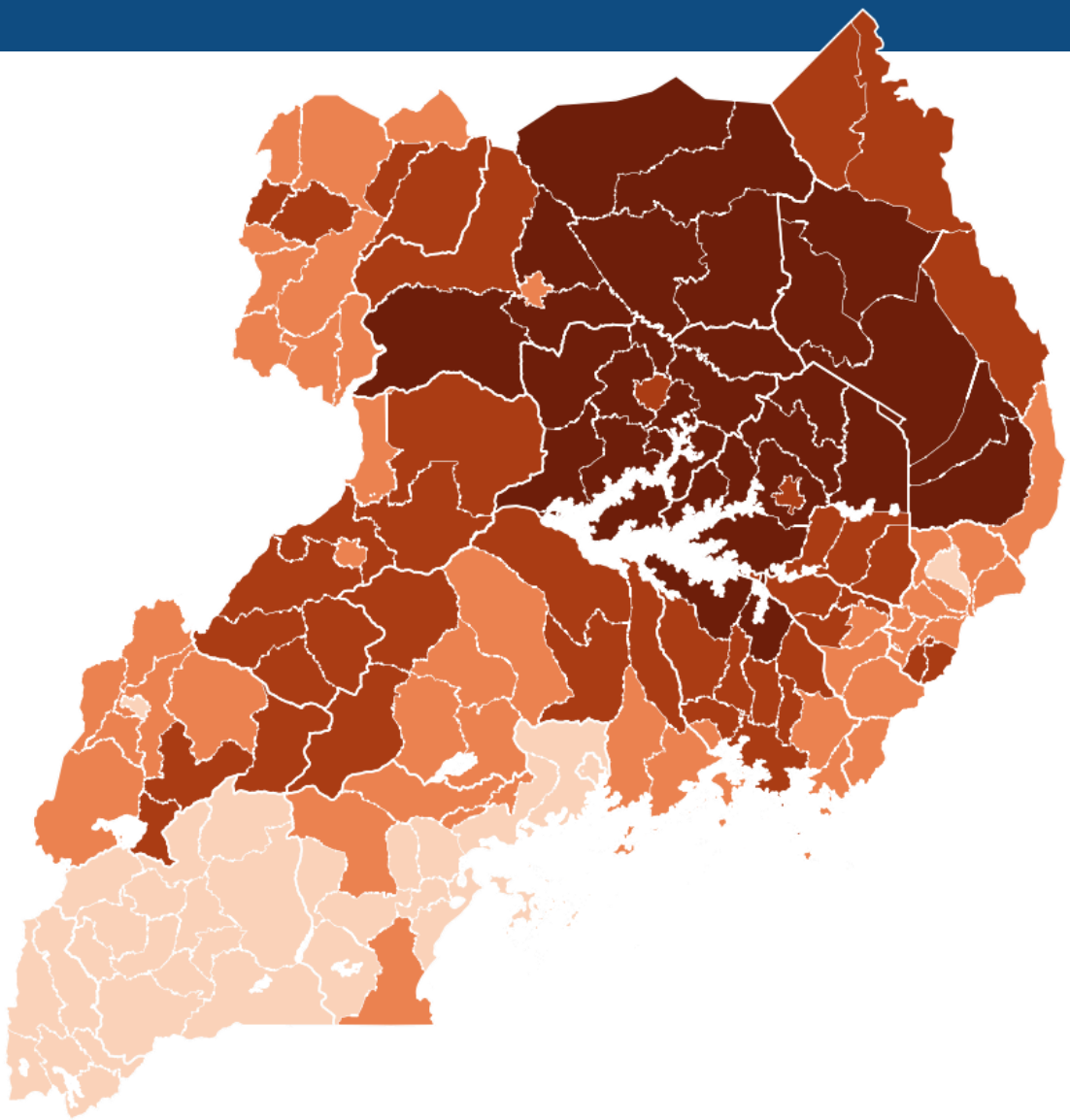
SPACEPILOTS spatial prediction & mapping

SPACEPILOTS RESEARCH REPORT · PREVIEW EDITION

Child Malaria in Uganda's Districts

Small area estimates of malaria infection among children aged 6–59 months for 146 districts, 2024–25



Prepared by **SpacePilots** · www.spacepilots.io

October 2026

Suggested citation. SpacePilots (2026). *Child Malaria in Uganda's Districts: Small Area Estimates of Malaria Infection among Children Aged 6–59 Months, 2024–25.* SpacePilots Research Report. Copenhagen: SpacePilots.

Disclaimer. The estimates in this report are model-based and were produced by SpacePilots. They are not official statistics of the Uganda Bureau of Statistics (UBOS) or the Ministry of Health. Official national and sub-region figures are those published from the Uganda Malaria Indicator Survey 2024–25. Every estimate should be read with its confidence interval and reliability class.

Data. Uganda Malaria Indicator Survey 2024–25 (The DHS Program); National Population and Housing Census 2024 10% sample and district and county boundaries (UBOS); Google Open Buildings 2.5D Temporal, Google Satellite Embedding, Landsat, MODIS, CHIRPS, SRTM, JRC Global Surface Water, ESA WorldCover and VIIRS (Google Earth Engine); Malaria Atlas Project surfaces; OpenStreetMap. Maps show the 2024 census boundaries and do not imply any judgement on the legal status of any territory.

Contact. Questions about this report, the estimates or the underlying files can be sent to SpacePilots at info@spacepilots.io.

Preview edition. This is an abridged edition of the full report. It shows the approach, the main findings and the quality of the estimates. District and county estimates with their confidence intervals, the full set of maps and the technical appendices are in the full report and data package (Chapter 8).

© SpacePilots, Copenhagen, 2026.

Contents

Executive summary 4

Abbreviations 5

1 Introduction 6

1.1 Background 6
1.2 The problem: too few children per district 6
1.3 Relation to earlier work 6
1.4 Objectives 6
1.5 Structure of the report 6

2 Data 7

2.1 Data sources 7
2.2 Survey outcomes 7
2.3 Census 7
2.4 Harmonizing survey and census variables 8
2.5 Survey cluster locations 8

3 Methodology 9

3.1 Overview of the estimators 9
3.2 How the data are used 9
3.3 Direct estimates 9
3.4 The main model: ML-grid 9
3.5 From model to district estimates 9
3.6 Uncertainty 10
3.7 Alternative estimators 10
3.8 Model selection 10

4 Results 11

4.1 Malaria prevalence by district 11
4.2 Differences within sub-regions 12
4.3 Number of malaria-positive children 12
4.4 Prevalence and numbers compared 13
4.5 Microscopy 14
4.6 Sub-regions and refugee settlements 14
4.7 County estimates (experimental) 15

5 Quality and validation 16

5.1 Precision 16
5.2 Gain over direct survey estimates 16
5.3 Independent accuracy test 16
5.4 Consistency with official figures 17

6 Discussion and limitations 18

6.1 What drives the results 18
6.2 Why the grid model performs well 18
6.3 Limitations 18

7 Conclusions and recommendations 19

7.1 Conclusions 19
7.2 Recommendations for users 19

References 20

8 The full report and data package 21

8.1 Data package 21
8.2 What the full report adds 21

LIST OF FIGURES

Figure 1 Malaria prevalence (RDT) among children aged 6–59 months by district, 2024–25, in four classes 11

Figure 2 District estimates of malaria prevalence (RDT) and MIS sub-region estimates, 2024–25 12

Figure 3 Estimated number of RDT-positive children aged 6–59 months by district, 2024–25, in four classes 13

Figure 4 Malaria prevalence (RDT) and number of RDT-positive children aged 6–59 months by district, 2024–25 14

Figure 5 Experimental malaria prevalence (RDT) estimates by county, 2024–25, in four classes 15

Figure 6 Split-sample validation error for district RDT prevalence, by method 16

LIST OF TABLES

Table 1 Data sources 7

Table 2 Estimators compared 9

Table 3 Priority districts: the ten with the highest malaria prevalence (RDT) and the ten with the most RDT-positive children aged 6–59 months, 2024–25 12

Table 4 Sub-region estimates: MIS 2024–25 figures and aggregated district estimates 14

Table 5 Precision of the estimates by level and indicator 16

Table 6 Split-sample validation results, district level (estimates fitted on half the survey) 17

Table 7 Contents of the data package 21

Executive summary

One in five young children in Uganda carried malaria parasites at the end of 2024, but the burden is extremely uneven: district prevalence ranges from below 1% in the south-western highlands to more than 60% in Lango and Acholi.

