Mapping routine measles vaccination in low- and middle-income countries

The safe, highly effective measles vaccine has been recommended globally since 1974, yet in 2017 there were more than 17 million cases of measles and 83,400 deaths in children under 5 years old, and more than 99% of both occurred in low- and middle-income countries (LMICs)\(^1\)–\(^4\). Globally comparable, annual, local estimates of routine first-dose measles-containing vaccine (MCV1) coverage are critical for understanding geographically precise immunity patterns, progress towards the targets of the Global Vaccine Action Plan (GVAP), and high-risk areas amid disruptions to vaccination programmes caused by coronavirus disease 2019 (COVID-19)\(^5\)–\(^8\). Here we generated annual estimates of routine childhood MCV1 coverage at 5 x 5-km\(^2\) pixel and second administrative levels from 2000 to 2019 in 101 LMICs, quantified geographical inequality and assessed vaccination status by geographical remoteness. After widespread MCV1 gains from 2000 to 2010, coverage regressed in more than half of the districts between 2010 and 2019, leaving many LMICs far from the GVAP goal of 80% coverage in all districts by 2019. MCV1 coverage was lower in rural than in urban locations, although a larger proportion of unvaccinated children overall lived in urban locations; strategies to provide essential vaccination services should address both geographical contexts. These results provide a tool for decision-makers to strengthen routine MCV1 immunization programmes and provide equitable disease protection for all children.

Subnational routine MCV1 coverage

Since 2016, the WHO and UNICEF have collected subnational coverage data through their annual Joint Reporting process, although poor data quality and biases currently limit the use of administrative data to track progress towards GVAP targets\(^18\)–\(^20\). For the 101 countries included in this study, 91 reported subnational data in 2018 in a total of 11,311 subnational geographical units. Of these countries, 71 reported MCV1 coverage greater than 100% in at least one unit and 55 reported such coverage in at least a quarter of units. Although researchers have estimated subnational MCV1 coverage in select countries or years for which there have been reliable surveys, to date, no comprehensive analysis of all available vaccine coverage data to produce subnational estimates of MCV1 coverage annually in all LMICs has been undertaken\(^21\)–\(^24\).

Building from our previous work mapping diphtheria–tetanus–pertussis vaccine coverage in Africa\(^24\), here we present mapped high-spatial-resolution estimates of routine MCV1 coverage across 101 LMICs from 2000 to 2019, aggregated to policy-relevant second-level administrative units (hereafter districts). Using geolocated data on MCV1 coverage from 354 household-based surveys representing approximately 1.70 million children and a suite of environmental, sociodemographic and health-related geospatial and national-level covariates, we extended model-based geostatistical methods that have been used to map child mortality and its main components and risk factors\(^25\)–\(^28\) to predict MCV1 coverage and uncertainty.
(Extended Data Figs. 1, 2), while calibrating estimates to results from GBD 2019. Using these estimates, we assessed trends in geographical inequality, progress towards global targets and differential vaccination status by geographical remoteness.

Tracking uneven progress
Despite marked global progress between 2000 and 2019, considerable inequalities in routine MCV1 coverage persist, both between and within countries (Fig. 1, Extended Data Figs. 3–7, see also our visualization tool [https://vizhub.healthdata.org/lbd/mcv/]). MCV1 coverage among children living in the 101 countries included in this study was 65.6% (95% uncertainty interval, 64.2–67.1%) in 2000 and 81.0% (95% uncertainty interval, 79.2–82.7%) in 2019. Coverage increased at the national level in 69.9% (95% uncertainty interval, 64.4–75.2%) of countries between 2000 and 2019 and in 57.4% (95% uncertainty interval, 50.4–64.6%) of districts (n = 20,795 districts).

The three districts with the lowest MCV1 coverage in 2000 were located in Hari Rasu, Ethiopia (4.0% (95% uncertainty interval, 1.1–9.7%)), Gabi Rasu, Ethiopia (4.8% (95% uncertainty interval, 1.4–11.4%)), and Isa, Nigeria (4.9% (95% uncertainty interval, 1.5–10.8%)). In 2000, 60 districts had a coverage below 10%; there was one such district in 2019. The three lowest-coverage districts in 2019 were all located in Afghanistan: Poruns (9.2% (95% uncertainty interval, 2.0–25.5%)), Wama (12.1% (95% uncertainty interval, 2.8–32.6%)) and Waygal (12.7% (95% uncertainty interval, 3.0–34.2%)).

In the period from 2000 to 2010, there were substantial increases in coverage and progress towards reducing subnational heterogeneity. The period from 2010 to 2019, however, showed slowing progress and, in some cases, regression of coverage compared to the 2000–2010 period (Fig.2). Between 2000 and 2010, 70.5% (95% uncertainty interval, 66.0–75.4%) of districts increased coverage, but between 2010 and 2019, coverage increased in only 40.1% (95% uncertainty interval, 34.2–46.9%) of districts (Extended Data Fig. 8). This finding persists even when accounting for the starting level of coverage (Supplementary Information section 2.3).

