Malaria micro-stratification using routine surveillance data in Western Kenya

Victor A Alegana[1,2,3*], Laurissa Suiyanka[1], Peter M Macharia[1], Grace Ikahu-Muchangi[4], Robert W Snow[1,5]

Author affiliations

1. Population Health Unit, Kenya Medical Research Institute - Wellcome Trust Research Programme, P.O. Box 43640-00100, Nairobi, Kenya
2. Geography and Environmental Science, University of Southampton, SO17 1BJ Southampton, UK
3. Faculty of Science and Technology, Lancaster University, Lancaster, LAI 4YW, UK.
4. National Malaria Control Program, Ministry of Health, P.O Box 30016-00100, Nairobi, Kenya
5. Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, OX3 7LJ, Oxford, UK

Corresponding author: Victor A Alegana

email: valegana@kemri-wellcome.org

Abstract

Background: There is an increasing need for finer spatial resolution data on malaria risk to provide micro-stratification to guide sub-national strategic plans. Here, spatial-statistical techniques are used to exploit routine data to depict sub-national heterogeneities in test positivity rate (TPR) for malaria among patients attending health facilities in Kenya.

Methods: Routine data from health facilities (*n*=1804) representing all ages over 24 months (2018-2019) was assembled across 8 countries (62 sub-counties) in Western Kenya. Statistical model-based approaches were used to quantify heterogeneities in TPR and uncertainty at fine spatial resolution adjusting for missingness, population distribution, spatial data structure, month, and type of the health facility.

Results: The overall monthly reporting rate was 78.7% (IQR 75.0 – 100.0) and public-based health facilities were more likely than private facilities to report ≥ 12 months (OR 5.7, 95% CI 4.3 – 7.5). There was marked heterogeneity in population-weighted TPR with sub-counties in the north of the lake-endemic region exhibiting the highest rates (exceedance probability >70% with 90% certainty) where approximately 2.7 million (28.5%) people reside. At micro-level the lowest rates were in 14 sub-counties (exceedance probability <30% with 90% certainty) where approximately 2.2 million (23.1%) people lived and indoor residual spraying had been conducted since 2017.

Conclusion: The value of routine health data on TPR can be enhanced when adjusting for underlying population and spatial structures of the data, highlighting small scale heterogeneities in malaria risk often masked in broad national stratifications. Future research should aim at relating these heterogeneities in TPR with traditional community-level prevalence to improve tailoring malaria control activities at sub-national levels.

Keywords: Malaria, Routine Data, Test Positivity Rate

Background

The highest public health burden posed by infection with *Plasmodium falciparum* continues to be borne by countries in sub-Saharan Africa (SSA) [1]. Infection prevalence and disease risks remain unevenly distributed between and within countries [2, 3]. This spatial heterogeneity requires strategies that facilitate targeting of limited resources for malaria control as outlined in the Global Technical Strategy (GTS) for malaria [4] and the High Burden-High Impact (HBHI) initiative [5]. Current national malaria strategic plans in SSA use a variety of metrics to depict sub-national variations in malaria risk ranging from modelled community-based parasite prevalence to crude estimates of clinical incidence from routine data [6]. The main challenge for National Malaria Control Programs (NMCPs) is to maximise all available malaria survey and routine data to provide increasingly finer spatial resolution robust data on malaria risk to provide micro-stratification to guide sub-national strategic plans.

Malaria routine data from District Health Information System 2 (DHIS2) summarised as test positivity rate (TPR) among patients attending health facilities is a simple metric, providing a means for micro-stratification and targeted responses [7-13]. Compared to cross-sectional community-based surveys of infection prevalence, TPR is more ubiquitous in time and space because data is collected continuously and across all treatment facilities in a locality.

Traditionally, NMCPs define TPR as a ratio of aggregated number of confirmed cases over parasitological tests undertaken within a single administrative unit. Such an approach does not adjust for a) the spatial and temporal heterogeneities in the data at a more granular scale; b) the populations who would use health facilities at the borders of administrative

units; or c) missingness of the reported data by health facility. Importantly, data use rarely considers uncertainty thresholds which are important metrics for decision making when choosing between malaria strategies [12, 14-18].

Here, the aim was to provide an example of quantifying the spatial heterogeneities in TPR using a Bayesian Model-Based framework [19, 20] adjusting for data missingness, Spatio-temporal dependencies and population density at fine-scale to guide malaria micro-stratification in Western Kenya.