1.26 million
children aged 6–59 months test positive for malaria by RDT

0.1–68%
range of RDT prevalence across the 146 districts

30 of 146
districts have a prevalence of 40% or more

60%
lower estimation error than district figures from the survey alone, in an independent test

Purpose. The Uganda Malaria Indicator Survey 2024–25 (MIS) gives reliable malaria figures for the nation and 15 sub-regions, but malaria control is planned and delivered by districts. With a median of 37 tested children and two survey clusters per district, direct survey estimates for districts are too imprecise to use (median standard error 11 percentage points; seven districts were not sampled). This report fills that gap.

Approach. Eight statistical and machine-learning methods were fitted, combining the survey with the 2024 census and with satellite and other geospatial data that cover every part of the country. The method used for the headline figures, a machine-learning model of survey-cluster prevalence predicted on a 500-metre grid, was the most accurate in an independent test. Its district estimates are rescaled so that they add up exactly to the official MIS sub-region figures.

KEY FINDINGS

- **A high-transmission belt in the north and east.** 30 districts exceed 40% RDT prevalence and 19 exceed 50%. All but two (Kaliro and Buyende in Busoga) are in Lango, Acholi, Teso and Karamoja. Kole (68%), Kwania (67%), Lira (65%), Oyam (64%) and Agago (64%) are highest.
- **Near elimination in the south-west and Kampala.** 32 districts are below 5% and nine below 1%, including Kabale, Rukiga and Rukungiri in Kigezi, Buhweju and Kiruhura in Ankole, and Kampala.
- **Large differences within sub-regions.** Districts in the same sub-region typically differ by 25 percentage points. Cities are much lower than their surroundings: Gulu City 18% against 46% for Acholi, Lira City 28% against 60% for Lango, Jinja City 7% against 30% for Busoga.
- **Where the positive children are.** Lango, Teso, Acholi and Karamoja hold 46% of RDT-positive children but only 21% of children aged 6–59 months. Oyam (43,000), Kamuli (29,000), Agago (27,000), Buyende (26,000) and Kole (25,000) have the largest numbers.

Reliability. Estimates are most precise where malaria is very high or very low. 38 of 146 district estimates are reliable, 52 should be used with caution and 56, all with a prevalence between 5% and 35%, are unreliable. The median standard error is 7.8 percentage points, against 11.0 for the survey alone. In an independent test, the mean squared error of the district estimates was 40% of that of direct survey estimates.

Recommendations

1. Use the district estimates to rank and target districts for malaria control and to size commodity needs by the number of positive children, always with their confidence intervals and reliability class.
2. Treat the ranking within the high-transmission belt as robust and the ranking among mid-prevalence districts (5–35%) as indicative only.
3. Use county figures only to guide closer investigation; they are experimental.
4. Validate the estimates against routine health data (test positivity and incidence from DHIS2).

Abbreviations

Abbreviation	Meaning
CI	Confidence interval
CV	Coefficient of variation (standard error divided by the estimate)
EBP	Empirical best predictor
FH	Fay–Herriot (area-level) model
GPS	Global Positioning System (survey cluster coordinates)
IRS	Indoor residual spraying
ITN	Insecticide-treated net
MAP	Malaria Atlas Project
MIS	Uganda Malaria Indicator Survey 2024–25
ML	Machine learning
MSE, RMSE	Mean squared error, root mean squared error
NPHC	National Population and Housing Census 2024
pp	Percentage points
RDT	Rapid diagnostic test for malaria
SAE	Small area estimation
SE	Standard error
UBOS	Uganda Bureau of Statistics

Names of the estimation methods

Name used	Method
ML-grid, benchmarked (main)	Cluster-level machine-learning model predicted on a ~500 m grid, rescaled to the MIS sub-region figures; source of the headline RDT estimates (Sections 3.4–3.6)
ML-grid	The same model without rescaling; source of the microscopy estimates
ML-unit	Child-level machine-learning model applied to census children (Section 3.7)
RF-unit	Child-level random forest applied to census children (Section 3.7)
EBP-unit	Unit-level census model: logistic mixed model of each child's test result (Section 3.7)
FH-census, FH-geo, FH-census+geo, FH-hybrid	Area-level Fay–Herriot models with census, geospatial, both, or both plus a machine-learning prediction as covariates (Section 3.7)
Direct estimate	Survey-weighted estimate from the MIS children in the area alone

1 Introduction

This report estimates malaria infection among young children for every district in Uganda in 2024–25, the level at which malaria control is planned and delivered.

1.1 Background

Malaria remains the leading cause of illness and death among young children in Uganda, and the country carries one of the largest malaria burdens in the world. Transmission is highly heterogeneous: it depends on altitude, rainfall, temperature, water bodies and land use, and on interventions such as insecticide-treated nets (ITNs), indoor residual spraying (IRS) and seasonal malaria chemoprevention, which are themselves targeted at districts.

Parasite prevalence among children aged 6–59 months, measured by a rapid diagnostic test (RDT) or microscopy in household surveys, is the standard population measure of malaria infection. The Malaria Indicator Survey 2024–25 tested 6,962 children in 342 clusters between November 2024 and January 2025 and found that 21.1% were positive by RDT and 12.5% by microscopy. The survey is designed to give reliable estimates for the nation, urban and rural areas and 15 sub-regions, with refugee settlements as a separate domain. It is not designed for the 146 districts.

1.2 The problem: too few children per district

In the host population (outside refugee settlements) the survey tested a median of 37 children in two clusters per district, and seven districts had no sampled cluster. Malaria clusters strongly by village, so the effective sample is smaller still: the survey's design effect for RDT prevalence is 3.3. Direct district estimates have a median standard error of 11 percentage points and a median coefficient of variation of 69%; 91% of districts fail the usual 20% reliability threshold.

Small area estimation (SAE) addresses this by borrowing strength from auxiliary data that cover every area: here the 2024 census, satellite imagery and other geospatial data on climate, terrain, water, settlement and access to health facilities, together with modelled malaria surfaces from the Malaria Atlas Project.

1.3 Relation to earlier work

Small area estimation was developed for poverty mapping (Elbers, Lanjouw and Lanjouw 2003; Molina and Rao 2010) and is standard practice in national statistics offices (Rao and Molina 2015; Corral et al. 2022). For malaria, most subnational maps come from geostatistical models that combine many surveys over many years with environmental covariates (Bhatt et al. 2015). This report uses a single, recent survey and selects among statistical and machine-learning estimators by their measured accuracy on independent data.

