Release Statement

High-resolution, modelled estimates of vaccination coverage for the Democratic Republic of Congo, including estimates of zero-dose- and under-vaccinated children, version 1.0

24 December 2025

Abstract

This data release provides gridded estimates (at a spatial resolution of 30 arc-seconds, approximately 1 km grid cells) of DTP1-3 and MCV1 vaccination coverage rates and numbers of zero-dose and under-vaccinated children for the Democratic Republic of Congo (DRC). The project team utilized the 2023 Enquête de Couverture Vaccinale (ECV) survey dataset, conducted by the Kinshasa School of Public Health (KSPH), along with settlement extents and geospatial covariates, to model and estimate vaccination coverage rates for children aged 12–23 months (at the time of the survey). Estimates were calculated for each grid cell within a Bayesian statistical modelling framework. The approach facilitated simultaneous accounting for the multiple levels of variability within the data. It also allowed the quantification of uncertainties in parameter estimates. These model-based coverage estimates can be considered as most accurately representing the year 2023. Although the methods were robust enough to explicitly account for key random biases within the datasets, it is noted that systematic biases, which may arise from sources other than random errors within the observed data collection process, are most likely to remain. The un- and under-vaccinated children estimation combined the new vaccination coverage estimates with existing high-resolution population estimates of children aged under one-year-old. The reference year of the un- and under-vaccinated children estimates is 2024.

These data were produced by the WorldPop Research Group at the University of Southampton. This work was part of the GRID3 – DRC-GAVI-EAF project, with funding by the Zero Dose Child Vaccination Project of the Equity Acceleration Fund (EAF) of Gavi, the Vaccine Alliance [grant number: FAE/GRID3/001/2024]. Project partners included United Nations Office for Project Services (UNOPS), GRID3 Inc, the Center for Integrated Earth System Information (CIESIN) within the Columbia Climate School at Columbia University, and WorldPop at the University of Southampton. The final statistical modelling was designed and developed by C.E. Utazi and implemented by K.S. Krishnaveni. H.R.

Chamberlain led on the geospatial data processing of the survey, with support from A. Cunningham, who led the geospatial covariate processing and map production. Project oversight was provided by Attila Lazar and Heather Chamberlain. The 2023 ECV data were collected, processed, anonymised and shared by the KSPH and its implementing partners. The settlement extent data was prepared and shared by CIESIN (2024).

The authors followed rigorous procedures designed to ensure that the used data, the applied method and thus the results are appropriate and of reasonable quality. If users encounter apparent errors or misstatements, they should contact WorldPop at release@worldpop.org.

RELEASE CONTENT

1. COD_Coverage_Unconstrained_Rasters.zip
2. COD_ZeroDose_UnderVaccinated_children_Rasters.zip
3. COD_shapefiles.zip
4. COD_pop_Global2_GRID3_v4_3_mosaiced_agg1km.tif

LICENSE

SUGGESTED CITATION

Krishnaveni K.S., Utazi, C.E., Cunningham, A., Chamberlain H., Lazar A. N., Tatem A. J. 2025. Modelled gridded vaccination coverage and zero-dose and under-vaccinated children estimates for the Democratic Republic of Congo version 1.0 (2023), WorldPop, University of Southampton. doi: https://dx.doi.org/10.5258/SOTON/WP00878

FILE DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984).

COD_Coverage_Unconstrained_Rasters.zip

These geotiff raster contains estimates of vaccination coverage rates for each approximately 1km grid cell (0.00898 decimal degrees grid) across the DRC for children aged 12-23 months. The grid cell values of the individual layers are:

- DTP1-3: mean, median, standard deviation, 95% credible interval lower and upper bounds, and the Probability of attaining 80% coverage
- MCV: mean, median, standard deviation, 95% credible interval lower and upper bounds, and the Probability of attaining 95% coverage
- Relative dropout rates between DTP1 and DTP2, DTP2 and DTP3, DTP1 and 3, for 12-23 months.

nodata value set to -99999.

COD_ZeroDose_UnderVaccinated_children_Rasters.zip

These geotiff raster images contain estimates of numbers of un- and under-vaccinated children for each approximately 1km grid cell (0.00898 decimal degrees grid) across the DRC for children aged under one year for DTP and MCV vaccines.

Zero-dose children are those who have not received the DTP1 vaccine. MCV zero-dose children are those who have not received the MCV1 vaccine. Under-vaccinated children here refers to the difference between DTP3 and DTP1 vaccinated children.

The raster layers provided include estimates of DTP zero-dose children, MCV zero-dose children, and DTP under-vaccinated children (children who received DTP1 but not DTP3).