Methods

Study setting

The present study used routine health facility data from eight counties in Western Kenya: Bungoma, Busia, Homa Bay, Kakamega, Kisumu, Migori, Siaya, and Vihiga. These counties represent devolved administrative units responsible for making sub-national decisions on the provision of health care, including malaria, and are administratively subdivided into 62 sub-counties (Figure 1). The NMCP provides overall national malaria policies, strategic direction and coordinates bi-lateral and multi-lateral support for national malaria control while counties are expected to adapt national policies to their local epidemiological context [21].

The eight Lake-endemic counties (Figure 1) and cover 19.4% (9.4 million people) of Kenya's population [22]. The area experiences two rainy seasons, March to May and October to December. Malaria transmission is intense throughout the year with community-based *Plasmodium falciparum* prevalence among children exceeding 30% in 2009 [23], and continue to experience the highest rates of malaria transmission in Kenya in 2015 [18].

Transmission is maintained by high biting rates from local vector populations including *An. funestus* s.s., *An. arabiensis* and *An. gambiae* s.s. [24, 25].

Between 2016 and 2019, 1.1 million long-lasting insecticide-treated nets were distributed routinely (antenatal clinic clients) across the 8 counties. Since 2017, there have been three rounds of indoor residual spraying (IRS) in Homa Bay and Migori counties using Actellic 300 CS and SumiShield 50 WG [26]. In September 2019, 23 sub-counties in Western Kenya were randomly allocated to receive the Food and Drug Administration (FDA) approved RTS,S/AS01 (RTS,S) vaccine and form part of an ongoing evaluation of safety and effectiveness [27].

Routine malaria data from DHIS2

An aggregate of monthly outpatient malaria cases representing all presentations among all ages to public and private health facilities was obtained from the DHIS2. Data were assembled for 24 months from January 2018 to December 2019. DHIS2 is the electronic routine health data platform for reporting, analysing, and dissemination of data for health programmes, piloted in 2010 and rolled out national-wide in Kenya in 2011 [28, 29]. Health facilities comprised of level 4 or level 5 (hospitals), level 3 (health centres) and levels 2 and 1 (primary care facilities, or dispensaries) [30].

Recent evidence shows that over 90% of suspected malaria cases are subjected to a malaria parasitological test in Western Kenya [31]. Malaria Rapid Diagnostic Tests (mRDTs) were introduced to scale-up fever testing of all age groups in 2012 in Kenya [32]. The focus of the present analysis was on the monthly number of patients suspected for malaria (the denominator) and the number of cases of positive mRDT or blood slide-confirmed malaria

cases (the numerator) excluding follow-up visits and referrals resulting in a TPR. The definition of TPR is not the strict definition of fever test positivity rates used as a historical metric of malaria risk, that aimed to test all fevers [33, 34], but a suspected malaria TPR, based on service provider perceptions of probable malaria.

Population data

Fine spatial resolution, 1 km gridded population data for Western Kenya was derived using the 2019 national census data available at sub-county levels [22], and distributions of populations at Enumeration Area (EA) levels used during the 2009 census as input data. Standardised dasymetric distributions were used to allocate population density weights using a Random Forest (RF) model [35]. The modelled EA population distribution was projected to 2019 using 2009-2019 inter-censual growth rates and matched to 2019 sub-county census population estimates. Population adjustments were modelled based on land cover using the random forest model to provide a continuous 1km gridded estimate of population in 2019 map (Supplementary Information 1).

Data pre-processing and geo-referencing

DHIS2 data completeness was checked based on the number of facility monthly reports recorded out of the expected number of facility-month reports from the universe of public and private facilities in the 62 sub-counties across 8 counties available from the master health facility list of operational facilities, that had been geocoded to provide spatial locations, described elsewhere [23, 36]. The missing values in the DHIS2 had been recorded as NAs distinguished from a reported zero.

A hierarchical Space-time geostatistical analysis of TPR

The geographic coordinates of the health facility combined with data indexed in time (month) allowed the prediction of TPR using a hierarchical Bayesian space-time modelling context adjusting for three broad levels of service provision (hospitals, health centres and dispensaries or clinics). The interest was to define the underlying spatial-temporal process of TPR. Since a universe of all facilities was available and geocoded, the space-time analysis aimed at predicting TPR in space at 1km by 1km pixels to match population distribution. This scale was used, rather than county or sub-county, to allow for the fact that facilities are located on administrative boundaries serving more than one administrative population and assumes that if a facility was located at each grid it would have TPR properties to those most proximal and temporal (month) to existing, reporting facilities. Fine-scale (1km x 1km) TPR predictions were then aggregated as the average, population-weighted area estimates at the sub-county level.

