Estimating Malaria Attributable Fractions With Changing Transmission Intensity: Bayesian Latent Class Vs Logistic Models

Kennedy Mwai (kmwai@kemri-wellcome.org)
University of the Witwatersrand

Irene Nkumama
Kenya Medical Research Institute-Wellcome Trust Research Programme

Amos Thairu
Kenya Medical Research Institute-Wellcome Trust Research Programme

James Mburu
Kenya Medical Research Institute-Wellcome Trust Research Programme

Dennis Odera
Kenya Medical Research Institute-Wellcome Trust Research Programme

Rinter Kimathi
Kenya Medical Research Institute-Wellcome Trust Research Programme

Lydia Nyamako
Kenya Medical Research Institute-Wellcome Trust Research Programme

James Tuju
Kenya Medical Research Institute-Wellcome Trust Research Programme

Samson Kinyanjui
Kenya Medical Research Institute-Wellcome Trust Research Programme

Eustasius Musenge
University of the Witwatersrand

Faith Osier
Kenya Medical Research Institute-Wellcome Trust Research Programme

Research Article

Keywords: case definition, malaria, attributable fractions, probability

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-916409/v1

Estimating malaria attributable fractions with changing transmission intensity: Bayesian latent class vs logistic models.

Kennedy Mwai [1,2*], Irene Nkumama [2,3], Amos Thairu [2], James Mburu [2], Dennis Odera [2,3], Rinter Kimathi [2], Lydia Nyamako [2], James Tuju [2,4], Samson Kinyanjui [2,4,5], Eustasius Musenge [1], Faith Osier [2,3,4,5]

[1]Epidemiology and Biostatistics Division, School of Public Health, University of the Witwatersrand, Johannesburg, South Africa.

[2]Centre for Geographic Medicine Research (Coast), Kenya Medical Research Institute-Wellcome Trust Research Programme, Kilifi, Kenya

[3]Centre of Infectious Diseases, Parasitology, Heidelberg University Hospital, Heidelberg, Germany

[4]Department of Biotechnology and Biochemistry, Pwani University, Kilifi, Kenya

[5]Centre for Tropical Medicine and Global Health, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, United Kingdom

Abstract:

Background

Asymptomatic carriage of malaria parasites is common in high transmission intensity areas and confounds clinical case definitions for research studies. This is important for investigations that aim to identify immune correlates of protection from clinical malaria. The proportion of fevers attributable to malaria parasites is widely used to define different thresholds of parasite density associated with febrile episodes. We investigated whether varying intensity of malaria transmission had a significant impact on parasite density thresholds. We used the same dataset to explore an alternative statistical approach using the probability of developing fevers as a choice over threshold cut-offs as the former has been reported to increase predictive power.

Methods

Data from children monitored longitudinally between 2005 and 2017 from Junju and Chonyi in Kilifi, Kenya were analysed. We compare the performance of Bayesian-latent class and logistic power models in estimating malaria attributable fractions and probabilities of having fever given a parasite density with changing malaria transmission intensity. Zero-inflated beta regressions were used to assess the impact of using probabilities to evaluate anti-merozoite antibodies as correlates of protection compared with multilevel binary regression.

Results

Malaria transmission intensity declined from over 49% to 5% between 2006 and 2017 respectively. During this period, malaria attributable fraction varied between 27%-59% using logistic regression compared to 10%-36% using the Bayesian latent class approach. Both models estimated similar patterns of fevers attributable to malaria with changing transmission intensities. The former performed well in estimating the probabilities of having fever, while the latter was efficient in determining the parasite density threshold. However, compared to the logistic power model, the Bayesian algorithm yielded lower estimates for both attributable fractions and probabilities of fever. In modelling the association of merozoite antibodies and clinical malaria, both approaches resulted in comparable estimates, but the utilization of probabilities had a better statistical fit.

Conclusions.

Malaria attributable fractions varied with an overall decline in the malaria transmission intensity in this setting but did not significantly impact the outcomes of analyses aimed at identifying immune correlates of protection. These data confirm the statistical advantage of using probabilities over binary data.