Although district-level MCV1 coverage generally increased between 2000 and 2019, further gains are required to reach both 80% and 95% key coverage targets (Extended Data Fig. 9). In 2000, 38.4% of districts had a high probability (>95% posterior probability) of reaching the GVAP target of 80% district-level coverage, which remained stagnant at 33.2% of districts in 2019. Only 15 countries had a high probability of reaching the GVAP target of greater than 80% district-level coverage in all districts.

Quantifying geographical inequalities
To further assess the effect of geographical heterogeneity in MCV1 coverage, we computed the absolute geographical inequality,
a Gini-coefficient-derived metric that ranges between zero (perfectly homogenous coverage) and one (perfectly heterogeneous coverage), at the 5 × 5-km² level. In 2000, nine countries (Cameroon, Democratic Republic of the Congo, Guinea, India, Laos, Madagascar, Mali, Nigeria and Yemen) had high absolute geographical inequality (above 0.15). In 2019, only five countries had high absolute geographical inequality (Angola, Ethiopia, Madagascar, Nigeria and Pakistan). Nigeria had absolute geographical inequality above or equal to 0.2 in both 2000 and 2019, and 25 countries had increased absolute geographical inequality in 2019 compared with 2000. Notably, absolute geographical inequality decreased considerably in India, from 0.23 in 2000 to 0.07 in 2019.

In general, and as expected, improvements in national-level coverage over time were accompanied by reductions in subnational absolute geographical inequality (Fig. 3). Changes in coverage were negatively correlated ($\rho = -0.47$, Pearson’s product moment test statistic = $-5.36$, $P < 0.001$) with changes in absolute inequality. India is a true exemplar in this trend, with sizeable reductions in inequality occurring as coverage increased. This improvement was not the only pathway for a country, however; some countries with increasing coverage also experienced increasing inequality, such as Chad and Ethiopia. Other countries experienced decreasing coverage alongside increasing inequality, such as Angola.

**Coverage in urban and rural areas**

In a post hoc analysis, we compared vaccination status in urban and remote rural settings, using proxies of travel time of ≤30 min and ≥3 h, respectively, to the nearest major city or settlement\(^a\) and the number of children under 5 years old\(^b\) from gridded estimates. In 2019, MCV1 coverage was relatively lower in remote rural areas: in remote rural locations, 33.3% of children were MCV-unvaccinated, compared with 15.2% of children living in urban areas. Owing to the concentration of populations in urban centres, however, more unvaccinated children lived in urban locations (47.9% of all unvaccinated children) than remote rural areas (16.0% of all unvaccinated children) in 2019, although this pattern varied across countries and regions (Fig. 4). For example, in Chad, 19.3% of unvaccinated children lived in urban locations and 44.4% lived in remote rural locations in 2019. In Mexico in 2019, 72.3% of unvaccinated children lived in urban locations and 3.4% lived in remote rural locations (Extended Data Fig. 10).

Our results show the variability of urban and rural contributions to unvaccinated populations in each country and region. In Latin America and the Caribbean, for example, MCV1 coverage is generally similar between urban and rural settings (Fig. 4); the urban–remote rural distribution of unvaccinated children therefore largely reflects the underlying population distribution. In other regions, the interaction between population and coverage is more complex. In South Asia, for example, 21 times more unvaccinated children live in urban locations compared with remote rural locations. Strategies focused solely on reaching the most unvaccinated children possible in that region, therefore, might reasonably prioritize urban areas. Overall, however, MCV1 coverage in urban areas of South Asia averages 90.7%, compared to only 77.4% in remote rural areas. Approaches that focus primarily on reaching urban areas, therefore, would probably exacerbate existing urban–rural coverage inequalities.

**Discussion**

Our MCV1 coverage estimates show overall progress from 2000 to 2019, corresponding to the creation of benchmark targets from the Measles and Rubella Initiative and GVAP, as well as substantial financial support for comprehensive vaccination programming generated by the introduction of Gavi, the Vaccine Alliance\(^5,6,31–34\). Moreover, 62 out of 101 countries increased national-level MCV1 coverage while reducing subnational geographical inequalities over time, a noteworthy achievement.

This remarkable global progress should be celebrated, but this trend was not universal. Our results show a decline and stagnation in routine MCV1 coverage in certain locations, particularly since 2010, that may be related to conflict, vaccine scepticism, available funding support and supply disruptions\(^35\). Among countries with stagnant or declining coverage rates, the Central African Republic and Nigeria are experiencing...
ongoing political instability and conflict, which serve as major barriers to the success of vaccination programmes. Supply disruptions also present a major barrier to achieving and sustaining high levels of MCV coverage. For example, in 2018, the Philippines reported a national-level MCV stockout. The stockout, alongside waning public confidence in vaccination programmes stemming from misinformation related to risks of the Dengvaxia dengue vaccine, contributed to a national-level drop in coverage from 80% to 69% between 2017 and 2018. In Angola, economic turmoil led to a 28% decrease in governmental health spending per capita between 2010 and 2018, which may have also contributed to declines in vaccination coverage. While global immunization initiatives have often focused on low-income countries, districts in middle-income countries also experienced recent declines, emphasizing the need for reliable immunization programmes and monitoring in these nations. In Indonesia, for instance, 3 districts had coverage that reached 95% in 2000, increasing to 13 in 2010, but decreasing back to zero in 2019. In addition, countries with higher than average vaccine scepticism, such as Peru and Moldova, also experienced decreasing coverage rates and increasing within-country inequality.