To predict gridded estimates of TPR, the methodology exploits the spatial and temporal autocorrelation in outpatient case counts to predict the missing or unsampled values as weighted linear combinations of the data points close in space-time. Thus, using the health facility spatial location $s(s=1,....,n)$, the corresponding number of people visiting the health facility suspected with malaria $N(s,t)$, month (time) $t(t=1,....,T)$ and the number confirmed malaria cases $y(s,t)$, the modelling framework translates the discretised observations to a prediction of TPR. The important aspect of hierarchical Bayesian formulation is linking the observational data model to latent processes (the Spatio-temporal process and the parameters). A binomial likelihood was used (data likelihood), combined with prior information containing uncertainty in the data generating process resulting in a posterior probability distribution. The data likelihood function for observational data given the linear predictor $\eta(s,t)$was defined as:

$$y(s,t) \mid \eta(s,t) \sim Binomial(N(s,t), P(s,t))$$

where $\eta(s,t) = \log it(P(s,t))$. The spatiotemporal process, defined on the linear predictor as:

$$\eta(s,t) = \beta_0 + w(s,t) + \alpha_s(t) + e(s,t),$$

where β_0 is an intercept and for a generic location s, the process $w(s,t)$ is a mean-zero Spatio-temporal process and $e(s,t)$ are $i.i.d\ N(0,\sigma_e^2)$ and independent of other processes. The error $e(s,t)$ is the residual adjustment to the Spatio-temporal explanation. With $t = 1,2,....T$, $\alpha(t)$ represent monthly variables adjusting for seasonality specified using first-order random walk.

Modelling was implemented using the Integrated Nested Laplace Approximation (INLA) R-statistical package [37]. R-INLA uses both analytical approximation and numerical integration to perform approximate Bayesian inference for the class of latent Gaussian models, such as the spatio-temporal models [38]. The geostatistical implementation in R-INLA was implemented via the space-time stochastic partial differential equation (SPDE) approach [39]. The Bayesian specification was completed by assigning prior distribution for parameters of the random walk using the penalized complexity prior [40], SPDE, and fixed effect (flat priors) (Supplementary Information2).

Micro-stratification using exceedance probability

Micro-stratification within counties is a priority for the county ministries of health, to set priorities for malaria control investment. However, a degree of certainty is necessary to set priorities [16, 17]. As such, exceedance probabilistic methodology [14, 15] was used on the fitted population-weighted model for TPR. This probabilistic estimate identified locations where $p_c(s,t) = P\{\eta(s,t) > l\}$ with l as the threshold level of interest. A threshold of >70%

population-weighted TPR represented high burden sub-counties, while <30% represented sub-counties with low malaria risk. In previous studied 30% TPR was associated with low malaria prevalence estimated from community survey data [41, 42]. Thus, areas where $p_c(s,t)$ was closer in value to 100%, indicated the likelihood of location to be above the threshold l. Conversely, when $p_c(s,t)$ value was close to 0% indicated an increased likelihood of being below the threshold. For $p_c(s,t)$ equal to 50% correspond to sub-counties with the highest uncertainty, with an equal probability below or above the threshold l.

Model validation procedures

Cross-validation techniques were used to evaluate the predictive performance of the model. This was based on a 20% subset of data selected randomly and used in the computation of prediction error metrics namely: the mean absolute error, the mean prediction error (MPE), mean absolute error (MAE), the root mean square error (RMSE), and a Pearson's product-moment correlation coefficient that quantified the association between observed and predicted values.

Results

Data coverage and reporting rate

Figure 2 provides a summary of assembled data by the type of health facility among the expected 1,804, including 150 hospitals, 309 health centres, and 1,345 dispensaries and clinics. Only 160 health facilities (8.9%) did not report any data for the 24 months, with 147 being the lowest level of facility (dispensaries or clinics). The overall monthly reporting rate for the data period was 78.7% (IQR 75.0 – 100.0); 1,339 (74.2%) facilities reported data for 18/24 or more months, 264 (14.6%) reported for at least 12 months, and 41 (2.3%) facilities

reported data to the DHIS2 for 6 months or less. Analysis suggested that public-based health facilities were more likely than private facilities to report ≥ 12 months (OR 5.7, 95% CI 4.3 – 7.5), and dispensaries or clinics had lower odds of reporting ≥ 12 months (OR 0.4, 0.3 – 0.7). Lastly, there was no difference in TPR by age in the DHIS2 (for under-5 years 47.4% 95% CI 45.9 – 48.9 compared to the over-5 years 47.8% 95% CI 46.2-49.2, respectively). Therefore, all subsequent analysis of TPR was aggregated for all ages.

