Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Location, location, location: implementation-unit identity dominates reported schistosomiasis classification across four East African countries

Celestin Nzanzu Mudogo

celestin.mudogo@unikin.ac.cd

University of Kinshasa

Jean-Louis Mumbere Katembo

University of Kinshasa

Humphrey Deogratias Mazigo

Catholic University of Health and Allied Sciences

Johan G. L. Verheyden

Aries Consult Kenya

Article

Keywords:

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11533118/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Location, location, location: implementation-unit identity dominates reported schistosomiasis classification across four East African countries

Celestin Nzanzu Mudogo[1,*], Jean-Louis Mumbere Katembo[2,3], Humphrey Deogratias Mazigo[4,5] and Johan G. L. Verheyden[6]

[1] Department of Basic Sciences, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo
[2] Department of Pathology, Faculty of Medicine, University of Kinshasa, Kinshasa, Democratic Republic of the Congo
[3] Department of Pathology, Faculty of Medicine, Université Catholique du Graben, Butembo, Democratic Republic of the Congo
[4] Department of Parasitology, School of Medicine, Catholic University of Health and Allied Sciences, Mwanza, Tanzania
[5] School of Public Health, Catholic University of Health and Allied Sciences, Mwanza, Tanzania
[6] Aries Consult Kenya, Nairobi, Kenya
*Corresponding author: Celestin Nzanzu Mudogo, University of Kinshasa, Kinshasa, Democratic Republic of the Congo; e-mail: celestin.mudogo@unikin.ac.cd

Author e-mail: JGLV, jverheyden@ariesconsult.eu
ORCID iDs: CNM 0000-0003-0524-9481; JLMK 0009-0008-3120-4483; HDM 0000-0001-5219-6234; JGLV 0000-0002-6129-2794

Abstract

Climate information supports early warning for some East African diseases, but it is unclear whether it helps classify schistosomiasis endemicity once local context is known. We analysed 16,438 implementation-unit (IU) years from 1,754 IUs in Kenya, Uganda, mainland Tanzania and Ethiopia (2014–2025). A hierarchical Bayesian ordered-logistic model related the WHO ESPEN endemicity class to rainfall and temperature anomalies, El Niño–Southern Oscillation (ENSO) phase, treatment coverage and year-level variation. Most variability in the reported class lay between IUs (89.5%; 89% credible interval 85.0–93.2), with 9.3% between countries and 1.2% between years. Rainfall had a non-monotonic association with class. ENSO-phase estimates were imprecise, their credible intervals included zero, and the panel holds only three El Niño and two La Niña years. Adding the fitted climate terms changed predicted IU-level risk by less than one percentage point on average. Because classifications rarely change (5% of consecutive-year transitions), the results describe reported programme classification, not infection incidence. They suggest that country-level ENSO forecasts are a weak basis for changing IU-level decisions; whether finer climate data would do better needs separate evaluation.

Introduction

Climate information already supports early warning for some East African diseases. Satellite-based Rift Valley fever (RVF) risk mapping, which uses ENSO-related sea-surface temperature, rainfall and vegetation anomalies, predicted the 2006–07 Horn of Africa outbreak and gave a two-to-six-week warning before human cases were identified1, and El Niño has been linked to a shifted cholera burden in East Africa2. Whether schistosomiasis, a chronic disease transmitted through freshwater snails, shows a comparable climate signal is less clear.

A rainfall–schistosomiasis link is plausible. Flooding can disperse Bulinus and Biomphalaria snail habitat and increase human–water contact, whereas extreme flooding can flush snails out, which motivates testing a non-monotonic rainfall relationship; temperature affects snail development and cercarial shedding. But programme data are noisy. The WHO field protocol samples about 20 school-age children at each of 15 sites per district3, and the accuracy of reagent-strip diagnosis itself depends on infection intensity4, so the class assigned to a single implementation unit (IU) is unstable. This is a reason to model IUs hierarchically, with partial pooling. A second hazard is pseudoreplication: the same IU is observed across up to twelve years, while ENSO phase varies only between years, so without a year-level term a model can attribute to “El Niño” or “La Niña” what is simply ordinary year-to-year variation.

Bayesian geostatistical models are well established for schistosomiasis mapping, from West Africa5,7 to eastern Africa6. Three recent studies are closest to ours. Clements et al.8 fitted Bayesian geostatistical models to parasitological data from 143 schools in Tanzania; land surface temperature and rainfall predicted Schistosoma haematobium prevalence only until spatial correlation was modelled, whereas distance to water and minimum temperature remained significant for S. mansoni. Singer et al.9 reanalysed randomised trials covering 589 communities in five endemic countries to predict, at baseline and before treatment, which communities would remain hotspots under mass drug administration (MDA); baseline prevalence was a leading predictor in all models, and accuracy was lower in countries absent from the training data and for hotspots defined by relative prevalence reduction. Mathewson et al.10 combined a zero-inflated Poisson model with regression kriging on 2021 survey data from 17,400 schoolchildren in six Tanzanian regions to estimate ward-level prevalence, and showed that ward-level targeting would move treatment away from 193 wards that district-level MDA over-treats and towards 149 wards it would omit.

These studies address prevalence or hotspot status at surveyed locations. Less is known about the operational classification that programmes actually use, the IU endemicity band reported annually to

WHO: how its variability is divided between persistent local differences and country- and year-level shifts, and whether climate information adds anything to a baseline that knows only the IU and country. We asked three questions of the IU classifications reported to WHO ESPEN. RQ1: is there a detectable association between ENSO phase, rainfall or temperature anomalies and an IU's endemicity class, after accounting for persistent local differences and shared-year variation? RQ2: at which level (IU, country or year) does most of the variability in class lie? RQ3: does adding climate information change predicted risk materially relative to a baseline that knows only the IU and country? RQ1 motivated the analysis; RQ2 and RQ3 took shape as results emerged, are descriptive, and the analysis was not preregistered. The model does not represent spatial autocorrelation or predict prevalence for unsurveyed locations, so it complements rather than replaces geostatistical mapping.

