS.[15] Methodological details on the conduct of these five surveys is published elsewhere.[10 16–19] All five surveys employed stratified two-stage or three-stage cluster sampling. Each PSU for DHS 2003 was composed of one or more enumeration areas (EAs) developed for the 1991 Population Census. PSUs for other DHS and MICS from 2016 to 2017 consisted of one or more EAs developed for the Population and Household Census 2006. MICS 2016–2017 excluded 101 PSUs located in Borno, Yobe and Adamawa States due to insecurity in those regions. For the same reason, DHS 2018 excluded 11 of 27 local government areas (LGAs) in Borno State.

To protect confidentiality of personal geoinformation, the public versions of all five surveys' data randomly displaced Global Positioning System coordinates of the locations of respondents' PSUs. Geolocations of urban PSUs were randomly displaced within 2 km buffers, whereas rural PSUs were displaced within either 5 km buffers in 99% of cases or 10 km buffers in 1% of cases. The direction and distance of displacement for each household was randomly determined.[20 21] A previous study reported that errors in direction and distance due to random household displacement were sufficiently negligible to estimate measles vaccination coverage in a 10 km^2 grid.[22] Geolocation data for 16 of 3533 PSUs (0.5%) were missing across the four DHSs. Similarly, geolocation data for 1 of 2239 PSUs (0.0004%) was missing in MICS 2016/2017. After initial random displacement, 14 PSUs (1 in DHS 2008 and 13 in MICS 2016/2017) were 'located' either in the sea or outside country boundaries. We resampled those PSUs by conducting random displacement again within 5 km until they were within Nigeria's land boundaries. If there was no appropriate point within 5 km from the original PSUs' locations, we resampled those PSUs within 10 km. Of these 14 PSUs, 8 were successfully resampled, and 6 cases that could not be appropriately displaced across 10 000 attempts were discarded. Online supplemental file 1 shows the locations of PSUs with at least one target population for ANC and facility-based delivery.

Essential MCH services

Ten essential MCH services were included in this study: (1) ANC; (2) facility-based delivery; (3–8) childhood vaccinations (BCG, first and third DTP/pentavalent (Penta) and oral polio (OPV), and first dose of measles); and (9 and 10) treatments of childhood illnesses for fever/cough and diarrhoea. The definitions and study population of essential services are shown in table 1. The target population for ANC and facility-based delivery are women 15–49 years of age having given a live birth during the last 23 months. Children 12–23 months of age are the study population for childhood vaccinations.

The outcome variables for this study are the grid-level utilisation rates for each essential MCH service. First, we created a map with 1 km×1 km cells in urban areas and 5 km×5 km cells in rural areas. In this study, we employed a grid cell-based approach to classify urban and rural areas, adhering to the harmonised definition by Dijkstra and Poelman.[23] Urban areas were defined as grid cells with a population density of at least 300 inhabitants per km^2 and a total population exceeding 5000, ensuring a consistent classification across Nigeria. Grid cells not meeting these criteria were classified as rural. To accurately capture the heterogeneous nature of urban areas and manage computational efficiency across Nigeria's vast landscapes, we differentiated spatial resolutions: 1 km×1 km cells

Table 1 Definitions and target populations of essential health services

Health service	Definition	Target population
Four or more antenatal care visits	Antenatal care four or more times by trained health personnel (ie, doctor, nurse, midwife, auxiliary midwife) during pregnancy at the point of survey	Women aged 15–49 years with a last birth in the last 23 months
Facility-based delivery	Delivery at public or private health facility in the last 23 months at the point of survey	Women aged 15–49 years with live births in the last 23 months
Childhood immunisation	Children who had received one dose of BCG vaccine	Children aged 12–23 months
	Children who had received first and third dose of diphtheria-tetanus-pertussis vaccine or pentavalent vaccine	
	Children who had received first and third dose of oral polio vaccine	
	Children who had received first dose of measles vaccine	
Treatment for common childhood illness	Children under five with fever/cough and diarrhoea in the last 2 weeks for whom care was sought at a health facility	Children aged 0–59 months

were used for urban areas to detail the high variability in population distribution and access to services, whereas 5 km×5 km cells were applied in rural areas, balancing the need for detailed analysis and computational feasibility.

For each essential MCH service, survey year and grid cell, we constructed a utilisation rate using the count of children 12–23 months of age or women 15–49 years of age who used that service as the numerator, and the number of individuals categorised into the study population, children 12–23 months of age or women 15–49 years of age depending on the indicator, as the denominator.

Covariate data

Several earlier studies found geospatial socioeconomic, environmental and physical factors to be associated with the spatial distribution of under-five mortality cases and geographical heterogeneity in MCH service coverage.[2 24 25] Table 2 shows the geospatial covariates employed in this study, except for longitude and latitude. Household poverty and female and male education attainment by grid cell were extracted from high-resolution data published in earlier studies.[26 27] Population density and total population data were extracted from the WorldPop database.[28] Road density data were extracted from the Global Roads Inventory Project.[29] Meteorological data such as precipitation and evapotranspiration and nighttime light data were extracted from TerraClimate[30] and High Resolution Electricity Access.[31] We also created three spatial covariates: (1) travel time to the most accessible health facility; (2) travel time to the nearest city; and (3) the number of major roads crossing the household's grid cell. Travel time to the most accessible health facility was estimated by using the geolocations of health facilities run by state governments, community-based organisations and faith-based organisations in Nigeria[32] with the friction surface developed by Malaria Atlas Project.[33] Similarly, travel time to the nearest city was estimated by using geolocational data of cities, extracted from OpenStreetMap and the same friction surface.[34] We counted the number of major roads, such as primary, secondary and tertiary roads, motorway, and track grade 1, 2 and 3, in each cell using OpenStreetMap. In addition, we used high-resolution spatial data on the prevalence of *P. falciparum* malaria, lower respiratory infections and childhood diarrhoeal morbidity[6 35 36] as disease-related covariates.

