Predictive Malaria Risk and Uncertainty Mapping in Nchelenge District, Zambia: Evidence of Widespread, Persistent Risk and Implications for Targeted Interventions

Jessie Pinchoff,* Mike Chaponda, Timothy Shields, James Lupiya, Tamaki Kobayashi, Modest Mulenga, William J. Moss, and Frank C. Curriero for the Southern Africa International Centers of Excellence for Malaria Research
Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland;
Tropical Disease Research Centre, Ndola, Zambia

Abstract. Malaria risk maps may be used to guide policy decisions on whether vector control interventions should be targeted and, if so, where. Active surveillance for malaria was conducted through household surveys in Nchelenge District, Zambia from April 2012 through December 2014. Households were enumerated based on satellite imagery and randomly selected for study enrollment. At each visit, participants were administered a questionnaire and a malaria rapid diagnostic test (RDT). Logistic regression models were used to construct spatial prediction risk maps and maps of risk uncertainty. A total of 461 households were visited, comprising 1,725 participants, of whom 48% were RDT positive. Several environmental features were associated with increased household malaria risk in a multivariable logistic regression model adjusting for seasonal variation. The model was validated using both internal and external evaluation measures to generate and assess root mean square error, as well as sensitivity and specificity for predicted risk. The final, validated model was used to predict and map malaria risk including a measure of risk uncertainty. Malaria risk in a high, perennial transmission setting is widespread but heterogeneous at a local scale, with seasonal variation. Targeting malaria control interventions may not be appropriate in this epidemiological setting.

INTRODUCTION

Zambia is a malaria-endemic country in sub-Saharan Africa that has historically experienced a high burden of malaria morbidity and mortality. Between 2006 and 2011, several national malaria control interventions were scaled-up, including case management with rapid diagnostic tests (RDTs) and artemisinin combination therapy (ACT), distribution of long-lasting insecticide-treated nets (LLINs), indoor residual spraying (IRS), and intermittent preventive therapy (IPT) for pregnant women.¹ Despite substantial progress, malaria transmission remains high in parts of Zambia. New or enhanced methods are necessary to achieve malaria control in resource-constrained areas with persistently high malaria transmission. Generating maps of variation in malaria risk at different spatial scales may be used to guide targeted control interventions for maximal impact.

One obstacle to control is that malaria transmission is heterogeneous in its distribution across space and time.² Mathematical models predict that heterogeneity of transmission reduces the efficacy of disease control strategies.³ Advances in remote sensing and satellite imagery allow for highly accurate characterization of environmental and ecological features that may be associated with mosquito breeding sites,⁴ such as proximity to water,^{5,6} topography,⁷ vegetation,^{8,9} and anthropogenic features such as roads or irrigation systems.^{10,11} These features alter the spatial distribution of malaria by directly or indirectly influencing the development, density, and location of mosquito vectors and their breeding sites.¹² The ability to predict spatial variation in malaria risk has improved with these advanced technologies.

Spatial and seasonal heterogeneity of malaria transmission remains poorly characterized and many existing risk maps have limited operational use for malaria control activities

because they are at coarse spatial resolution (national level or higher), and are based on passive case detection at health centers with resultant biases or on national-level surveys with large unsampled areas.^{13–15} Extrapolating policy and programmatic decisions to these unsampled areas may be negatively impacted by statistical uncertainty. Some but not all malaria risk maps are associated with measures of spatial risk uncertainty, which is critical to provide valuable information on model fit, accuracy, and interpretation. Recent global maps created by the Malaria Atlas Project and Swiss Tropical Public Health Institute include uncertainty maps and use sophisticated statistical techniques to account for autocorrelation.^{16,17} High-resolution maps may be useful for efficient and cost-effective targeting of interventions to the highest risk areas^{8,12,18–22} or in deciding to implement blanket coverage.

The aims of this study were to generate and validate a high-resolution empirical risk map for household malaria risk in Nchelenge District, Luapula Province, Zambia. This is a region with poorly controlled, perennial malaria transmission. According to the National Malaria Indicator Survey in 2012, Luapula Province had the highest malaria prevalence in the country.²³ In Nchelenge District, malaria parasitemia increased from 38% to 56% between 2006 and 2012.²⁴ The risk map approach identified environmental features predictive of malaria risk, information that can be used to guide policy decisions on whether malaria control activities should be targeted to specific geographical areas within the district.