Results

Model convergence

The model converged on all parameters (R-hat 1.00–1.01, no divergent transitions across 4,800 post-warmup draws). The fitted sample comprised 16,438 IU-years from 1,754 IUs.

Variance decomposition

Of the between-group variance in the model, 89.5% was attributable to IU identity, 9.3% to country and 1.2% to year (Fig. 1). Between-IU differences therefore account for roughly ten times more of the variation in classification than between-country differences, and roughly seventy times more than year-to-year differences. The 89% credible intervals were 85.0–93.2% (IU), 5.6–13.8% (country) and 0.5–2.3% (year). Computing the shares instead from the realised group-level effects gave 78% (IU), 21% (country) and 0.8% (year): the ranking is unchanged, although country matters more under this alternative. These shares reflect the persistence of ESPEN classifications as much as underlying risk: in the fitted sample an IU's band changed in only 5.2% (766 of 14,684) of consecutive observed-year transitions, and 1,166 of the 1,754 IUs (66.5%) never changed band. The year share is small partly because the outcome rarely moves, not only because climate-driven year-to-year variation is small. In practical terms, knowing which IU a record belongs to tells far more about its classification than knowing its country or year.

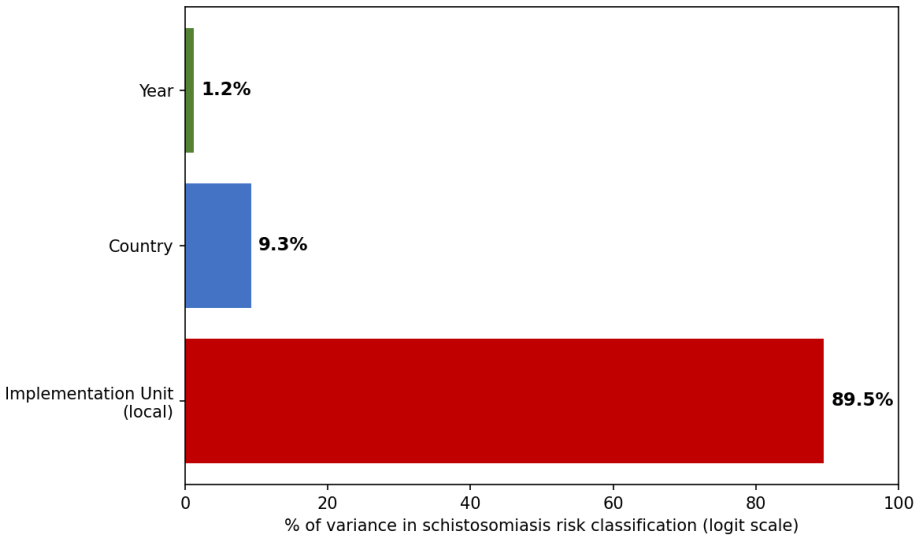

Fig. 1 Variance decomposition of schistosomiasis classification across implementation unit, country and year. Share of the model's between-group variance attributable to each level (posterior means). IU accounts for 89.5% (89% credible interval 85.0–93.2), country for 9.3% (5.6–13.8) and year for 1.2% (0.5–2.3). Shares are computed from the group-level standard deviations and exclude the residual variance of the logistic link.

Climate and treatment effects

Two covariates had credible intervals that excluded zero. The quadratic rainfall term was negative (posterior mean −0.069, 89% credible interval −0.098 to −0.04, posterior probability of a negative effect >99.9%), indicating a non-monotonic relationship: predicted class is highest under near-normal rainfall and lower under both wetter- and drier-than-normal conditions. Treatment coverage was positive (0.180, 89% credible interval 0.13–0.23, posterior probability >99.9%), which we interpret as a treatment-targeting signal and not as treatment failure. Neither the linear rainfall term, the temperature main effect, nor either ENSO-phase interaction term had an interval that excluded zero (Fig. 2; annual rainfall and temperature anomalies by country are shown in Figs 3 and 4).

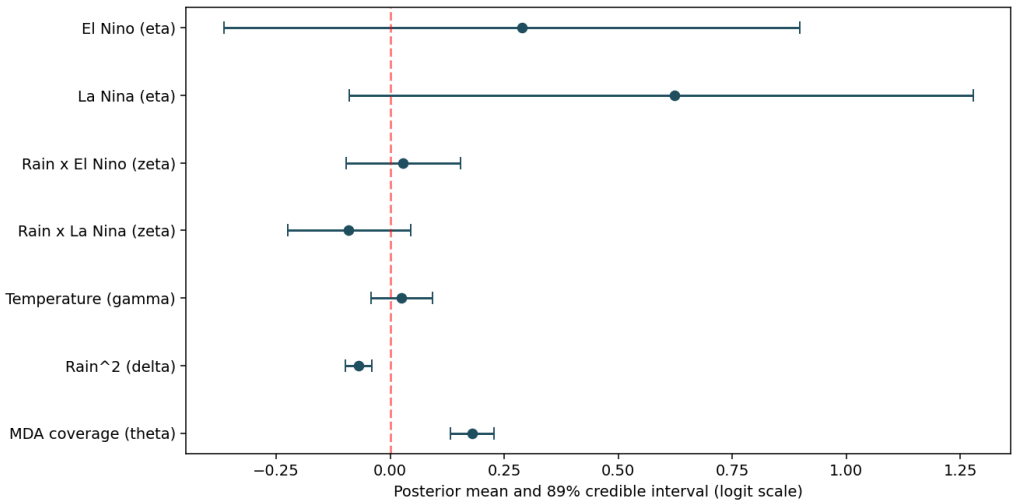

Fig. 2 Fixed-effect estimates from the hierarchical ordered-logistic model. Points show posterior means and bars show 89% credible intervals on the logit scale (16,438 IU-years from 1,754 IUs). Only the quadratic rainfall term and MDA coverage have intervals that exclude zero. The dashed vertical line marks zero.