1.4 Objectives

1. Estimate malaria prevalence by RDT and by microscopy and the number of RDT-positive children aged 6–59 months for all 146 districts, with standard errors, confidence intervals and reliability classes.
2. Choose the estimation method by measured accuracy, comparing area-level, unit-level and machine-learning approaches on independent data.
3. Make the district estimates consistent with the official MIS sub-region figures.
4. Assess whether counties (312) and a ~500 m grid can be supported.

1.5 Structure of the report

Chapter 2 describes the data and Chapter 3 the methods. Chapter 4 presents the results and Chapter 5 their quality and validation. Chapter 6 discusses limitations and Chapter 7 concludes. Chapter 8 describes the full report and data package, including the technical appendices that are not part of this preview edition.

2 Data

The analysis combines one survey with malaria tests, the 2024 census and a large set of geospatial layers, most of them from 2022–2025.

2.1 Data sources

Table 1. Data sources

Source	Period	Used for
Uganda Malaria Indicator Survey 2024–25 (MIS)	Nov 2024 – Jan 2025	Malaria test results of 6,962 children aged 6–59 months; household characteristics; displaced cluster coordinates (342 clusters)
National Population and Housing Census 2024, 10% sample (UBOS)	2024	1.07 million households and 4.7 million persons: harmonized household variables, area averages, children for prediction
District and county boundaries 2024 (UBOS)	2024	146 districts, 312 counties
Google Open Buildings 2.5D Temporal	2023 (2018 for growth only)	Building density, height and built-up share; placing census children on the grid; building growth since 2018
Google Satellite Embedding	2024	64-dimensional summary of satellite imagery
CHIRPS rainfall	Jun 2024 – Jan 2025 (monthly); 1995–2024 (mean)	Rainfall before and during the survey; long-term rainfall
MODIS land surface temperature and EVI	2024	Day and night temperature, vegetation (annual and September–December)
Landsat 8/9 NDVI	2024	Vegetation
SRTM elevation	2000	Altitude and slope
JRC Global Surface Water	1984–2021 summary; 2021 for seasonality	Distance to permanent water (water in at least half of all 1984–2021 observations); months with water in 2021
ESA WorldCover	2021	Land cover: tree, grass, cropland, built-up, wetland
VIIRS night lights	2024	Electrification and urbanization
Malaria Atlas Project surfaces (June 2024 release)	2022	Modelled parasite prevalence, incidence, ITN access and use, IRS coverage
Malaria Atlas Project travel time	2019 (health facilities); 2015 (cities)	Travel time to health facilities (motorized and walking) and to cities

All geospatial layers were exported or resampled to a common grid of 0.0045° (about 500 m).

2.2 Survey outcomes

The outcomes are the share of de facto children aged 6–59 months who tested positive by RDT (the headline indicator) and by microscopy. RDTs detect a parasite antigen that persists for some weeks after an infection is cleared, so RDT prevalence is higher than microscopy prevalence (21.1% against 12.5% nationally). The survey-weighted national and sub-region estimates reproduce all 16 published values within 0.1 percentage points.

Host population and refugee settlements. The MIS sampled refugee settlements as a separate stratum with its own weight; the national weight is zero there. The national figures, and all district and county estimates in this report, therefore refer to the host population. Refugee settlements are reported as a separate domain (15.9% RDT, 466 children in 20 clusters).

2.3 Census

The NPHC 2024 10% sample contains 1,069,767 households and 4,693,190 persons, linked by a household identifier, with geography down to district, county, sub-county and parish. It records housing, water, sanitation, energy, assets, mosquito-net ownership, food insecurity and hand-washing for households, and education, fertility, vaccination and birth registration for persons. Seventeen sub-counties that are refugee settlements, identified by name (for example Bidi Bidi, Nakivale, Kyangwali and Kyaka II), are excluded from the host prediction frame. The frame contains 650,987 children under 5, or 5.86 million children aged 6–59 months when scaled up (children under 1 are counted at half, since only those aged 6–11 months are in the target group).

2.4 Harmonizing survey and census variables

Household variables enter the child-level models only if they are measured in the same way in both sources, as the World Bank guidelines for small area estimation recommend checking before survey and census covariates are combined (Corral et al. 2022). 24 household variables were constructed from both questionnaires and their survey and census means compared against a pre-specified rule. 18 pass, including mosquito-net ownership (87% in both), and enter the child-level models. The others, and education, which the MIS household roster does not record, are used only as census area averages.

2.5 Survey cluster locations

To protect confidentiality, the DHS Program displaces cluster coordinates by up to 2 km in urban areas and up to 5 km in rural areas (10 km for 1% of rural clusters), keeping each point within its true district (Burgert et al. 2013). District assignment is therefore exact: all 342 clusters fall within a 2024 district, and every district's clusters belong to a single MIS sub-region. County assignment is not exact, because 68% of clusters lie within the displacement distance of a county boundary. Geospatial features at cluster locations are computed as expectations over the possible true locations within the district, weighted by buildings.

3 Methodology

Eight estimators were fitted and compared on independent data; the most accurate one provides the headline figures, rescaled to the official sub-region estimates.

3.1 Overview of the estimators

Table 2. Estimators compared

Estimator	Unit of analysis	Main idea
Direct	District	Survey-weighted share of positive children in the district
FH-census, FH-geo, FH-census+geo, FH-hybrid	District	Fay–Herriot model: shrinks each direct estimate towards a regression prediction from district covariates, in proportion to its precision
EBP-unit	Child	Logistic mixed model of each child's test result with district and cluster effects, applied to every census child
ML-unit	Child	Gradient boosting or regularized logistic regression of each child's result, applied to every census child
RF-unit	Child	Random forest of each child's result, applied to every census child
ML-grid	Survey cluster, then 500 m grid	Regression of cluster prevalence on geospatial and census features, predicted for every populated grid cell and averaged over each district's children

3.2 How the data are used

The survey provides the outcome. The census provides household characteristics for the child-level models, area averages for all models, and the number and location of children used to aggregate predictions. The geospatial layers describe every 500 m cell of the country. Census children under 5 are placed on the grid within their county in proportion to building density, so that grid predictions can be averaged over where children live.

3.3 Direct estimates

Direct estimates are survey-weighted proportions. Their variances are computed by Taylor linearization with the survey's 29 host strata and clusters as primary sampling units, and smoothed with a generalized variance function (design effect 3.25 for RDT, 2.73 for microscopy) so that districts with few clusters do not have unstable or zero variances.

3.4 The main model: ML-grid

The main model[1] predicts the RDT prevalence of each survey cluster from geospatial features (building density and growth, satellite imagery, rainfall before and during the survey, temperature, vegetation, terrain, water and wetlands, land cover, travel time, night lights and modelled malaria surfaces) and district census averages (housing, water, sanitation, energy, assets, nets, food insecurity, women's education, fertility and others). It is a penalized regression, weighted by the number of children tested per cluster, chosen from a set of ridge regressions and gradient-boosting machines by district-grouped cross-validation.

[1] Similar applications: malaria prevalence mapped on a grid with machine-learning models of survey data (Bhatt et al. 2017); poverty and wealth predicted from satellite data at survey clusters and mapped on a grid (Jean et al. 2016; Chi et al. 2022); satellite-based poverty mapping to guide social protection targeting (Fisker et al. 2022).

3.5 From model to district estimates

The fitted model is applied to all 565,409 populated grid cells. A district's estimate is the average of its cells' predictions, weighted by the number of children under 5 in each cell. The model does not use the district's own direct estimate or a district random effect, so districts with and without survey data are treated alike.