METHODS

Study site. The study was conducted in Nchelenge District, Luapula Province, Zambia between 2012 and 2014, and is currently ongoing. Luapula Province experiences hyperendemic, perennial transmission of *Plasmodium falciparum* malaria. The study site is located along Lake Mweru, and shares a border with the Democratic Republic of Congo. Nchelenge District is located at an altitude of approximately 807 m above sea level in a habitat characterized as marsh. There is a single rainy season from approximately November through April, followed by a dry

*Address correspondence to Jessie Pinchoff, Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, 615 N Wolfe Street, Baltimore, MD 21205. E-mail: jpincho1@jhu.edu

season from May to October; rainfall follows a seasonal pattern, with about 2,700 mm in the rainy months and close to 0 mm during the dry months. The catchment area is heterogeneous, with a densely populated urban area along the lake and farmland located further inland. *Anopheles funestus* and *Anopheles gambiae* are the primary vectors for malaria transmission. LLINs and IRS programs have been implemented in Nchelenge District since 2006 and 2007, respectively. However, transmission in Nchelenge District is persistently high despite the scale-up of malaria control interventions.²⁴

Study participants and procedures. Satellite images were used to generate a sampling frame for the random selection of households to enroll in a prospective community cohort study.²⁵ Longitudinal (households visited repeatedly) and cross-sectional (households visited once) surveys were conducted, alternating by month. A grid with $1 \times 1 \text{ km}^2$ cells was drawn over the satellite image of Nchelenge District and grid cells were selected to obtain a representative sample based on geographic location. Households within each selected grid cell were enumerated, assigned a global positioning system coordinate and randomly selected for enrollment. Satellite imagery was obtained from DigitalGlobe Services, Inc. (Denver, CO). The image was imported into ArcGIS version 10.2 (Esri, ArcGIS, Redlands, CA) and locations of households were identified and enumerated manually. All households were eligible to participate.

A field team was provided coordinates of the selected households to contact for enrollment. If a household was not found or refused participation, a household was selected from a backup list of randomly selected households. After obtaining permission from the local chief and head of household, as well as informed consent from each individual, a questionnaire was administered to each participant residing within the household and a blood sample was collected by finger prick. For children under the age of 16 years, the questionnaire was directed to their caregiver who provided parental permission. RDTs were used to detect *P. falciparum* histidine-rich protein 2. Based on national policies, ICT RDTs (ICT Diagnostics, Cape Town, South Africa) were used from April 2012 to May 2013. First-response RDTs (Premier Medical Corporation Ltd., Mumbai, India) were used from June 2013 to September 2013, and SD Bioline RDTs (Standard Diagnostics, Kyonggi, Republic of Korea) were used subsequently. All individuals with positive RDTs were offered treatment with artemether-lumefantrine (Coartem®; Novartis, Basel, Switzerland).

The study was approved by the Institutional Review Board of the Tropical Diseases Research Center, Ndola, Zambia and the Institutional Review Board of the Johns Hopkins University Bloomberg School of Public Health. Informed consent forms were translated into Bemba and administered to adult participants and the parents or guardians of children younger than 16 years.

Landscape characterization. Environmental variables were generated for the study area and integrated with household level RDT data using ArcGIS version 10.2 (Esri, ArcGIS). A handheld Android tablet was used to record the household coordinates. A digital elevation model (DEM) for the area with 90-m resolution was obtained from the Shuttle Radar Topography Mission (SRTM) version 3. Each pixel represents a 90-m average elevation around each pixel's center. The DEM was processed in ERDAS Imagine 2011 software

(ERDAS Inc., Bethesda, MD) and imported into ArcGIS. The ArcHydro Tools module of ArcGIS,²⁶ based on elevation and degree of slope derived from the SRTM image to determine water flow direction and accumulation, was used to build a stream network with corresponding Strahler stream classification.²⁷ The Strahler classification assigns ordered values 1, 2, 3, and so on based on the hierarchy of tributaries, with the beginning of each stream or river segment a first order or category 1 stream. Category 2 streams are formed when two category 1 streams come together, category 3 streams formed when two category 2 streams come together and so on. Population density was calculated as the sum of enumerated structures located within 500 m of a study household. The distance from each enrolled household to the nearest road, nearest health facility, and Lake Mweru were also calculated. A binary variable denoting households near the lake or interior was generated based on spatial location, with households less than 3 km from the lake considered near.