Prediction modelling and high-resolution map creation

We compared the prediction performances of two models, GAMs and GB models, using a pooled dataset of five nationally representative cross-sectional surveys. These models are described in the Prediction models section. For each service, we identify the best prediction model based on each model's performance on four indicators described in the section of Model performance indicators and selection of final models. Finally, we create high-resolution coverage maps for each essential MCH service using the best prediction model for that service. We also estimated the aggregated service coverage at various administrative levels and in rural and urban areas. The High-resolution map creation section provides methods of creating high-resolution maps and estimating aggregated service coverage.

Prediction models: GAMs and GB models

We consider two modelling strategies from machine learning—GAMs and GB models—for predicting essential MCH service coverage. Both models rely on the same set of covariates but employ different estimation techniques: GAMs emphasise non-parametric estimation of flexible functional forms, whereas GB involves prediction from a weighed ensemble of sequentially constructed models, each of which attempts to better predict cases

Table 2 Description of spatial covariates for modelling in this study

Category	Year*	Data type	Data description	Grid size	Data source
Poverty	2010	Continuous	Estimates of proportion of people per grid square living in poverty, as defined by \$1.25 a day	1 km^2	https://www.worldpop.org/geodata/listing?id=23
Female education	2003–2017	Continuous	Pixel-level estimates of mean educational attainment among female 15–49	5 km^2	https://cloud.ihme.washington.edu/index.php/s/CTnfWYaZxc7ZENc?path=2FData205BGeoTIFF5D
Male education			Pixel-level estimates of mean educational attainment among male 15–49		
Population density	2003–2018	Continuous	Estimated population density per grid-cell (1 km)	1 km^2	https://www.worldpop.org/geodata/listing?id=77
Total population	2003–2018	Continuous	Estimated total population per grid-cell (1 km)	1 km^2	https://www.worldpop.org/geodata/listing?id=75
Time to health facility	2015	Continuous	Time to the most accessible health facility using the health facility locations and the friction surface	5 km^2 for rural and 1 km^2 for urban	NA
Time to city	2015	Continuous	Time to the most accessible city using the city locations and the friction surface		NA
Number of main roads crossing a grid	2021	Continuous	Number of main roads crossing a grid using OpenStreetMap data		OpenStreetMap; http://download.geofabrik.de/africa.html
Road density	2018	Continuous	Total road density	8 km^2	https://www.globio.info/download-grip-dataset
Precipitation	2003–2018	Continuous	High-spatial resolution (1/24°, ~4 km) map of cumulative precipitation	4 km^2	http://thredds.northwestknowledge.net:8080/thredds/catalog/TERRACLIMATE_ALL/data/catalog.html
Evapotranspiration	2003–2018	Continuous	High-spatial resolution (1/24°, ~4 km) map of cumulative evapotranspiration		
Nighttime lights	2012–2018	Continuous	Nighttime light annual composite	450 m^2	http://www-personal.umich.edu/~brianmin/HREA/data.html
Malaria prevalence	2003–2018	Continuous	Pixel-level Age-standardised parasite prevalence rate for *Plasmodium falciparum* malaria for children 2–10 years of age	5 km^2	https://cloud.ihme.washington.edu/s/teDKnPGcJnBjJ5F?path=2F520-20Prevalence3A20Plasmodium20falciparum205BGeoTIFF5D2FRate
Lower respiratory infection	2003–2017	Continuous	Pixel-level prevalence of lower respiratory infections among children under 5	5 km^2	http://ghdx.healthdata.org/record/ihme-data/africa-under-5-lri-incidence-prevalence-mortality-geospatial-estimates-2000-2017
Childhood diarrhoea	2003–2017	Continuous	Pixel-level estimates of under-5 diarrhoea prevalence rate	5 km^2	https://cloud.ihme.washington.edu/s/fGpEZwJetEJjdxG

*Data availability between 2003 and 2018.

that were poorly predicted by the previous model in the sequence.

Generalised additive models

We estimated grid-level quasi-binomial GAMs for each essential MCH service using the mgcv package in R.[37] The functional form of the models allows for non-linear relationships between the outcome and covariates, and is given by:

$$\begin{aligned}\log\left(\frac{y_{ijt}}{\mathrm{Pop}_{ijt}}\right) = {} & \beta_0 + \beta_1 \mathrm{Urban}_{ij} + \mathrm{State}_{ij}\gamma + s\left(\mathrm{year}_t\right) + s\left(\mathrm{long}_{ij}, lat_{ij}\right) + s\left(\mathrm{Poverty}_{ij}\right) + \\ & s\left(\mathrm{Female\_edu}_{ijt}\right) + s\left(\mathrm{Male\_edu}_{ijt}\right) + s\left(\mathrm{Pop\_den}_{ijt}\right) + s\left(\mathrm{Pop\_total}_{ijt}\right) + \\ & s\left(\mathrm{No\_road}_{ij}\right) + s\left(\mathrm{Road\_den}_{ij}\right) + s\left(\mathrm{Time\_city}_{ij}\right) + s\left(\mathrm{Time\_HF}_{ij}\right) + \\ & s\left(\mathrm{Rain}_{ijt}\right) + s\left(\mathrm{Dry}_{ijt}\right) + s\left(\mathrm{Light}_{ijt}\right) + \left(\mathrm{Malaria}_{ijt}\right) + s\left(\mathrm{LRI}_{ijt}\right) + \\ & s\left(\mathrm{Diarrhoea}_{ijt}\right)\end{aligned}$$

y_{ijt} represents the number of eligible individuals using essential MCH services in the *i*th grid of the *j*th state in year *t*. Pop_{ijt} indicates the number of eligible individuals in the *i*th grid of the *j*th state in year *t*. Urban_{ij} is a binary indicator for whether the *i*th grid of the *j*th state is urban or rural. State_{ij} is a vector of dummy variables indicating states of Nigeria. *s*(year) is a two-dimensional thin plate regression spline of years. *s*(lat, long) is an isotropic smooth of latitude and longitude on the sphere with second derivative penalty.[37] In all other cases, $s(\cdot)$ indicates a cubic spline smoothing function of the given covariate.