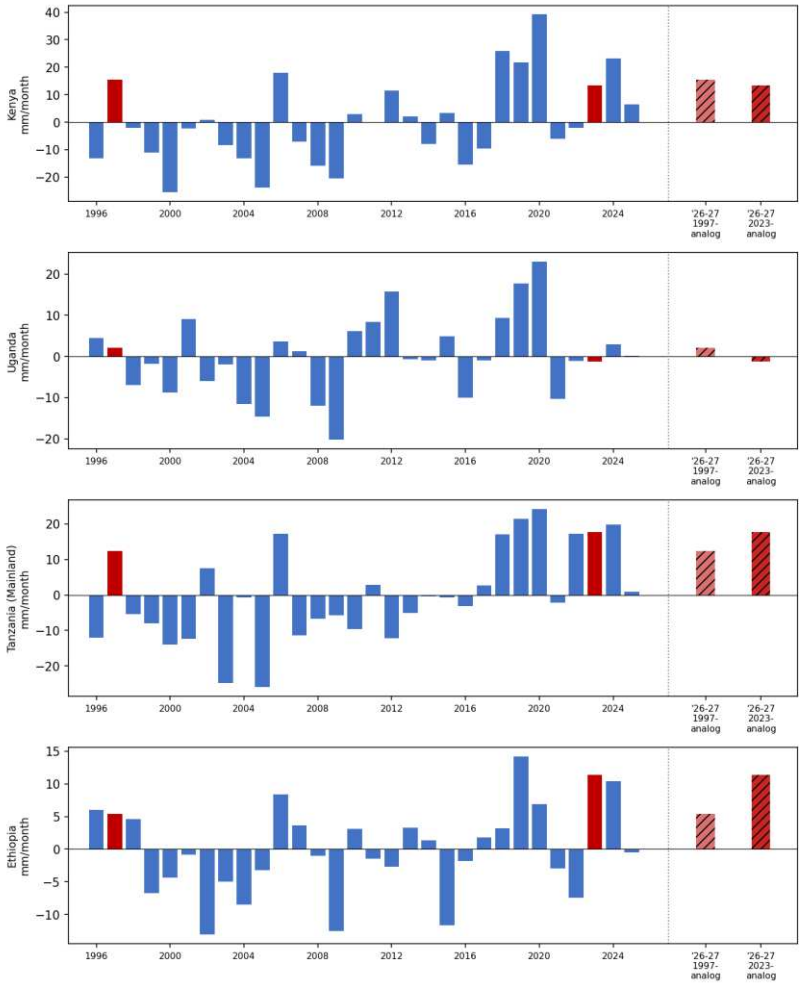

Fig. 3 Historical annual rainfall anomaly by country, 1996–2025. Bars show the annual rainfall anomaly (mm per month) for Kenya, Uganda, Tanzania (Mainland) and Ethiopia. Red bars mark the two analog years (1997 and 2023) used for the illustrative scenario; hatched bars repeat their observed values as scenario inputs. No forecast is shown.

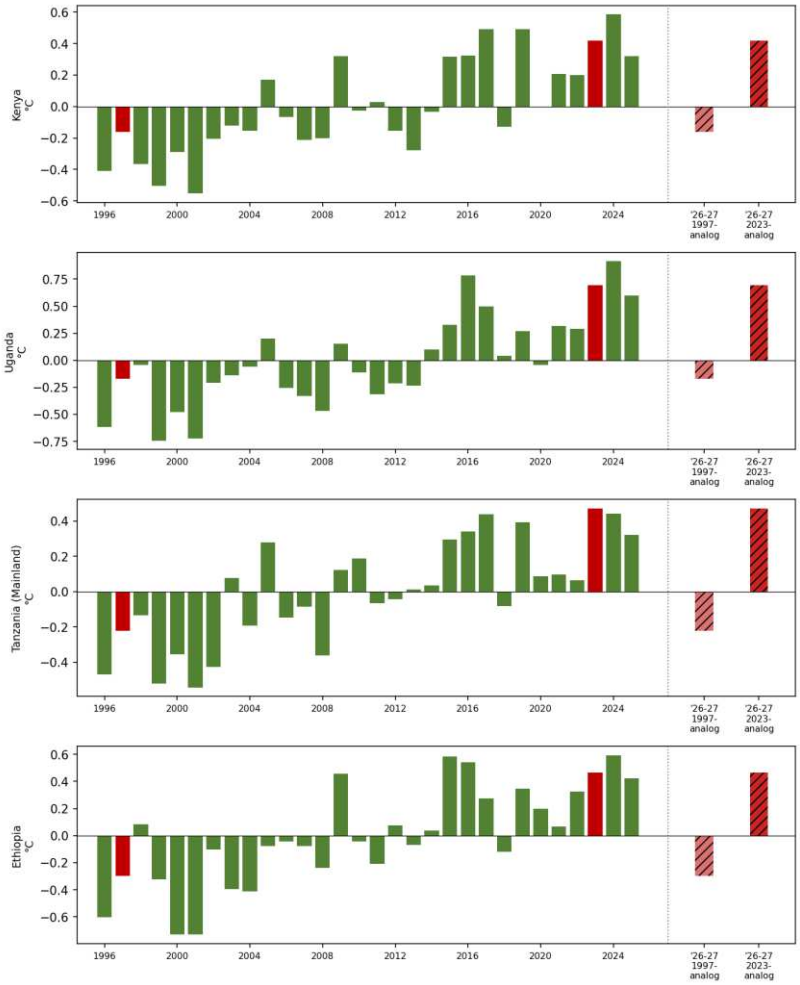

Fig. 4 Historical annual temperature anomaly by country, 1996–2025. Same format as Fig. 3, for temperature (°C). A warming trend from 2014 onward is apparent in all four countries.