The normalized difference vegetation index (NDVI) was used to assess ground cover. NDVI is derived from Landsat 5 Moderate Resolution Imaging Spectroradiometer (USGS Earth Resources Observation and Science [EROS] Center, Sioux Falls, SD) from the U.S. Geological Survey Land Processes Distributed Active Archive Center. Values for NDVI range from -1 to +1. Negative values represent bodies of water, values near zero represent asphalt, and increasing values correspond to increasing abundance of actively photosynthesizing vegetation or "greenness."⁹

Rainfall data from a weather-monitoring tool was used to generate a variable for season, including historical data and the HOBO Micro Station (Onset Computer Corporation, Bourne, MA). The HOBO Micro Station is a four-sensor data logger designed to take measurements of rainfall, dew point, temperature, and relative humidity at hourly intervals. The HOBO Micro Station was placed within the study area. A binary variable for season was created and used in analyses.

Statistical analyses. Prevalent malaria infections were identified by RDT using data from the cross-sectional surveys and first visit to the longitudinal survey households. Logistic regression was used to identify environmental features associated with the proportion of individuals in a household who were RDT positive. Backward variable selection was implemented; a full model was created and only statistically significant variables ($P \leq 0.1$) included in the final model. Results of univariate models are also presented. Logistic regression inference was based on the quasi-binomial distribution to account for overdispersion. Model fit was evaluated using the Hosmer-Lemeshow goodness of fit test. Semivariogram plots based on regression standardized residuals were used to assess residual spatial variation (spatial variation in the proportion of RDT-positive individuals per household not accounted for by the regression variables).²⁸

Prediction performance of the final model was evaluated internally and with an external data set. For internal evaluation, a Monte Carlo scheme was designed as follows: for each of 1,000 iterations, the data were randomly split into a 25% prediction data set ($N = 115$) and a 75% training data set ($N = 346$) of the total 461 sampled households from 2012 to 2014. The final regression model was refit in each iteration based on data in the training set and used to predict the number of RDT-positive household members in the prediction data set. Root mean squared error (RMSE) comparing

the predicted number of RDT positives (predicted logistic regression probability times the known number of household members) to the true number of RDT positives per household was the performance metric used. Results were summarized by the average RMSE and corresponding 95% prediction interval (taken as the 2.5th and 97.5th percentiles) from the distribution of 1,000 Monte Carlo RMSEs. RMSE was also stratified and calculated by season (rainy and dry) to determine seasonal difference in predictions.

Although RMSE was used to evaluate prediction based on the number of RDT-positive household members, sensitivity and specificity were calculated to evaluate prediction based on any (or at least one) RDT-positive household member compared with no RDT-positive household members. Identifying positive and negative households in this manner may be useful from a programmatic perspective, such as identifying households to receive IRS or target for active case finding at the household level. Sensitivity and specificity were calculated within the Monte Carlo scheme described above as follows. Predicted number of RDT-positive household members were dichotomized into two groups: at least one predicted RDT positive household member versus zero predicted RDT positive household members. This new variable was compared with the observed binary outcome of at least one-predicted RDT-positive household member versus no RDT-positive household members. Monte Carlo sensitivity and specificity results were summarized similarly as with RMSE, reporting the mean and 95% prediction intervals based on the Monte Carlo distribution and stratified by season.

For external evaluation, the final regression model was run on the data restricted to 2012–2013 and used to predict the number of RDT-positive household members for the 2014 sampled households ($N = 164$). In the external evaluation of the model's predictive performance, RMSE, sensitivity, and specificity were calculated for the predictions and stratified by rainy versus dry season. Prediction intervals in the internal evaluation were based on the Monte Carlo results predicting the repeatedly varying holdout samples. No such design was implemented for the external evaluation; hence, the results are reported without Monte Carlo prediction intervals. Statistical analyses were conducted using R statistical software packages (version 3.1.1; R Core Team, Vienna, Austria).²⁹ As a final step in the analysis, environmental variables matching those in the final model were generated for a fine mesh of 500 m² grid cells over the study area. The final model fit using

the full 2012–2014 data was used to predict the risk of RDT household positivity at each grid cell and then mapped. A map of prediction uncertainties was also produced based on the standard error from the prediction model. Risk maps and risk uncertainty maps are displayed for the rainy and dry season, as well as a map showing the difference in risk across these two seasons to highlight the seasonal change in risk.

RESULTS

A total of 461 households comprising 1,725 individuals were enrolled between April 2012 and December 2014, of whom 48% were RDT positive. The median age of the study participants was 14 years (interquartile range [IQR]: 5, 32) and 53.8% were female. Forty-eight percent of study visits occurred during the rainy season. Seventy-four percent of households had at least one RDT-positive resident, with 75% of households having an RDT-positive resident in the rainy season and 72% in the dry season. Households had an average of 3.7 household members (minimum of 1 and maximum of 13 individuals). The median proportion of RDT-positive individuals in a household was 45% (IQR: 0%, 70%) (Table 1).