$\mathrm{Poverty}_{ij}$ denotes the estimated proportion of people living in poverty, defined as living on less than US\$1.25 a day, in the *i*th cluster of the *j*th state. $\mathrm{Female\_edu}_{ijt}$ and $\mathrm{Male\_edu}_{ijt}$ represent mean educational attainment among females and males aged 15–49 years in the *i*th cluster of the *j*th state in year *t*, respectively. $\mathrm{Pop\_den}_{ijt}$ and $\mathrm{Pop\_total}_{ijt}$ represent population density and total population in the *i*th cluster of the *j*th state in year *t*. $\mathrm{No\_road}_{ij}$ and $\mathrm{Road\_den}_{ij}$ denote the number of main roads crossing and total road density in the *i*th cluster of the *j*th state. $\mathrm{Time\_city}_{ij}$ and $\mathrm{Time\_HF}_{ij}$ denote time to the most accessible health facility and city in the *i*th cluster of the *j*th state. Rain_{ijt}, Dry_{ijt} and Light_{ijt} represent cumulative precipitation, cumulative evapotranspiration and annual composite nighttime light, respectively, in the *i*th cluster of the *j*th state in year *t*. $\mathrm{Malaria}_{ijt}$, LRI_{ijt} and $\mathrm{Diarrhoea}_{ijt}$ represent prevalence of *P. falciparum* malaria, lower respiratory infections and diarrhoea in the *i*th cluster of the *j*th state in year *t*.

We employed a shrinkage approach to smoothing functions in the GAM model with all covariates, since the best-subset selection approach was not computationally feasible. A simulation study found that the shrinkage approach performed better in terms of predictive ability than other methods, such as backward selection.[38]

Gradient boosting

We also estimated GB regression models for each essential MCH service using the XGBoost package in R.[39] We used the same set of the covariates as listed in the Covariate data section, and assumed service coverage outcomes followed the Tweedie distribution. We tested a sequence of variance power between 1 and 2 by running iterative models and selected the best power based on the negative log-likelihood. Moreover, we conducted a grid search to select the best set of hyper-parameters based on negative log-likelihood for Tweedie regression. In the grid search, we tested the following parameters and their ranges: (1) the maximum depth of a tree, 2–10 by increments of 2; (2) the minimum sum of instance weight needed in a child, 1–3 by increments of 0.5; (3) the subsample ratio of columns when constructing each tree, 0.5–1.0 by increments of 0.1; (4) the subsample ratio of the training instances, 0.5–1.0 by increments of 0.1; and (5) step size shrinkage used in update to prevent overfitting, 0.01–0.3 (specifically, we tested 0.01–0.1 by increments of 0.01, as well as 0.2 and 0.3).

Model performance indicators and selection of final models

We compared the performance of the GAM and GB models for each MCH service on four performance criteria, each of which was computing using 5-fold cross-validation:

weighted root-mean-squared error (WRMSE) of proportions of essential MCH service utilisations,

$$\mathrm{WRMSE} = \sum\left(d_i \times \sqrt{\left(\left(P_i - \hat{P}_i\right)^2\right) / \sum d_i}\right);$$

mean bias error,

$$\mathrm{MBE} = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - \hat{P}_i\right);$$

mean squared error,

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(P_i - \hat{P}_i\right)^2; \text{ and}$$

nominal coverage of 95% prediction intervals,

$$\mathrm{nCoverage} = 100 \times \sum_{i=1}^{n} I\left(l_i \le \hat{y}_i\right) \le u_i\right) / n.$$

P_i and $\hat{P}_i$ denote observed and predicted proportions at cluster *i*, d_i is the size of the study population at cluster *i*, and *n* is the count of cluster locations with non-zero study populations. l_i *and* u_i are the lower and upper limits of the prediction interval and $I(\cdot)$ is an indicator function. Nominal coverage indicates the proportion of the estimated numbers of study population having used essential maternal and child services (y_i) within 95% CIs of predicted services coverage. Due to erratic behaviour around the endpoints (ie, 0 and 1) using binomial probabilities,[40] whenever observed values were exactly 0 or 1, we set the estimated lower and upper CIs in GAMs to either 0 or 1, respectively. We employed methods employed in an earlier study on cluster-level estimates of measles vaccinations[2] to calculate the nominal coverage of 95% prediction intervals.

To determine a final model for each MCH service, we first checked whether either the GAM or GB model was dominant on all the four measures of goodness of fit. If either model was uniformly superior on these four metrics, we selected it as the prediction model for the service. When the results of the four metrics disagreed on the best model, we further examined whether one of the models had WRMSE at least 0.01 units lower, MSE at least 0.005 units lower, MBE at least 0.01 units lower, or nCoverage at least 1% closer to 95%. Whichever model met a greater number of these criteria than the others was selected as the prediction model for that service. There was no case that both the GAM and GB model met the same number of these criteria.

High-resolution map creation

Using the best predicting model, either GAM or GB, for each service, we created high-resolution coverage maps for essential MCH services from 2003 to 2018. We also computed 95% CIs of predicted grid-level service coverage using the estimated SE for each predicted value for GAMs or bootstrapped SEs for GB models. In addition, to ensure the logical sequence of vaccine coverage across vaccination doses, we checked the number of grids that have a higher coverage for a subsequent dose (Penta3/OPV3) than that for an initial dose (Penta1/OPV1). We have not found such grids over years.

