

Multi-Indicator and Multistep Assessment of Malaria Transmission Risks in Western Kenya

Guofa Zhou,^{1*} Daibin Zhong,¹ Ming-Chieh Lee,¹ Xiaoming Wang,¹ Harrysone E. Atieli,² John I. Githure,³ Andrew K. Githeko,⁴ James Kazura,⁵ and Guiyun Yan¹

¹Program in Public Health, University of California, Irvine, California; ²School of Public Health and Community Development, Maseno University, Kisumu, Kenya; ³International Center of Excellence in Malaria Research, Tom Mboya University College, Homabay, Kenya; ⁴Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya; ⁵Center for Global Health and Diseases, Case Western Reserve University, Cleveland, Ohio

Abstract. Malaria risk factor assessment is a critical step in determining cost-effective intervention strategies and operational plans in a regional setting. We develop a multi-indicator multistep approach to model the malaria risks at the population level in western Kenya. We used a combination of cross-sectional seasonal malaria infection prevalence, vector density, and cohort surveillance of malaria incidence at the village level to classify villages into malaria risk groups through unsupervised classification. Generalized boosted multinomial logistics regression analysis was performed to determine village-level risk factors using environmental, biological, socioeconomic, and climatic features. Thirty-six villages in western Kenya were first classified into two to five operational groups based on different combinations of malaria risk indicators. Risk assessment indicated that altitude accounted for 45–65% of all importance value relative to all other factors; all other variable importance values were < 6% in all models. After adjusting by altitude, villages were classified into three groups within distinct geographic areas regardless of the combination of risk indicators. Risk analysis based on altitude-adjusted classification indicated that factors related to larval habitat abundance accounted for 63% of all importance value, followed by geographic features related to the ponding effect (17%), vegetation cover or greenness (15%), and the number of bed nets combined with February temperature (5%). These results suggest that altitude is the intrinsic factor in determining malaria transmission risk in western Kenya. Malaria vector larval habitat management, such as habitat reduction and larviciding, may be an important supplement to the current first-line vector control tools in the study area.

INTRODUCTION

Since the year 2000, malaria burden has declined substantially throughout sub-Saharan Africa because of scaled-up mass distribution of long-lasting insecticidal nets (LLINs), indoor residual spraying (IRS), artemisinin-based combination therapies (ACTs), and improved diagnostic and clinical case management.¹ Nonetheless, malaria remains a major public health challenge in Africa and other malaria-endemic regions.^{1–4} The scaled-up interventions have caused changes in vector biology and malaria epidemiology. Malaria vectors have developed widespread resistance to multiple insecticides including pyrethroids, the key component of LLINs.^{3,5} Outdoor transmission has become an increasingly important issue facing malaria control.^{1,3,6,7} The reduction in malaria transmission is highly heterogeneous; some areas have low or very low transmission, whereas others still have high transmission levels.^{1,2,4} The changes in malaria epidemiology and vector biology require improved intervention strategies.^{3,6,7} This is especially important from an operational planning point of view.^{7–9}

Kenya is one of the malaria-endemic countries in Africa, and malaria burden there has been substantially reduced through the use of first-line interventions, including LLINs, IRS, and ACTs.¹⁰ However, malaria transmission in western Kenya remains high,^{8,10–12} coupled with high vector insecticide resistance and outdoor transmission.^{13–15} New interventions using pirimiphos-methyl (Actellic 300CS, Syngenta, Basel, Switzerland) IRS have been implemented in some counties in western Kenya by the President's Malaria Initiative,^{16–18} and preliminary results are encouraging.^{18,19} The government of

Kenya has also proposed to implement piperonyl butoxid-treated LLINs and larval source management (LSM) in targeted areas in future interventions.^{7,8} All of these new interventions are being or will be implemented regionally, for example, at the county or subcounty level, according to the government's 2019–2023 vector control strategies.⁸ However, determining the right intervention for each area and the operational plan that minimizes logistics and operational costs will depend on area-wide risk assessment and risk stratification in addition to prior effectiveness assessment.^{8,20} For example, in western Kenya, interventions in lowland areas with high transmission may differ from those in highland areas with unstable transmission because of differences in epidemiology, entomology, and environment.^{8,10,11,15,21,22} Highland regions are classified as fringe areas of malaria transmission and are therefore considered possible targets for elimination; thus, they may need the kind of targeted interventions proposed in the government's malaria control plans.^{8,11,14,15,18,23,24} Establishing risk factors based on past observations of infections and vector populations may help to inform the optimal operational plans and facilitate resource allocation.^{7,8,11,21,24} This is also part of the Kenya government's 2019–2023 malaria surveillance plan to improve decision-making for program performance, including mapping the mosquito breeding sites in malaria hotspot, strength health facility surveys, supporting community surveys, and conducting entomological surveillance.⁸ In this context, multi-observation multi-indicator-based surveys are essential for finding the risk factors to support decision-making while better using the surveillance data available.

Risk factor analysis is an important step in determining the appropriate intervention strategies and operational planning for a given ecological and epidemiological setting.²⁵ Previous studies of risk factor assessment have focused mainly on risks at the individual, household, or village level.^{11,19,21,25–27} For example, Cook et al.¹¹ assessed factors related to individual

* Address correspondence to Guofa Zhou, Program in Public Health, University of California, Rm. 3066, Hewitt Hall, Irvine, CA 92697. E-mail: zhoug@uci.edu

malaria infections, including gender, age, household socio-economic status, altitude, malaria prevention measures used (e.g., LLIN and IRS), and travel out of the study area. All of these are individual- or household-level risks; that is, they focus on the risk of acquiring infections. Abong'o and others looked at the effects of house structures (e.g., open eaves) and livestock-keeping on indoor vector density.¹⁹ These studies are important for informing the public about ways to reduce malaria transmission—for example, closing or screening eaves to deter mosquitoes from entering the house. Some of these household-level risk factors can be incorporated into policies, such as encouraging locals to use bed nets; however, such policies can be difficult to implement from an operational point of view. At the village level, most studies have assessed malaria risks using environmental factors (including climatic parameters) mainly observed from satellite images.^{25–31} For example, in a typical study, Solano-Villarreal et al.²⁷ assessed the factors associated with high malaria risk (annual parasite index [API] > 10 or API > 50 cases/1,000 people) in different villages. These factors included cumulative annual rainfall, forest coverage, annual forest loss, annual mean land surface temperature, normalized difference vegetation index (NDVI), normalized difference water index (NDWI), shortest distance to rivers, time to populated villages, and population density.²⁷ In their study, distance to the nearest rivers and time to populated villages were among the important risk factors.²⁷ This type of risk assessment study, that is, one using environmental data, may be more relevant for informing operational planning and intervention strategies,^{32,33} for example, to determine suitable interventions for different risk areas.^{8,34} In nearly all of these risk factor assessment studies, whether at the individual or village level, only one transmission indicator

was considered. In most cases, only observation at one time point was recorded (e.g., one observation of malaria infection prevalence or API).²⁵ This may not be enough, however. For example, malaria infection prevalence varies seasonally in many places, using infection prevalence from different seasons will likely end up with different sets of risk factors.^{22,35,36} A more appealing approach might be to include multiple observations so that uncertainties such as seasonal variation can be reduced or minimized, or to use multiple transmission indicators such as combined epidemiological indicators with entomological observations,^{20–22,37} which reduces uncertainty in risk assessments. Because the government already planned the surveillances, this approach will better use the surveillance data available.

This study aimed to measure malaria risks at the village level and to predict risks using environmental, biological, socio-economic, and climatic features, and then to further classify villages into groups for intervention operational purposes based on risk factors using a multistep approach. We measured malaria risks using a combination of three malaria metrics: cross-sectional seasonal malaria infection prevalence and vector density, and malaria incidence based on cohort surveillance at the village level.