Several environmental features were significantly associated with the proportion of individuals testing RDT positive within households in both univariate and the final multivariate logistic regression model (Table 2). Malaria increased with proximity to streams and during the rainy season. Household malaria risk was 50% higher during the rainy season compared with the dry season (odds ratio [OR]: 1.59, 95% confidence interval [CI]: 1.05, 2.42); however, the risk remained relatively high during the dry season. The household risk of malaria increased 12% (OR: 1.12, 95% CI: 1.05, 1.19) for every 250-m decrease in distance from a category 1 stream. Although proximity to a category 2 stream was not associated with RDT positivity, a significant interaction was identified between proximity to category 2 streams and season. In the rainy season, household risk of malaria increased 5% (OR: 1.05, 95% CI: 1.00, 1.10) for every 250-m decrease in distance from a category 2 stream (Table 2). Household malaria risk was also associated with terrain, as measured by slope. The degree of slope is the angle at which the terrain lies. The range for the entire study area was between 0 and 30 degrees and the range for sampled households was between 0 and 10 degrees. For each increase in the degree

TABLE 1
Characteristics of sampled households and environmental variables in Nchelenge District, 2012–2014

	Households with ≥ 1 RDT-positive resident	Households with no RDT-positive residents	All households
Number (%)	339 (74%)	122 (26%)	461
Median number of household members (IQR)	4 (3, 5)	2 (1, 3)	4 (2, 5)
Median distance to nearest category 1 stream in meters (IQR)	620 (329, 1,072)	676 (292, 906)	627 (322, 1,042)
Median distance to nearest category 2 stream in meters (IQR)	2,296 (589, 2,849)	2,023 (539, 2,855)	2,140 (574, 2,857)
Median distance to nearest category 3 stream in meters (IQR)	1,056 (342, 1,447)	1,621 (428, 2,196)	1,248 (352, 2,152)
Median number of structures within 500 m (IQR)	157 (75, 450.5)	313 (129, 609)	181 (76, 497)
Median distance to nearest road in meters (IQR)	74 (40, 130)	96 (55, 174)	80 (41, 140)
Median degree of slope (IQR)	1.8 (1.1, 2.5)	1.5 (0.9, 2.1)	1.6 (1.1, 2.4)
Median elevation in meters (IQR)	957 (947, 962)	960 (952, 971)	958 (948, 969)
Median normalized difference vegetation index (IQR)	0.5 (0.5, 0.6)	0.5 (0.4, 0.6)	0.5 (0.5, 0.6)
Median distance to nearest health facility in meters (IQR)	3,569 (1,258, 5,808)	3,345 (1,007, 5,142)	3,533 (1,155, 5,425)
Median distance to Lake Mweru in meters (IQR)	838 (476, 5,981)	773 (526, 5,884)	815 (492, 5,955)

IQR = interquartile range; RDT = rapid diagnostic test.

TABLE 2

Univariate and multivariable logistic regression models of environmental factors associated with RDT positivity in Nchelenge District, 2012–2014

	Univariate Models*			Multivariable Model		
	OR	95% CI	P value	OR	95% CI	P value
Distance to category 1 stream (per 250 m)	1.06	1.00, 1.13	0.04	1.12	1.05, 1.19	0.001
Distance to category 2 stream (per 250 m)	1.00	0.98, 1.02	0.96	0.94	0.90, 0.98	0.01
Distance to category 3 stream (per 250 m)	1.00	0.98, 1.02	0.88	1.01	0.99, 1.03	0.47
Rainy season	1.08	0.86, 1.36	0.49	1.59	1.05, 2.42	0.03
Interaction: category 2 stream and rainy season	1.06	1.01, 1.11	0.01	1.05	1.00, 1.10	0.05
Number of structures within 500 m	0.99	0.98, 0.99	0.01	0.99	0.98, 0.99	0.0001
Distance to nearest road (per 250 m)	1.36	1.16, 1.64	0.001	1.42	1.20, 1.70	0.0001
Degree of slope	1.06	0.99, 1.13	0.11	1.11	1.03, 1.19	0.01
Elevation (per 10 m)	1.07	0.82, 1.39	0.63	–	–	–
NDVI value ≥ 0.5	1.32	1.05, 1.66	0.02	–	–	–
Distance to health facility (per 250 m)	1.00	0.99, 1.01	0.82	–	–	–
Distance to Lake Mweru (per 250 m)	1.00	0.99, 1.01	0.81	–	–	–
Lake vs. interior	0.97	0.76, 1.23	0.8	–	–	–

CI = confidence interval; NDVI = normalized difference vegetation index; RDT = rapid diagnostic test.

of slope, household risk of malaria increased 11% (OR: 1.11, 95% CI: 1.03, 1.19). Household risk of malaria increased by 42% (OR: 1.42, 95% CI: 0.120, 1.70) for every 250-m distance closer to the nearest road.