Keywords: case definition, malaria, attributable fractions, probability

Background

Asymptomatic carriage of malaria parasites is highly prevalent in areas with high malaria transmission as a result of naturally acquired immunity (1). It is therefore likely that, in such areas, an individual with a non-malarial fever has coincidental parasitemia. Since the likelihood of having fever generally increases with parasite density (2–4) the assumption is that fever in the presence of parasitemia necessarily constitutes a clinical malaria. However, in high transmission settings (5) parasitemia accompanied by fever may not be adequate to define an episode of clinical malaria and may lead to differential misclassification. Besides causing an over-estimation of malaria burden in an area (2,6), the misclassification complicates immunological and clinical trials where clinical malaria cases are an endpoint or one of the outcome variable. As an outcome variable it is particularly important for identifying correlates of protection from clinical episodes to inform vaccine development.

To overcome this problem of misdiagnosis, different studies have based the case definition of malaria on fever with parasite density above a locally defined threshold. Computation of malaria attributable fractions (MAF) or the proportion of fevers due to malaria parasites has been used to define different thresholds for parasitemia (3,4).

The classical method for deriving the attributable fraction is a simple numerator denominator approach (7) which is prone to bias when applied in high malaria transmission areas (6). In high transmission settings, individuals may have parasites and not show clinical signs of malaria. Logistic regression models are typically used to handle

this bias. They determine the risk of the outcome as a continuous function of parasite density (2,3), and have been widely used to obtain attributable fractions against a range of outcomes with parasitemia as the exposure variable (2,3,8–10). Additionally, a Bayesian latent class model of two-component mixture distributions was proposed to improve the estimation of attributable fractions (4). The latent class model was developed to handle the limitation of imprecise or negative attributable fractions occasionally observed in standard logistic regression models (6).

Malaria transmission intensity has been found to strongly influence the attributable fractions. In a study conducted in two areas with different transmission intensities in Kilifi at the coast of Kenya, a MAF of 50.2% derived for Ngerenya, the low transmission site and 47.9% for Chonyi, the high transmission site. In this study, the logistic regression method was applied and derived a parasite density threshold of 2500 parasites/μL of blood as the most appropriate to distinguish malaria attributable fevers from fevers due to other causes in both settings. Following this study, this threshold has been widely applied in the definition of malaria cases in various studies conducted along the Kenyan coast (8,11–14).

Significant reductions in malaria transmission and admissions have been reported over the last decade in endemic countries in Africa (15) and in particular at the Kenyan coast (12,16,17). Based on this observed reduction in transmission and the influence of transmission intensity on the malaria attributable fractions, we conducted the present study to determine the variation of malaria attributable fractions over time.

We also estimated the probability of fever as a function of parasite density and the optimal parasite thresholds using logistic regression (4,18). The estimated probabilities of fever have been used in determining risk of developing clinical episodes in malaria vaccine trials. In these trials, the probabilities estimated from a Bayesian latent class model were proposed as a better approach to compare the placebo and control groups (19,20).

Several articles (21–25) have pointed out problems associated with the categorization of data. These include not only the loss of information on variation and statistical power, but also an increased risk of type I errors and poor predictive performance (22,23,25). In this study, we explore the utilization of probability estimates from Bayesian latent class models as an alternative to dichotomizing individuals using a selected parasite density threshold.

Methods

Study area and population

We used cohort data from Junju and Chonyi sub-counties in Kilifi County, which is part of the Kilifi Health and Demographic Surveillance System (KHDSS) on the coastal region of Kenya, Figure 1 (26). The area has two malaria transmission seasons May-July and November-December. For Junju the data were prospectively collected from participants aged 1 to 15 years old between 2005 and 2017 (inclusive) who were initially recruited into a malaria vaccine trial (FP9 ME-TRAP and MVA-ME TRAP, Bejon et al. 2006).The

Chonyi dataset had 286 children aged between 0-10years collected in October 2000 and was only used for model comparison (28).