MATERIALS AND METHODS

Study area. The study area is located in Nyakach sub-county of Kisumu County, western Kenya (34°49' E to 34°59' E, 0°15' S to 0°22' S) (Figure 1). The study area covering ~300 km² spans from the lowlands along the Lake Victoria shoreline (altitude 1,120 m above sea level [a.s.l.]) to the highlands (altitude 1,670 m a.s.l.), and the study sites included

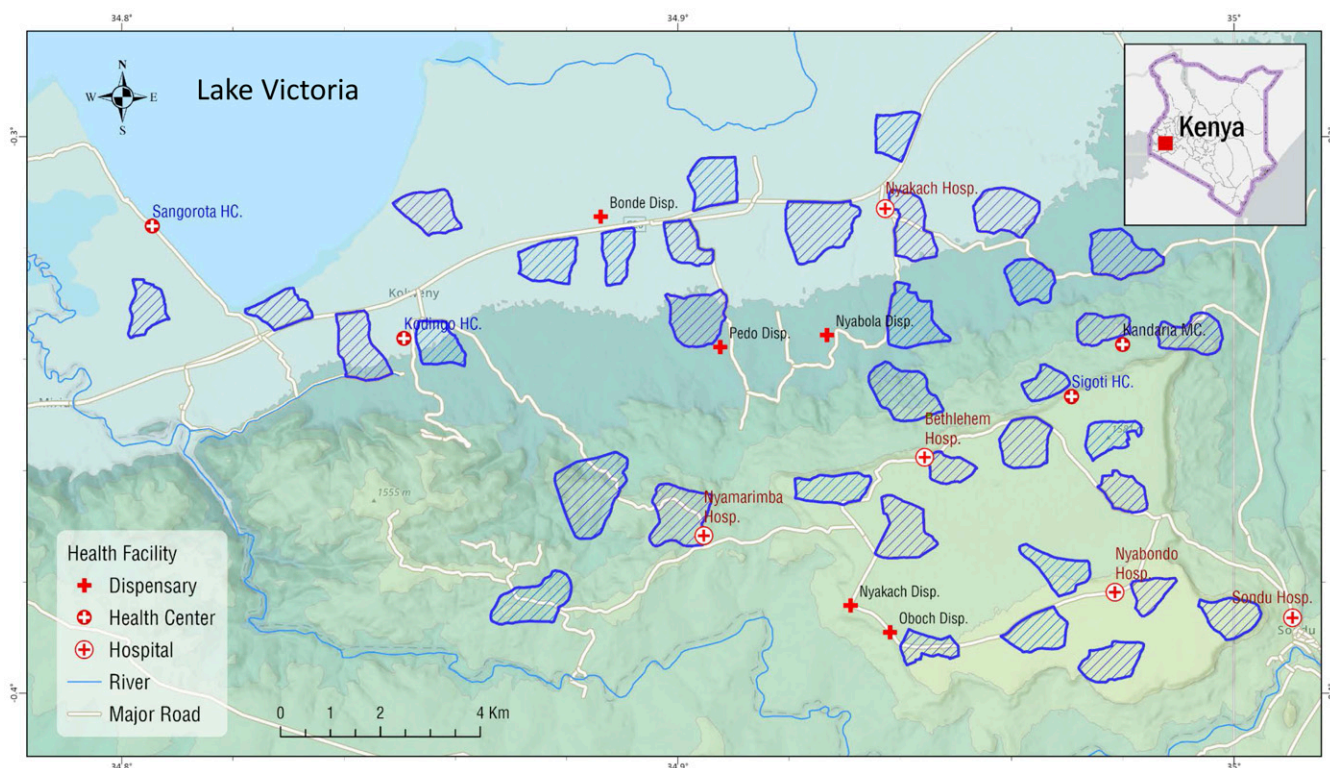


FIGURE 1. Study area and locations of clusters. This figure appears in color at www.ajtmh.org.

36 randomly selected clusters from four wards of Nyakach subcounty. A cluster is a village or several neighboring villages, typically covering an area of approximately 2 km² and comprising 200–300 households and about 600–1,000 inhabitants. Local residents are predominantly farmers who depend on crop farming and cattle/goat herding for subsistence. Fishing is carried out by communities living along the Lake Victoria shoreline. Malaria transmission is seasonal, with two peaks in vector abundance reflecting the bimodal rainfall pattern: a major peak between April and July and a minor peak between October and November.²² Annual precipitation is about 1,500 mm. Malaria is predominantly caused by *Plasmodium falciparum*.^{22,38} The main malaria vectors in the area are *Anopheles gambiae* sensu stricto, *Anopheles arabiensis*, and *Anopheles funestus* sensu lato.^{14,18,22} In the study area, a 2017–2018 survey found that *An. funestus* accounted for 45% of all *Anopheles* captured, followed by *An. gambiae* sensu lato (37%) and other *Anopheles* species (18%), and that 89% of *An. gambiae* s.l. were *An. arabiensis*.¹⁸ Mosquito resistance to different insecticides, including pyrethroids, has been reported in the study area.¹⁸ Household LLIN ownership was > 90% in the study area in 2017.¹⁸

Malaria transmission indicator surveys. Malaria transmission indicators surveyed included malaria cross-sectional seasonal infection prevalence and vector density, and cohort surveys of malaria incidence.

Asymptomatic parasite infection prevalence. Cross-sectional malaria parasite infection surveys were conducted in June 2019 and November 2019, which, respectively, represented the high and low transmission seasons. In each cluster, 100 participants were randomly selected. Based on previous study conducted in the nearby Kendu Bay area, parasite prevalence was about 50% in 2015.³⁸ Using Poisson regression, 36 clusters with a sample size of 100 and population size of 1,000 per cluster would be able to detect a 5% variation in parasite prevalence with 80% power, based on a significance level of 5% and assuming the coefficient of variation in true proportions between clusters was 0.01. House location and altitude were determined for each participant using a handheld GPS.^{22,36} On signing of the informed consent/assent (for minors younger than 18 years) forms, blood samples were collected by the standard finger-prick method.³⁶ Thin and thick blood smears were prepared for laboratory microscopy examination and filter paper blood dots were prepared for PCR detection of parasite infection status and species.^{36,38} All slides were examined by two experienced laboratory technicians at the International Center of Excellence for Malaria Research laboratory at Tom Mboya University College, Homabay, Kenya, to identify the parasite species. For quality control purposes, a third technician randomly selected 5% of the slides for reexamination.³⁶

Active malaria case detection. Clinical malaria incidence was identified through active case detection conducted from January to April 2020. A cohort of 150 households, comprising about 500 residents, was selected randomly from each of the 36 clusters, and all residents in the selected households were recruited for participation based on the inclusion and exclusion criteria. House location and altitude for each participating household were determined using a handheld GPS.³⁶ Written informed consent/assent (for minors younger than 18 years) for study participation was obtained from all consenting heads

of households and from each individual who was willing to participate in the study. The inclusion criteria were as follows: provision of informed consent/assent and no reported chronic or acute illness except malaria or fever. The exclusion criteria were as follows: unwillingness to participate or being younger than 6 months. Participants were visited every other week by a team of government trained and certified local community health volunteers (CHVs) and screened for clinical malaria. A clinical malaria case is defined as an individual with fever (axillary temperature of 37.5°C or higher) and other related symptoms such as chills, severe malaise, headache, or vomiting at the time of examination or 1–2 days before the examination, with a *Plasmodium*-positive blood smear or rapid diagnosis test (RDT).³⁶ For all fever and suspected malaria cases, blood samples were taken, malaria infection was tested using RDT, thin and thick smears prepared on a labeled slide, and blood dots prepared on a labeled filter paper. Body temperature and symptoms and signs of illness were recorded on a case report form. Blood testing was only performed with individuals who currently had fever (based on testing by the CHV) and other malaria symptoms, or who reported having had fever and other malaria symptoms in the prior 2 days. Clinical cases were referred to the nearest government-run hospital or health center for free treatment. The incidence rate (cases/1,000 people/year) estimated from this active case surveillance was used as one of the risk indicators in this study.

Malaria vector abundance samplings. Vector abundance is a potential measure of malaria transmission intensity. Indoor-resting mosquitoes were collected from all clusters using the pyrethrum spraying collection (PSC) method in July 2019 and January 2020. In each cluster, 20–25 houses were randomly selected for vector sampling. House location and altitude were determined using a handheld GPS.^{13,36} Mosquito species were identified morphologically,¹³ and female anopheline mosquitoes were classed as unfed, blood fed, half-gravid, and gravid.³⁹ The number of sleepers in each sampling house was recorded during the surveys. Specimens of *An. gambiae* s.l. were further analyzed by rDNA-PCR for species identification.^{22,40}

Malaria risk factor surveys. *Aquatic habitat and mosquito larval survey.* Aquatic habitats were searched within all clusters and within 1.0 km from the boundary of all clusters. All aquatic habitats were located using GPS and examined for *Anopheles* larval and pupal abundance using standard dippers.⁴¹ Aquatic habitat surveys were conducted in July 2019 and March 2020. In addition to seasonal cross-sectional surveys, “large” water ponds were identified using Google Maps and their latitude/longitude was recorded. These ponds were likely permanent habitats, regardless of season and rainfall.

Climatic data. Climatic data were obtained from the National Aeronautics and Space Administration Earth Observation satellite images (<https://earthobservatory.nasa.gov>). Variables included monthly and annual mean daytime/nighttime LSTs in 2018 and 2019 (MOD11A2, MOD11A3 MODIS/Terra LST/Emissivity 1 km SIN Grid), monthly and annual cumulative precipitation in 2018 and 2019 (global precipitation measurement IMERG Final Precipitation L3 1 month 0.1° × 0.1° V06), and monthly and annual mean net evapotranspiration in 2018 and 2019 (MOD16A3GF v006 MODIS/Terra Net Evapotranspiration Gap-Filled Yearly L4 Global 500 m SIN Grid).