We estimated aggregate predicted proportions aggregated at various administrative levels by calculating weighted mean values and using as weights in each grid location the proxy target population from WorldPop.[28] For ANC and facility-based delivery, we used the sum of the number of children 0–12 months of age in the year of estimation and in the prior year as a proxy for the number of women 15–49 years of age with a live birth during the last 23 months. For childhood vaccinations, we used the total number of children 0–12 months of age in the prior year as a proxy for the number of children aged 12–23 months. To estimate gaps in MCH service coverage between rural and urban areas over time, the difference-in-difference (DiD) estimator for each essential MCH service was calculated by taking the difference in an MCH service coverage of urban areas between 2018 and 2003 and subtracting the difference for rural areas between 2018 and 2003. 95% CIs for the DiD estimates were computed using bootstrapped SEs. A bivariate choropleth map is used to display mean MCH coverage.

Following previous work,[35] we treated as missing values of MCH service coverage for the grid locations which include either lakes or very low population (less than 10 people per km^2). We did not estimate predicted values for grid locations lacking covariate data.

Patient and public involvement

Patients or the public were not involved in the design, conduct, reporting or dissemination plans of our research.

RESULTS

Model selection and validation

Online supplemental file 2 reports the performance of the GAMs and GB models. Across a range of essential MCH services, GAMs matched or outperformed GB models based on all four performance indicators, as shown in online supplemental file 2. In particular, for ANC, facility-based delivery, BCG, first and third Penta, GAMs performed better than GB models on all four measures: MBEs, MSEs, WRMSEs and nominal 95% coverage. For third OPV and measles vaccinations, GAMs outperformed GB models in terms of MBE and the nominal coverage, while obtaining similar MSEs and WRMSEs. For first OPV, GAMs outperformed GB models on MBE, while MSE, WRMSE and nominal coverage were similar across both approaches. Finally, due to low predictive power and wide prediction intervals in both GAM and GB models, we did not select final models or create coverage maps for treatments of childhood illnesses.

The final predictive models had uniformly low systematic prediction error, with a range of MBE from 0.001 to 0.016 across 10 essential MCH services. The range of MSEs was similarly narrow (0.057–0.068).

To check model fitting, we further compared the observed and estimated share of study populations using essential MCH services (see online supplemental file 3a). Also, online supplemental file 3b includes in-sample and out-of-sample plots for observed vs predicted probabilities for each essential MCH service. Online supplemental file 4 provides plots of smoothed functional forms from the GAMs. GB model estimates of the importance of each feature for predicting each MCH service based are provided in online supplemental file 5. As shown in online supplemental file 4, smoothers fitted by the GAMs show different slopes by MCH service types. Online supplemental file 6 compares high-resolution maps of coverage of six essential MCH services based on GAMs and GB models.

Geographic inequality in essential MCH service coverage

High-resolution maps of coverage from 2003 to 2018 of 10 essential MCH services, excepting treatments of childhood illnesses, have been stored in Figshare. High-resolution maps of 6 of 10 essential MCH services (ANC, facility-based delivery, BCG, first Penta, third Penta and measles vaccination) as of 2018 are shown in figure 1 to highlight geographic inequality in service coverage across states and LGAs. Overall, MCH service coverage in rural areas was lower than urban areas across service types. Likewise, MCH service coverage was also lower in northern parts of Nigeria, though higher coverage can be found in some major cities and towns in those states.

Figure 2 shows LGA-level coverage estimates for 8 of 10 essential MCH services by state as of 2018. Online supplemental file 7 also provides mean, minimum and maximum LGA-level coverage of eight essential MCH services by state. Overall, the LGA-level inequalities of third OPV, third Penta and measles were narrower, compared with BCG, first OPV and first Penta. States with either high or low service coverage exhibited reduced IQRs at the LGA level, suggesting diminished intra-state disparities. Conversely, states with

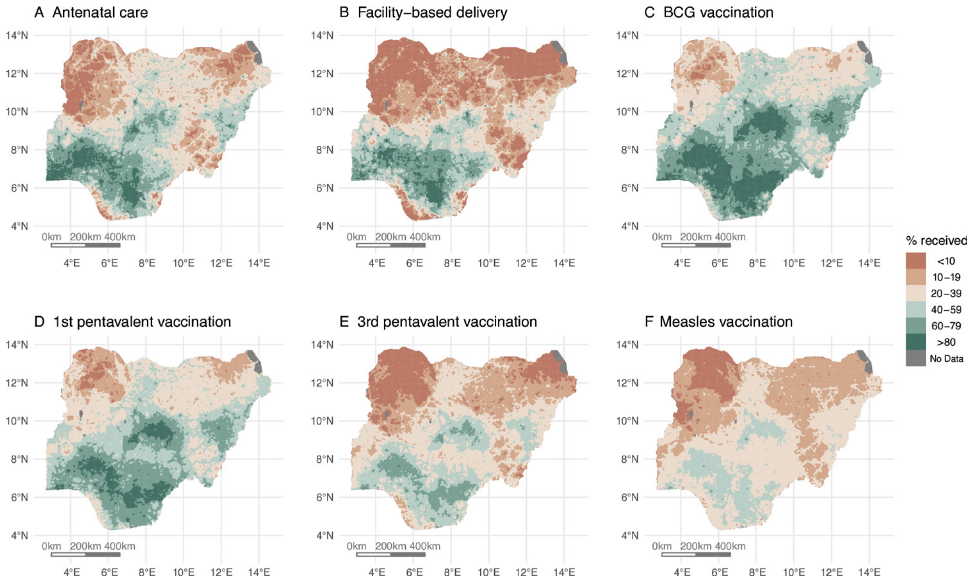

Figure 1 Predicted service coverage of (A) antenatal care, (B) facility-based delivery, (C) BCG vaccination, (D), first pentavalent vaccination, (E) third pentavalent vaccination and (F) measles vaccination in 2018.

moderate coverage revealed more substantial IQRs, indicative of increased variability in service utilisation across LGAs. The three states with the greatest LGA-level inequality in terms of ANC were Kwara (IQR: 54.2%–82.6%), Borno (IQR: 16.6%–40.8%) and Bayelsa (IQR: 23.4%–46.3%), whereas the top three states for LGA-level inequality in BCG vaccination were Kaduna (IQR: 53.4%–88.1%), Gombe (IQR: 45.3%–78.4%) and Borno (IQR: 35.7%–60.4%). Overall, we find that Borno, Gombe, Kaduna, Kwara and Niger states showed consistently high inequality across LGAs, with each of these states ranking among the top three states in terms of IQR for at least four of eight MCH services. In terms of inequality across the 10 MCH services shown in figure 2, we find the highest LGA-level IQR for facility-based delivery coverage, followed by BCG, first OPV and ANC. The LGA-level IQR for measles vaccination was the lowest of the eight services, although its mean coverage was the lowest, as well.