A total of 6.0 million outpatient malaria cases were confirmed at health facilities among 12.8 million suspected malaria out-patient cases over 24 months. The 24-month mean for hospitals was 688 confirmed cases compared to 309 cases at the primary-level facilities over the 24 months. The number of confirmed cases varied by month ranging from 0.2 million confirmed cases among 0.6 million suspected cases in December and highest in July, 0.8 million confirmed among 1.7 million suspected.

TPR Model sensitivity analysis

Model validation was assessed using the MAE as well as an assessment of prediction performance based on the 20% validation sample. The MPE summarising the difference between predicted and observed values was 0.01 while the MAE was 0.12 and the RMSE 0.21 (Table 2). Pearson's correlation between observed and predicted values was 0.64 ($p<0.0001$). The analysis of residuals showed minimum spatial autocorrelation after modelling as depicted in the semi-variogram of the residuals. The residual variogram was within the 95% interval suggesting that the spatial structure in the data was accounted for by the space-time modelling (Supplementary information 2). The model spatial range was approximately 20.7Km (95% Bayesian Credible Interval in Km 16.4 – 25.7). Supplementary Information 2 (Table S2) lists the posterior summaries of the parameters of space-time

modelling representing the fixed effects and the temporal and spatial parameters. From these parameters, adjusting for the year of data point and facility type (level) was important in the model estimation at the 95% Bayesian credible interval.

Spatial heterogeneity in TPR at sub-county level

Figure 3A shows the crude aggregated TPR, compared to Figure 3B which shows the modelled population-weighted TPR for each of the 62 sub-counties. Both crude and adjusted sub-county TPRs highlight the marked heterogeneity across the region. The crude estimates, however, do not adjust for population density or missingness of the uncertainty in data. Figure 4 is a comparison of crude TPR to the modelled population-weighted estimates. There were differences between crude and modelled estimates particularly in the sub-counties in the north of the lake e.g. Bungoma county. Higher modelled population-weighted TPR areas were also located in these northern sub-counties. At a county level, the highest mean predicted TPR was in Busia county mean 70.6% (95% Bayesian credible interval 68.1% - 72.8); 6 sub-counties in Busia (Bunyala, Butula, Samia, Matayos and Teso South) and Kakamega (Butere) had TRP greater than 70% (Supplementary Information 3). Homa Bay county had the lowest population-weighted TPR 33.2% (30.4% - 36.0). Rachuonyo East sub-county in Homa Bay was the lowest 23.6% (22.1% - 25.1%).

The population at risk and micro-stratification at sub-county level

Of the 9.4 million residents of Western Kenya in 2019, 2.7 million (28.5%) lived in 19 sub-counties where the probability of TPR exceeded 70% at 10% chance of a Type I error occurring (Figure 5). These were predominantly in the north of Lake Victoria and for the two sub-counties in Migori. 3.1 million (32.6%) lived in areas where TPR was likely to be ≥40% and <70% (19 sub-counties), and 1.5 million (15.8%) lived in 10 sub-counties where TPR

was ≥ 30% but less than 40%. Finally, approximately 2.2 million (23.1%) lived in 14 sub-counties where TPR was likely to be <30% at 90% certainty (Supplementary Information 3) corresponding to low-risk sub-counties where IRS was recently implemented.

Discussion

Routine data for micro-stratification in stable, malaria endemic settings should increasingly form the basis for tailoring malaria control and monitoring the impact of intervention [6, 8]. Here a geostatistical approach was applied to routine data from eight counties of Western Kenya to explore heterogeneities in TPR to inform a micro-stratification at the sub-county level (n=62). These outputs have immediate potential to enhance the capacity of decision-makers for malaria control within the devolved national structure. The western Kenya region has high coverage of health facilities congruent to population density and with good reporting rates (79%) of malaria out-patient data to the national surveillance system (DHIS2). However, simple aggregation of TPR data (Figure 3A) does not account for the underlying spatio-temporal structure of the data nor the underlying heterogeneous population distributions within each sub-county (Figure 3B; Figure 4).