Landscape, vegetation cover, and topographic data. These variables included three major categories related to the formation of larval habitats: geomorphometry, vegetation cover from spectral analysis of satellite images, and soil surveys (Figure 2). Geomorphometric variables included 30-m resolution digital elevation model and derived variables, that is, aspect, slope (in degrees and percentage rise), curvature, curvature profile, curvature plan, topographic wetness index, module multi-resolution index of valley bottom flatness, ground surface roughness, and solar radiation measures (monthly global radiation, direct radiation, diffused radiation, and duration of radiation). Vegetation cover observed from Landsat 8 includes 30-m resolution monthly and annual averages of the NDVI, enhanced vegetation index (EVI), normalized difference built-up index, NDWI, soil-adjusted vegetation index (SAVI), and modified SAVI for 2018 and 2019 (Supplemental Table S1). Soil properties (LandGIS Soil Water contents at 33 kPa and 1500 kPa) included absolute depth to bedrock, bulk density (fine earth), soil water capacity, and texture class. Additional parameters include land cover types such as water body, grassland, and cropland (MCD12Q1v006 MODIS/Terra + Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid), and tree cover values (type, percentage coverage, etc.).

Socioeconomic data. A demographic survey was conducted in 2019 including all participating households and individuals in all clusters.³⁷ Parameters recorded from the demographic survey included household demographics

(number of residents, age, and gender), house type, and malaria prevention measures (number of nets, net usage, IRS, repellent, and others). Other parameters included distance to major roads and healthcare facilities. House type is an indirect estimate of household economic status based on roof, wall, and ground material types. Distance to a major road where public transportation is available is an indirect measure of healthcare availability and socioeconomics. Distance to health facilities is a good indicator of the availability of healthcare services.

Other derived risk factors. Other variables included distance to the nearest aquatic habitat (including water ponds), number of habitats within 1.0 km of the cluster, weighted number of habitats, distance to the nearest river, and distance to Lake Victoria. The weighted number of habitats was calculated using an exponential decay weight with a maximum of one (100%) and a minimum of 0.05 (5%) at a distance of 1 km; that is, habitat adjacent to the house was assigned a 100% contribution to malaria transmission, and the contribution declined to 5% by 1 km. Previous studies have used only the distances to rivers/habitats, neither the number of habitats nor the distance weighted number of habitats.^{25,28–30} However, having one habitat versus 10 similarly sized habitats in a village can make a big difference in the adult mosquito density in the village. In addition, since distance determines dispersal effects, the number of habitats and distance-weighted number of habitats may be important risk factors.

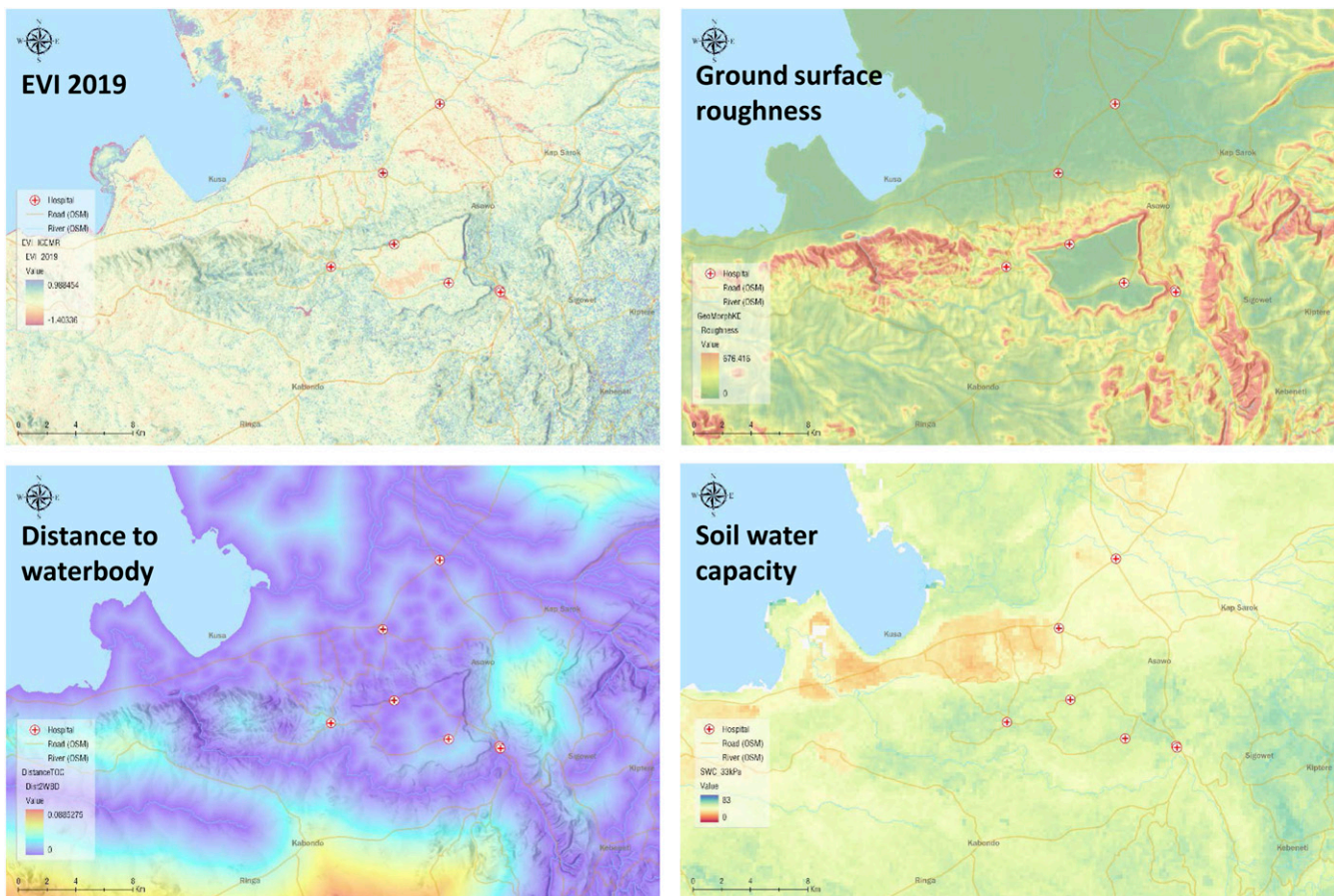


FIGURE 2. Examples of satellite images. This figure appears in color at www.ajtmh.org.

Malaria risk modeling. This study used a multistep approach to assess malaria transmission risks, analyze risk factors, and eventually classify villages into risk groups for operational planning purposes (Figure 3). First, we used field observations of malaria transmission indicators from each cluster to classify clusters into groups (i.e., observation-based risk groups) using unsupervised classification methods. Then, we applied risk factor analysis to the risk groups to determine key risk factors using machine learning multinomial logistic regression analysis. Principal component analysis (PCA) was used to reduce the redundancy in risk factor data before risk analysis. We then used unsupervised classification to reclassify the clusters into new, risk factor-adjusted groups based on the risk indicators adjusted by key risk factors. Finally, risk factors were reanalyzed based on risk factor-adjusted groups (Figure 3).

Risk grouping by classification of clusters based on risk indicators. First, clusters were classified into risk groups based on field observations of the three malaria transmission indicators: parasite infection prevalence for each cluster, malaria vector abundance, and malaria case incidence (Figure 3). These indicators are all on the Kenya government’s surveillance list.⁸ To reduce uncertainty due to seasonality or variations in each indicator, we selected three different combinations of risk indicators: prevalence + vector density, prevalence + incidence, and prevalence + vector density + incidence. We included parasite infection prevalence in all cases because it has been included in all malaria indicator surveys in malaria-endemic African countries.¹² The original data were normalized for each indicator at each sampling occasion before performing the clustering analysis. To further reduce uncertainty due to classification method, we tested both hierarchical and *k*-means clustering methods.⁴² For the

k-mean clustering, we selected all *k* values that satisfied the following criteria: 1) all indicator variables must be significantly different in their mean values between groups after classification, and 2) each group must include at least five members (clusters). The purpose of using different combinations of transmission indicators and different methods to assess cluster-level risks was to 1) evaluate the consistency of the classification results, 2) reduce data selection bias due to specific indicator selected, and 3) determine the common key risk factors.