ENSO phase

El Niño and La Niña main effects were both directionally positive but their 89% credible intervals included zero (El Niño: 0.29, −0.37 to 0.90; La Niña: 0.62, −0.09 to 1.30). The posterior probability of elevated risk was 92.2% for La Niña and 78.2% for El Niño. These estimates rest on very few years: the panel contains three El Niño years (2015, 2019, 2023) and two La Niña years (2021 and 2022, a single episode), and because the year random effect absorbs shared-year variation, the ENSO terms are informed only by the contrast between these years and the seven Neutral years. The year 2016, which included the decay of the 2015–16 El Niño (five months at or above +0.5 °C), is coded Neutral under the modal-phase rule. Taken together, these results do not show a reliable association between ENSO phase and classification; the El Niño and La Niña estimates are suggestive only.

No individual country's rainfall slope had a credible interval that excluded zero (Table 1, Fig. 5). Kenya's rainfall slope exceeded Uganda's with 83.5% posterior probability and Tanzania's with 78.4%.

Table 1 Country-level rainfall slopes. Posterior mean slope, 89% credible interval (CI) and posterior probability that the slope is positive, P(slope > 0). No interval excludes zero.

Country	Rainfall slope	89% CI	P(slope > 0)
Kenya	0.009	−0.084 to 0.103	55.2%
Tanzania (Mainland)	−0.038	−0.142 to 0.065	27.6%
Ethiopia	−0.055	−0.161 to 0.043	20.0%
Uganda	−0.074	−0.221 to 0.049	18.5%

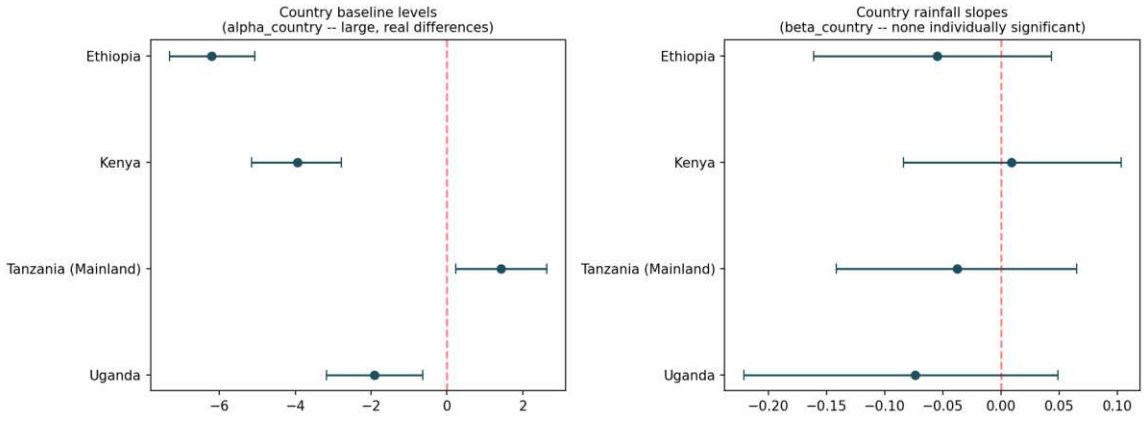

Fig. 5 Country-level baseline levels and rainfall slopes. Left: country baseline levels, which differ substantially between countries. Right: country-specific rainfall slopes, none of whose 89% credible intervals excludes zero. Points are posterior means and bars are 89% credible intervals.

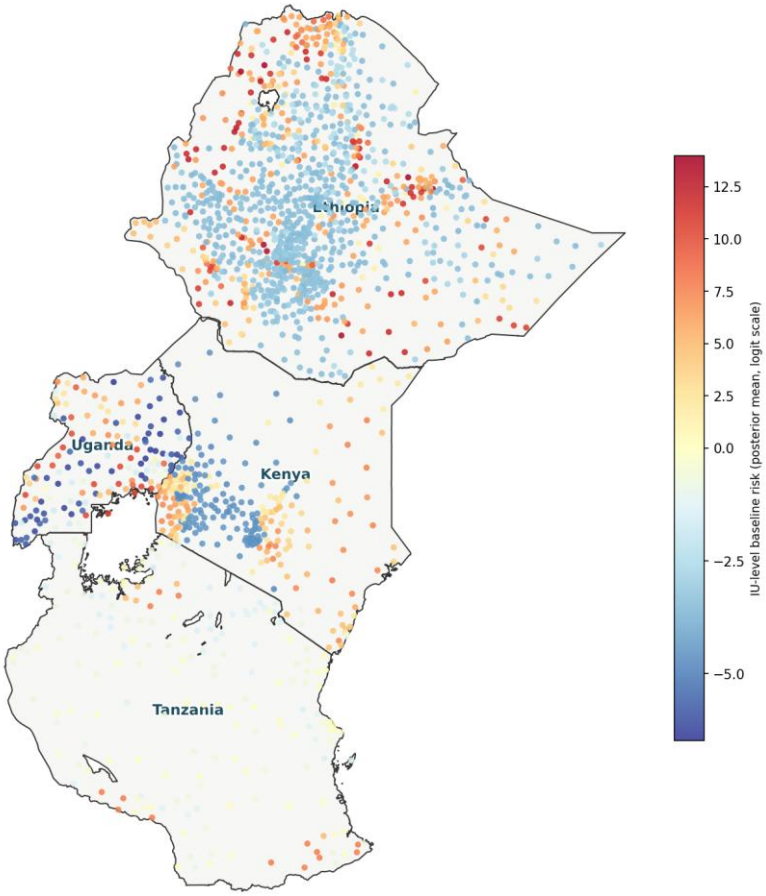

Fig. 6 IU-level baseline schistosomiasis risk across four East African countries. Each point is one IU, coloured by its posterior mean baseline risk from the fitted model (16,438 IU-years from 1,754 IUs; Kenya, Uganda, Tanzania Mainland and Ethiopia).