There was marked heterogeneity in TPR with sub-counties in the north of the lake exhibiting the highest TPR (exceedance probability >70% with 90% certainty) where approximately 2.7 million (28.5%) people reside. The regions with the highest malaria burden would require concerted effort to increase vector control and other interventions to reduce transmission and consequent morbidity. Evaluating the probability of TPR exceeding a certain threshold promotes a policy-relevant dialogue on uncertainties related to estimates. For example, if one was to include chemo-prophylactic initiatives [43] to accelerate a reduction in transmission, the targeting of these interventions and the added costs would require some level of certainty.

The Bayesian credible intervals presented for TPR account for the 21% missing data and account for the uncertainty introduced by the need to predict these missing data in time. In Kenya, national stratification was based on a threshold of parasite prevalence without further consideration of related uncertainty [23].

Carefully assembled TPR results also serve as a means to track the impact of malaria interventions. For example, the effect of IRS in the two counties of Homa Bay and Migori where 3 rounds of IRS had been implemented since 2017 all showed considerably lower predicted TPR than neighbouring sub-counties. Importantly, the two sub-counties in Migori (Kuria West and Kuria East) where IRS was not implemented had much higher population-weighted, adjusted TPR values between 2018-19, 60.8% (59.0% - 62.6%) and 54.5% (52.7% - 56.3%), respectively. Future applications of these routine data might include the possible impact of the pilot RTS,S vaccine programme. However, it is notable that aggregated DHIS2 do not currently allow for finer age-structured data beyond under and above five years of age which would limit a closer understanding of vaccine effectiveness among children <2 years.

Combining metrics from routine data (DHIS2) with community parasite prevalence could potentially improve estimates of disease burden at the population level [44-46]. The underlying assumption is that the spatiotemporal correlation in TPR is usually driven by the underlying PR spatio-temporal structure. This, however, requires further investigation over large regions and possible interaction with malaria co-infections. However, the immediate application of such a hybrid approach is dependent upon a better understanding of the relationship between TPR to PR at varying endemicity, which is not always linear [41, 42].

The analysis presented here was limited to a two-year time-series data and could potentially be improved by the inclusion of longer space-time data sets to extract long term trends [47, 48]. However, the data before 2018 were influenced by nationwide medical staff strikes [49]. There could be biases introduced due to the type of diagnosis at the facility level by using mRDT or microscopy with varying sensitivities [50, 51]. RDTs are the most common diagnostic tools at the lower tier-facilities without a laboratory technician. However, no information was recorded on the type of RDT used. For microscopy, information on the quality of slide and reading was unavailable. We have not accounted for the quality of diagnosis at the facility levels or the differences in fever testing rates, only possible through direct observational audits [31, 52, 53]. Finally, the quality of DHIS2 documentation is known to vary [54], and the reliability of individual records cannot be quantified without substantial health facility audits.

Conclusion

Adjusting for population distributions, data missingness and building in statistical uncertainty can improve the value of routine data for malaria micro-stratification. These approaches can identify impacts of local-scale vector control and allow sub-national county ministries of health to tailor existing national recommendations for control. Future research should aim at relating these heterogeneities in TPR with traditional community-level prevalence to improve micro-stratification or and at granular and specific levels to improve our ability to track the impact of vaccination interventions targeted for young children below 2 years.

Abbreviations

DHIS2: District Health Information System 2; EA: Enumeration Area; EP: Exceedance Probability; GTS: Global Technical Strategy; HB-HI: High Burden-High Impact;

INLA: Integrated Nested Laplace Approximation; IRS: Indoor Residual Spraying; LLIN: Long-Lasting Insecticide Treated Nets; MOH: Ministry of Health; MPE: Mean Prediction Error; NMCP: National Malaria Control Programs; RDT: Rapid Diagnostic Tests; RF: Random Forest; RMSE: Root Mean Square Error; SPDE: Stochastic Partial Differential Equation; SSA: Sub-Saharan Africa; TPR: Test Positivity rate; WHO: World Health Organization

Acknowledgements

The authors are grateful to James Kiare, Jacinta Omariba, Beatrice Kemunto and other NMCP personnel for reviewing the early analysis of DHIS2 data for western Kenya. We are also grateful to Abdisalan Mohammed Noor and Emelda Okiro for comments on the draft of the manuscript.

Authors contributions

VAA and RWS conceived the study; VAA analysed the data. LS provided help in spatial reconciliation of the DHIS2 data. PMM and GIM provided assistance in the interpretation of the data. VAA and RWS drafted the first version of the manuscript. All authors reviewed the final manuscript.