COD_shapefiles.zip

These GIS shapefiles contain

- estimated vaccination coverage rates (mean, median, standard deviation, lower 95% credible interval, upper 95% credible interval),
- estimated total number of children under the age of one,
- estimated numbers of children NOT vaccinated for DTP1(dtp zero dose), and MCV1(mcv zero dose),
- DTP under-vaccinated children (i.e., those who received DTP1 but not DTP3).

These results are summarized at various administrative levels:

- admin 1 (province)
- admin 2 (health zone)

- admin 3 (health area)

COD_pop_Global2_GRID3_v4_3_mosaiced_agg1km.tif
This geotiff raster contains estimates of total population size for each approximately 1km grid cell (0.00898 decimal degrees grid) across the DRC. This is a composite layer consisting of the mean of the posterior probability distribution for the predicted population size of the GRID3 population estimates for 15 provinces (version 3) and the WorldPop Global 2 population estimates for 2024, spatially aggregated to 30 arc-second grid cells.

RELEASE HISTORY

Version 1.0 (24 December 2025) [doi: 10.5258/SOTON/WP00878]

- Original release

ASSUMPTIONS AND LIMITATIONS

These vaccination coverage estimates represent the year 2023, because the ECV data reference year was 2023; however, the zero-dose children estimate represents the year 2024 due to the reference year of the population estimates. It is assumed that coverage rates observed in 2023 would not have changed substantially in 2024. The geospatial covariates data were collected over different time periods between 2020 and 2024; however, most covariates had a reference year of 2023 or 2024. Similarly, the CIESIN settlement layers were produced in 2024. The inherent heterogeneity in the temporal alignment of these datasets used in the modelling may introduce uncertainties and potential inaccuracies in the modelling process.

The model does not account for external factors, such as migration, displacement, or sudden demographic changes, which could significantly influence population dynamics and, consequently, vaccination coverage rates.

SOURCE DATA

The key datasets used to produce the modelled vaccination coverage and population estimates are:

ECV data

ECV2023 survey data shared through a data sharing agreement with KSPH, via CIESIN. The ECV 2023 survey used a cross-sectional, mixed-methods approach that combined quantitative and qualitative elements to assess routine infant vaccination coverage and identify obstacles to service utilization for DTP1-3 and MCV1 vaccine doses. The target age group for the survey was children aged 6 to 23 months, with a specific emphasis on those aged 12 to 23 months to evaluate their complete immunisation status.

The total sample size comprised 81,868 households, encompassing 83,414 eligible children, and achieved an exceptionally high response rate of 99.8%. Data geolocated to households, using GPS-enabled devices. The primary sampling unit (PSU) was the health area, with further segmentation within each health area. We did NOT have access to digitised cluster boundaries.

Data cleaning was needed to address: (i) Inconsistencies in tabulated data relating to individual children and (ii) Errors in household geolocation.

- Using the combinations of HHS_KEY, MKEY, and CKEY values, the likely duplicated child records were removed.
- We also aimed to identify households with implausible locations, considering the location of remaining households (HHs) in the survey cluster (identified through the q103105 variable). We compared the HH GPS coordinate with the cluster centroid. If the distance was greater than 3 times the median absolute deviation, this HH was flagged as an outlier. In addition, if (i) the HH-centroid cluster distance is greater than 10km (rural) or 1.5km (urban) (n=1,922), or (ii) survey supervisor (q112 variable) is in a subset with disproportionately high number of flagged HH locations (n=2,791), the cluster was flagged as potentially erroneous. These flagged locations were manually reviewed and it was decided whether the whole cluster is dropped from the modelling, or the HH location is updated with the cluster centroid.

Following the cleaning of HH locations, 2,656 clusters and 79,404 households were included in the survey dataset used for modelling.

Health area and health zone boundaries

GRID3 v3.0 boundaries for 14 provinces (CIESIN 2025), supplemented by boundaries from OSM (shared from BlueSquare via GRID3) to create national coverage.

Province boundaries
Derived from health zone boundaries (dissolved based on attributes)

Gridded population estimates
To estimate the under-1 population across all provinces of the DRC, we created a mosaicked raster combining two complementary population datasets: the GRID3 v4.3 modelled population estimates (covering 15 provinces – Yankey et al 2025) and the WorldPop Global2 gridded population estimates (covering 11 provinces and adjusted to match the UN World Population Prospects 2024 national totals – Bondarenko et al 2025). Both input rasters had an approximate resolution of 100 meters. Estimates of children under 1 year old were generated by multiplying the total population in each raster cell by the under-1 population derived from WorldPop Global2 demographic data.

Settlement Data
We used the GRID3 COD - Settlement Extents v3.1 (CIESIN, 2024) as the input settlement layer to constrain the estimates to settled extents.