Baseline risk versus a climate-adjusted scenario

To test whether the fitted climate terms materially change predicted risk, the posterior predictive probability of the highest (≥50%) band was computed for every IU in two ways: from the IU and country baseline terms alone, and from the full model under an illustrative climate scenario (rainfall and temperature from the observed 1997 and 2023 analog years, with El Niño phase). Fig. 7 maps both. The two are nearly indistinguishable: Pearson correlation 0.997, Spearman rank correlation 0.992, mean absolute difference in predicted probability 0.008. Adding the full climate specification shifted predicted risk by less than one percentage point on average; only 2.9% of IUs shifted by more than five percentage points and 1.0% by more than ten. This near-identity largely re-expresses the fitted coefficients rather than providing an independent test: across IUs the climate-driven shift in log-odds had a standard deviation of about 0.2, against about 5.3 for the between-IU baseline spread. Its value is in expressing

the small coefficients on the predicted-probability scale. The two analog years also differ in temperature relative to the 1996–2025 baseline (mean standardised anomaly −1.8 in 1997 and +0.4 in 2023), so the scenario reflects the regional warming trend as well as ENSO.

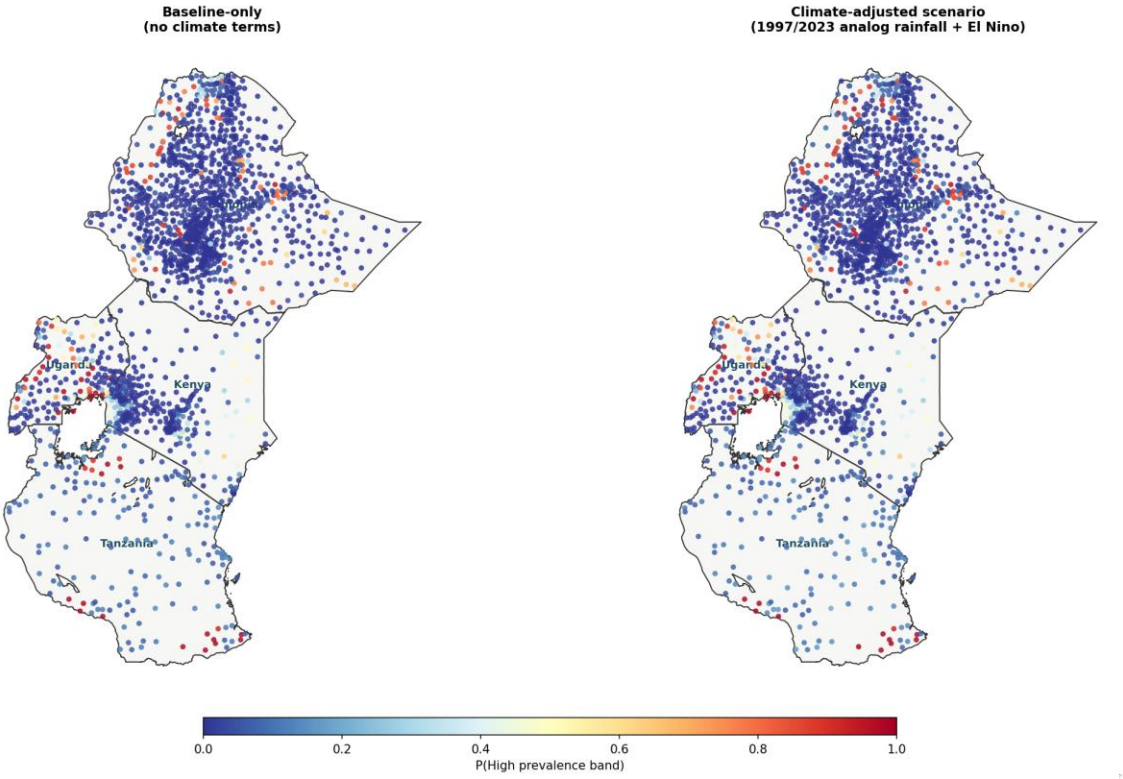

Fig. 7 Baseline-only versus climate-adjusted scenario maps. Predicted probability of the highest-prevalence band (≥50%) for each IU under the baseline-only model (left) and under the climate-adjusted scenario (analog 1997/2023 rainfall plus El Niño; right). Pearson r = 0.997; Spearman rank correlation = 0.992.

Discussion

In this panel, most variability in reported schistosomiasis classification lay between IUs (89.5%), with 9.3% between countries and 1.2% between years, and adding the fitted climate terms changed predicted IU-level risk by less than one percentage point on average. Rainfall had a non-monotonic association with class, whereas the ENSO-phase estimates were imprecise. These results describe the information content of a programme classification that rarely changes: 5.2% of consecutive-year transitions, and 66.5% of IUs never changed band. A variance decomposition of such an outcome assigns most variance to the IU almost by construction, and cannot separate stable underlying risk from classifications carried forward between survey rounds. The findings therefore concern reported classification, not the response of transmission to climate. A weak climate signal in infection intensity or

incidence, which a classification would not register, is not excluded; testing it would need survey-level prevalence matched to survey dates.