Funding

VAA is funded as a Wellcome Trust Training Fellow (number 211208) that also provided support to LS. PMM is supported by the DELTAS Africa Initiative (DEL-15-003). The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa and supported by the New Partnership for Africa's Development Planning and Coordinating Agency with

funding from the Wellcome Trust (number 107769) and the UK government. RWS is supported as a Wellcome Trust Principal Fellow (numbers 103602 and 212176) that also provided support for PMM. VAA, LS, PMM, and RWS are grateful to the support of the Wellcome Trust to the Kenya Major Overseas Programme (number 203077).

Availability of data and materials

Aggregated DHIS2 data is available online with access provided by Ministry of Health https://hiskenya.org/dhis-web-commons /security/login.action. The datasets used and/or analysed during the current study are also available from the corresponding author on reasonable request.

Ethics approval and consent to participate

None required.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

References

1. World Health Organization: **World Malaria Report 2019.** Geneva2019. [Available: https://www.who.int/publications-detail/world-malaria-report-2019]. Accessed: *February, 2016*

2. Snow RW, Sartorius B, Kyalo D, Maina J, Amratia P, Mundia CW, Bejon P, Noor AM: **The prevalence of Plasmodium falciparum in sub-Saharan Africa since 1900.** *Nature* 2017, **550:**515.

3. Weiss DJ, Lucas TCD, Nguyen M, Nandi AK, Bisanzio D, Battle KE, Cameron E, Twohig KA, Pfeffer DA, Rozier JA, et al: **Mapping the global prevalence, incidence, and mortality of *Plasmodium falciparum*, 2000-17: a spatial and temporal modelling study.** *Lancet* 2019, **394:**322-331.

4. WHO: **Global technical strategy for malaria 2016-2030.** Geneva2015. [Available: https://www.who.int/malaria/publications/atoz/9789241564991/en/]. Accessed: *June, 2016*

5. WHO, RBM: **High burden to high impact: a targeted malaria response.** Geneva: Organization WH; 2018. [Available: https://www.who.int/malaria/publications/atoz/high-impact-response/en/]. Accessed: *June 2019*

6. Alegana VA, Okiro EA, Snow RW: **Routine data for malaria morbidity estimation in Africa: challenges and prospects.** *BMC Medicine* 2020, **18:**121.

7. Afrane YA, Zhou G, Githeko AK, Yan G: **Utility of health facility-based malaria data for malaria surveillance.** *PLOS ONE* 2013, **8:**e54305.

8. Ashton RA, Bennett A, Yukich J, Bhattarai A, Keating J, Eisele TP: **Methodological considerations for use of routine health information system data to evaluate malaria**

program impact in an era of declining malaria transmission. *American journal of tropical medicine and hygiene* 2017, **97:**46-57.

9. Githinji S, Noor AM, Malinga J, Macharia PM, Kiptui R, Omar A, Njagi K, Waqo E, Snow RW: **A national health facility survey of malaria infection among febrile patients in Kenya, 2014.** *Malaria Journal* 2016, **15:**591.

10. Oduro AR, Bojang KA, Conway DJ, Corrah T, Greenwood BM, Schellenberg D: **Health centre surveys as a potential tool for monitoring malaria epidemiology by area and over time.** *PLoS One* 2011, **6:**e26305.

11. Oduro AR, Maya ET, Akazili J, Baiden F, Koram K, Bojang K: **Monitoring malaria using health facility based surveys: challenges and limitations.** *BMC Public Health* 2016, **16:**1-9.

12. Plucinski MM, Candrinho B, Dimene M, Smith T, Thwing J, Colborn J, Rogier E, Zulliger R: **Estimation of Malaria-Attributable Fever in Malaria Test-Positive Febrile Outpatients in Three Provinces of Mozambique, 2018.** *Am J Trop Med Hyg* 2020, **102:**151-155.

13. Thawer SG, Chacky F, Runge M, Reaves E, Mandike R, Lazaro S, Mkude S, Rumisha SF, Kumalija C, Lengeler C, et al: **Sub-national stratification of malaria risk in mainland Tanzania: a simplified assembly of survey and routine data.** *Malaria Journal* 2020, **19:**177.

14. Diggle PJ, Giorgi E: **Model-based geostatistics for prevalence mapping in low-resource settings.** *Journal of the American Statistical Association* 2016, **111:**1096-1120.