Principal component analysis for dimension reduction and removal of redundancy. We started with 200+ risk variables, some of which were significantly linearly correlated. To reduce the dimension (thus reducing the complexity of the risk assessment and avoiding overfitting) and remove redundancy due to linear correlation, PCA was performed on risk factor variables.^{43,44} Principal component analysis is one of the most popular linear dimension reduction algorithms. It maps data in a higher-dimensional (more variable) space to data in a lower-dimensional (less variable) space while maximizing the variance of the data in the lower-dimensional space, that is, minimizing the loss of information.⁴⁵ We applied PCA to all risk factors based on normalized data. Dimension reduction was performed in two steps: selection of the top contributing principal components and selection of the top contributing variables.⁴⁵ Principal components were selected by applying two criteria: 1) total variance explained > 80% and 2) all eigenvalues after the sharp bend in the scree plot are rejected. The criterion that produced fewer principal components was chosen. The top contributing variables were selected from the previously selected principal components.⁴⁵ Selection of the top contributing variables was based on three criteria: 1) the variable occurs before the sharp bend in the variable scree

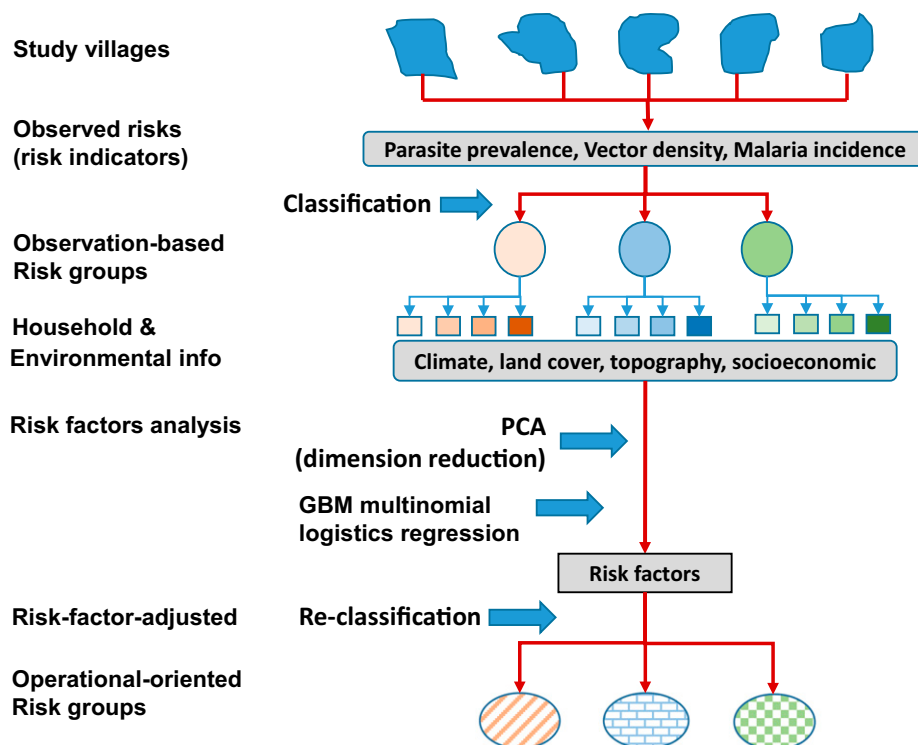


FIGURE 3. Flowchart of data analysis process. This figure appears in color at www.ajtmh.org.

plot, 2) the contribution exceeds the expected average contribution, and 3) there is no overlap in variables selected from different principal components (i.e., no collinearity).⁴⁵ Only the selected variables were used in the subsequent risk factor analysis (Figure 3).

Risk factor analysis with multinomial logistic regression. Based on observed risks, clusters were assigned to different observation-based risk categories obtained from “Risk grouping” classification analysis. These risk groups were used as multinomial dependent variables against risk factors (at each household in different clusters) selected by PCA (Figure 3). A total of 1,938 observations (households) for all risk factors were used in the risk analysis. A generalized boosted regression model (GBM) with multinomial logistic regression analysis was used to identify the risk factors that significantly associated with different risk levels.^{27,46,47} Ten-fold cross-validation was applied to validate the model performance. Model performance was evaluated using classification accuracy, sensitivity, and specificity, both overall and for each observation-based risk group. Variable relative importance was calculated as the percentage of variance explained.

Altitude-adjusted reclassification and reclassification-based risk factors. In this step, we wanted to test if moving certain key risk factors into the “risk indicator” category, but use it as a control variable to do reclassification, would improve the grouping results for operational purposes. This step is not seen in conventional risk assessment analysis. The principle behind it is simple. In Kenya, malaria transmission intensity is strongly affected by altitude. The highlands of western Kenya, where malaria transmission is seasonal and unstable, are considered an epidemic fringe area, whereas the lowlands of western Kenya, where malaria is perennial, are considered moderate-to-intense endemic.^{11,12,21,48} Altitude is like an intrinsic driving force that affects malaria transmission fundamentally.⁸ We wanted to test to see if altitude was the single standout risk factor for malaria transmission. If the answer was yes, we wanted to see how the variable importance and cluster regrouping changed when we reanalyzed the risk factors using altitude-adjusted risk classifications, that is, using altitude as control variable. Therefore, this step involved reclassification of clusters based on adjusted risk indicators, and risk factor analysis based on the reclassification results (Figure 3).

Mapping and satellite images were processed using ArcGIS 10.0 (ESRI, Redlands, CA), and all other data analysis was performed using R \times 64 3.6.3 (The R Foundation for Statistical Computing, Vienna, Austria).

Informed consent and ethical clearance. Ethical clearance was obtained from the Ethical Review Committee of Maseno University, Kenya (MSU/DRPI/MUERC/00778/19), and the Institutional Review Board of the University of California, Irvine, USA (HS# 2017-3512). Written consent was obtained from all adult participants. Written assent for children (< 18 years of age) was obtained from the participants and their parents or guardians. Inclusion criteria were provision of informed consent (assent for children) and no reported chronic or acute illness other than malaria. Exclusion criteria were unwillingness to participate in the study, reported chronic or acute illness other than malaria, or age < 6 months.

RESULTS

Field observations. Table 1 shows a summary of field observations of parasite infection prevalence (parasite identified by PCR), mosquito abundance (adult and larval), and malaria incidence based on cohort active case surveillance. The results indicate that parasite prevalence and larval density varied substantially between seasons, although indoor-resting mosquito density did not change much between seasons (Table 1). Malaria incidence was estimated as 351.2 ± 61.5 cases/1,000 people/year. Overall, the study area can probably be classified as having moderate-to-intense endemic transmission.^{8,48,49}

Classification of transmission intensity. Using different transmission indicators (Table 1), study clusters were classified into two to five groups depending on the variables used, with each group representing different risk levels (Figure 4). Using a combination of parasite infection prevalence and indoor-resting vector density, clusters could be classified into two, three, or five groups with a significant difference in risk indicators among groups; using parasite infection prevalence, vector density, and malaria incidence could classify clusters into two or three groups; and using parasite infection prevalence and malaria incidence could classify clusters into two or three groups (Supplemental Text S1). Overall, seven grouping results were produced (Supplemental Text S1). Different combinations of risk indicators led to different classification results (i.e., inconsistent assignments of cluster members), as shown by the example in Figure 4. This was no surprise because different risk indicators can be affected by different factors. This highlighted the importance of indicators selection process.

Dimension reduction of potential risk factors. Principal component analysis results illustrated that the top 14

TABLE 1

Descriptive statistics of field observations of parasite infection prevalence, mosquito density, and malaria incidence (calculated based on active case surveillance)

Items	Survey month/year	Sample size (95% CI)
Asymptomatic malaria infection (individuals, PCR prevalence %)	June 2019	3,514 (12.3 \pm 3.4)
	November 2019	3,933 (26.3 \pm 4.1)
Indoor-resting mosquito pyrethrum spraying collection (house-nights, density in females/house)	July 2019	720 (0.65 \pm 0.27)
	January 2020	848 (0.52 \pm 0.18)
Larval breeding habitat (habitats, abundance in larvae/10 dips)	July 2019	1,134 (3.03 \pm 0.56)
	March 2020	715 (0.82 \pm 0.15)
	Water ponds*	208 (n.a.)*
Malaria incidence (visits, incidence rate in cases/1,000 people/year)	January–April 2020	45,791 (351.2 \pm 61.5)

n.a. = not available.

* Large water pond identified from satellite image and larval density not available.