The Hosmer–Lemeshow *P* value of 0.44 indicated strong model fit. Residual semivariograms based on Pearson standardized regression residuals and maximum likelihood–fitted spherical semivariogram functions were calculated for both the null (intercept only) and final regression models (results not shown). Comparing the null model to the final regression model revealed that the included regression covariates substantially accounted for spatial variation in the regression outcome. However, regression inference and prediction variances were still adjusted for overdispersion to provide more conservative estimates.

Model predictive performance was evaluated based on the internal and external evaluation (Table 3). For the internal evaluation based on the 2012–2014 data, the average RMSE was 1.22, suggesting that on average, when applied to the total number of household members, the model prediction was within 1.22 individuals (95% prediction interval: 1.05, 1.40) of predicting the correct number of RDT-positive household members. Model prediction in the rainy season showed an average 4% improvement compared with the dry season; the average ratio of the rainy to dry season RMSE was 0.96 (95% prediction interval: 0.82, 1.11).

The sensitivity and specificity were strong for the model to predict negative (no RDT-positive members) and positive (at least one RDT-positive member) households. The overall sensitivity of the model was 92% (95% CI: 87%, 96%) and overall specificity was 69% (95% CI: 63%, 78%), summarized from the Monte Carlo results (Table 4). Predictions

were about the same in the rainy season (sensitivity 91% [95% CI: 84%, 98%] and specificity 67% [95% CI: 57%, 79%]) compared with the dry season (sensitivity 92% [95% CI: 86%, 98%] and specificity 72% [95% CI: 62%, 84%]).

In the external evaluation, the model was generated using 2012–2013 data and used to predict the 164 households visited during 2014 (Table 3). The external validation produced an RMSE of 1.28, indicating that on average the model predicted risk was within 1.28 individuals of predicting the correct number of RDT-positive household members, slightly higher than for the internal validation. The effect of season on prediction accuracy was more pronounced in this external evaluation compared with the internal evaluation; the ratio for rainy to dry season RMSE was 1.05. For the external evaluation, the overall sensitivity was 90% (95% CI: 83%, 95%) and specificity was 60% (95% CI: 44%, 74%). Predictions were not as strong in the rainy season (sensitivity 67% [95% CI: 53%, 79%] and specificity 61% [95% CI: 39%, 80%]) compared with the dry season (sensitivity 97% [95% CI: 90%, 100%] and specificity 58% [95% CI: 37%, 78%]) (Table 3).

After evaluation of the final model, a predictive malaria risk map and map of risk uncertainty were generated for the rainy and dry seasons based on the complete data set (Figure 1). The maps indicate increased household malaria risk near roads and category 1 streams, and in proximity to category 3 streams only during the rainy season. Predicted risk (probability of infection) per household ranged from 0.02 to 0.75, with the highest values in the rainy season. A predicted risk of 0.75 means 75% of household members are predicted to have

TABLE 3

RMSE values and ratios calculated for an internal and external validation of final multivariable models for malaria risk in Nchelenge District, 2012–2014

	RMSE values from internal validation (95% prediction interval)	RMSE values from external validation using data collected in 2014
RMSE	1.22 (1.05, 1.40)	1.28
RMSE ratio rainy vs. dry	0.96 (0.82, 1.11)	1.05

RMSE = root mean squared error.

TABLE 4

Sensitivity and specificity of households correctly identified as positive (any RDT-positive resident) or negative (no RDT-positive resident)

Characteristic	Internal validation (95% prediction interval)		External validation using data collected in 2014	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Overall	92 (87, 96)	70 (63, 78)	90	60
Rainy	91 (84, 98)	67 (58, 79)	67	61
Dry	92 (86, 98)	72 (63, 84)	97	58

CI = confidence interval; RDT = rapid diagnostic test.