Table 2 sets our study beside the three closest comparators. They differ from ours in outcome (prevalence or hotspot status in surveyed schools or communities, rather than the annual IU classification), in scale (schools, communities or wards, rather than IUs) and in time (one survey or a treatment follow-up of a few years, rather than a twelve-year panel). Their findings on environmental covariates are consistent with ours: in Clements et al.8 the rainfall and temperature effects on S. haematobium prevalence disappeared once spatial correlation was modelled, and in Singer et al.9 baseline prevalence, a locality-specific quantity, was a leading predictor in every model. Mathewson et al.10 show what geographically targeted control looks like when prevalence data are scarce.

Table 2 Comparison with three related studies. IU, implementation unit; MDA, mass drug administration. Studies differ in outcome, scale and period, so the comparison is qualitative.

Study	Question and outcome	Setting and scale	Principal findings	Programme implication
Clements et al. 2006 (ref. 8)	Where S. haematobium and S. mansoni occur, to plan praziquantel distribution; school prevalence	143 schools in Tanzania; Bayesian geostatistical models	Temperature and rainfall predicted S. haematobium prevalence only until spatial correlation was modelled; distance to water and minimum temperature stayed significant for S. mansoni	Identify high-prevalence areas and areas needing more data
Singer et al. 2024 (ref. 9)	Which communities will remain hotspots after MDA, predicted at baseline; three hotspot definitions	589 communities in five countries; reanalysis of randomised trials; outcomes at year 5	Baseline prevalence was a top predictor in all models; accuracy was lower in countries absent from training data and for persistent hotspots	Prioritise high-risk communities for more frequent surveillance or intervention
Mathewson et al. 2024 (ref. 10)	Ward-level prevalence of intestinal and urogenital schistosomiasis to support targeted MDA	17400 schoolchildren in six Tanzanian regions; 2021 survey; regression kriging	Adjusted R-squared 0.51 (intestinal) and 0.32 (urogenital); ward-level targeting would move treatment from 193 over-treated wards to 149 wards omitted by district-level MDA	Target MDA at ward rather than district level where prevalence data are scarce
This study	Association of ENSO and climate anomalies with annual IU endemicity class; where variability lies; added value of climate information	1754 IUs in four East African countries; 2014 to 2025; hierarchical ordered-logistic model	89.5% of variance between IUs, 9.3% between countries, 1.2% between years; climate terms changed predicted risk by under one percentage point	Country-level ENSO forecasts alone do not support changing IU decisions; other applications need evaluation

Local heterogeneity in schistosomiasis prevalence, and the limited added value of environmental covariates once spatial structure is modelled, are established; we do not claim to have discovered them. Our contribution is narrower and concerns what is measured. First, the outcome is the programme classification at IU level, the unit at which treatment decisions are taken, so the variance partition describes the information available to programme managers and not prevalence at surveyed sites. Second, because IUs are observed in several years, variability can be divided at once between persistent IU differences, country-level differences and year-level shifts shared across IUs, which a single-survey geostatistical map cannot do. Third, repeated observation allowed a direct test of whether IU-year climate anomalies and ENSO phase add information beyond an IU- and country-only baseline; they added little. Because classifications change rarely, the partition partly reflects the persistence of the classification process and should not be read as evidence that underlying transmission is static.

The non-monotonic rainfall association is statistically clear in this model and is consistent with a balance between snail-habitat expansion and flushing, but the data cannot confirm that mechanism. It is estimated from annual anomalies matched to a slowly changing classification, and other explanations, such as differences in which IUs were reassessed after high-rainfall years, cannot be excluded. The La Niña and El Niño estimates (posterior probabilities 92% and 78%) are suggestive only: a twelve-year panel contains at most a handful of independent episodes, so a longer panel is needed to resolve them. The positive association with treatment coverage should not be read as evidence that MDA is ineffective; WHO preventive-chemotherapy guidance directs more intensive treatment towards higher-prevalence IUs, and the coefficient most plausibly reflects that targeting rule.

The classification analysed here is the IU endemicity band (below 10%, 10–49%, 50% or more) that guides treatment frequency under WHO preventive-chemotherapy guidance, so the findings bear on programme decisions only as far as that classification does. What this study supports is limited. Country- or region-level ENSO forecasts did not materially change predicted IU classification (mean absolute change 0.008 in the predicted probability of the highest band), which does not support using such forecasts alone to change IU treatment or survey schedules in these four countries. Country explained only 9.3% of the variability, so country-level aggregates would hide most of the variation in reported classification, which is consistent with managing treatment at IU rather than country level. Several applications would need further evaluation. Using rainfall anomalies to prioritise IUs for earlier re-survey has a plausible mechanism but a modest, untested association, and would need survey-date-matched prevalence and a prospective comparison with routine re-survey. Flagging IUs after La Niña years should

wait for more years of data. Targeting below IU level is outside what this study can address; the geostatistical approach of Mathewson et al.10 is the appropriate route. IU-resolution environmental, water-source and surveillance data may be valuable, but we did not evaluate them, so that is a hypothesis and not a finding.