Benchmarking. Unadjusted, the ML-grid estimates overstate prevalence in the low-transmission sub-regions (for example 4.6% against the survey's 0.9% in Kigezi). The headline RDT estimates are therefore rescaled within each sub-region so that the population-weighted district estimates reproduce the MIS sub-region figure exactly. This was tested in the validation (Section 5.3): benchmarking did not measurably change the accuracy of district RDT estimates, so the consistent version is preferred. For microscopy, benchmarking reduced accuracy significantly, so microscopy estimates are not benchmarked.

3.6 Uncertainty

The sampling uncertainty of the ML-grid estimates is measured by a cluster bootstrap within sub-regions (50 replicates, refitting the model each time). This captures estimation variance only, not the error of the model itself: the bootstrap variance was about a quarter of the error measured in the validation. A model-error term, calibrated on the validation and proportional to p(1 − p), is therefore added to each district's variance. For benchmarked estimates the variance also includes the sampling variance of the sub-region figure. Confidence intervals are taken from a beta distribution with the estimate's mean and variance, widened to the normal interval where that is wider; they stay within 0–100% and are asymmetric near zero.

Reliability classes. A coefficient of variation alone labels every low-prevalence district unreliable (an estimate of 1% with a standard error of 0.5 points has a CV of 50%). An estimate is therefore classed as reliable if its CV is below 20% *or* its 95% confidence interval is no wider than ±5 percentage points, use with caution if its CV is below 33% or the interval is no wider than ±10 points, and unreliable otherwise.

3.7 Alternative estimators

Fay–Herriot models[2] work on the arcsine-square-root scale of the district direct estimates, with covariates chosen by a precision-weighted LASSO (at most 10), and a parametric bootstrap for the mean squared error.

[2] Similar applications: income of small places in the United States (Fay and Herriot 1979); county and school-district child poverty in the US Small Area Income and Poverty Estimates (US Census Bureau); regional child mortality from DHS surveys (Mercer et al. 2015).

EBP-unit[3] is a logistic mixed model with child age and sex, the 18 harmonized household variables, district context variables and sub-region effects, and random effects for districts and clusters, fitted by Laplace approximation; predictions integrate over the unknown cluster effect of each census child. The district context variables are selected by the model itself, from the district census and geospatial variables, so that only information not already carried by the household variables and sub-region is added.

[3] Similar applications: poverty rates in Spanish provinces (Molina and Rao 2010); poverty proportions in Valencian counties with a unit-level logit model (Hobza and Morales 2016); child stunting and underweight in Ethiopian woredas from the DHS and the census (Sohnesen et al. 2017).

ML-unit[4] is a child-level gradient-boosting model with the same household variables and all district features. **RF-unit** is a child-level random forest (Breiman 2001): 300 regression trees, each grown on a bootstrap sample of children and choosing each split from a random third of the child, household, district and sub-region features, with survey weights. The trees' average predicted probability is averaged over each area's census children. The minimum leaf size was chosen by district-grouped cross-validation.

[4] Similar applications: household welfare in several developing countries, comparing gradient boosting and other machine-learning methods with EBP (Merfeld and Newhouse 2023); household income in Mexico with mixed effects random forests (Krennmair and Schmid 2022); health, housing and well-being in Dutch neighbourhoods with gradient boosting (Viljanen et al. 2022).

3.8 Model selection

The method is chosen by accuracy in a split-sample validation (Section 5.3): the most accurate method for RDT prevalence provides the headline estimates, and its benchmarked version is used when it is not significantly less accurate (paired t below 2). The same method is used for microscopy unless another is significantly better.

4 Results

Malaria is concentrated in a belt across Lango, Acholi, Teso and Karamoja, while the south-west and Kampala are close to elimination.

> **Preview edition.** This chapter shows summary maps in four classes and the priority districts in rank order. The full report has detailed maps of every indicator, ranked tables with estimates and confidence intervals, and the data package gives every district and county estimate (Chapter 8).

4.1 Malaria prevalence by district

District RDT prevalence ranges from 0.1% (Buhweju) to 68.3% (Kole). 30 districts exceed 40% and 19 exceed 50%; 28 of the 30 are in Lango, Teso, Acholi and Karamoja, and the other two, Kaliro and Buyende, in Busoga (Figure 1, Table 3). At the other end, 32 districts are below 5% and nine below 1%: Buhweju, Kabale, Kalangala, Rukiga, Kiruhura, Rukungiri, Kampala, Mbarara City and Rubanda.

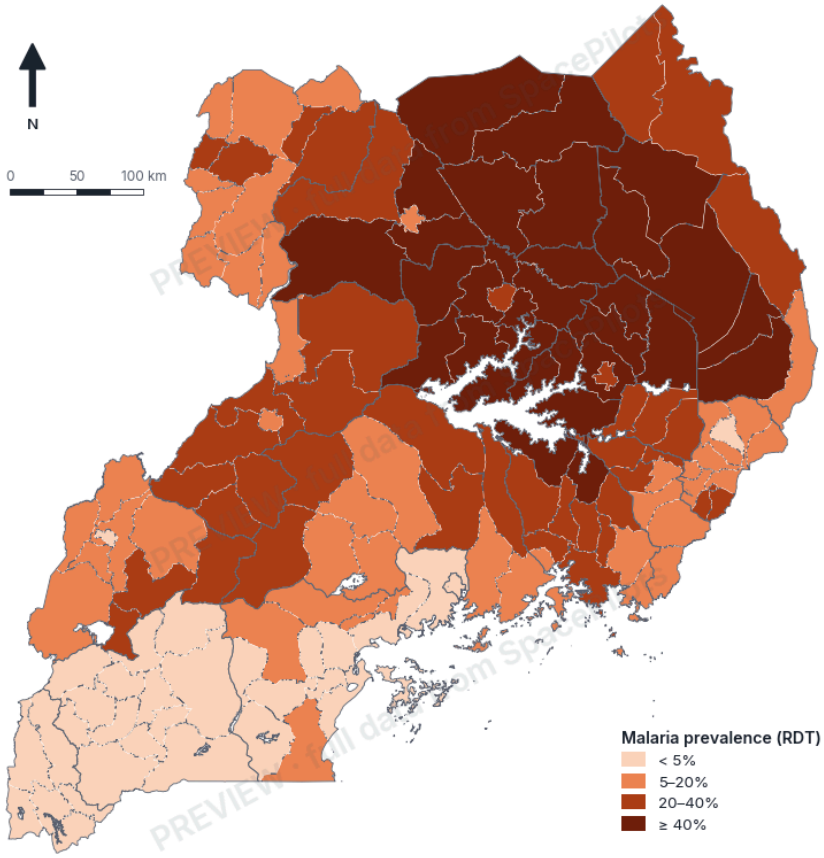


Figure 1. Malaria prevalence exceeds 40% in 30 districts, mostly in Lango, Teso, Acholi and Karamoja. Malaria prevalence (RDT) among children aged 6–59 months by district, 2024–25, in four classes.

Source: SpacePilots small area estimates (ML-grid, benchmarked) from MIS 2024–25, NPHC 2024 and geospatial data. Grey lines are MIS sub-regions; lakes are shown in white. The data package gives each district's estimate with its standard error, 95% confidence interval, CV and reliability class.