Overlaid on these persistent challenges, the ongoing COVID-19 pandemic has caused the cancellation of supplementary measles immunization campaigns and puts the delivery of critical routine immunization services at risk. Baseline subnational estimates of routine MCV1 immunization in LMICs can help to define the geographical areas of highest vulnerability to pandemic-associated disruptions. To mitigate the risk of measles outbreaks in the setting of the COVID-19 pandemic, the maintenance of routine immunization services is crucial—particularly in areas with pre-existing routine immunization weaknesses.

Even before the current pandemic, few countries reached the GVAP target of 80% coverage in all districts by 2019. Stagnant progress between 2010 and 2019 further suggests that new approaches are needed to reach unvaccinated children and resolve geographical inequalities. As the era of GVAP draws to a close and the Immunization Agenda 2030 begins, these results provide a platform to identify successes and inform strategies for the next decade. India, for instance, saw exemplary improvement in both national-level coverage and geographical equality over time. This may be attributable to specific interventions such as Mission Indradhanush, launched in 2014 with the goal of targeting underserved unvaccinated populations. In addition, India introduced a second dose of MCV (MCV2) in select subnational geographies with MCV1 coverage below 80% starting in 2008, and expanded MCV2 to cover all districts in 2010 through the strengthening of both routine and supplementary immunization programmes. The introduction of MCV2 into the national schedule may provide a second opportunity for first-dose vaccination among children who missed the scheduled MCV1 dose. Understanding the specific drivers of simultaneous coverage and equality gains may provide critical insights for the immunization agenda in countries and regions that have fallen behind.

The Equity Reference Group for Immunization highlights the need for increased attention on vaccinating vulnerable children who live in remote rural, urban poor and conflict settings, as well as for equality in coverage by gender. These recommendations suggest that the agenda to leave no child unvaccinated, set by global partners and the Sustainable Development Goals, should transcend geography types and aim to eliminate coverage gaps among children who live in both urban and remote rural areas. These geographically resolved MCV1 estimates provide a tool for decision-makers to plan supplementary immunization activities and routine immunization strengthening programmes, to reach both the urban and remote rural communities where unvaccinated children live.

Despite large improvements made in MCV1 coverage from routine immunization programmes between 2000 and 2019, stalling progress and substantial subnational variation remain in many LMICs, leaving children at risk of preventable death. Policymakers should note where progress is most critically needed to successfully meet global immunization targets and protect the most-vulnerable children against measles. Our subnational estimates of routine MCV1 coverage at policy-relevant scales provide a tool for decision-makers to use in advocating for strong, sustainable immunization programmes that provide equitable protection for all children.
Methods

Data reporting
As this is a modelling study, no statistical methods were used to pre-determine sample size, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Overview
Building from our previous study of diphtheria–tetanus–pertussis vaccination coverage in Africa, we fitted a geostatistical model with correlated errors across space and time to predict 5 × 5 km2 level estimates of MCV1 coverage from 2000 to 2019 using a suite of geospatial and national-level covariates for 101 LMICs. This overall process has been summarized in Extended Data Fig. 1. We spatially aggregated estimates using population-weighted averages to second administrative units from a modified version of the Database of Global Administrative Units (GADM), referred to as districts, and performed post hoc analyses to assess geographical inequality to examine progress towards GVAP (GADM), referred to as districts, and performed post hoc analyses to summarize in Extended Data Fig. 1. We spatially aggregated estimates national-level covariates for 101 LMICs. This overall process has been of MCV1 coverage from 2000 to 2019 using a suite of geospatial and related errors across space and time to predict 5 × 5-km2 level estimates.

Building from our previous study of diphtheria–tetanus–pertussis vac-

cipovax (SIAs) (Supplementary Information section 1.3.4), there is no explicit need for the vaccines to differ by country. In this initial fitting step, there were no explicitly defined temporal or spatial

Geostatistical model
First, stacked generalization was used to capture potential nonlinear and complex relationships between covariates and vaccination coverage. This approach has previously been shown to increase the predictive accuracy of geospatial models. Using the optimized set of covariates selected for each region by the VIF algorithm, three different child models—generalized additive models, lasso regression and boosted regression trees—were fit, with each model predicting MCV1 coverage as the outcome of interest. When fitting boosted regression trees, country-level fixed effects were included to allow relationships between coverage and covariates to differ by country. In this initial modelling step, there were no explicitly defined temporal or spatial
effects included in the models beyond those inherently present in the covariate patterns and correlations between covariates.