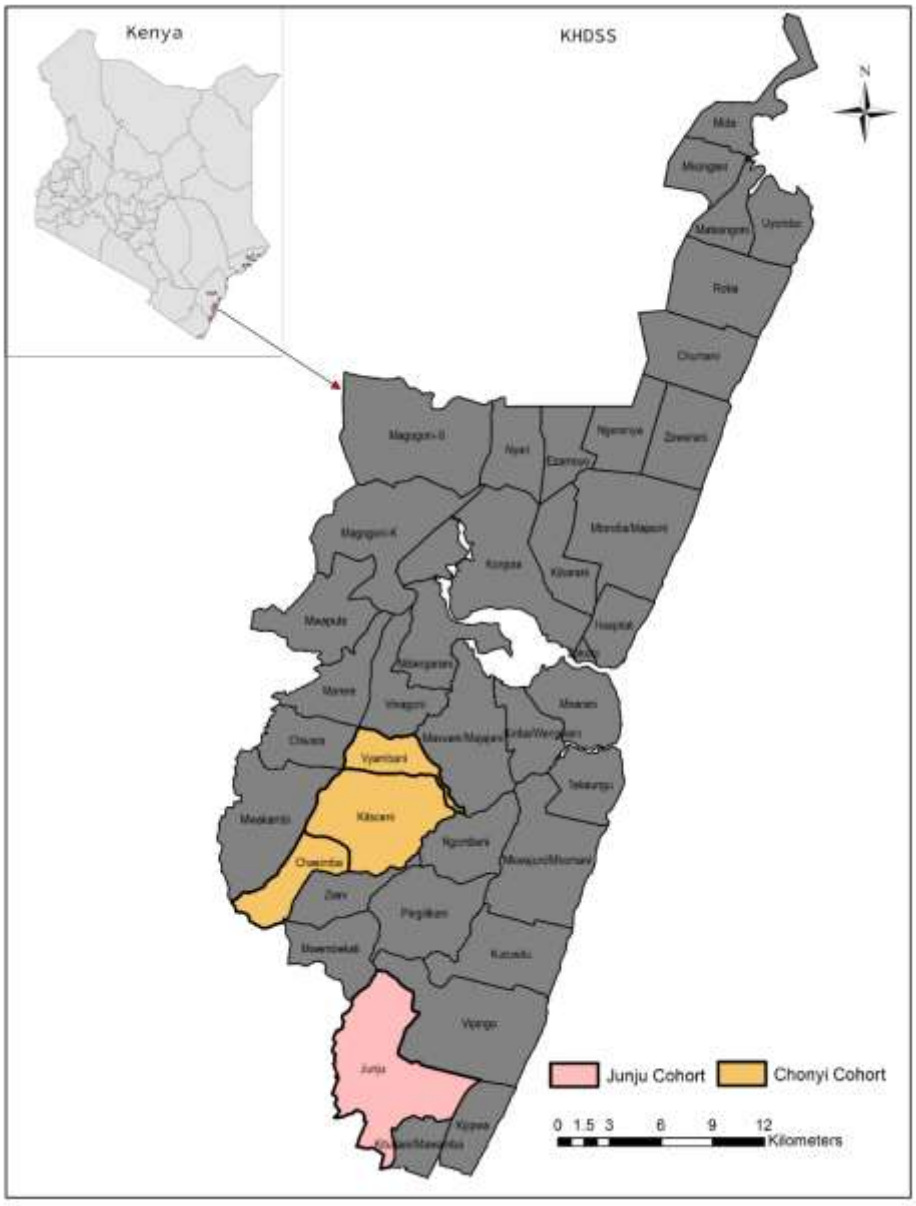

Figure 1: Junju and Chonyi study sites in the Kilifi Health and Demographic Surveillance System (KHDSS)

Malaria parasite prevalence cross-sectional survey

A cross-sectional malaria survey was done every year at the beginning of the malaria season (March-May) for the Junju cohort as shown in Table 1 except for 2005 and 2006 where the surveys were done during the malaria season for a vaccine trial. For the Chonyi cohort, the cross-sectional malaria survey was done in October 2000. Parasitemia was determined by thin smear microscopy. In both studies, the participants were followed up both actively with weekly home visits by trained field workers and passively at health facilities to identify clinical episodes of malaria. Blood smears were prepared to determine parasite densities for any child who had a fever (axillary temp≥37.5°C) for the cross-section surveys and follow-up surveillance, respectively. The Government of Kenya-recommended first-line treatment was used for treatment episodes of malaria.

For the parasitemia determination by microscopy, the number of asexual-stage parasites/200 leukocytes was counted, and parasitemia estimated based on actual or assumed (8,000 leukocytes/μL) leukocyte count measured for each blood smear.

Statistical analysis

A case-control analysis of the relationship between the risk of fever and parasite density was carried out separately for each year using logistic regression and Bayesian latent class models using R (29) and OpenBugs (30) respectively. Additionally, the analysis was repeated in each age group using the overall dataset. Cases corresponded to fever episodes detected by passive or active case detection and controls were those collected

during the cross-sectional surveys or the individuals with parasites and no fever during follow-up.