FIGURE 1. Predictive malaria risk map and uncertainty maps by season for Nchelenge District, 2012–2014.

malaria. The uncertainty maps highlight the variation in these relationships; uncertainty patterns are similar across the seasons but slightly higher (more uncertain) in the rainy season. The ratio of risk predictions between rainy and dry season ranges from 0.75 to 1.52. When the ratio is below 1, the risk is higher in the dry season than the rainy season, when it is above 1, the risk is higher in the rainy season. The risk ratio map highlights areas where the values are higher than 1, mainly the increased

risk areas along category 2 streams during the rainy season (Figure 2).

DISCUSSION

Malaria risk maps based on active case detection for Nchelenge District, Zambia identified significant spatial and seasonal variation in malaria risk within a small geographic

FIGURE 2. Map of ratio of risk (rainy over dry prediction risk values) between seasons for Nchelenge District, 2012–2014 to identify areas of increased change between seasons.

area in a region with high, perennial malaria transmission. To date, ecological analyses to guide malaria control have been limited. Overall, models are complex and use data sources of variable quality, resulting in variation in the effects of ecological features in different areas or times. Marked residual variation of malaria continues to be detected, signifying that models are not controlling for all aspects of malaria transmission.⁸ This inconsistency and high level of complexity make ecological models and maps often not readily available to malaria control programs. Even if they are available, most countries fail to use them in planning their national malaria control strategies.²² Potentially, capacity building or other steps to make them more accessible to local policy makers and programs will be critical to ensuring the data are understood and effectively used.

This model was constructed based on high-resolution satellite imagery and malaria prevalence data collected through active case detection in a cohort of randomly selected households, thus distinguishing the quality and spatial resolution of these risk maps from maps based on passive case detection at health-care facilities. Health-care facility data are often inaccurate due to poor reporting, RDT stock outs at the clinic, and inaccurately identified cases if RDT or microscopy is not used for confirmation.^{30–32} The use of data based on passive case detection makes assumptions about health seeking behaviors and data quality at the clinic that can introduce spatial biases as these facilities are not randomly distributed.

Within the ecological setting of Nchelenge District, characterized by a long border along Lake Mweru and a network of streams and swamps, the risk of malaria was high throughout the district but was particularly concentrated adjacent to the category 1 streams and category 2 streams during the rainy season. This spatial distribution of risk does not lend itself readily to targeted interventions other than targeting households within a specified distance from category 1 or 2 streams. Although the risk of malaria was significantly higher during the rainy season, the risk remained elevated during the dry season despite seasonal rainfall.

At finer spatial resolution, the model accurately identified environmental features associated with increased household malaria risk and characterized variation in prediction uncertainty. Evaluating the approach to determining risk and the inclusion of a risk uncertainty map are important components in spatial analyses of malaria transmission. Spatial variation in prediction uncertainty is critical to identify areas for additional surveillance, ensure appropriate use of risk maps, and address ways to improve model predictions.

The risk map identified several high-risk areas based in proximity to environmental features. Streams and marshlands are highly associated with vector breeding sites. Category 1 streams are small and the water flows slowly; these streams may provide ideal breeding sites for anopheline mosquitoes year-round in Nchelenge District. Category 2 and 3 streams may be too large and fast moving to provide vector breeding sites. However, in the rainy season, category 2 streams may flood, creating marsh-like conditions ideal for anopheline breeding sites. This may increase the abundance of anopheline vectors during the rainy season. Additional potential sources of year-round breeding sites are highlighted by the increased risk of malaria along roads.^{33,34}

The risk maps generated in this analysis have some limitations. The semivariogram, a measure of residual spatial variation, indicates that the model accounts for most but not all spatial variation. Unexplained spatial variation may be from individual and household level factors such as age distribution, use of protective measures (IRS and LLINs), and roof material of the household.⁸ These variables cannot be included in models predicting from sampled to unsampled areas because data in unsampled areas are not known. Nevertheless, environmental data can be derived in unsampled areas from satellite imagery and remotely sensed data and used to predict the outcome of malaria risk. The risk maps are also based on RDT positivity of individuals within enrolled households. In a high-transmission setting, some people may have persistent hrp2 antigen for several weeks after parasite clearance, leading to overestimation of parasitemia by RDT.

CONCLUSIONS

In Nchelenge District, the prevalence of malaria remains high and more effective strategies are necessary to reduce the burden of disease. The malaria risk map accurately characterized the fine-scale, focal heterogeneity of widespread malaria transmission. There are many uses for malaria risk maps by national malaria control programs if they can be generated at the appropriate level of spatial resolution. Generating high-resolution, predictive risk maps that highlight heterogeneity of malaria can help target limited resources more efficiently.^{12,15} Using risk maps to guide targeting may be particularly applicable to IRS, which is an expensive yet effective method of vector control that is increasingly used despite few recommendations for where and how to target. For risk maps to be useful spatial decision support tools for targeted malaria surveillance and intervention delivery, the maps must be at high spatial resolution and based on data and models accessible to national malaria control programs. These findings may be expanded and used to build improved risk maps over a larger scale for Zambia. If significant variation can be accurately detected, then interventions may be targeted. However, the epidemiological setting in Nchelenge District is characterized by

widespread, persistent transmission, limiting the effectiveness of focal interventions.