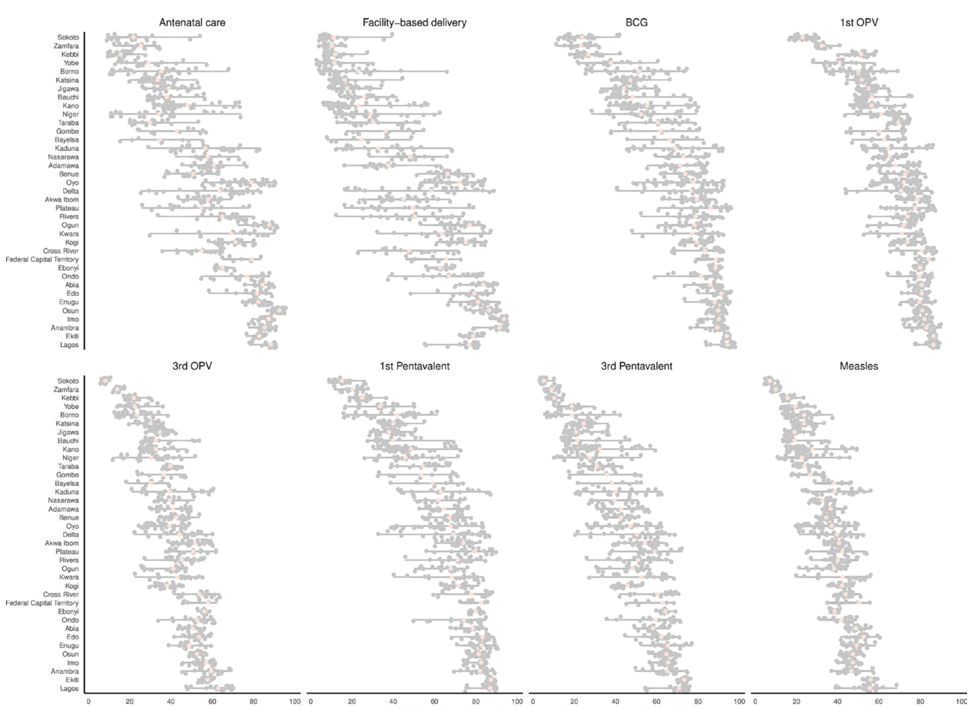

Figure 2 Predicted local government area (LGA)-level coverage of eight essential maternal and child health services in 2018 by state. Orange dots mark the mean coverage within each state. Grey dots show service coverage for each LGA within the state. The distribution of these grey dots is summarised by a violin plot.

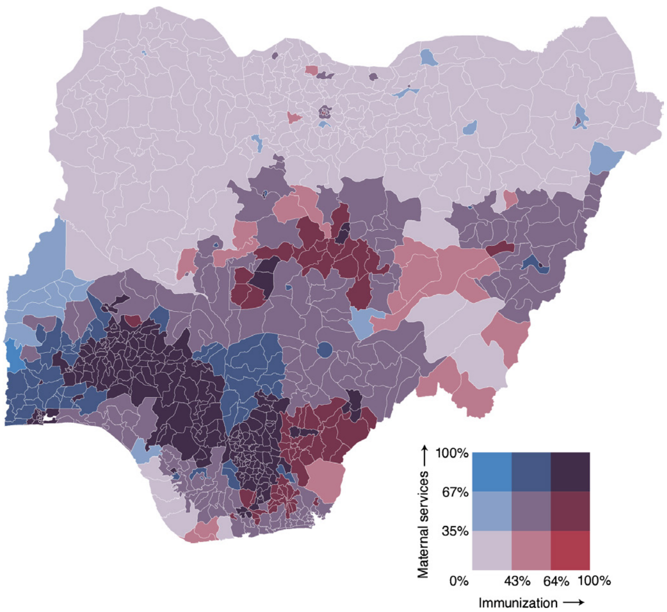

Figure 3 Service coverage distribution of essential maternal health services (antenatal care and facility-based delivery) and immunisation services in 2018 as a bivariate choropleth map. Mixtures of the colours red and blue indicate coverage levels in each local government area. Areas with high levels of maternal service coverage but low levels of immunisation coverage are shown in red. Areas with high levels of immunisation coverage, but low levels of maternal service coverage are shown in blue. Areas with similar levels of coverage across both services are shown in shades of purple ranging from lilac (low coverage of both services) to plum (high coverage of both services).

Trends in essential MCH service coverage

The mean LGA-level service coverage for essential MCH services increased during the period from 2003 to 2018. The annual mean increase in service coverage was higher for third Penta (1.4%), first Penta (1.3%), BCG (1.3%), and third OPV (1.2%), and lower for measles vaccination (0.7%), first OPV (0.7%), facility-based delivery (0.6%) and ANC (0.5%). However, trends in LGA-level essential MCH service coverage between 2003 and 2018 varied by type of service and by state. Although most LGAs are trending towards better essential MCH service coverage, some are stagnating or even decreasing over time.