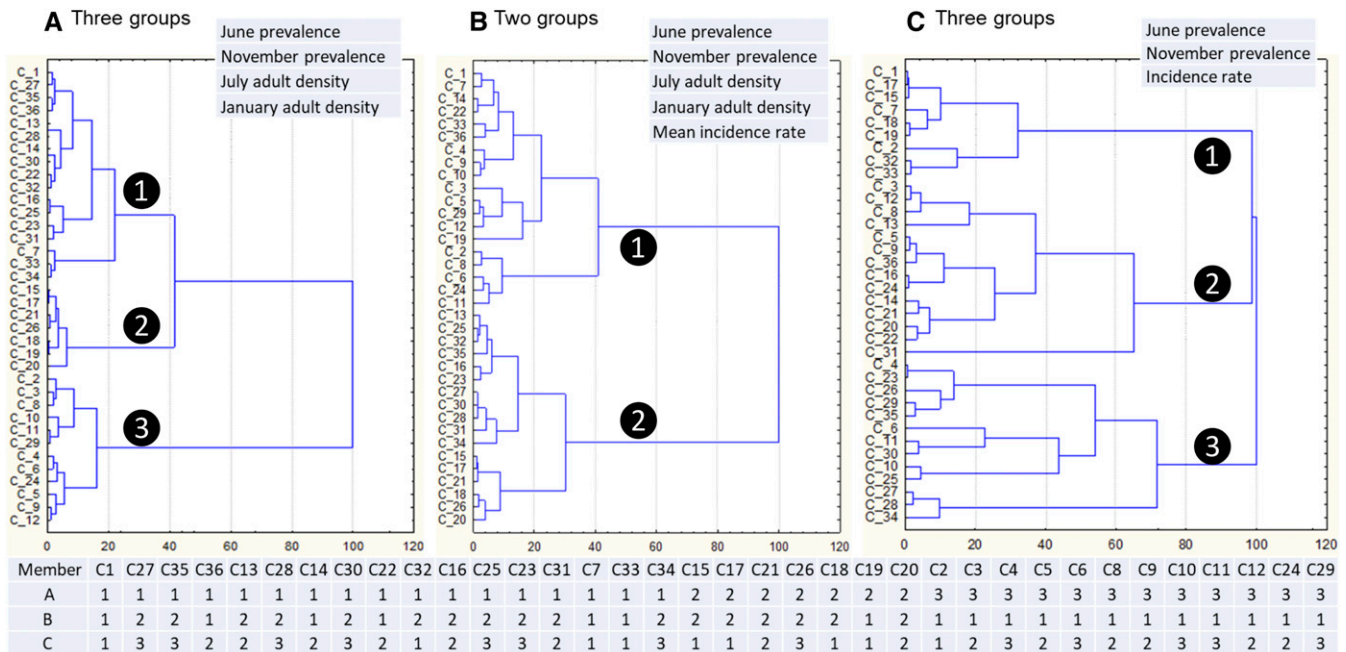


FIGURE 4. Examples of classification (top) and member assignment (bottom) based on observed risk indicators. The numbers corresponding to each member village within the table at the bottom represent the group number (white number inside the black circle) in each classification result. C1, C2, . . . , etc. are the labels of each village. This figure appears in color at www.ajtmh.org.

components explained 95.5% of data variance and the top five components explained 80.6%. The scree plot showed a sharp bend between the third and fourth components. To avoid over-reduction of risk factors, the top five components were selected for dimension reduction. After evaluation of the variables contributing to each principal component, 53 risk factors were selected for further risk analysis.

Risk factor analysis. Initial risk factor assessment was conducted based on the classification results and the 53 selected risk factors. There were seven classification results (Supplemental Text S1), and a GBM multinomial logistic model was run on each of them. Results showed that all models produced predictions with > 0.88 accuracy, sensitivity, and specificity, except one model that predicted group-level sensitivities at < 0.8 (Supplemental Text S2). These results indicated that model predictions were reasonably good.

The results of variable importance (measured as accounted for % of the variance) were especially interesting. In all models, altitude was the only variable that accounted for 44.5–64.5% of the variance. All other variable importance values were < 6% of the variance in all models for all variables (Supplemental Text S2); that is, there was at least a 7-fold difference in variable importance between altitude and all other factors. Other important risk factors included number of habitats, distance to river/lake, vegetation index, wetness index, and topographic features (Supplemental Text S2).

Reclassification of clusters adjusted by altitude. Because altitude accounted for about half of all risk in all cases, we moved altitude from the risk factor group to the “risk indicator” group. Now, we had four types of risk indicators: parasite infection prevalence, vector density, malaria incidence, and altitude. Very interestingly, once altitude was added as a risk indicator, the unsupervised classification of study clusters yielded the same results, that is, the three

groups shown in Figure 5, regardless of the combinations of variables and classification methods. The group on top (Figure 5 right, group 1), where altitude was low, included 13 clusters, with 12 clusters significantly below the average altitude (Figure 5 right). Many clusters in this group were located in the Lake Victoria shore plain area (Figure 6, Supplemental Figure S1), with low altitude but high transmission intensity measured in all transmission indicators (Figure 5 left). This group can be called lake plain zone. The middle group (Figure 5 right, group 2) had medium but varying altitude, and most clusters in this group had lower transmission intensity (Figure 5 left). This group can be called middle slope zone. This group formed a long zone separating the other two groups on the map (Figure 6, Supplemental Figure S1). The bottom group (Figure 5 right, group 3) had significantly higher than average altitude, with low incidence and varied prevalence but low-to-medium vector density (Figure 5 left). This group formed the highland plateau zone (Figure 6, Supplemental Figure S1). Operationally, this grouping is probably the best because it will be much easier to implement interventions assuming different risk zones need different interventions. It was noted that altitude alone might classify the clusters into 2–5 groups with significantly different mean altitudes; however, malaria indicators in most cases were not significantly different among different groups (results not shown). In other words, altitude itself was not enough to be treated as a transmission indicator.

Model-predicted risk probability was similar for each of the three groups (Supplemental S3). For the lake plain zone (Figure 6, Supplemental S3), the average risk probability was 0.61, ranging from 0.51 to 0.67. For the middle slope zone (Figure 6, Supplemental S3), the average risk probability was 0.65, ranging from 0.54 to 0.73. The highland plateau had the highest risk probability (Figure 6, Supplemental S3), and the average risk probability was 0.67, ranging from 0.59 to 0.72.

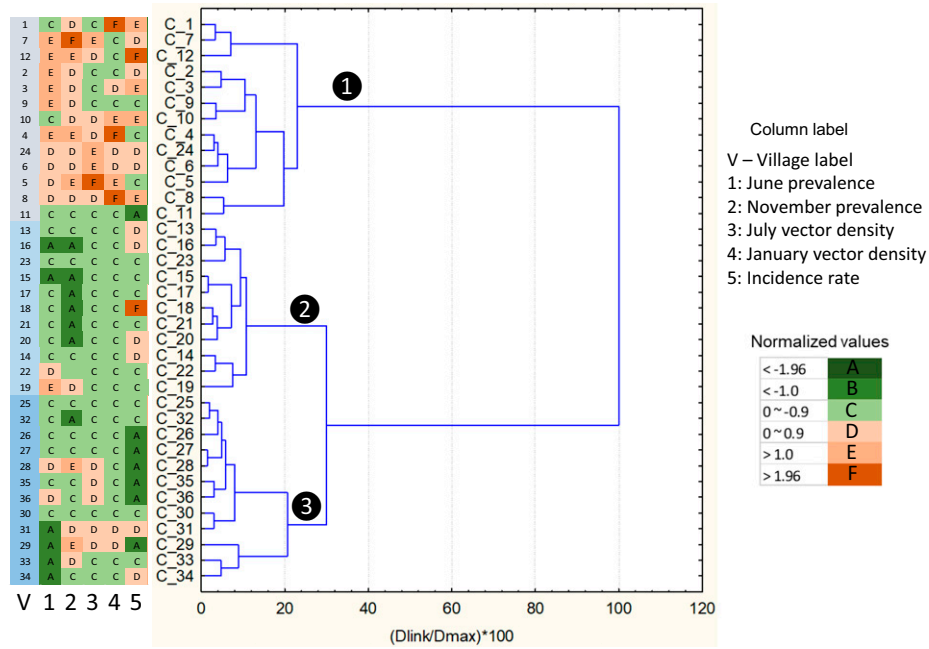


FIGURE 5. Reclassification of clusters after altitude was treated as a risk indicator. Numbers in white color within the black circle represent number of clusters. This figure appears in color at www.ajtmh.org.

Important risk factors other than altitude: Risk reassessment. Based on the results of the cluster reclassification, we reassessed the risk factors other than altitude. The same GBM multinomial logistic analysis was used to build the model based on the risk variables previously selected by PCA excluded altitude. The overall model prediction showed an accuracy of 80.6%, sensitivity of 71.1%, and specificity of

85.6%. For each group, the sensitivity ranged from 63% to 79% and specificity ranged from 84% to 87%.