Each child model was fit using fivefold cross-validation to avoid overfitting. This generated out-of-sample predictions of coverage for each location and year per region. Each model in each region was also fit using the full set of vaccine coverage outcome data, which yielded a corresponding set of in-sample predictions. The predictions of MCV1 coverage obtained from each child model were in turn used as predictors in the second-step geostatistical model described below. Out-of-sample predictions from each child model were used as explanatory covariates when fitting the geostatistical model. In-sample predictions from each model and region were used when generating predictions from the fitted geostatistical model.

After the first step (stacked generalization), a second-step Bayesian geostatistical modelling framework was used to model vaccination coverage as counts in a binomial space with a logit link through a generalized linear regression with explicit spatial and temporal terms. This second step leverages the covariate relationships estimated through stacked generalization while also accounting for additional correlation in coverage across space and time.

A separate model of MCV1 coverage was fit for each of the 13 regions as defined below:

$$C_d[p_{id}, t|d] \sim \text{Binomial}(\rho_{id}, t|d) \sim N(\gamma) \forall \text{ observed clusters } d$$

$$\logit(p_{it}) = \beta_0 + X_i \beta + Z_i t + \epsilon_{\text{country}, i} + \epsilon_{i,t} \forall i \in \text{ spatial domain } \forall t \in \text{ time domain}$$

$$\sum_{h=1}^{3} \beta_h = 1 = \epsilon_{c,t} \sim N(0, \sigma^2_{\text{nugget}})$$

$$Z \sim \text{GP}(0, \Sigma_{\text{space}} \otimes \Sigma_{\text{time}})$$

$$\Sigma_{\text{space}} = \frac{2^{-1} \sigma^2_{\text{space}}}{I(v)} \times \left[ \frac{\sqrt{8}}{\delta} D \right]^v \times K_{ij} \left[ \frac{\sqrt{8}}{\delta} D \right]$$

$$\Sigma_{\text{time}} = \rho^{v-1}$$

This model, adopted from widely used Bayesian hierarchical models\cite{9,10}, has been described in detail in other work\cite{11,12,13,14}. In brief, this method estimates the number of children, $C$, in cluster $d$ at location $i$ and time $t$ with sample size $N$ that have been vaccinated with a specific antigen-dose combination. $p_{id, t, d}$ is the proportion of children vaccinated with MCV1 among the target age population in cluster $d$. Each child model generates a prediction $X_i$, for each child model $h$. Residual terms $\epsilon$ are unique to each particular location in space and time across all modelled geographies and years.

In this generalized linear regression framework, the proportion of children vaccinated $p_{id, t}$ is modelled using the out-of-sample predictions of vaccine coverage $X_i$ from each of three stacked generalization child models ($h$) as explanatory variables. The $\beta$ coefficients are constrained to sum to 1, via the `extrastrn' R-INLA parameter\cite{15}, to improve computational tractability\cite{16} and are representative of the predictive weighting used in the stacking process.

$\epsilon_{\text{country}, i}$ represents a country-level random effect. $\epsilon_{i,t}$ represents an independent nugget effect for irreducible error for a given observation, which accounts for true variation that is unable to be captured by the model and variation from measurement error. $Z_i$ represents a correlated spatiotemporal error term, for any residual autocorrelation across space and time that remains after accounting for the predictive capacity of the stacked-modelled covariates, country-specific variation in vaccine coverage and observation-specific irreducible error.

These additional spatiotemporal residuals $Z_i$ were modelled as a three-dimensional spatiotemporal Gaussian process with a mean of zero and a covariance matrix formed from the Kronecker product of spatial and temporal covariance kernels. The temporal covariance $\Sigma_{\text{time}}$ was modelled via an annual autoregressive order 1 function from all study years from 2000 to 2019, where $p$ is the autocorrelation function and $k$ and $j$ are points in the annual time series. The spatial covariance $\Sigma_{\text{space}}$ was assumed to be an isotropic, stationary Matérn function, where $I$ is the gamma function, $K_i$ is the modified Bessel function of the second kind of order $\nu > 0$, $\sigma^2_{\text{space}}$ is the marginal variance, $\nu$ is a scaling constant, $\delta$ is a range parameter with a penalized complexity prior, and $D$ is a spatial distance matrix\cite{16}, measured along the great circle in kilometres. The generalized linear model was fitted using an integrated nested Laplace approximation in R-INLA with a stochastic partial differential equation (SPDE) solver in package SPDE\cite{17}. Additional detailed information on priors, spatial mesh construction and model fitting is provided in Supplementary Information sections S4.2–S4.6 and Supplementary Fig. 5. This process produces a set of 1,000 posterior draws, each representing an estimate of vaccine coverage for each location and year — in other words, a set of 1000 candidate maps of coverage from 2000 to 2019.