Table 3. Priority districts: the ten with the highest malaria prevalence (RDT) and the ten with the most RDT-positive children aged 6–59 months, 2024–25

Rank	Highest prevalence	Most positive children
1	**Kole (Lango)**	**Oyam (Lango)**
2	Kwania (Lango)	Kamuli (Busoga)
3	Lira (Lango)	**Agago (Acholi)**
4	**Oyam (Lango)**	Buyende (Busoga)
5	**Agago (Acholi)**	**Kole (Lango)**
6	**Alebtong (Lango)**	Serere (Teso)
7	Otuke (Lango)	Kayunga (North Buganda)
8	Dokolo (Lango)	Yumbe (West Nile)
9	Pader (Acholi)	**Alebtong (Lango)**
10	Kapelebyong (Teso)	Kotido (Karamoja)

Districts in rank order, with their sub-region. Bold: in the top ten on both measures. Estimates, confidence intervals and the ranking of all 146 districts are in the data package.

4.2 Differences within sub-regions

Sub-region figures hide large differences. Within a sub-region, the highest and lowest district typically differ by 25 percentage points, and by more than 40 points in Lango, Busoga, Acholi and Karamoja (Figure 2). Cities stand out: Gulu City, Lira City, Soroti City and Jinja City are far below the rest of their sub-regions. In Karamoja, Amudat is far below Kotido and Nakapiripirit. Only the low-transmission sub-regions (Kampala, Kigezi, Ankole and South Buganda) are uniform.

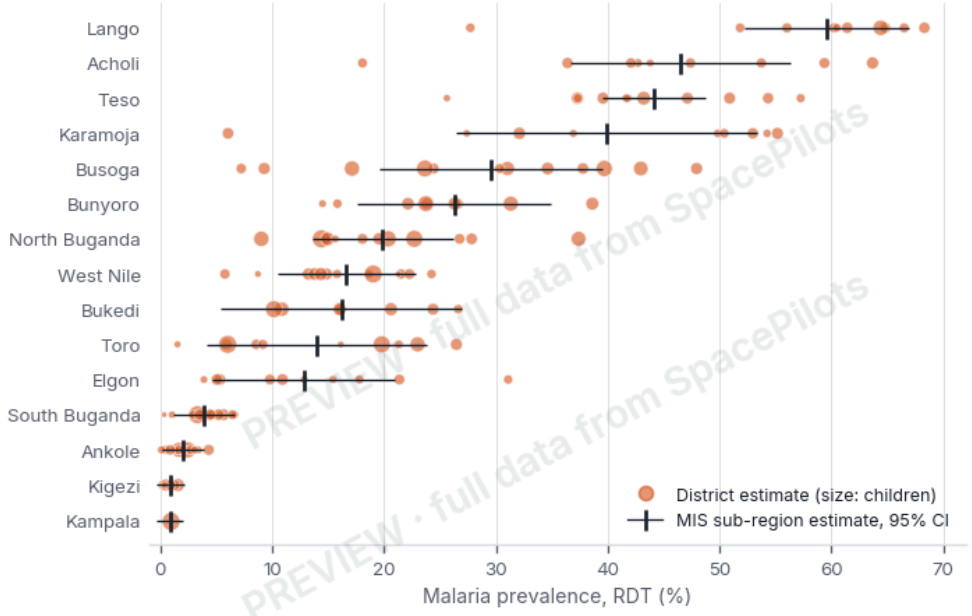


Figure 2. Districts in the same sub-region typically differ by 25 percentage points. District estimates of malaria prevalence (RDT) and MIS sub-region estimates, 2024–25.

Dots: district estimates, sized by the number of children aged 6–59 months. Bars: MIS 2024–25 sub-region estimates with 95% confidence intervals.

4.3 Number of malaria-positive children

An estimated 1.26 million children aged 6–59 months test positive by RDT. Lango, Teso, Acholi and Karamoja hold 46% of them but only 21% of the children. Oyam has the most, followed by Kamuli, Agago, Buyende and Kole (Figure 3). Four districts, Oyam, Agago, Kole and Alebtong, are among the ten highest on both prevalence and numbers (Table 3). The 20 districts with the most positive children hold 37% of the total, and the 40 with the most hold 61%.

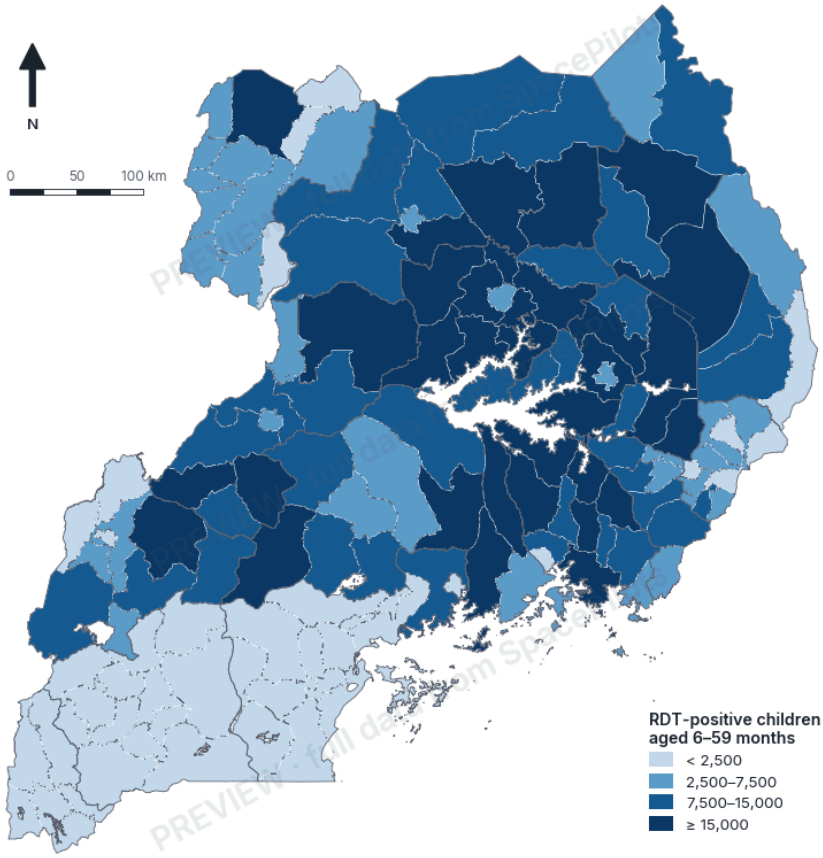


Figure 3. 37% of the 1.26 million RDT-positive children live in 20 of the 146 districts. Estimated number of RDT-positive children aged 6–59 months by district, 2024–25, in four classes.

Number = prevalence × children aged 6–59 months from the 2024 census (10% sample × 10), host population. The data package gives each district's number with its confidence interval.

4.4 Prevalence and numbers compared

Prevalence and numbers largely point to the same districts. Dividing districts at the national prevalence (21.5%) and the median number of positive children (about 6,900) gives four groups (Figure 4):

- **Higher prevalence, more positive children** (60 districts, 74% of positive children): the northern and eastern belt and the populous Busoga districts, led by Oyam, Kamuli, Agago and Buyende.
- **Higher prevalence, fewer positive children** (5 districts, 2%): small districts such as Karenga, Maracha, Adjumani and Moroto.
- **Lower prevalence, more positive children** (13 districts, 12%): populous districts with moderate prevalence, led by Yumbe, Mukono, Kyenjojo and Mubende.
- **Lower prevalence, fewer positive children** (68 districts, 12%).

The 20 districts with the highest prevalence and the 20 with the most positive children share 11 districts. Programmes that aim to reduce the national number of infections still need to include the populous moderate-prevalence districts of the third group.