The likelihood-ratio χ^2 -test indicated 19 significant variables in the model (Supplementary S4). Their relative importance is shown in Table 2. Seven variables had $\geq 5\%$ importance in the model and two had $> 10\%$ importance (distance to lake at 24.4% and distance to river at 10.6%)

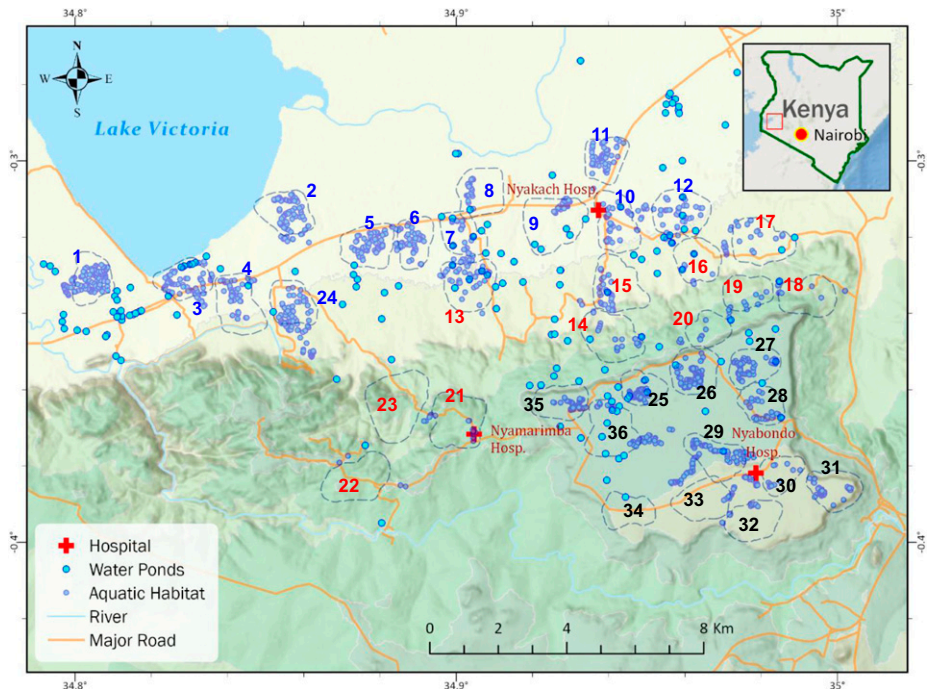


FIGURE 6. Distribution of villages and clusters based on reclassification result. Numbers in different colors indicated different risk groups. Each number represented a cluster as shown in Figure 5. The dash line illustrated the 500-m buffer boundary for each cluster. Blue dots and light blue circles, respectively, represent aquatic habitats and large water ponds found from on satellite images. This figure appears in color at www.ajtmh.org.

TABLE 2
Risk factors and their relative importance, grouped by variable types

Variable type	Variable (group)*	Importance (%)
Hydrology and habitat	Distance to lake	24.4
	Distance to river	10.6
	Weighted <i>N.</i> water ponds	9.2
	Weighted <i>N.</i> habitats, dry season	8.6
	Weighted <i>N.</i> habitats, rainy season	6.9
	<i>N.</i> water ponds	3.0
	Subtotal	62.7
Ground and soil property	Ground surface roughness	7.9
	Valley bottom flatness	5.0
	Soil coarse fragments volumetric	2.7
	Soil water capacity	1.7
	Subtotal	17.3
Vegetation coverage	EVI, February	1.2
	EVI, May	3.7
	EVI, December	3.3
	NDVI, April	1.2
	NDVI, May	1.3
	NDVI, September	1.8
	NDVI, October	2.8
	Subtotal	15.3
Others	No. of nets per household	2.6
	Land surface temperature daytime, February	2.2

* EVI = enhanced vegetation index; NDVI = normalized difference vegetation index.

(Table 2). Weighted numbers of habitats and water ponds were among the top five most important factors (Table 2, Figure 6). Although its importance value was low (2.6%), the number of bed nets owned per household was among the important factors affecting malaria transmission (Table 2). Risk factors could be roughly divided into four groups (Table 2). The first group, which accounted for 62.7% of total importance, was directly linked to habitats: distance to lake, number of habitats, etc. (Table 2). The second group, comprising 17.3% of importance, was associated with ground ponding: valley bottom flatness, soil water capacity, etc. (Table 2). The third group was vegetation cover, indirectly linked to soil wetness, which made up 15.3% of total importance (Table 2). Only about 5% of the variable importance came from the number of bed nets owned (2.6%) and February temperature (2.2%) (Table 2). February is usually the hottest and driest month in the study area. Overall, ground wetness and habitat abundance were the most important risk factors in the study area.

DISCUSSION

Malaria remains a public health threat, especially in sub-Saharan Africa.¹ Malaria risk factor assessment provides useful information for malaria prevention and disease management.^{11,21,32} Individual- and household-level risk assessment provides crucial information for public awareness and personal preventive measures.^{11,21,32} Village and area risk assessment provide critical information for next-level decision making, that is, regional-scale intervention planning and implementation.^{26,27} which is particularly important for guiding governmental malaria control policy. In this study, we used a multi-stage and multi-indicator approach to assess malaria risks in an endemic area in western Kenya. We found that altitude was the single most important risk factor and could potentially be treated as a malaria transmission indicator. When altitude was used as a malaria risk indicator, classification of villages produced near perfect results; villages were

classified clearly into separate areas, which is beneficial for intervention operational purposes at county or sub-county level. Our results are supported by other studies, which have found a distinct difference in malaria transmission between lowlands and highlands.^{50,51} More importantly, the reclassification results are very similar to the current government guidelines; that is, altitude is the decisive factor determining endemicity and thus impacts intervention strategies.^{8,12,50}

Factors related to larval habitats have high relative importance. In addition to identifying altitude as the most important risk factor, we found that the distance to rivers/lakes and (weighted) number of habitats accounted for 2/3 of all risk once altitude was removed. In nearly all previous studies, only the distance to the nearest habitats/river was used as a risk factor^{25-27,29}; the number of habitats was not considered. It's easy to understand the importance of the number of habitats: more habitats can produce more mosquitoes. The weighted number of habitats (including water ponds), which accounts for the impact of distance on the likelihood of dispersal, therefore provides a better measure of risk than distance alone. Other important malaria risk factors include landscape and topographic features, such as ground surface roughness and vegetation cover, derived from satellite observations such as NDVI and EVI in different months. Almost all of these factors relate to habitat formation and the ponding effect.^{52,53} Valley bottom flatness and ground surface roughness determine potential locations of ponds once water is available. Soil water capacity determines the sustainability of aquatic habitats, which in turn determines mosquito larval development.⁵³ Normalized difference vegetation index and EVI are measures of greenness or ground wetness and indicators of precipitation and potential habitat availability.^{54,55} The contributions of preventive measures and climatic factors are limited, which is not surprising given the high temperatures and high coverage of LLIN use in the study area. The study area has an annual average temperature of 25°C with no clear cold period except in the highlands, where temperatures are lower

because of high altitude. During this study, LLIN ownership varied from 82% to 100% in different villages, and usage was between 93% and 100%,¹⁸ so the overall effective coverage (population usage) is between 75% and 100%, which is rather high.⁵⁶ Therefore, the effect on malaria risk of not using LLIN is relatively low.

A multi-indicator approach is important. All previous studies have explored only one malaria transmission indicator (e.g., parasite infection prevalence, confirmed malaria cases, clinical positivity rates, or vector abundance).^{25–27,29} These are malaria transmission indicators, but these indicators are affected by climatic factors, which usually show strong seasonality. In addition, different indicators are potentially affected by different ecological factors. For example, vector adult density, which has been evaluated in many studies,²⁵ is affected by larval habitat availability and productivity, whereas parasite infection is affected by human behaviors such as the use of preventive measures.^{14,57} Therefore, a combination of different transmission indicators may provide a better picture of risk factors and reduced bias of using single indicator to assess risks.^{20–22,35,37}

One way to combine multiple indicators is to create a weighted integrated index.³⁷ However, selection of weight for each indicator is very subjective, which potentially brings additional uncertainty. In this study, we used three indicators with observations in different seasons and multivariate unsupervised classification. Then, we analyzed risk factors using multilevel logistic regression analysis based on the multinomial classification results. This is similar to the single indicator cutoff method^{26,27}; our study can better use the collected integrated information and reduces uncertainty due to the selection of indicators. Although many previous studies have surveyed multiple transmission indicators,²⁵ our study provides a different approach to better use field survey data.

Risk factor dimension reduction is necessary. The conventional method of risk analysis is to directly build a model using transmission indicator survey results against climatic/environmental and socioeconomic factors.^{27,29} There are many potential risk factors, which may be linearly correlated. For example, seasonal variation in vegetation (NDVI) is strongly affected by precipitation (likely time-lagged), although vegetation may change little when rainfall exceeds threshold amounts, which is especially true in some extremely wet areas.^{58,59} Using too many factors may lead to model overfitting, and including correlated variables in a linear model may produce biased results, thus increasing predictive uncertainty.^{43,60–62} Carefully selected predictive variables are especially important when using non-stepwise methods (e.g., non-stepwise linear regression), geographically weighted regression, and certain machine learning methods such as gradient-boosted regression (linear or logistic)^{47,60,62,63}; these are the methods frequently used for risk assessment modeling.^{27,29} In our study, we used PCA before risk analysis to reduce redundant variables.⁴⁴ Indeed, in our risk model, NDVI and EVI were important predictors of transmission intensity, whereas the contribution of precipitation was minimal.