Geospatial Covariates
A wide variety of geospatial covariates, related to population distribution, were considered in the modelling. These geospatial covariates include land use and land cover data, climate variables such as temperature and rainfall, physical features and infrastructure, including roads and schools, as well as conflict data. Model covariates were selected using a generalized linear model (GLM) based stepwise selection method. The selected covariates were further assessed for multi-collinearity and statistical significance (see Annex).

METHODS OVERVIEW
We fitted Bayesian binomial geostatistical models using the INLA-SPDE approach implemented in the R-INLA package. We used the ratio-based approach (Utazi et al., 2022) to ensure that $p(DTP1) \geq p(DTP2) \geq p(DTP3)$.

K-fold cross-validation was used to evaluate the accuracy of the modelled estimates at the cluster-level. The modelled estimates were also compared with direct survey estimates at the provincial level.

Table 2: Out-of-sample prediction based on K-fold cross-validation (Cluster level)

Indicator	CORR	RMSE	MAE	AVG_BIAS
DTP1	0.760	0.146	0.108	0.004
DTP2/1	0.531	0.151	0.115	0.001
DTP3/2	0.496	0.188	0.140	-0.002
MCV1	0.657	0.195	0.156	0.002

Notes: CORR: Correlation; RMSE: Root mean square error; MAE: Mean absolute error; AVG_BIAS: Average bias. These k-fold results are reported for modelled indicators only, from which the remaining target indicators were derived (see Utazi et al., 2022)

Table 3: National level direct vs modelled estimates (correlation)

Source	DTP1	DTP2	DTP3	MCV1
2023 ECV Report	80.5	69.0	57.6	52.2
Modelled	81.1	71.8	60.3	52.7

Details of model fitting, validation, and prediction can be found in the following publications: Utazi et al. (2022) and Utazi et al. (2021).

ACKNOWLEDGEMENTS

We thank KSPH and its implementing partners for providing access to the anonymized EPC data, in accordance with the relevant data-sharing agreements. The WorldPop group is acknowledged for overall project support.

WORKS CITED

Armed Conflict Location & Event Data Project (ACLED) 2024, available from: https://acleddata.com/ [Accessed 28 Aug 2024]

Bondarenko M., Priyatikanto R., Tejedor-Garavito N., Zhang W., McKeen T., Cunningham A., Woods T., Hilton J., Cihan D., Nosatiuk B., Brinkhoff T., Tatem A., Sorichetta A. 2025 Constrained estimates of 2015-2030 total number of people per grid square at a resolution of 3 arc (approximately 100m at the equator) R2025A version v1. Global Demographic Data Project - Funded by The Bill and Melinda Gates Foundation (INV-045237). WorldPop - School of Geography and Environmental Science, University of Southampton. DOI:10.5258/SOTON/WP00839

Center for Integrated Earth System Information (CIESIN), Columbia University, Ministère de la Santé Publique, Hygiène et Prévention, Democratic Republic of the Congo, and GRID3. 2025. GRID3 COD - Health Areas v6.0. New York: Columbia University. https://doi.org/10.7916/k2zk-2j78

Center for International Earth Science Information Network (CIESIN), Columbia University. 2024. GRID3 COD - Settlement Extents v3.1. New York: GRID3. https://doi.org/10.7916/d6gy-yh28. Accessed [22nd August 2025].

Didan, K. (2021). MODIS/Terra Vegetation Indices Monthly L3 Global 1km SIN Grid V061 [Data set]. NASA Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MOD13A3.061 Date Accessed: 2024-08-28.

Food and Agriculture Organization of the United Nations (FAO) - GLW4. (2020) https://data.apps.fao.org/catalog//iso/9d1e149b-d63f-4213-978b-317a8eb42d02

Harris I, Osborn TJ, Jones P and Lister D (2020) Version 4 of the CRU TS Monthly High-Resolution Gridded Multivariate Climate Dataset. Sci Data 7, 109 (2020). https://doi.org/10.1038/s41597-020-0453-3.

Microsoft (2023) Road detection. https://github.com/microsoft/RoadDetections

OSMF (OpenStreetMap Foundation) and Contributors: OpenStreetMap (OSM) (Geofabrik) (2024). https://download.geofabrik.de/

Utazi CE, Nilsen K, Pannell O, Dotse-Gborgbortsi W, Tatem AJ. District-level estimation of vaccination coverage: Discrete vs continuous spatial models. Statistics in Medicine. 2021; 40: 2197–2211. https://doi.org/10.1002/sim.8897

Utazi CE, Aheto JMK, Chan HMT, Tatem AJ, Sahu SK. Conditional probability and ratio-based approaches for mapping the coverage of multi-dose vaccines. Statistics in Medicine. 2022; 41(29): 5662-5678. doi:10.1002/sim.9586

Yankey O., Nnanatu C., Boo G., Chaudhuri S., Krishnaveni K S, Megheib M, Bonnie A., Chamberlain H., Lazar A. N., Tatem A. J. 2025. Bottom-up gridded population estimates for the Democratic Republic of Congo, version 4.3. WorldPop. https://wopr.worldpop.org/?COD/Population/v4.3

Weiss DJ, Lucas TCD, Nguyen M, et al. (2019) Mapping the global prevalence, incidence, and mortality of Plasmodium falciparum, 2000–17: a spatial and temporal modelling study. Lancet 2019; published online June 19. DOI: 10.1016/S0140-6736(19)31097-9.