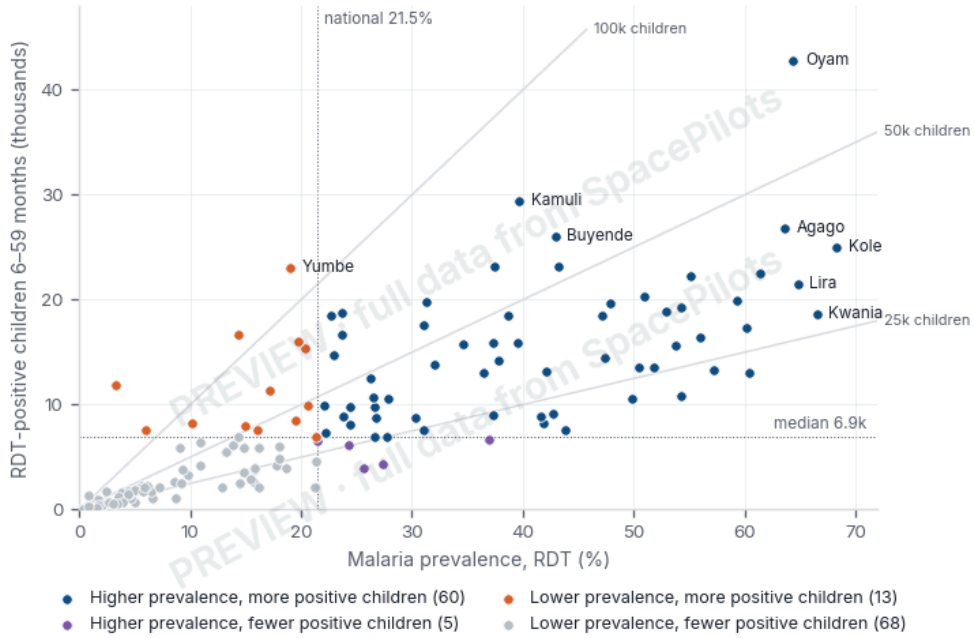


Figure 4. Three quarters of positive children live in high-prevalence districts with large numbers. Malaria prevalence (RDT) and number of RDT-positive children aged 6–59 months by district, 2024–25.

Dotted lines: national prevalence from the district estimates (21.5%) and median number per district. Grey lines: equal numbers of children aged 6–59 months.

4.5 Microscopy

Microscopy prevalence is 12.5% nationally, about 0.73 million children. District estimates follow the same geography as RDT prevalence (correlation 0.91) and are highest in Alebtong, Nabilatuk, Napak, Kaliro and Lira. Because the microscopy estimates are not benchmarked, their sub-region aggregates differ somewhat from the MIS figures: they are 1–2 points higher in Kampala, Kigezi and Ankole, where the survey found almost no infections (Table 4).

4.6 Sub-regions and refugee settlements

Table 4 compares the official MIS sub-region figures with the aggregated district estimates. Benchmarking makes the RDT aggregates equal to the MIS figures. Refugee settlements, a separate MIS domain, have an RDT prevalence of 15.9% (±13.1 points); they are not part of the district estimates, and the survey sample (20 clusters) is too small to estimate individual settlements.

Table 4. Sub-region estimates: MIS 2024–25 figures and aggregated district estimates

Sub-region	RDT, MIS (95% CI ±)	Microscopy, MIS	Microscopy, districts aggregated	Positive children (RDT)	Children 6–59 months	MIS clusters
Lango	59.6% (±7.4)	31.7%	29.8%	197,600	331,600	21
Acholi	46.5% (±9.8)	23.3%	24.1%	124,700	268,400	20
Teso	44.1% (±4.6)	22.9%	22.1%	156,500	355,100	20
Karamoja	39.9% (±13.4)	26.3%	23.6%	103,300	258,900	20
Busoga	29.5% (±9.9)	21.0%	18.1%	175,500	594,400	23
Bunyoro	26.3% (±8.7)	17.6%	15.3%	102,500	390,200	22
North Buganda	19.9% (±6.3)	11.4%	9.3%	127,600	642,100	25
West Nile	16.6% (±6.2)	8.0%	11.3%	80,400	483,300	22
Bukedi	16.2% (±10.8)	11.9%	14.8%	54,700	338,100	19
Toro	13.9% (±9.8)	9.2%	7.4%	63,100	453,600	22
Elgon	12.8% (±8.2)	6.1%	6.2%	35,800	279,800	20
South Buganda	3.9% (±2.7)	2.3%	3.7%	27,700	714,200	25
Ankole	2.0% (±1.9)	1.5%	3.2%	8,000	394,700	22
Kigezi	0.9% (±1.3)	0.6%	2.2%	1,800	201,200	20
Kampala	0.9% (±1.2)	0.4%	1.5%	1,300	154,500	21

Sub-region	RDT, MIS (95% CI ±)	Microscopy, MIS	Microscopy, districts aggregated	Positive children (RDT)	Children 6–59 months	MIS clusters
Refugee settlements	15.9% (±13.1)	9.8%	—	—	—	20

Sorted by RDT prevalence. District RDT estimates are benchmarked, so they aggregate exactly to the MIS figure. Refugee settlements are a separate MIS domain and are not included in the district estimates.

4.7 County estimates (experimental)

> **Experimental.** County results show where to look more closely within a district. They should not be used for allocating resources without further evidence.

The same model, averaged over each county's grid cells, gives estimates for all 312 counties. They range from 0% to 71%; 59 counties exceed 40%. Within a district, counties typically differ by only 3 percentage points, and by more than 20 points in Bulambuli, Kitgum and Kween. Because survey clusters cannot be assigned to counties reliably, county differences within a district come from the geospatial and census features alone. Figure 5 shows one sub-region, Busoga, as a sample.

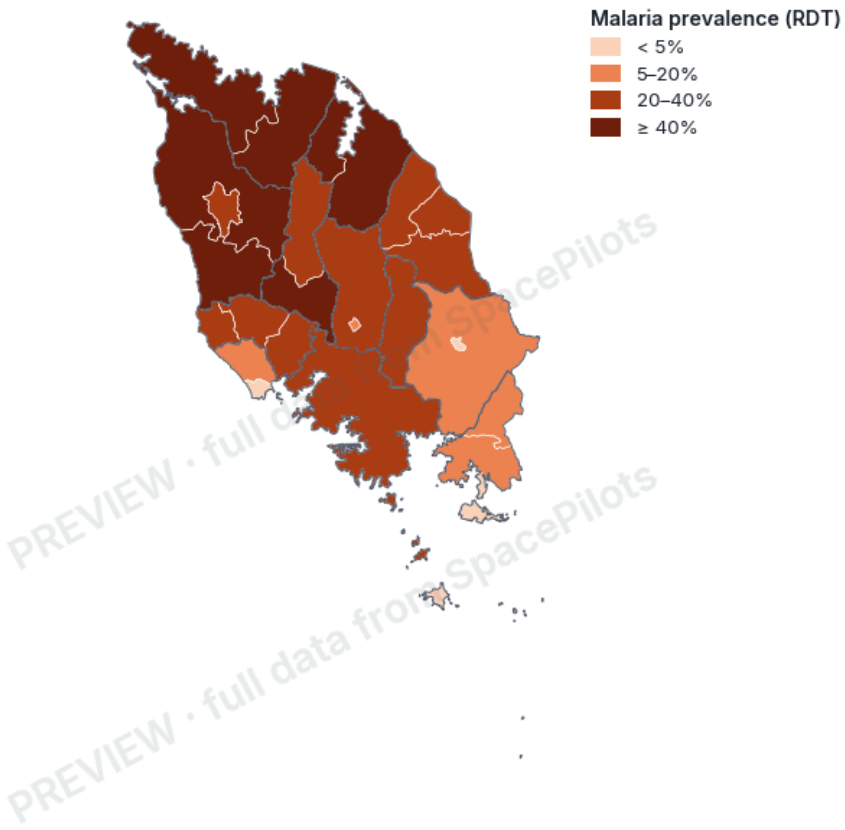


Figure 5. Sample of the county-level product: Busoga sub-region. Experimental malaria prevalence (RDT) estimates by county, 2024–25, in four classes. Within-district variation comes from geospatial and census features only. Use with great caution. Dark lines are district boundaries. The data package covers all 312 counties.