Trends in mean MCH coverage for urban and rural LGAs are shown in online supplemental file 8. On average, service coverage in both urban and rural areas increased over time. DiD estimates of the change in the urban-rural gap find no significant evidence that this gap has changed over time for any service. Specifically, between 2003 and 2018, the gap between urban and rural areas changed by −2.3% (95% CI −14.3% to 10.6%) for ANC, −3.8% (−16.2% to 8.6%) for facility-based delivery, 7.6% (95% CI −22.9% to 7.0%) for BCG, −6.4% (−19.3% to 6.3%) for first pentavalent, −2.6% (−17.2% to 11.9%) for first OPV and −1.8% (−13.5% to 9.5%) for measles vaccination. In no case was the DiD statistically significant.

Priority LGAs with lower coverage of essential MCH services

Figure 3 is an LGA-level bivariate choropleth map that two-dimensionally shows service coverage of childhood vaccinations (BCG, first and third DTP/pentavalent (Penta) and OPV, and first dose of measles) and maternal health (ANC and facility-based delivery). Coverage level was classified by tertile (ie, higher 33%, middle 33% and lower 33%). The map indicates that a majority of LGAs located in northern part of the country had lower childhood vaccination and maternal health service coverage, excepting some urban LGAs which had higher coverage for either or both types of services. The southwest, including districts in Ogun and Oyo states, had higher maternal health service coverage and lower childhood vaccination coverage. In contrast, the southeastern and central areas of the country had lower maternal service coverage and higher childhood vaccination coverage. Finally, those states in which 80% of LGAs were categorised as high coverage for both types of services. The dark purple areas in figure 3, comprising Abia, Anambra, Edo, Ekiti, Enugu, Imo, Lagos and Osun, can be found in the southern part of Nigeria and had relatively greater population sizes and density than other areas. Online supplemental file 9 provides LGA-level mean coverage of childhood vaccinations and maternal health services as of 2018, by state.

DISCUSSION

The global movement for UHC highlights the importance of identifying subnational inequities in health service coverage in order to better understand the characteristics of underserved populations and thereby correct maldistribution of health resources and services. This movement has led researchers to conduct studies on the heterogeneity of health service coverage within and between subnational levels and to identify the most vulnerable areas for priority setting. Responding to this need, this study estimated grid-level coverage of essential MCH services in Nigeria using publicly available geographic datasets and comparing the performance of two statistical approaches (GAMs and GB).

Cross-validation performance confirms the strong predictive power of the final models of essential service coverage, with the exception of treatments for childhood illnesses. MSE and MBE values of the final model for measles vaccination were slightly better than those reported in an earlier study on grid-level coverage estimates of measles vaccination that employed Bayesian multivariate spatial-temporal modelling.[2] GAM with a shrinkage approach might be applied to a multicounty or global model, to simultaneously account for differing drivers and non-linear relationships. GAM estimation is also less computationally intensive than GB models, particularly to find the best set of hyper-parameters.

The importance of each feature for predicting MCH service coverage, shown in online supplemental file 5, varied by MCH service types. This suggests that further study may need to select an appropriate set of geospatial covariates for predicting coverage of specific health services. Additionally, the results for some predictors should be interpreted with caution. For example, education attainment is a model-based estimate generated using geospatial covariates such as night light, access to roads, population and aridity. The consistent importance of male and female education attainment in predicting coverage in GB models might therefore reflect the influence of the other geospatial covariates that were used to generate education status estimates, rather than educational status itself.

One avenue for increasing predictive power is to use health administrative data on essential MCH service utilisations reported from health facilities. Almost all countries have health management information systems that store time-series data on the delivery of key health services from public and private health facilities. For example, Nigeria launched the web-based software District Health Information System version 2 (DHIS2) in 2010. Childhood vaccination data have been collected on a monthly basis since 2014.[41] National-level health administrative data were used for estimating national coverage of childhood vaccination services by the WHO and UNICEF[42] and by a recent publication in the *Lancet*.[43] In addition, DHIS2 contains the aggregated count of clients/patients who receive respective health services at the lowest administrative level every month. Using these lower-level health administrative data may increase predictive power. However, the quality of health administrative data from health facilities is questionable in many countries at present.[44]

This study clarifies that national and subnational MCH service coverage measures mask significant spatial heterogeneity within Nigeria, which may result in poor decision-making on geographical prioritisation at the LGA and local levels. Likewise, inequities in MCH service coverage between LGAs varied by state and type of service. However, the distribution and use of health resources in Nigeria remain suboptimal in terms of demography, disease burden and pre-existing endowments of health resources.[45 46] The high-resolution maps generated by this study could be an input into the decision-making process for better geographical prioritisation of health resources. Maps could help identify LGAs that may be significantly lagging behind in multiple indicators, thus requiring broad-scale interventions. In addition, the granular-level estimate may illuminate pockets of underservice within LGAs, thereby offer a more nuanced understanding that would otherwise be masked in broader LGA-level data. To operationalise the use of our high-resolution maps in decision-making, a collaborative framework with government health agencies is essential. We recommend the development of an interactive, user-friendly interface for real-time exploration of data, coupled with regular updates to maintain relevance. Training sessions for local health officials can further facilitate the accurate interpretation and application of these maps. Pilot studies may serve as an initial platform to assess the utility and effectiveness of this approach in guiding resource allocation and healthcare planning.

Whereas prioritising based on a single indicator may make sense for achieving specific goals, such as the Immunization Agenda 2030, broader agendas, such as the Sustainable Development Goals, require prioritisation based on multiple indicators. Accordingly, we recommend that criteria for prioritising geographic locations should not rely on a single indicator of coverage, but rather consider coverage rates for multiple health services in order to ensure optimal provision. In particular, the study highlights within-locality differences in coverage of childhood vaccinations and maternal health services, with some localities performing markedly better in one area or the other.