Received April 10, 2015. Accepted for publication June 30, 2015.

Published online September 28, 2015.

Acknowledgments: We would like to thank the TDRC field team for their assistance in data collection, and the community in Nchelenge District for participating in this study, and would also like to thank Andre Hackman for data management, and Doug Norris and Jenny Stevenson for their thoughtful guidance in mosquito ecology and how this may impact malaria risk maps.

Financial support: This work was supported by the Johns Hopkins Malaria Research Institute, the Bloomberg Family Foundation, and the Division of Microbiology and Infectious Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health as part of the International Centers of Excellence for Malaria Research (U19 AI089680).

Authors' addresses: Jessie Pinchoff, Timothy Shields, Tamaki Kobayashi, William J. Moss, and Frank C. Curriero, Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, E-mails: jpinchof@jhu.edu, tshield2@jhu.edu, tkobaya2@jhu.edu, wmoss1@jhu.edu, and fcurriero@jhu.edu. Mike Chaponda, James Lupiya, and Modest Mulenga, Tropical Disease Research Centre, Ndola, Zambia, E-mails: chapondam@tdrc.org.zm, jamlupiya@gmail.com, and mulengam@tdrc.org.zm.

REFERENCES

- Kamuliwo M, Chanda E, Haque U, Mwanza-Ingwe M, Sikaala C, Katebe-Sakala C, Mukonka V, Norris D, Smith D, Glass G, Moss W, 2013. The changing burden of malaria and association with vector control interventions in Zambia using district-level surveillance data, 2006–2011. *Malar J* 12: 437.
- Coleman C, Coleman M, Mabuza A, Kok G, Coetzee M, Durrheim D, 2009. Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes. *Malar J* 8: 68.
- Smith D, McKenzie F, Snow R, Hay S, 2007. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PLoS Biol* 5: 531–542.
- Clennon JAKA, Musapa M, Shiff C, Glass GE, 2010. Identifying malaria vector breeding habitats with remote sensing data and terrain-based landscape indices in Zambia. *Int J Health Geogr* 9: 58.
- Moss W, Hamapumbu H, Kobayashi T, Shields T, Kamanga A, Clennon J, Mharakurwa S, Thuma P, Glass G, 2011. Use of remote sensing to identify spatial risk factors for malaria in a region of declining transmission: a cross-sectional and longitudinal community. *Malar J* 10: 163.
- Ghebreyesus T, Haile M, Witten K, Getachew A, Yohannes AM, Yohannes M, Teklehaimanot HD, Lindsay SW, Byass P, 1999. Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey. *BMJ* 319: 663–666.
- Nmor J, Sunahara T, Goto K, Futami K, Sonye G, Akewya P, Dida G, Minakawa N, 2013. Topographic models for predicting malaria vector breeding habitats: potential tools for vector control managers. *Parasit Vectors* 6: 14.
- Bejon P, Williams T, Liljander A, Noor A, Wambua J, Ogada E, Olotu A, Osier F, Hay S, Farnert A, Marsh K, 2010. Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med* 7: e1000304.
- Eisele T, Keating J, Swalm C, Mbogo C, Githeko A, Regens J, Githure J, Andrews L, Beier J, 2003. Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya. *Malar J* 2: 44.
- Pinault L, Hunter F, 2012. Larval habitat associations with human land uses, roads, rivers, and land cover for *Anopheles albimanus*, *A. pseudopunctipennis*, and *A. punctimacula* (Diptera: Culicidae) in coastal and highland Ecuador. *Front Physiol* 3: 59.
- Omukunda E, Githeko A, Ndong'a M, Mushinzimana E, Yan G, 2012. Effect of swamp cultivation on distribution of anopheline larval habitats in western Kenya. *J Vector Borne Dis* 49: 61–71.
- Raso G, Schur N, Utzinger J, Koudou B, Tchicaya E, Rohner F, N'Goran E, Silue K, Matthys B, Assi S, Tanner M, Vounatsou P, 2012. Mapping malaria risk among children in Cote d'Ivoire using Bayesian geo-statistical models. *Malar J* 11: 160.
- Craig M, Sharp B, Mabaso M, Kleinschmidt I, 2007. Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. *Int J Health Geogr* 6: 44.