5 Quality and validation

The district estimates are 2.5 times as accurate as direct survey estimates in mean squared error, but estimates for mid-prevalence districts remain uncertain.

5.1 Precision

Table 5. Precision of the estimates by level and indicator

Level	Indicator	Median SE (pp)	Median 95% CI width (pp)	Reliable	Use with caution	Unreliable
District (146)	Malaria (rdt)	7.8	32.6	38	52	56
District (146)	Malaria (microscopy)	4.0	16.9	43	98	5
County (experimental) (312)	Malaria (rdt)	7.6	31.7	89	106	117
County (experimental) (312)	Malaria (microscopy)	4.0	16.8	102	191	19

Reliable: CV below 20% or 95% CI half-width at most 5 percentage points (pp); use with caution: CV below 33% or half-width at most 10 pp; otherwise unreliable.

The median standard error of district RDT estimates is 7.8 percentage points. Precision depends strongly on the level of prevalence. Districts with very high prevalence (above 50%) are mostly reliable, because their CV is small; most districts below 5% are reliable because their absolute uncertainty is small. All 56 unreliable districts have a prevalence between 5% and 35%, where the binomial variance is large and the model distinguishes districts less well. Microscopy estimates are more precise in absolute terms (median standard error 4.0 points); only five are unreliable.

Districts with no survey data. Kalangala, Masaka, Lyantonde, Kalungu (South Buganda), Bukwo (Elgon), Luuka (Busoga) and Obongi (West Nile) had no sampled cluster. Their estimates rest entirely on geospatial and census information. Bukwo and Obongi are unreliable; the others are reliable or use with caution.

5.2 Gain over direct survey estimates

The model reduces the median standard error from 11.0 to 7.8 percentage points and gives estimates for the seven districts without survey data. It moves imprecise direct estimates towards values supported by the geospatial and census data. 26 districts have a direct estimate of 0%, because none of their few tested children was positive; the model gives them small but non-zero values.

5.3 Independent accuracy test

Model-based standard errors assume the model is correct, so accuracy was also measured directly. In each of 20 replicates, the survey clusters in each district with at least two clusters (90 districts) were split at random into two halves. Every method was refitted on one half, and its district estimates were compared with direct estimates from the other, independent half. Subtracting the other half's sampling variance from the squared differences gives each method's mean squared error. Benchmarked versions were benchmarked to the first half's sub-region estimates (Figure 6, Table 6).

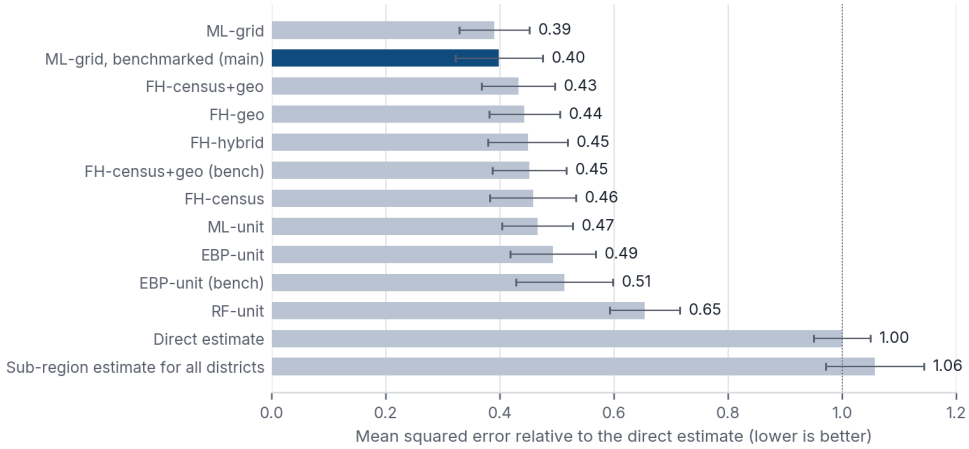


Figure 6. Model-based methods cut the error of direct estimates by 35–61%; ML-grid is the most accurate. Split-sample validation error for district RDT prevalence, by method.

Bars: mean squared error relative to the direct estimate, 20 replicates; whiskers: 95% intervals.

Table 6. Split-sample validation results, district level (estimates fitted on half the survey)

Method	RDT RMSE (pp)	MSE relative to direct	Paired t vs ML-grid, RDT	Microscopy RMSE (pp)
ML-grid	8.0	0.39	—	5.3
ML-grid, benchmarked (main)	**8.1**	**0.40**	**0.5**	**5.9**
FH-census+geo	8.4	0.43	1.5	5.3
FH-geo	8.5	0.44	1.8	5.0
FH-hybrid	8.6	0.45	2.2	5.4
FH-census+geo (bench)	8.6	0.45	2.0	6.1
FH-census	8.7	0.46	2.4	5.9
ML-unit	8.7	0.47	2.2	5.7
EBP-unit	9.0	0.49	3.1	6.2
EBP-unit (bench)	9.2	0.51	3.5	6.4
RF-unit	10.3	0.65	7.0	6.3
Direct survey estimate	12.8	1.00	14.0	9.0
Sub-region estimate for all districts	13.2	1.06	18.7	8.8

RMSE of half-sample estimates (90 districts with at least two survey clusters, 20 replicates). Positive t means a larger error than unbenchmarked ML-grid, the most accurate method; differences are paired within replicates.

All model-based methods except the random forest reduce the mean squared error of direct estimates by 49–61%, and their differences are modest; the random forest reduces it by 35%. ML-grid is the most accurate (−61%); its benchmarked version is statistically indistinguishable (paired t 0.5). FH-census+geo is not significantly worse (t 1.5), while ML-unit (t 2.2), EBP-unit (t 3.1) and RF-unit (t 7.0) are significantly less accurate. Assigning every district its sub-region's estimate is no better than the direct estimate itself, which shows that the gain comes from the district-level covariates, not from pooling within sub-regions. In root mean squared error, the main estimates are 37% smaller than direct estimates (8.1 against 12.8 points for half-sample estimates).

5.4 Consistency with official figures

The benchmarked RDT estimates aggregate exactly to the 15 MIS sub-region figures, and their population-weighted national figure (21.5%) is within half a standard error of the survey's 21.1%; the difference arises because the census and the survey weights distribute children slightly differently across sub-regions.

6 Discussion and limitations

The estimates are precise enough to identify the highest- and lowest-burden districts, but they describe one season and inherit the limits of the predictors.

6.1 What drives the results

The geography is driven mainly by environment and settlement. The most important predictors in the cluster model are cropland cover, district electricity and housing quality, the first satellite embedding component, the 2022 Malaria Atlas Project prevalence surface, building height, slope and elevation, and night-time temperature. High transmission coincides with the low-lying, cultivated, rural plains around Lake Kyoga; low transmission with the cooler south-western highlands and with cities. Rainfall timing also matters: prevalence correlates positively with rainfall in June–August 2024 and negatively with rainfall in December–January, reflecting the contrast between the north's single long rainy season and the south's two shorter ones.