Post-estimation

To leverage data from additional national-level sources, including administrative data, and maintain internal consistency, the set of candidate maps was calibrated to MCV1 coverage estimates produced for GBD 2019. This post hoc calibration preserves the overall spatial variation of estimates, while ensuring that the population-weighted averages of the geospatial estimates are equivalent to those produced by GBD\cite{18}. This step allows for the calibrated estimates to reflect information from data sources that are only available at the national level, such as surveys for which no subnational data are available, which are included in the GBD estimates but excluded from the geospatial model described in the `Geostatistical model' section. A description of the estimation of MCV1 coverage for GBD 2019 can be found in Supplementary Information section S1.5.1.

In this calibration process, each $5 \times 5$-km$^2$ pixel in each modelled region was first assigned to a second-level administrative unit. In locations in which boundary definitions transect a given pixel, the fraction of area of that pixel belonging to each overlapping second-level administrative unit was calculated. Because of the nested hierarchy of administrative units, this additionally allowed for the assignment of pixels and partial pixels to first administrative units and countries. Assuming that the population density within each pixel was uniform, WorldPop population values of children under 5 years old were divided for each whole or partial pixel proportional to fractional area. After pixel and partial pixel populations were assigned, population-level estimates were calibrated to GBD population estimates for each country and year.

Calibration methods similar to those used in this study have been described previously\cite{19}. To ensure vaccination coverage estimates post-calibration remained between 0 and 100%, calibration was performed in logit space such that for each country $c$ and year $t$, national-level estimates of coverage from GBD ($V_{\text{GBD}, c}$) and population-weighted national averages of coverage from the model-based geostatistical (MBG) model ($V_{\text{MBG}, c}$) can be related via a country-year-specific calibration factor ($k_{c,t}$) in the following equation:

$$\logit(V_{\text{GBD}, c,t}) = \logit(V_{\text{MBG}, c,t}) + k_{c,t}$$
Calibration factors were applied to each $5 \times 5\text{ km}^2$ pixel and partial pixel per draw per country-year. Pixels that were fractionally assigned to multiple countries were combined using a weighted average proportional to the fraction of each area. This process resulted in a set of calibrated draw-level estimates of vaccination coverage, which were used for all subsequent analyses.

Population-weighted averages of coverage for each pixel or partial pixel within a first or second administrative unit were then calculated. Fractional pixel membership was determined as described above. This process was repeated for each of the 1,000 posterior pixel-level draws, which yielded 1,000 posterior draws of MCV1 coverage per administrative unit per year. Estimates for first and second administrative units with uncertainty were derived from mean, 2.5th and 97.5th percentiles.

Model validation
We assessed the predictive performance of the models using fivefold out-of-sample cross-validation. We stratified data by first and second administrative units and ran models leaving out one-fifth of the spatially stratified data at a time. Predicted estimates of MCV1 coverage were then compared to the withheld observed data by calculating the mean error, root mean square error, correlation and other predictive validity metrics for all years for which survey data were available (2000–2018). Fitted model parameters can be found in Supplementary Table 8. Metrics and validity figures can be found in Supplementary Tables 9–12 and Supplementary Figs. 6–13, respectively. Additional information regarding uncertainty of estimates can be found in Supplementary Figs. 14–17.

Post hoc geospatial inequality analyses
Lorenz curves were generated using the relationship between the number of children and the number of vaccinated children for each pixel. Pixel-level Gini coefficients were calculated for 2000 and 2019 from corresponding Lorenz curves (Supplementary Table 13). Absolute geographical inequality per country was calculated from the national-level Gini coefficients and national MCV1 coverage using the following formula:

\[
\text{Absolute geographical inequality} = 2 \times \text{coverage} \times \text{Gini}
\]

We chose to use the absolute geographical inequality metric to represent inequality over the Gini coefficient alone. As the mean is related to Gini, we wanted to account for this relationship. Estimates are scaled by 2 as this puts the absolute geographical inequality coefficient back to the same scale as the mean.

Additionally, we assessed vaccination status as a function of geographical remoteness. Using a gridded surface of travel time to major cities or settlements, we classified each $5 \times 5\text{ km}^2$ pixel as remote rural, urban or neither. Pixels with travel times of less than 30 min were classified as urban, and pixels with travel times greater than 3 h were classified as remote rural. Overlaid with a gridded population surface from WorldPop, the number of unvaccinated children per pixel was also calculated.

We constructed concentration curves of the cumulative proportion of unvaccinated children as well as plots of MCV1 coverage by travel time to assess patterns across countries and regions. Country-specific concentration curves of the cumulative proportion of unvaccinated children as a function of travel time for select countries are shown in Extended Data Fig. 10. Summary metrics, such as the proportion of unvaccinated individuals living in each urban and remote rural location, were computed.