15. Giorgi E, Osman AA, Hassan AH, Ali AA, Ibrahim F, Amran JGH, Noor AM, Snow RW: **Using non-exceedance probabilities of policy-relevant malaria prevalence thresholds to identify areas of low transmission in Somalia.** *Malaria Journal* 2018, **17:**88.

16. Gething PW, Patil AP, Hay SI: **Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation.** *PLoS Comput Biol* 2010, **6:**e1000724.

17. Sullivan D: **Uncertainty in mapping malaria epidemiology: implications for control.** *Epidemiologic Reviews* 2010, **32:**175-187.

18. Macharia PM, Giorgi E, Noor AM, Waqo E, Kiptui R, Okiro EA, Snow RW: **Spatio-temporal analysis of Plasmodium falciparum prevalence to understand the past and chart the future of malaria control in Kenya.** *Malaria Journal* 2018, **17:**340.

19. Diggle PJ, Tawn JA, Moyeed RA: **Model-based geostatistics.** *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 1998, **47:**299-350.

20. Banerjee S, Carling PB, Gelfand AE: *Hierarchical modeling and analysis for spatial data.* London: Chapman & Hall/CRC; 2004.

21. NMCP K: **Kenya national malaria strategy 2019-2023.** Nairobi2019. [Available: http://www.nmcp.or.ke]. Accessed: *January, 2020*

22. KNBS: **2019 Kenya population and housing census volume I: population by county and sub-county.** Nairobi, Kenya2019. [Available: https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county]. Accessed: *November, 2019*

23. Noor AM, Gething PW, Alegana VA, Patil AP, Hay SI, Muchiri E, Juma E, Snow RW: **The risks of malaria infection in Kenya in 2009.** *BMC Infect Dis* 2009, **9:**180.

24. Ndenga B, Githeko A, Omukunda E, Munyekenye G, Atieli H, Wamai P, Mbogo C, Minakawa N, Zhou G, Yan G: **Population dynamics of malaria vectors in western Kenya highlands.** *Journal of Medical Entomology* 2014, **43:**200-206.

25. Okara R, Sinka M, Minakawa N, Mbogo C, Hay S, Snow R: **Distribution of the main malaria vectors in Kenya.** *Malaria Journal* 2010, **9:**69.

26. Abong'o B, Gimnig JE, Torr SJ, Longman B, Omoke D, Muchoki M, ter Kuile F, Ochomo E, Munga S, Samuels AM, et al: **Impact of indoor residual spraying with pirimiphos-methyl (Actellic 300CS) on entomological indicators of transmission and malaria case burden in Migori County, western Kenya.** *Scientific Reports* 2020, **10:**4518.

27. **Q&A on the malaria vaccine implementation programme (MVIP)** [https://www.who.int/malaria/media/malaria-vaccine-implementation-qa/en/] Accessed: *June, 2020*

28. Karuri J, Waiganjo P, Manya A: **DHIS2: The tool to improve health data demand and use in Kenya.** *Journal of Health Informatics in Developing Countries* 2014, **8**.

29. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, Khajehpour N, Khajeh Z, Mirshekari N, Hasani M, et al: **The District Health Information System (DHIS2): A literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries.** *Health Information Management Journal* 2019, **48:**62-75.

30. MOH K: **Kenya Health Sector Strategic and Investment Plan (KHSSP) 2014-2018.** Nairobi, Kenya2019. [Available: http://www.nmcp.or.ke]. Accessed: *June, 2019*

31. Amboko B, Stepniewska K, Macharia P, Machini B, Bejon P, Snow R, Zurovac D: **Trends in health workers' compliance with outpatient malaria case-management guidelines across malaria epidemiological zones in Kenya, 2010 - 2016.** *Under review* 2020.

32. NMCP K: **Kenya national malaria strategy 2009-2014 (Revised 2014).** Nairobi2014. [Available: http://www.nmcp.or.ke]. Accessed: *July, 2016*

33. Hay SI, Smith DL, Snow RW: **Measuring malaria endemicity from intense to interrupted transmission.** *Lancet Infect Dis* 2008, **8:**369-378.

34. Macdonald G, Goeckel GW: **The malaria parasite rate and interruption of transmission.** *Bulletin of the World Health Organization* 1964, **31:**365-377.

35. Stevens FR, Gaughan AE, Linard C, Tatem AJ: **Disaggregating census data for population mapping using Random Forests with remotely-sensed and ancillary data.** *PLoS One* 2015, **10:**e0107042.