6.2 Why the grid model performs well

Malaria varies sharply over short distances with altitude, water and settlement, which satellite data capture at 500 m but district averages do not. Household characteristics, by contrast, predict a child's infection only weakly once location is known: in the unit-level model, piped water and crowding are the only household variables with clear effects, and owning a net has a small, non-significant association. This is why the child-level models do not outperform the grid model.

6.3 Limitations

- **One season.** The survey took place from November 2024 to January 2025, during and after the second rainy season. Malaria prevalence varies by season, and the timing of fieldwork varied across the country.
- **Mid-prevalence districts.** 56 district estimates, all between 5% and 35%, are unreliable. Their ranking should not be over-interpreted.
- **Interventions.** District-level data on indoor residual spraying, net campaigns and chemoprevention enter only through modelled surfaces from 2022 and through the census net question. Up-to-date programme data would likely improve the estimates.
- **Malaria Atlas Project surfaces.** Only the 2022 surfaces, which predate the survey, are used. A newer release may already include the MIS 2024–25 data and was therefore not used as a covariate.
- **Refugee settlements.** Settlements are excluded from district estimates; their separate estimate is imprecise.
- **Counties and grid.** County and grid estimates rely on covariates for within-district differences and are not validated against independent data.

7 Conclusions and recommendations

District malaria estimates for Uganda can be produced from a single survey, the census and open geospatial data, with a measured gain in accuracy over the survey alone.

7.1 Conclusions

- Malaria among young children is concentrated in a belt across Lango, Acholi, Teso and Karamoja, where 30 districts exceed 40% prevalence, and is close to elimination in the south-west and Kampala.
- Sub-region averages hide large differences: districts in the same sub-region typically differ by 25 points, and cities are far below their surroundings.
- A machine-learning model on a 500 m grid, rescaled to the official sub-region figures, was the most accurate of the methods tested; it cuts the mean squared error of direct estimates by 60%.
- Estimates are reliable at the extremes of prevalence and uncertain in the middle.

7.2 Recommendations for users

1. Use the district estimates with their confidence intervals and reliability classes; give most weight to the "reliable" and "use with caution" classes.
2. Use the number of positive children together with prevalence when allocating commodities and campaigns.
3. Treat county estimates as experimental.

References

Bhatt, S. et al. (2015). The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* 526: 207–211.

Bhatt, S., Cameron, E., Flaxman, S.R., Weiss, D.J., Smith, D.L. and Gething, P.W. (2017). Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of the Royal Society Interface* 14: 20170520.

Breiman, L. (2001). Random forests. *Machine Learning* 45(1): 5–32.

Burgert, C.R., Colston, J., Roy, T. and Zachary, B. (2013). *Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys.* DHS Spatial Analysis Reports No. 7. Calverton, Maryland: ICF International.

Chi, G., Fang, H., Chatterjee, S. and Blumenstock, J.E. (2022). Microestimates of wealth for all low- and middle-income countries. *Proceedings of the National Academy of Sciences* 119(3): e2113658119.

Corral, P., Molina, I., Cojocaru, A. and Segovia, S. (2022). *Guidelines to Small Area Estimation for Poverty Mapping.* Washington, DC: World Bank.

Elbers, C., Lanjouw, J.O. and Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica* 71(1): 355–364.

Fay, R.E. and Herriot, R.A. (1979). Estimates of income for small places: an application of James–Stein procedures to census data. *Journal of the American Statistical Association* 74: 269–277.

Fisker, P., Gallego-Ayala, J., Malmgren-Hansen, D., Sohnesen, T.P. and Murrugarra, E. (2022). *Guiding Social Protection Targeting Through Satellite Data in São Tomé and Príncipe.* Social Protection & Jobs Discussion Paper No. 2212. Washington, DC: World Bank. https://openknowledge.worldbank.org/entities/publication/b0cb02cc-4484-5f3c-a226-fe69b4934210

Hobza, T. and Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of Official Statistics* 32(3): 661–692.

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B. and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301): 790–794.

Krennmair, P. and Schmid, T. (2022). Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society Series C* 71(5): 1865–1894.

Mercer, L.D., Wakefield, J., Pantazis, A., Lutambi, A.M., Masanja, H. and Clark, S. (2015). Space–time smoothing of complex survey data: small area estimation for child mortality. *Annals of Applied Statistics* 9(4): 1889–1905.

Merfeld, J.D. and Newhouse, D. (2023). *Improving Estimates of Mean Welfare and Uncertainty in Developing Countries.* Policy Research Working Paper 10348. Washington, DC: World Bank.

Molina, I. and Rao, J.N.K. (2010). Small area estimation of poverty indicators. *Canadian Journal of Statistics* 38(3): 369–385.

Rao, J.N.K. and Molina, I. (2015). *Small Area Estimation.* 2nd edition. Hoboken, NJ: Wiley.

Sohnesen, T.P., Ambel, A.A., Fisker, P., Andrews, C. and Khan, Q. (2017). Small area estimation of child undernutrition in Ethiopian woredas. *PLoS ONE* 12(4): e0175445.

Uganda Bureau of Statistics (2024). *National Population and Housing Census 2024: Final Report, Volume 1 (Main).* Kampala: UBOS.

Uganda Malaria Indicator Survey 2024–25 [datasets UGHR91FL, UGPR91FL, UGGE91FL, UGGC91FL]. Rockville, Maryland: ICF (The DHS Program).

US Census Bureau. Small Area Income and Poverty Estimates (SAIPE) Program. https://www.census.gov/programs-surveys/saipe.html

Viljanen, M., Meijerink, L., Zwakhals, L. and van de Kassteele, J. (2022). A machine learning approach to small area estimation: predicting the health, housing and well-being of the population of Netherlands. *International Journal of Health Geographics* 21: 4.

8 The full report and data package

This preview is an abridged edition. The full report and the data package give the estimates for every area, the complete set of maps and the technical documentation needed for review.

8.1 Data package

Table 7. Contents of the data package

Component	Coverage	Contents
District estimates	146 districts	Malaria prevalence by RDT and by microscopy, each with standard error, 95% confidence interval, CV and reliability class; number of RDT-positive children; children aged 6–59 months; survey sample sizes and direct estimates
Sub-region estimates	15 sub-regions and refugee settlements	MIS figures with confidence intervals, aggregated district estimates and consistency checks
County estimates (experimental)	312 counties	Malaria prevalence (RDT), its uncertainty and the number of RDT-positive children
Alternative estimates	Districts	The estimates before benchmarking to the sub-region figures, and the alternative estimators for sensitivity analysis
Maps and charts	All levels	High-resolution maps of every indicator and all charts in the full report

Tables are supplied as CSV files that join to the 2024 census boundaries.

8.2 What the full report adds

- **Results.** Detailed district maps (seven classes) of malaria prevalence by RDT and by microscopy, the numbers of RDT-positive children and the reliability of each estimate; ranked tables of the districts with the highest prevalence and the most positive children, with estimates, confidence intervals and numbers; the full county map; and exploratory estimates on the 500 m grid.
- **Methods.** The comparison of every household variable in the survey and the census; the features, settings and most important variables of the main model; the specification, coefficients and configuration of the alternative models.
- **Quality.** The map of reliability classes and the comparison of model and direct estimates for every district with survey data.
- **Technical appendices.** Data preparation; the main model; the alternative models; the validation design and replicate-level results; agreement between methods; benchmarking to the sub-region figures; and the exploratory grid estimates.

For the full report and data package, contact SpacePilots at info@spacepilots.io.