Research impacts policy. The purpose of scientific research is to inform policy, and in fact malaria control policy is strongly influenced by research results.^{8,9,12,18} The purpose of risk assessment is not only to identify the risk factors; here, we went a step further by using risk factors to reclassify villages into risk groups. This method may be called risk factor-adjusted classification. The reclassification of villages clearly

provides a better risk map for operational purposes. None of the previous risk assessment studies have included this step.^{25–27} Other than altitude, which has been frequently used as a key factor defining the malaria transmission intensity in Kenya and elsewhere,^{8,12,18,25} we found that the (weighted) number of habitats and related factors (e.g., distance to nearest river/lake) accounted for 2/3 of malaria transmission risk. This is another step not taken by most previous risk assessment studies.^{25–27} Although the distance to the nearest habitat is important, more habitats within and surrounding a village will clearly increase the malaria risk in the village.⁶⁴ Therefore, habitat management practices such as habitat reduction and larviciding may be an important supplement to the current first-line vector control tools, that is, LLIN and IRS.^{3,8,10,65}

CONCLUSION

Malaria risk analysis is a useful way to inform malaria control policies. This study used multi-indicators for the risk analysis and provided additional information on the operational use of risk assessment results. Policy-makers in malaria-endemic countries should take into account health and environmental factors on determining intervention strategies in areas with different risks, as well as the cost-effectiveness in implementing these strategies. This is especially important in LSM, as larval source was one of the key transmission driving factors in this study.

Received September 16, 2020. Accepted for publication December 24, 2020.

Published online February 8, 2021.

Note: Supplemental text, tables, and figures appear at www.ajtmh.org.

Acknowledgments: We thank all who participated in the field surveys for their assistance with fieldwork. We thank the communities and hospitals/health centers for their willingness to participate in this study. We thank the two anonymous reviewers for their constructive comments and suggestions to make this a better article.

Financial support: This study is funded by the National Institutes of Health (R01 A1050243, D43 TW01505, and U19 AI129326).

Disclosure: The funder has no role in study design; collection, management, analysis, and interpretation of data; writing of the report; or the decision to submit the report for publication. The datasets supporting the conclusions of this article are included within the article (and its additional files). Ethical clearance was obtained from the Ethical Review Committee of Maseno University, Kenya (MSU/DRPI/MUERC/00778/19), and the Institutional Review Board (IRB) of the University of California, Irvine, USA (HS# 2017-3512). Written consent was obtained from all adult participants. Written assent for children (< 18 years of age) was obtained from the participants and their parents or guardians. Inclusion criteria were provision of informed consent (assent for children) and no reported chronic or acute illness other than malaria. Exclusion criteria were unwillingness to participate in the study, reported chronic or acute illness other than malaria, or age < 6 months.

Authors' addresses: Guofa Zhou, Daibin Zhong, Ming-Chieh Lee, Xiaoming Wang, and Guiyun Yan, Program in Public Health, University of California at Irvine, Irvine, CA, E-mails: zhoug@uci.edu, dzhong@uci.edu, mingchi1@uci.edu, xiaomiw1@hs.uci.edu, and guiyuny@uci.edu. Harrysone E. Atieli, School of Public Health and Community Development, Maseno University, Kisumu, Kenya, E-mail: etemesi2012@yahoo.com. John I. Githure, International Center of Excellence in Malaria Research, Tom Mboya University College, Homabay, Kenya, E-mail: jgithure@gmail.com. Andrew K. Githeko, Centre for Global Health Research, Kenya Medical Research Institute,

Kisumu, Kenya, E-mail: githeko@yahoo.com. James Kazura, Center for Global Health and Diseases, Case Western Reserve University, Cleveland, OH, E-mail: jxk14@case.edu.

REFERENCES

- WHO, 2019. *World Malaria Report 2019*. Geneva, Switzerland: World Health Organization.
- Nkumama IN, O'Meara WP, Osier FHA, 2017. Changes in malaria epidemiology in Africa and new challenges for elimination. *Trends Parasitol* 33: 128–140.
- Rabinovich RN et al., 2017. malERA: an updated research agenda for malaria elimination and eradication. *PLoS Med* 14: e1002456.
- Vitor-Silva S et al., 2016. Declining malaria transmission in rural Amazon: changing epidemiology and challenges to achieve elimination. *Malar J* 15: 266.
- WHO, 2018. *Global Report on Insecticide Resistance in Malaria Vectors: 2010–2016*. Geneva, Switzerland: World Health Organization.
- WHO, 2015. *Global Technical Strategy for Malaria 2016–2030*. Geneva, Switzerland: World Health Organization.
- WHO, 2017. *A Framework for Malaria Elimination*. Geneva, Switzerland: World Health Organization.
- National Malaria Control Programme, 2018. *Kenya Malaria Strategy 2019–2023*. Nairobi, Kenya: Ministry of Health.
- Mutero CM, Kramer RA, Paul C, Lesser A, Miranda ML, Mboera LE, Kiptui R, Kabatereine N, Ameshewa B, 2014. Factors influencing malaria control policy-making in Kenya, Uganda and Tanzania. *Malar J* 13: 305.
- Zhou G, Lee MC, Githeko AK, Atieli HE, Yan G, 2016. Insecticide-treated net campaign and malaria transmission in western Kenya: 2003–2015. *Front Public Health* 4: 153.
- Cook J, Owaga C, Marube E, Baidjoe A, Stresman G, Migiro R, Cox J, Drakeley C, Stevenson JC, 2019. Risk factors for *Plasmodium falciparum* infection in the Kenyan highlands: a cohort study. *Trans R Soc Trop Med Hyg* 113: 152–159.
- National Malaria Control Programme (NMCP), Kenya National Bureau of Statistics (KNBS), and ICF International, 2016. *Kenya Malaria Indicator Survey 2015*. Nairobi, Kenya, and Rockville, MD: NMCP, KNBS, and ICF International.
- Degefa T, Yewhalaw D, Zhou G, Lee MC, Atieli H, Githeko AK, Yan G, 2017. Indoor and outdoor malaria vector surveillance in western Kenya: implications for better understanding of residual transmission. *Malar J* 16: 443.
- Ototo EN, Mbugi JP, Wanjala CL, Zhou G, Githeko AK, Yan G, 2015. Surveillance of malaria vector population density and biting behaviour in western Kenya. *Malar J* 14: 244.
- Wanjala CL, Mbugi JP, Ototo E, Gesuge M, Afrane YA, Atieli HE, Zhou G, Githeko AK, Yan G, 2015. Pyrethroid and DDT resistance and organophosphate susceptibility among *Anopheles* spp. mosquitoes, western Kenya. *Emerg Infect Dis* 21: 2178–2181.
- PMI, 2017. *Kenya Malaria Operational Plan FY 2017*. USAID, US President's Malaria Initiative.
- PMI, 2018. *Kenya Malaria Operational Plan FY 2018*. USAID, US President's Malaria Initiative.
- PMI, 2019. *VectorLink Project Kenya, Annual Entomological Monitoring Report. October 2017–September 2018*. Rockville, MD: The PMI VectorLink Project, Abt Associates Inc.
- Abong'o B et al., 2020. Impact of indoor residual spraying with pirimiphos-methyl (Actellic 300CS) on entomological indicators of transmission and malaria case burden in Migori county, western Kenya. *Sci Rep* 10: 4518.
- Killeen GF, Kiware SS, Seyoum A, Gimnig JE, Corliss GF, Stevenson J, Drakeley CJ, Chitnis N, 2014. Comparative assessment of diverse strategies for malaria vector population control based on measured rates at which mosquitoes utilize targeted resource subsets. *Malar J* 13: 338.
- Stevenson JC, Stresman GH, Baidjoe A, Okoth A, Oriango R, Owaga C, Marube E, Bousema T, Cox J, Drakeley C, 2015. Use of different transmission metrics to describe malaria epidemiology in the highlands of western Kenya. *Malar J* 14: 418.
- Zhou G, Afrane YA, Vardo-Zalik AM, Atieli H, Zhong D, Wamae P, Himeidan YE, Minakawa N, Githeko AK, Yan G, 2011. Changing patterns of malaria epidemiology between 2002 and 2010 in western Kenya: the fall and rise of malaria. *PLoS One* 6: e20318.
- Wanjala CL, Waitumbi J, Zhou G, Githeko AK, 2011. Identification of malaria transmission and epidemic hotspots in the western Kenya highlands: its application to malaria epidemic prediction. *Parasit Vectors* 4: 81.
- Bousema T et al., 2016. The impact of hotspot-targeted interventions on malaria transmission in Rachuonyo south district in the Western Kenyan highlands: a cluster-randomized controlled trial. *PLoS Med* 13: e1001993.
- Parselia E, Kontoes C, Tsouni A, Hadjichristodoulou C, Kioutsoukias I, Magiorkinis G, Stilianakis NI, 2019. Satellite earth observation data in epidemiological modeling of malaria, dengue and west Nile virus: a scoping review. *Remote Sens* 11: 1862.
- Bui QT, Nguyen QH, Pham VM, Pham MH, Tran AT, 2018. Understanding spatial variations of malaria in Vietnam using remotely sensed data integrated into GIS and machine learning classifiers. *Geocarto Int* 6049: 1300–1314.
- Solano-Villarreal E et al., 2019. Malaria risk assessment and mapping using satellite imagery and boosted regression trees in the Peruvian Amazon. *Sci Rep* 9: 15173.
- Giardina F, Franke J, Vounatsou P, 2015. Geostatistical modelling of the malaria risk in Mozambique: effect of the spatial resolution when using remotely-sensed imagery. *Geospat Health* 10: 232–238.
- Homan T et al., 2016. Spatially variable risk factors for malaria in a geographically heterogeneous landscape, western Kenya: an explorative study. *Malar J* 15: 1.
- Kabaria CW, Molteni F, Mandike R, Chacky F, Noor AM, Snow RW, Linard C, 2016. Mapping intra-urban malaria risk using high resolution satellite imagery: a case study of Dar es Salaam. *Int J Health Geogr* 15: 26.
- Valle D, Lima JM, 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.
- Alimi TO, Fuller DO, Quinones ML, Xue RD, Herrera SV, Arevalo-Herrera M, Ulrich JN, Qualls WA, Beier JC, 2015. Prospects and recommendations for risk mapping to improve strategies for effective malaria vector control interventions in Latin America. *Malar J* 14: 519.
- Rijal KR et al., 2019. Micro-stratification of malaria risk in Nepal: implications for malaria control and elimination. *Trop Med Health* 47: 21.
- Taffese HS, Hemming-Schroeder E, Koepfli C, Tesfaye G, Lee MC, Kazura J, Yan G, Zhou G, 2018. Malaria epidemiology and interventions in Ethiopia from 2001 to 2016. *Infect Dis Poverty* 7: 103.
- Drakeley C et al., 2017. Longitudinal estimation of *Plasmodium falciparum* prevalence in relation to malaria prevention measures in six sub-Saharan African countries. *Malar J* 16: 433.
- Zhou G, Afrane YA, Malla S, Githeko AK, Yan G, 2015. Active case surveillance, passive case surveillance and asymptomatic malaria parasite screening illustrate different age distribution, spatial clustering and seasonality in western Kenya. *Malar J* 14: 41.
- Zhou G, Wiseman V, Atieli HE, Lee MC, Githeko AK, Yan G, 2016. The impact of long-lasting microbial larvicides in reducing malaria transmission and clinical malaria incidence: study protocol for a cluster randomized controlled trial. *Trials* 17: 423.
- Lo E, Nguyen K, Nguyen J, Hemming-Schroeder E, Xu J, Etemesi H, Githeko A, Yan G, 2017. *Plasmodium malariae* prevalence and csp gene diversity, Kenya, 2014 and 2015. *Emerg Infect Dis* 23: 601–610.
- Gillies MT, Coetzee M, 1987. *A Supplement to the Anophelinae of Africa South of the Sahara (Afro-Tropical Region)*, Vol. 55. Johannesburg, South Africa: South African Institute for Medical Research.
- Ndenga B, Githeko A, Omukunda E, Munyekenye G, Atieli H, Wamai P, Mbogo C, Minakawa N, Zhou G, Yan G, 2006. Population dynamics of malaria vectors in western Kenya highlands. *J Med Entomol* 43: 200–206.