Several limitations apply. The outcome is a three-level classification, not site-level counts, and climate anomalies are matched to the calendar year of each report because survey dates are not available. The panel contains three El Niño years and two La Niña years, so the ENSO terms rest on few contrasts. DR Congo and Tanzania-Zanzibar are excluded, and 1,052 IU-years (285 IUs without a usable centroid coordinate, mainly in Kenya and Ethiopia) lacked climate data and 870 further IU-years had an unknown or unreported status; if these units differ systematically the results may not generalise to them. Country-level variance is estimated from only four countries and Ethiopia contributes 63% of fitted IU-years. IU effects are treated as exchangeable and spatial autocorrelation between neighbouring IUs is not modelled, so the model cannot map prevalence for unsurveyed locations or be compared directly with geostatistical predictions. The comparison with earlier studies is qualitative because outcomes, scales and periods differ. The analysis was not preregistered, RQ2 and RQ3 are descriptive, and the analog-year scenario is illustrative, with temperature inputs that mix ENSO with the regional warming trend.

In four East African countries, most variability in reported schistosomiasis classification lay between implementation units, little between countries or years, and adding fitted climate terms changed predicted IU-level risk by less than one percentage point. Rainfall had a non-monotonic association with class and the ENSO-phase estimates were imprecise. Together with earlier geostatistical and hotspot-prediction studies, these results suggest that IU-level information, rather than country- or region-level climate forecasts, should anchor decisions about reported classification. They do not show that climate is unimportant for transmission, because the outcome is a slowly changing programme classification. Evaluating IU-resolution climate and environmental data with survey-date-matched prevalence is the logical next step.

Methods

Study design and data

We fitted a single hierarchical Bayesian ordered-logistic model to a twelve-year panel (2014–2025) of schistosomiasis endemicity classifications for IUs in Kenya, Uganda, Tanzania (Mainland only) and

Ethiopia. The Democratic Republic of the Congo is the subject of a separate analysis with different source data, and the available Zanzibar export lacked the coverage fields needed for the model's covariates. The analysis was not preregistered.

Schistosomiasis endemicity classifications were obtained from the WHO ESPEN platform11. ESPEN's raw per-country export reports each IU-year four times, once per population stratum (Adults, PreSAC, SAC, Total), because these strata carry different treatment-coverage figures; the endemicity classification itself does not vary by stratum. Only the "Total" stratum row was retained, giving one observation per IU-year. The resulting panel contains 18,360 unique IU-years from 2,040 IUs. Of these, 17,308 have matching climate data; the other 1,052 IU-years (285 IUs without a usable centroid coordinate, concentrated in Kenya, 10.4% of its IU-years, and Ethiopia, 6.2%) were excluded. A further 870 IU-years have an endemicity status of "Unknown" or "Not reported" and cannot be placed on the ordinal scale. The fitted model therefore uses 16,438 IU-years from 1,754 IUs (Ethiopia 10,393; Kenya 2,416; Tanzania Mainland 2,177; Uganda 1,452). ESPEN's two endemicity classification schemes were harmonised onto a single three-level ordinal scale (<10%, 10–49%, ≥50%), anchored on the WHO preventive-chemotherapy decision thresholds that both schemes are built around.

Monthly rainfall (CHIRPS v2.012, via Google Earth Engine) and monthly 2 m temperature (ERA5 reanalysis13, Copernicus Climate Data Store) were extracted at each IU's centroid coordinate (from the ESPEN Cartography shapefiles) for 1996 to mid-2026. Anomalies were computed relative to each IU's own 1996–2025 mean for the corresponding calendar month and averaged to an annual anomaly per IU-year. Monthly Oceanic Niño Index (ONI) values from the NOAA Climate Prediction Center14 classified months as El Niño (≥ +0.5 °C), La Niña (≤ −0.5 °C) or Neutral, aggregated to an annual modal phase: three El Niño years (2015, 2019, 2023), two La Niña years (2021, 2022) and seven Neutral years (2014, 2016–2018, 2020, 2024, 2025).

Statistical model

The ordinal outcome y_i (band 0/1/2) for IU i, country c(i) and year t(i) follows an ordered-logistic regression with linear predictor:

$$\eta_i = \alpha_IU[i] + \alpha_country[c(i)] + \alpha_year[t(i)] \quad \text{(IU, country and year baseline levels)}$$
$$+ \beta_country[c(i)]\cdot rain_i + \gamma\cdot temp_i + \delta\cdot rain_i^2 \quad \text{(climate main effects)}$$
$$+ \eta_ElNiño\cdot ElNiño_i + \eta_LaNiña\cdot LaNiña_i \quad \text{(ENSO main effects)}$$
$$+ \zeta_ElNiño\cdot(rain_i \times ElNiño_i) + \zeta_LaNiña\cdot(rain_i \times LaNiña_i) \quad \text{(ENSO × rainfall interactions)}$$
$$+ \theta\cdot mda_i \quad \text{(treatment coverage)}$$

Here $\alpha_{IU}[i] \sim \text{Normal}(0, \sigma_{IU})$, $\alpha_{country}[c] \sim \text{Normal}(\mu_{region}, \sigma_{country})$ and $\alpha_{year}[t] \sim \text{Normal}(0, \sigma_{year})$. The explicit year random effect is the main design choice: it absorbs ordinary shared-year variation before any of it can be attributed to the ENSO-phase covariates. Country-varying rainfall slopes $\beta_{country}[c] \sim \text{Normal}(\beta_{region}, \sigma_{\beta})$ allow hydrology to modulate the rainfall relationship differently by country. Continuous covariates were standardised before fitting. Fixed effects had weakly informative Normal(0, 1) priors on the logit scale and group-level standard deviations had Half-Normal(0, 0.5) priors.