Woods D., McKeen T., Cunningham A., Priyatikanto R., Sorichetta A., Tatem A.J. and Bondarenko M. 2024 "WorldPop high resolution, harmonised annual global geospatial covariates. Version 1.0." University of Southampton: Southampton, UK. DOI:10.5258/SOTON/WP00772

Annex: Selected geospatial covariates for modelling

	Name	Covariate Description	Year	Data Source
1.	Distance to coastline	Distance to open-water coastline	2023	Woods et al (2024)
2	Distance to education facilities	Euclidean distance to education facilities, defined by OpenStreetMap per 100m	2023	OSMF (2024)
3	Distance to violence against civilians	Distance to violence against civilians conflict point events with filter applied to exclude data newer than 31/12/2022 to align with input survey data	2022	ACLED (2024)
4	Distance to battles	Distance to battle conflict point events, with filter applied to exclude data newer than 31/12/2022 to align with input survey data	2022	ACLED (2024)
5	Distance to protests	Distance to protest conflict point events, with filter applied to exclude data newer than 31/12/2022 to align with input survey data	2022	ACLED (2024)
6	Distance to strategic and remote violence	Distance to strategic and remote violence conflict point events, with a filter applied to exclude data newer than 31/12/2022 to align with input survey data	2022	ACLED (2024)
7	Distance to Herbaceous Cover, Grassland, Mosses edges	Distance to herbaceous cover, grassland, mosses edges. The values of the raster are from the cell centres to the nearest feature	2022	Woods et al (2024)
8	Distance to Urban Areas edges	Distance to the edges of urban areas. The values of the raster are from the cell centres to the nearest feature	2022	Woods et al (2024)
9	Near-infrared radiation reflectance 2022	Near-infrared radiation reflectance (unstated units) collated from monthly datasets into an annual total	2022	MODIS 'MOD13A3' Didan (2021)
10	Malaria prevalence (latest year)	Proportion of Children 2 to 10 years of age showing, on a given year, detectable Plasmodium falciparum parasite 2000-2022	2022	Malaria Atlas Project: https://malariaatlas.org/explorer/#/ Weiss et al (2019)
11	Daily max temp 2021-2023	Daily max temperature measured as a monthly average in degrees Celsius, averaged across a 3-year period. (version 4.08)	2021-2023	CRU TS 4.08. https://crudata.uea.ac.uk/cru/data/hrg/#info Harris et al (2020)
12	Distance to explosions and remote violence	Distance to explosions and remote violence	2022	ACLED (2024)

	Name	Covariate Description	Year	Data Source
13	Cattle density (average)	Cattle density per 10km pixel	2020	GLW 4: Gridded Livestock Density (Global - 2020 - 10 km), FAO (2020)
14	Distance to roads	Distance to roads. https://github.com/microsoft/RoadDetections	2023	Microsoft (2023)
15	Distance to Places of Worship	Euclidean distance to Places of worship, defined by OpenStreetMap per 100m	2023	Africa stack (OSM)
16	Distance to shrubland edges	Distance to shrubland edges. The values of the raster are from the cell centres to the nearest feature	2022	Woods et al (2024)
17	Roads density	Road density. https://github.com/microsoft/RoadDetections	2023	WorldPop (Microsoft)
18	Proximity to protected areas (Category 1)	Distance to IUCN strict nature reserve and wilderness area edges 2015-2022	2022	Woods et al (2024)
19	Walking Travel time to health facilities	Travel time based on only walking transport methods to the nearest Health Facility, as defined by the Malaria Atlas Project. The values of the raster are the from the cell centres to the nearest feature and are calculated in minutes	2020	Malaria Atlas Project: https://malariaatlas.org/research-project/accessibility-to-healthcare/ Weiss et al (2019)
20	Rural/Urban stratification (1=urban, 2=rural)	GHS-SMOD classes 23 and 30 have been recoded to 1 (urban), with all remaining classes considered to be rural and recoded to have a value of 2	2024	Global Human Settlement - GHS-SMOD_GLOBE_R2023A - European Commission