36. Maina J, Ouma PO, Macharia PM, Alegana VA, Mitto B, Fall IS, Noor AM, Snow RW, Okiro EA: **A spatial database of health facilities managed by the public health sector in sub Saharan Africa.** *Scientific Data* 2019, **6:**134.

37. Rue H, Martino S, Chopin N: **Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009, **71:**319-392.

38. Blangiardo M, Cameletti M, Baio G, Rue Hv: **Spatial and spatio-temporal models with R-INLA.** *Spatial and Spatio-temporal Epidemiology* 2013, **4:**33-49.

39. Lindgren F, Rue H, Lindström J: **An explicit link between Gaussian fields and Gaussian markov random fields: the stochastic partial differential equation approach.** *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011, **73:**423-498.

40. Fuglstad G-A, Simpson D, Lindgren F, Rue H: **Constructing priors that penalize the complexity of gaussian random fields.** *Journal of the American Statistical Association* 2018**:**1-8.

41. Boyce RM, Reyes R, Matte M, Ntaro M, Mulogo E, Lin F-C, Siedner MJ: **Practical implications of the non-linear relationship between the test positivity rate and malaria incidence.** *PLoS One* 2016, **11:**e0152410.

42. Kamau A, Mtanje G, Mataza C, Bejon P, Snow R: **The relationship between facility-based fever test positivity rate and community-based parasite prevalence.** *PLoS One* 2020, **In Press**.

43. Alonso PL: **The role of mass drug administration of antimalarials.** *American Society of Tropical Medicine and Hygiene* 2020.

44. Ouma J, Jeffery C, Valadez JJ, Wanyenze RK, Todd J, Levin J: **Combining national survey with facility-based HIV testing data to obtain more accurate estimate of HIV prevalence in districts in Uganda.** *BMC Public Health* 2020, **20:**379.

45. Bi Y, Hu W, Liu H, Xiao Y, Guo Y, Chen S, Zhao L, Tong S: **Can slide positivity rates predict malaria transmission?** *Malaria Journal* 2012, **11:**117.

46. Jensen T, Bukirwa H, Njama-Meya D, Francis D, Kamya M, Rosenthal P, Dorsey G: **Use of the slide positivity rate to estimate changes in malaria incidence in a cohort of Ugandan children.** *Malaria Journal* 2009, **8:**213.

47. Awine T, Malm K, Peprah NY, Silal SP: **Spatio-temporal heterogeneity of malaria morbidity in Ghana: Analysis of routine health facility data.** *PLoS One* 2018, **13:**e0191707.

48. Nguyen M, Howes RE, Lucas TCD, Battle KE, Cameron E, Gibson HS, Rozier J, Keddie S, Collins E, Arambepola R, et al: **Mapping malaria seasonality in Madagascar using health facility data.** *BMC Medicine* 2020, **18:**26.

49. Irimu G, Ogero M, Mbevi G, Kariuki C, Gathara D, Akech S, Barasa E, Tsofa B, English M: **Tackling health professionals' strikes: an essential part of health system strengthening in Kenya.** *BMJ Glob Health* 2018, **3:**e001136.

50. Boyce MR, O'Meara WP: **Use of malaria RDTs in various health contexts across sub-Saharan Africa: a systematic review.** *BMC Public Health* 2017, **17:**470.

51. WHO-FIND: **Malaria rapid diagnostic test performance: results of WHO product testing of malaria RDTs: round 1 (2008).** Geneva2009. [Available: https://www.who.int/malaria/publications/atoz/9789241514965/en/]. Accessed: *January, 2010*

52. Colborn JM, Zulliger R, Da Silva M, Mathe G, Chico AR, Castel-Branco AC, Brito F, Andela M, de Leon GP, Saifodine A, et al: **Quality of malaria data in public health facilities in three provinces of Mozambique.** *PLOS ONE* 2020, **15:**e0231358.

53. Plucinski MM, Candrinho B, Dimene M, Smith T, Thwing J, Colborn J, Rogier E, Zulliger R: **Estimation of malaria-attributable fever in malaria test–positive febrile outpatients in three provinces of Mozambique, 2018.** *American Journal of Tropical Medicine and Hygiene* 2019.

54. Okello G, Molyneux S, Zakayo S, Gerrets R, Jones C: **Producing routine malaria data: an exploration of the micro-practices and processes shaping routine malaria data quality in frontline health facilities in Kenya.** *Malaria Journal* 2019, **18:**420.