Implementation, diagnostics and summaries

The model was implemented in PyMC with the NumPyro No-U-Turn Sampler, run as four chains of 1,200 tuning and 1,200 post-warmup draws (target acceptance 0.95). Convergence was assessed with R-hat (target <1.01) and effective sample size, and a data-integrity check confirming zero duplicate IU-years was run immediately before fitting. Variance shares were computed draw by draw from the three group-level standard deviations as the variance of each level divided by the sum of the three (IU, country, year) and summarised by the posterior mean and 89% equal-tailed interval. They describe latent (logit-scale) variability among group-level intercepts and exclude the residual variance implied by the ordered-logistic link. As a sensitivity check, shares were also computed from the realised group-level effects. To describe outcome persistence we counted how often an IU's band differed between consecutive observed years. Credible intervals are 89% throughout. This is an association analysis and not a validated forecast; the analog-year scenario (1997 and 2023) is used diagnostically, to test how much climate information changes predicted risk relative to baseline.

Ethics

This study used only publicly available, aggregated and de-identified schistosomiasis programme classifications and public climate and ocean-index archives. It did not involve recruitment, intervention or access to individual-level records, so ethics approval and consent were not required.

Use of AI assistance

The research questions, the choice of data sources and outcome definition, the model specification, the interpretation of the results and the conclusions were decided by the authors. An AI assistant (Claude, Anthropic) was used as a coding assitantunder the authors' direction. It prepared data-processing and model-fitting code and analysis notebooks to the authors' specifications, recomputed reported figures and performed arithmetic checks. The authors reviewed and ran the code and outputs independently,

checked the reported numbers against the saved posterior and the source data files. The AI assistant is not an author, and the authors take full responsibility for the content of the paper. Because the code, notebooks and saved model output are public (see Code availability), the analysis can be re-run without the assistant.

Data availability

The datasets generated and analysed during this study, with the fitted posterior, are available in the Zenodo repository (https://doi.org/10.5281/zenodo.23191723)15. The underlying public data are available from WHO ESPEN (https://espen.afro.who.int)11, CHIRPS v2.012, the Copernicus Climate Data Store (ERA5)13 and the NOAA Climate Prediction Center (Oceanic Niño Index)14. Sources, versions, access dates and processing steps are listed in Supplementary Information.

Code availability

All data-acquisition, data-processing, model-fitting and figure scripts and notebooks are archived in the same Zenodo repository15 (code MIT licence; data and documentation CC BY 4.0, with upstream data subject to the terms of the original providers). The saved posterior allows all reported analyses to be reproduced without refitting. Software: Python with PyMC, NumPyro, ArviZ, xarray and pandas; a GPU is recommended for model fitting.

References

1. Anyamba, A. et al. Prediction of a Rift Valley fever outbreak. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 955–959 (2009).

2. Moore, S. M. et al. El Niño and the shifting geography of cholera in Africa. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4436–4441 (2017).

3. World Health Organization. Schistosomiasis Practical and Precision Assessment (SPPA) manual: field protocol for mapping and precision assessment surveys, version 3 (World Health Organization, undated).

4. Krauth, S. J. et al. Infection intensity-dependent accuracy of reagent strip for the diagnosis of Schistosoma haematobium and estimation of treatment prevalence thresholds. *PLoS Negl. Trop. Dis.* **16** (2022).

5. Schur, N. et al. Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged ≤20 years in West Africa. *PLoS Negl. Trop. Dis.* **5**, e1194 (2011).

6. Schur, N. et al. Spatially explicit Schistosoma infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Trop.* **128**, 365–377 (2013).

7. Clements, A. C. A. et al. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. *Emerg. Infect. Dis.* **14** (2008). https://doi.org/10.3201/eid1410.080366.

8. Clements, A. C. A. et al. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop. Med. Int. Health* **11**, 490–503 (2006).

9. Singer, B. J. et al. Development of prediction models to identify hotspots of schistosomiasis in endemic regions to guide mass drug administration. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2315463120 (2024).

10. Mathewson, J. D. et al. Enabling targeted mass drug administration for schistosomiasis in north-western Tanzania: exploring the use of geostatistical modeling to inform planning at sub-district level. *PLoS Negl. Trop. Dis.* **18**, e0011896 (2024).

11. World Health Organization Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN). ESPEN data portal: schistosomiasis implementation-unit-level data. https://espen.afro.who.int (accessed 5 September 2026).

12. UCSB Climate Hazards Center. Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), version 2.0. https://www.chc.ucsb.edu/data/chirps (accessed 5 September 2026).

13. Copernicus Climate Data Store. ERA5 hourly and monthly reanalysis data on single levels (European Centre for Medium-Range Weather Forecasts). https://cds.climate.copernicus.eu (accessed 5 September 2026).

14. NOAA Climate Prediction Center. Oceanic Niño Index (ONI), ERSST.v5-based, monthly values. https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php (accessed 5 September 2026).

15. Mudogo, C. N., Katembo, J.-L. M., Mazigo, H. D. & Verheyden, J. G. L. Data and code for: Location, location, location: implementation-unit identity dominates reported schistosomiasis classification across four East African countries. Zenodo https://doi.org/10.5281/zenodo.23191723 (2026).

Acknowledgements

The authors received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Author contributions

CNM contributed to conceptualisation, provided resources (prior schistosomiasis fieldwork and disease background), contributed to validation and visualisation, and contributed to writing (original draft; review and editing). JGLV contributed to conceptualisation, data curation, formal analysis, investigation, methodology, software, validation and visualisation, and to writing (original draft; review and editing); data-processing and model-fitting code was prepared with AI assistance as described in Methods. JLMK

contributed to formal analysis and validation and to writing (original draft). HDM contributed to validation and to writing (review and editing). All authors reviewed and approved the final manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Correspondence: Correspondence and requests for materials should be addressed to C.N.M. (celestin.mudogo@unikin.ac.cd).

Supplementary information: Supplementary Information (source manifest and glossary of statistical and disease terms) accompanies this paper.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.docx
- Additionalfile2Glossary.docx