Limitations
This work is subject to several limitations. First, the primary data used in this analysis came from child-level survey data with varying degrees of representativeness, consistency, accuracy and comparability, from both HBR and parental recall. The magnitude and direction of recall bias varies, and we therefore were unable to correct for it. We estimate coverage using data from children aged 12–39 months, and while we accounted for target age at vaccination, this does not fully account for differential mortality due to vaccine status or catch-up vaccination. We aim to estimate routine coverage and have excluded doses delivered via SIAs from the analysed survey data wherever possible (Supplementary Information section 1.3.4), but misclassification of SIA doses is still likely, particularly in cases of parental recall—especially for older children—and in cases in which survey methodology does not distinguish clearly between SIA and routine doses.

In data-sparse areas for which covariate relationships may not fully capture coverage patterns, results may be biased. Additionally, data representativeness among vulnerable populations, such as those living in urban slums or migrant populations, might vary due to data collection in survey design. We include as much data on MCV1 coverage as possible, including data that are only geo-resolved to areal locations. The methodology that we used to assign areal data to specific locations for modelling could lead to oversmoothing in final estimates, obscure relationships between coverage and covariates, and underestimate uncertainty, but this method has been shown to have a higher predictive validity compared with the exclusion of the data. Limitations due to data availability should not be taken lightly and should reinforce to stakeholders and policymakers the need for additional resources to collect high-quality data that are representative of all populations, especially the most vulnerable for being unvaccinated, and to increase the quality of routinely collected subnational administrative data.

Because the estimates that we used to assess geographical remoteness in post hoc analyses were also used as spatial covariates in the geospatial model, these results are limited by the possibility of circularity and subsequent confounding. In addition, we used a stacked generalization method to allow for complex and nonlinear relationships between covariates and vaccination coverage. These methods are optimized for prediction, not causal inference. For that reason, these results cannot be used to identify the specific effect of any particular covariate on MCV1 coverage. In addition, owing to limitations in the underlying data and computational feasibility, we were unable to incorporate several potentially important sources of uncertainty into this analysis, including from covariates, population estimates, the incorporation of areal data and the stacked generalization process.

We fitted our geostatistical models using R-INLA, as opposed to a full Markov chain Monte Carlo sampler. Although using a more traditional Bayesian model fitting approach that takes true samples from the posterior typically results in increased parameter identifiability, the Laplace approximation approach used by R-INLA is more computationally feasible given our current modelling scale. Our model is separable, yet symmetric, across time and space. This approach assumes that, for each region, the covariance has the same functional form between years and locations regardless of the locations themselves; the use of a non-separable covariance function could relax these assumptions. However, owing to the additional computational challenges associated with fitting a non-separable model, as well as data sparsity in several regions throughout space and time, we determined that fitting a non-separable model would be challenging and complex, and would probably yield little benefit compared to our current modelling approach.

In some settings with high levels of natural immunity (derived from previous infection), greater than 95% vaccination coverage may not be required to prevent disease transmission. These estimates only focus on the first routine dose of MCV, and immunity can also be obtained through later vaccination via SIA or natural infection. In an ideal long-term measles elimination scenario, all immunity would be vaccine-derived, and no natural infections would occur. A 95% coverage target for routine immunization, therefore, still has practical programmatic relevance.
Article

Finally, our study describes spatial heterogeneity in coverage, but not pockets of low coverage within social or age groupings that can facilitate ongoing disease transmission, particularly in densely populated areas, despite nominally high average vaccine coverage. Although these results provide a powerful tool for policymakers to identify weaknesses in routine immunization systems and plan for SIA, they should be used in conjunction with other data sources that can be used to make decisions about vaccine policy, including analyses of cost effectiveness, determinants of high or low coverage, and specific coverage initiatives to reduce disease burden.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
The findings of this study are supported by data available in public online repositories and data publicly available upon request from the data provider. A detailed table of data sources and availability can be found in Supplementary Table 4 and at http://gbdx.healthdata.org/lbd-publication-data-input-sources. Administrative boundaries were modified from the Database for Global Administrative Areas (GADM) dataset. Populations were retrieved from WorldPop, and gridded estimates of traveltime to the nearest city or settlement were available online from a previously published study. This study complies with the GATHER recommendations.

Code availability
This study is compliant with the GATHER recommendations; as such, all computer code is available from GitHub (https://github.com/ihmewu/lbd/tree/mcv1-linic-2020). All maps and figures presented in this study are generated by the authors using RStudio (R version 3.6.1), ArcGIS Desktop 10.6 and Python 2.7.

Acknowledgements
This work was primarily supported by grants from the Bill & Melinda Gates Foundation (OPP1182474, OPP11069011 and OPP1332415). S.H. is funded by additional grants from the Melinda Gates Foundation (OPP1119407 and OPP1306961). All authors expressed their opinions in this paper are those of the authors and not necessarily those of the World Health Organization. J.-W.D.N. was supported by the Alexander von Humboldt Foundation. C.H. is partially supported by a grant of the Romanian National Authority for Scientific Research and Excellence for Epidemiological Modelling and Analysis (SAGEMA). C.A.T.A. reports personal fees from Johnson & Johnson (The Philippines), outside the submitted work. M.L.B. reports grants from the US Environmental Protection Agency, the National Institutes of Health (NIH) and the Wellcome Trust Foundation, during the conduct of the study. M.L.B. also reports honoraria/travel reimbursement from the NIH for the NHI (for the NHI). P.E. acknowledges support from the NIH Oxford Biomedical Research Centre and the BHF Centre of Research Excellence, Oxford. M.A.M. acknowledges NIGER and NIMAD grants. A. Sheikh acknowledges support by Health Data Research UK. S.B.Z. acknowledges support from the Australian Government research training program (RTP) for his academic career.