41. Munga S, Minakawa N, Zhou G, Githeko AK, Yan G, 2007. Survivorship of immature stages of *Anopheles gambiae* s.l. (Diptera: Culicidae) in natural habitats in western Kenya highlands. *J Med Entomol* 44: 758–764.
42. Witten IH, Frank E, Hall MA, Pal CJ, 2016. *Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems)*, 4th edition. Cambridge, MA: Morgan Kaufmann.
43. Abdi H, Williams LJ, 2010. Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2: 433–459.
44. Jolliffe IT, Cadima J, 2016. Principal component analysis: a review and recent development. *Phil Trans R Soc* 374: 20150202.
45. Kassambara A, 2017. *Practical Guide to Principal Component Methods in R*. CreateSpace Independent Publishing Platform, Kindle online Edition.
46. Martínez-Rincón PO, Ortega-García S, Vaca-Rodríguez JG, 2012. Comparative performance of generalized additive models and boosted regression trees for statistical modeling of incidental catch of wahoo (*Acanthocybium solandri*) in the Mexican tuna purse-seine fishery. *Ecol Model* 233: 20–25.
47. Ridgeway G, 2010. *Gbm: Generalized Boosted Regression Models. R Package Version 1.6–3.1*. Available at: <http://CRAN.R-project.org/package=gbm>. Accessed May 10, 2020.
48. Hay SI, Smith DL, Snow RW, 2008. Measuring malaria endemicity from intense to interrupted transmission. *Lancet Infect Dis* 8: 369–378.
49. Snow RW et al., 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *Lancet* 349: 1650–1654.
50. Macharia PM, Giorgi E, Noor AM, Waqo E, Kiptui R, Okiro EA, Snow RW, 2018. Spatio-temporal analysis of *Plasmodium falciparum* prevalence to understand the past and chart the future of malaria control in Kenya. *Malar J* 17: 340.
51. Kapesa A, Kweka EJ, Atieli H, Afrane YA, Kamugisha E, Lee MC, Zhou G, Githeko AK, Yan G, 2018. The current malaria morbidity and mortality in different transmission settings in Western Kenya. *PLoS One* 13: e0202031.
52. Hansen B, Schjønning P, Sibbesen E, 1999. Roughness indices for estimation of depression storage capacity of tilled soil surfaces. *Soil Tillage Res* 52: 103–111.
53. Zhao L, Hou R, Wu F, Keesstra S, 2018. Effect of soil surface roughness on infiltration water, ponding and runoff on tilled soils under rainfall simulation experiments. *Soil Tillage Res* 179: 47–53.
54. Amadi JA, Olago DO, Ong'amo GO, Oriaso SO, Nanyingi M, Nyamongo IK, Estambale BBA, 2018. Sensitivity of vegetation to climate variability and its implications for malaria risk in Baringo, Kenya. *PLoS One* 13: e0199357.
55. Rueda LM, Brown TL, Kim HC, Chong ST, Klein TA, Foley DH, Anyamba A, Smith M, Pak EP, Wilkerson RC, 2010. Species composition, larval habitats, seasonal occurrence and distribution of potential malaria vectors and associated species of *Anopheles* (Diptera: Culicidae) from the Republic of Korea. *Malar J* 9: 55.
56. Zhou G, Li JS, Ototo EN, Atieli HE, Githeko AK, Yan G, 2014. Evaluation of universal coverage of insecticide-treated nets in western Kenya: field surveys. *Malar J* 13: 351.
57. Atieli HE, Zhou G, Afrane Y, Lee MC, Mwanzo I, Githeko AK, Yan G, 2011. Insecticide-treated net (ITN) ownership, usage, and malaria transmission in the highlands of western Kenya. *Parasit Vectors* 4: 113.
58. Hu MQ, Mao F, Sun H, Hou YY, 2011. Study of normalized difference vegetation index variation and its correlation with climate factors in the three-river-source region. *Int J Appl Earth Obs Geoinfo* 13: 24–33.
59. Schultz PA, Halpert MS, 1993. Global correlation of temperature, NDVI and precipitation. *Adv Space Res* 13: 277–280.
60. Chicco D, 2017. Ten quick tips for machine learning in computational biology. *BioData Mining* 10: 35.
61. Freedman DA, 2009. *Statistical Models: Theory and Practice*. Cambridge, MA: Cambridge University Press.
62. Tibshirani R, 1996. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B* 58: 267–288.
63. Warne RT, 2011. Beyond multiple regression: using commonality analysis to better understand R^2 results. *Gifted Child Quart* 55: 313–318.
64. Zhou G, Munga S, Minakawa N, Githeko AK, Yan G, 2007. Spatial relationship between adult malaria vector abundance and environmental factors in western Kenya highlands. *Am J Trop Med Hyg* 77: 29–35.
65. Kahindi SC, Muriu S, Derua YA, Wang X, Zhou G, Lee MC, Mwangangi J, Atieli H, Githeko AK, Yan G, 2018. Efficacy and persistence of long-lasting microbial larvicides against malaria vectors in western Kenya highlands. *Parasit Vectors* 11: 438.