According to the Lancet Commissions,[46] Nigeria needs to distribute available health resources more equitably through increasing resource management and strategic purchasing capacities. Routinely updated high-resolution maps could also support micro-planning of supplementary activities such as outreach services and childhood vaccination, especially if such mapping is connected with a health information management system such as DHIS2. Specifically, mapping can help estimate the number of individuals in need of services at the community level and thereby inform budgeting and resource allocation.

There are several limitations to this study. Our modelling approach is less applicable to estimating the coverage of health services like treatments of childhood illnesses, which are more associated with individual and contextual characteristics than spatial covariates.[47–49] Moreover, several other factors likely contribute to the models' weaker performance in this area. For instance, treatment-seeking behaviours for childhood illnesses can be highly variable, influenced by cultural and religious beliefs, gender dynamics and financial constraints. Additionally, health awareness and severity and frequency of disease can vary substantially, making it challenging to predict coverage effectively. These complexities are not easily captured by spatial covariates alone, resulting in reduced model performance for this service. Inequalities between relatively affluent urban areas and their adjacent peri-urban areas suffering higher poverty rates were not captured in the maps we created due to limitations of the available input data. Moreover, our study investigated inequality not by income, educational status, and other factors, but by geospatial conditions only. Due to the data availability issue, we used some temporal mismatched data in our modelling. While the temporal mismatch is not ideal, the alternative of excluding these variables could result in a less robust model, losing potentially significant predictors.

CONCLUSION

This study visualised significant geographical inequities of essential MCH service utilisations in Nigeria.

The high-resolution maps herein provide health policymakers/planners with guidance for geographic prioritisation of specific MCH services. These estimates serve as a resource to further develop implementation strategies for maximising limited resources. Strengthening routine MCH service delivery and its supplementary activities should be implemented in the priority areas with low coverage of essential MCH services.

Author affiliations

[1]Department of Global Health, University of Washington, Seattle, Washington, USA
[2]Health Metrics Sciences, University of Washington, Seattle, Washington, USA
[3]Department of Political Science, University of Washington, Seattle, Washington, USA
[4]UNICEF, New York, New York, USA
[5]School of Tropical Medicine and Global Health, Nagasaki University, Nagasaki, Japan
[6]Department of Global Health, George Washington University School of Public Health and Health Services, Washington, DC, USA
[7]Department of Medicine, University of Washington, Seattle, Seattle, Washington, USA

Acknowledgements We would like to acknowledge ICF Macro, Calverton USA, who collected the original survey data used in this study, and DHS Archive and IPUMS-DHS, where the collected data were stored. Map data is copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org.

Contributors As a guarantor, YK performed and interpreted data analyses and visualisations, and wrote the manuscript. JFM, CA, DAW and KHS contributed to the study concept and design. JFM and CA also verified data analysis and visualisation. PB and FC provided technical insights on data interpretation. JFM, CA and HA contributed to the critical revision of the manuscript for intellectual content. KHS supervised the process of this manuscript. All authors read and approved the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available in DHS and MICS website.

ORCID iDs

Yoshito Kawakatsu http://orcid.org/0000-0003-1853-533X
Jonathan F Mosser http://orcid.org/0000-0001-9093-3186
D A Watkins http://orcid.org/0000-0001-6341-9595

REFERENCES

1. Mosser JF, Gagne-Maynard W, Rao PC, *et al*. Mapping diphtheria-pertussis-tetanus vaccine coverage in Africa, 2000–2016: a spatial and temporal modelling study. *Lancet* 2019;393:1843–55.
2. Utazi CE, Thorley J, Alegana VA, *et al*. High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine* 2018;36:1583–91.
3. Utazi CE, Thorley J, Alegana VA, *et al*. A spatial regression model for the disaggregation of areal unit based data to high-resolution Grids with application to vaccination coverage mapping. *Stat Methods Med Res* 2019;28:3226–41.
4. Bosco C, Alegana V, Bird T, *et al*. Exploring the high-resolution mapping of gender-disaggregated development indicators. *J R Soc Interface* 2017;14:20160825.
5. Dwyer-Lindgren L, Cork MA, Sligar A, *et al*. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature* 2019;570:189–93.
6. Weiss DJ, Lucas TCD, Nguyen M, *et al*. Mapping the global prevalence, incidence, and mortality of Plasmodium Falciparum, 2000-17: a spatial and temporal modelling study. *Lancet* 2019;394:322–31.
7. Gething P, Tatem A, Bird T. Burgert-Brucker CR: creating spatial interpolation surfaces with DHS data. In: *DHS spatial analysis reports No 11*. Rockville, Maryland, USA: ICF International, 2015.
8. Burgert-Brucker CR, Dontamsetti T, Mashall A. Guidance for use of the DHS program modeled map surfaces. In: *DHS spatial analysis reports No 14*. Rockville, Maryland, USA: ICF International, 2016.
9. Sharrow D, Hug L, You D, *et al*. Global, regional, and national trends in Under-5 mortality between 1990 and 2019 with scenario-based projections until 2030: a systematic analysis by the UN inter-agency group for child mortality estimation. *Lancet Glob Health* 2022;10:e195–206.
10. National Population Commission - NPC, ICF. *Nigeria demographic and health survey 2018 - final report*. Abuja, Nigeria: NPC and ICF, 2019.
11. Burstein R, Henry NJ, Collison ML, *et al*. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature* 2019;574:353–8.
12. Oleribe O, Kumar V, Awosika-Olumo A, *et al*. Individual and socioeconomic factors associated with childhood immunization coverage in Nigeria. *Pan Afr Med J* 2017;26:220.
13. Matsuoka S, Kawakatsu Y, Koga S, *et al*. Underlying causes of underutilization of maternal, neonatal and child health (MNCH) services in Africa: a survey from Lagos state, Nigeria. *Glob Health Med* 2020;2:184–9.
14. Kawakatsu Y, Aiga H, Ubani O, *et al*. Socioeconomic barriers to facility-based delivery in urban poor communities of Lagos: wealth, linguistic capacity, and residential area. *Health Policy OPEN* 2020;1:100019.
15. Boyle EH, King M, Matthew S. *IPUMS demographic and health surveys version 8*. Minneapolis, MN: ICF, 2020. Available: https://doi.org/10.18128/D080.V8
16. National Population Commission - NPC/Nigeria. ORC macro: Nigeria demographic and health survey 2003. 2004.
17. National Population Commission - NPC/Nigeria, ICF Macro. *Nigeria demographic and health survey 2008*. Abuja, Nigeria: NPC/Nigeria and ICF Macro, 2009.
18. National Population Commission (NPC) N, International I. *Nigeria demographic and health survey 2013*. Abuja, Nigeria, and Rockville, Maryland, USA: NPC and ICF International, 2014.
19. National Bureau of Statistics (NBS), United Nations Children's Fund (UNICEF). *Nigeria - multiple indicator cluster survey 2016-2017*. Abuja, Nigeria, 2018.
20. Carolina P-H, Warren JL, Burgert CR, *et al*. Guidelines on the use of DHS GPS data. In: *Spatial analysis reports No. 8*. Calverton, Maryland, USA: ICF International, 2013.
21. Biostat Global Consulting, NBS/Nigeria NBoS. *MICS/NICS 2016-17 cluster GPS coordinates - displaced for public use*. Abuja, Nigeria: National Bureau of Statistics - NBS/Nigeria, 2019.
22. Takahashi S, Metcalf CJE, Ferrari MJ, *et al*. The geography of measles vaccination in the African great lakes region. *Nat Commun* 2017;8:15585.