Author contributions
J.F.M., A.N.S., S.S.L. and S.I.H. conceived and planned the study. A.N.S., J.F.M., J. Quiroz-Narvarte, N.G. and N.O.G. identified and vetted data for this analysis. A.N.S. extracted, processed and geo-positioned the data. A.N.S. carried out the statistical analyses. J.F.M., A.N.S., S.S.L. and S.I.H. conceived and planned the study. A.N.S., L.E., S.S.L. and S.I.H. provided intellectual input into aspects of this study.

Competing interests
This study was funded by the Bill & Melinda Gates Foundation. Authors employed by the Bill & Melinda Gates Foundation provided feedback on initial maps and drafts of this manuscript. Otherwise, the funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the final report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication. O.O.A. is supported by DSI-NRF Centre of Excellence for Epidemiological Modelling and Analysis (SAGEMA). C.A.T.A. reports personal fees from Johnson & Johnson (The Philippines), outside the submitted work. M.L.B. reports grants from the US Environmental Protection Agency, the National Institutes of Health (NIH) and the Wellcome Trust Foundation, during the conduct of the study. M.L.B. also reports honoraria/travel reimbursement from the NIH for the NHI (for the NHI). P.E. acknowledges support from the NIH Oxford Biomedical Research Centre and the BHF Centre of Research Excellence, Oxford. M.A.M. acknowledges NIGER and NIMAD grants. A. Sheikh acknowledges support by Health Data Research UK. S.B.Z. acknowledges support from the Australian Government research training program (RTP) for his academic career.

Code availability
This study is compliant with the GATHER recommendations; as such, all computer code is available from GitHub (https://github.com/ihmewu/lbd/tree/mcv1-linic-2020). All maps and figures presented in this study are generated by the authors using RStudio (R version 3.6.1), ArcGIS Desktop 10.6 and Python 2.7.
for clinical trials and receives arm's length funding from 12 pharmaceutical companies, outside of the submitted work. J. A. Singh serves on the FDA Arthritis Advisory Committee, is a member of the Veterans Affairs Rheumatology Field Advisory Committee, and is the editor and the Director of the UAB Cochrane Musculoskeletal Group Satellite Center on Network Meta-analysis, all outside the submitted work. R.U. reports other financial activities from Deakin University, outside the submitted work. J.F.M. reports grants from the Bill and Melinda Gates Foundation, during the conduct of the study.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-020-03043-4.
Correspondence and requests for materials should be addressed to S.I.H., S.S.L. or J.F.M.
Peer review information Nature thanks C. Edson Utazi and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.
Survey data and the suite of covariates used in modelling are first compiled and processed (orange and grey). The modelling process (purple) consists of data being used in a stacked generalization ensemble modelling process via boosted regression tree, lasso and generalized additive models, fitting the second-stage spatiotemporal model using integrated nested Laplace approximation, and finally calibration to GBD estimates (blue). Steps in dark green and outputs in yellow indicate the post-estimation process when the full posterior distribution of predictions is transformed to both $5 \times 5$-km$^2$ and first and second administrative unit-level maps and their various final results. Intermediate outputs throughout the process are shown in blue and overall processes are shown in light green.
**Extended Data Fig. 2 | Regions of countries used in modelling.** Analyses were divided into 13 regions based on the GBD super-regions to allow for locations similar in data availability and patterns of vaccine coverage to be analysed using similar covariate and modelling relationships. Each colour represents a different region, as described in the legend.
Extended Data Fig. 3 | National, first- and second-administrative-unit level, and pixel-level MCV1 coverage, 2000. a–d, Posterior means are represented at the national (a), first-administrative-unit (b), second-administrative-unit (c) and 5 × 5 km² pixel (d) levels. Pixels that are grey in colour are either not included in the analysis, or have been classified as being ‘barren or sparsely vegetated’ or had fewer than 10 people per 1 × 1 km² pixel.30,50
Extended Data Fig. 4 | National, first- and second-administrative level, and pixel-level MCV1 coverage, 2005. a–d. Posterior means are represented at the national (a), first-administrative-unit (b), second-administrative-unit (c) and 5 × 5-km² pixel (d) levels. Pixels that are grey in colour are either not included in the analysis, or have been classified as being ‘barren or sparsely vegetated’ or had fewer than 10 people per 1 × 1-km² pixel\textsuperscript{30,50}.
Extended Data Fig. 5 | National, first- and second-administrative-unit level, and pixel-level MCV1 coverage, 2010. a–d. Posterior means are represented at the national (a), first-administrative-unit (b), second-administrative-unit (c) and 5 × 5-km² pixel (d) levels. Pixels that are grey in colour are either not included in the analysis, or have been classified as being ‘barren or sparsely vegetated’ or had fewer than 10 people per 1 × 1-km² pixel.\(^{30,30}\)
Extended Data Fig. 6 | National, first- and second-administrative-unit level, and pixel-level MCV1 coverage, 2015. a–d, Posterior means are represented at the national (a), first-administrative-unit (b), second-administrative-unit (c) and 5 × 5-km² pixel (d) levels. Pixels that are grey in colour are either not included in the analysis, or have been classified as being ‘barren or sparsely vegetated’ or had fewer than 10 people per 1 × 1-km² pixel.
Extended Data Fig. 7 | National, first- and second-administrative-unit level, and pixel-level MCV1 coverage, 2019. a–d. Posterior means are represented at the national (a), first-administrative-unit (b), second-administrative-unit (c) and 5 × 5-km² pixel (d) levels. Pixels that are grey in colour are either not included in the analysis, or have been classified as being ‘barren or sparsely vegetated’ or had fewer than 10 people per 1 × 1-km² pixel\textsuperscript{30,50}. 
Extended Data Fig. 8 | Probability of increased or decreased coverage from 2000 to 2010 and 2010 to 2019. a–d, Probability of an increase in coverage in each district (a, b) and probability of decrease in coverage in each district (c, d) from 2000 to 2010 (a, c) and 2010 to 2019 (b, d).
Extended Data Fig. 9 | Estimated district-level probabilities of reaching MCV1 coverage targets in 2019. a. Probability of districts having achieved 80% GVAP and Measles Rubella Initiative targets (a) and 95% critical proportion to immunize coverage targets to reach herd immunity (b). Countries excluded from the analysis and pixels classified as ‘barren or sparsely vegetated’ based on ESA-CCI satellite data or with fewer than 10 people per 1 × 1 km² pixel based on WorldPop estimates are masked in grey.30,31
Extended Data Fig. 10 | Country examples of concentration curves. Concentration curves of the cumulative proportion of unvaccinated children as a function of travel time (in hours) in Chad (a), Madagascar (b), India (c) and Mexico (d). Curves for both the indicated countries (blue) and all LMICs (grey) are shown.
### Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- **n/a**