- Yeshiwondim A, Gopal S, Hailemariam A, Dengela D, Patel H, 2009. Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. *Int J Health Geogr* 8: 5.
- Riedel N, Vounatsou P, Miller J, Gosni L, Chizema-Kawesha E, Mukonka V, Steketee R, 2010. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). *Malar J* 9: 37.
- Gemperli A, Sogoba N, Fondjo E, Mabaso M, Bagayoko M, Briet O, Anderegg D, Liebe J, Smith T, Vounatsou P, 2006. Mapping malaria transmission in west and central Africa. *Trop Med Int Health* 11: 1032–1046.
- Hay S, Guerra C, Gething P, Patil A, Tatem A, Noor A, Kabaria C, Manh B, Elyazar I, Brooker S, Smith D, Moyeed R, Snow R, 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med* 6: e100048.
- Cohen J, Dlamini S, Novotny J, Kandula D, Kunene S, Tatem A, 2013. Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malar J* 12: 61.
- Noor A, Moloney G, Borle M, Fegan G, Shewchuk T, Snow R, 2008. The use of mosquito nets and the prevalence of *Plasmodium falciparum* infection in rural south central Somalia. *PLoS One* 3: e2081.
- Rulisa S, Kateera F, Bizimana J, Agaba S, Dukuzumuremyi J, Baas L, Harelimana J, Mens P, Boer K, de Vries PJ, 2012. Malaria prevalence, spatial clustering and risk factors in a low endemic area of eastern Rwanda: a cross sectional study. *PLoS One* 8: e69443.
- Magalhaes R, Langa A, Sousa-Figueiredo J, Clements A, Nery SV, 2012. Finding malaria hot-spots in northern Angola: the role of individual, household and environmental factors within a meso-endemic area. *Malar J* 11: 385.
- Noor A, ElMardi K, Abdelgader T, Patil A, Amine A, Bakhier S, Mukhtar M, Snow R, 2012. Malaria risk mapping for control in the Republic of Sudan. *Am J Trop Med Hyg* 87: 1012–1021.
- MOH, 2012. *Zambia National Malaria Indicator Survey 2012*. Lusaka, Zambia: Ministry of Health Government of the Republic of Zambia.
- Mukonka V, Chanda E, Haque U, Kamuliwo M, Mushinge G, Chileshe J, Chibwe K, Norris D, Mulenga M, Chaponda M, Muleba M, Glass G, Moss W, 2014. High burden of malaria following scale-up of control interventions in Nchelenge district, Luapula province, Zambia. *Malar J* 13: 153.
- Lowther S, Curriero F, Shields T, Ahmed S, Monze M, Moss W, 2009. Feasibility of satellite image-based sampling for a health survey among urban townships of Lusaka, Zambia. *Trop Med Int Health* 14: 7–18.
- Maidment D, 2002. *Arc Hydro: Gis for Water Resources*. Press E, ed. Redlands, CA: ESRI.
- Tarboton D, Bras R, Rodriguez-Iturbe I, 1991. On the extraction of channel networks from digital elevation data. *Hydrol Processes* 5: 81–100.
- Waller LA, Gotway CA, 2004. *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: Wiley.
- R Core Team, 2013. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical computing.

30. Hahn D, Wanjala P, Marx M, 2013. Where is information quality lost at clinical level? A mixed-method study on information systems and data quality in three urban Kenyan ANC clinics? *Glob Health Action* 6: 21424.
31. Chandler C, Jones C, Boniface G, Juma K, Reyburn H, Whitty C, 2008. Guidelines and mindlines: why do clinical staff over-diagnose malaria in Tanzania? A qualitative study. *Malar J* 7: 53–66.
32. Githinji S, Kigen S, Memusi D, Nyandigisi A, Wamari A, Muturi A, Jagoe G, Ziegler R, Snow R, Zurovac D, 2014. Using mobile phone text messaging for malaria surveillance in rural Kenya. *Malar J* 13: 107–116.
33. Valle D, Lima JT, 2014. Large-scale drivers of malaria and priority areas for prevention and control in the Brazilian amazon region using a novel multi-pathogen geospatial model. *Malar J* 13: 443–457.
34. Vittor A, Gilman R, Tielsch J, Glass G, Shields T, Lozano W, Pinedo-Cancino V, Patz J, 2006. The effects of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian amazon. *Am J Trop Med Hyg* 74: 3–11.