23 Dijkstra L, Poelman H. A Harmonised definition of cities and rural areas: the new degree of Urbanisation. In: *Regional working paper 2014*. European Commission, 2014.
24 Fagbamigbe AF, Nnanatu CC. Modelling the spatial distribution and the factors associated with under-five mortality in Nigeria. *Spat Demogr* 2022;10:255–82.
25 Metcalf CJE, Tatem A, Bjornstad ON, *et al*. Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. *Epidemiol Infect* 2015;143:1457–66.
26 Tatem AJ, Bhatt S, Weiss D, *et al*. *Pilot high resolution poverty maps*. Southampton/Oxford Uo, 2013.
27 Graetz N, Woyczynski L, Wilson KF, *et al*. Mapping disparities in education across low- and middle-income countries. *Nature* 2020;577:235–8.
28 WorldPop. n.d. Available: https://www.worldpop.org/
29 Meijer JR, Huijbregts MAJ, Schotten KCGJ, *et al*. Global patterns of current and future road infrastructure. *Environ Res Lett* 2018;13:064006.
30 Abatzoglou JT, Dobrowski SZ, Parks SA, *et al*. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci Data* 2018;5:170191.
31 BaOK M. *High resolution electricity access indicators dataset*. Center for Political Studies UoM, 2021.
32 Maina J, Ouma PO, Macharia PM, *et al*. A spatial database of health facilities managed by the public health sector in sub Saharan Africa. *Sci Data* 2019;6:134.
33 Weiss DJ, Nelson A, Gibson HS, *et al*. A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* 2018;553:333–6.
34 Download Openstreetmap data for this region: Africa. n.d. Available: http://download.geofabrik.de/africa.html
35 Local Burden of Disease Diarrhoea Collaborators. Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000–17: analysis for the Global Burden of Disease Study 2017. *Lancet* 2020;395:1779–801.
36 Reiner RC, Welgan CA, Casey DC, *et al*. Identifying residual Hotspots and mapping lower respiratory infection morbidity and mortality in African children from 2000 to 2017. *Nat Microbiol* 2019;4:2310–8.
37 Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R Stat Soc Series B Stat Methodol* 2011;73:3–36.
38 Marra G, Wood SN. Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis* 2011;55:2372–87.
39 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016:785–94.
40 Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. *Statist Sci* 2001;16:133.
41 Shuaib F, Garba AB, Meribole E, *et al*. Implementing the routine immunisation data module and dashboard of DHIS2 in Nigeria, 2014–2019. *BMJ Glob Health* 2020;5:e002203.
42 Burton A, Monasch R, Lautenbach B, *et al*. WHO and UNICEF estimates of national infant immunization coverage: methods and processes. *Bull World Health Organ* 2009;87:535–41.
43 GBD 2020, Release 1, Vaccine Coverage Collaborators. Measuring routine childhood vaccination coverage in 204 countries and territories, 1980–2019: a systematic analysis for the Global Burden of Disease Study 2020, release 1. *Lancet* 2021;398:503–21.
44 Maïga A, Jiwani SS, Mutua MK, *et al*. Generating statistics from health facility data: the state of routine health information systems in Eastern and Southern Africa. *BMJ Glob Health* 2019;4:e001849.
45 Onwujekwe O, Ezumah N, Mbachu C, *et al*. Exploring effectiveness of different health financing mechanisms in Nigeria; what needs to change and how can it happen? *BMC Health Serv Res* 2019;19:661.
46 Abubakar I, Dalglish SL, Angell B, *et al*. The lancet Nigeria commission: investing in health and the future of the nation. *The Lancet* 2022;399:1155–200.
47 Kawakatsu Y, Tanaka J, Ogawa K, *et al*. Community unit performance: factors associated with childhood diarrhea and appropriate treatment in Nyanza province, Kenya. *BMC Public Health* 2017;17:202.
48 Dougherty L, Gilroy K, Olayemi A, *et al*. Understanding factors influencing care seeking for sick children in Ebonyi and Kogi States, Nigeria. *BMC Public Health* 2020;20:746.
49 Ugboko HU, Nwinyi OC, Oranusi SU, *et al*. Risk factors of diarrhoea among children under five years in Southwest Nigeria. *Int J Microbiol* 2021;2021:8868543.