- [x] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

- [x] The statistical test(s) used AND whether they are one- or two-sided

- [x] Only common tests should be described solely by name; describe more complex techniques in the Methods section.

- [ ] A description of all covariates tested

- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

- [x] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

- [x] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.

- [x] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

- [x] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

- [x] Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

### Software and code

**Policy information about availability of computer code**

| Data collection | No primary data collection was carried out for this analysis. |
|-----------------|---------------------------------------------------------------|
| Data analysis   | This analysis was carried out using R version 3.6.1 and using R-INLA v.20.01.29.9000. Maps were produced using ArcGIS Desktop 10.6 and Python 2.7. All code used for these analyses will be made publicly available upon publication. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

### Data

**Policy information about availability of data**

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The findings of this study are supported by data available in public online repositories and data publicly available upon request of the data provider. A detailed table of data sources and availability can be found in Supplementary Table 4 and http://ghdx.healthdata.org/lbd-publication-data-input-sources. Administrative boundaries were modified from the Database for Global Administrative Areas (GADM) dataset. Populations were retrieved from WorldPop and gridded estimates of travel time to nearest city or settlement were available online from work by Weiss, et al 2018. This study complies with the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations. All maps and figures presented in this study are generated by the authors; no permissions are required for publication.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size                  | This observational study incorporated all available survey data sources that met the inclusion criteria as described in detail in the manuscript and supplementary information. The combined dataset from 354 household based surveys contained information on vaccination status from 1.70 million individual children. |
|------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions              | Surveys were excluded due to unrealistic national or geographic trends compared to other surveys in nearby country-years, inability to match the microdata to geographic locations, or non-standard methodology. These criteria were pre-established prior to reviewing the data. A full list of excluded surveys is included in Supplementary Table 5. |
| Replication                  | All code and data are available publicly for reproducibility.                                                                                                                                                                                                 |
| Randomization                | As this work is an observational mapping study, there were no experimental groups.                                                                                                                                                                          |
| Blinding                     | As this work is an observational mapping study, there was no need for blinding.                                                                                                                                                                           |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a  | Involved in the study |
|------|-----------------------|
| ☒    | Antibodies            |
| ☒    | Eukaryotic cell lines |
| ☒    | Palaeontology         |
| ☒    | Animals and other organisms |
| ☒    | Human research participants |
| ☒    | Clinical data         |

Methods

| n/a  | Involved in the study |
|------|-----------------------|
| ☒    | ChIP-seq              |
| ☒    | Flow cytometry        |
| ☒    | MRI-based neuroimaging |