Logistic regression

A logistic regression model was fit to the data modelling the risk of fever as a continuous function of the parasite density. The model was of the form $\text{logit}(\pi_i) = \alpha + f(x_i)$ where π_i is the probability that observation i with parasite density x_i is a (fever) case. Along with $f(x_i) = \beta x_i^{\tau}$, a smooth monotonic function of x^{τ} where τ is the power transformation of the parasite density. This power function τ was tested at different values between 0.10 and 0.90 with a precision of 0.01 and the value that maximized the log-likelihood best was chosen. The malaria attributable fraction (MAF), λ , was estimated using the slope coefficient of the logistic regression; $\lambda = (1/N)\sum_{1}^{i}(R_i - 1) \Big/ R_i$ where $R_i = \exp[f(x_i)]$ and the standard error estimated using the bootstrap approach with 1000 bootstrap samples. The logistic model was also used to estimate the probability of each individual case of fever being attributable to malaria using $\lambda_{ind} = (R_i - 1)/R_i$ (2).

Bayesian latent class

For the Bayesian latent class model, the parasite density was resolved to a mixture of two multinomial distributions. One component $g_1(.)$ corresponds to non-malaria fever episodes and the other component, $g_2(.)$ to children with clinical malaria episodes (fever and parasites). Parasite levels during the cross-sectional bleed were available and were

used as the training sample, i.e., a sample that comes from the component of the mixture corresponding to children without fever but who may have parasites. The data was then divided into K ordered categories over the range of the parasite density $\mathbf{X}$. This was followed by counting the of test samples $\mathbf{n}=(n_0,n_1,...,n_{k-1})$ and control samples (non-fever cases) $\mathbf{m}=(m_0,m_1,...,m_{k-1})$. Then the MAF, λ, was then estimated from the two multinomial distributions,

$$\begin{aligned}\theta_i &= P(x \in \text{ category i} \mid P_1),\\ \phi_i &= P(x \in \text{ category i} \mid P_2), \quad (1.1)\\ \lambda &= P(x \in P_2)\end{aligned}$$

The parameters P_1 and P_2 are the distributions functions of the components $g_1(.)$ and $g_2(.)$ respectively. The category-specific attributable fractions were obtained using,

$$\lambda_i = P(x \in P_2 \mid x \in \text{category i}) = \frac{\lambda\phi_i}{(1-\lambda)\theta_i + \lambda\phi_i}. \quad (1.2)$$

To estimate the probability, λ_{ind}, of each individual case of fever being attributable to malaria local and piece-wise cubic polynomial models were used. The models were fitted using category-specific MAF, λ_c, together with the category-specific midpoint of parasite density. This was followed by predicting the individual λ_{ind} using their parasite density measurements from the results of various model fitting functions.

Sensitivity and specificity of various cut-off values for parasite density were estimated by $n_c\lambda_c \big/ N\lambda$ and $1-n_c\left(1-\lambda_c\right) \big/ N\left(1-\lambda\right)$ respectively where $n_c = \sum_{i=c}^{K} n_i$, $\lambda_c = \left(\sum_{i=c}^{K} \lambda_i n_i\right) \big/ n_c$, n_i the number of fever cases in the category i and c represents the parasite density category of which it is the selected cut-off in logistic regression or the lower bound for the category in latent class models.

Association with protection.

Multi-level logistic and zero-inflated models were used to investigate the association between high versus low merozoite antibodies and clinical malaria. Various antibody concentrations were applied as cutoffs to define the high and low responders (28). The results were used to compare the performance of probability and binary outcomes. The zero inflated approach was used since we had a probability mass at zero due to the non-febrile participants. Specifically, for the probability outcome, we compared results from the Maximum Likelihood (MLE) and Bayesian inference estimations (28,31,32).

Results

Study population

A total of 4,722 participants were recruited in Junju, Kilifi County from 2005 to 2017. Approximately 300 or more participants were followed up each year with average recruitment age of 6.5 years (ranging between 1 month old to 16 years) as shown in

Table 1. Each child had on average 2.94 test occurrences during follow-up giving rise to a total of 14,404 events during the entire study period.

Temporal distribution of fever and parasite density

Table 1 shows the distribution of fevers (axillary temperature of $\geq 37.5^{o}C$) at all the cross-sectional surveys. Approximately 1,034 (2.19%) occasions of fever were reported during the cross-sectional surveys. A decreasing trend of fevers was observed over the study period except for 2005 and 2006 where the samples were collected specifically for a vaccine trial (33). The prevalence of *P. falciparum* was also artificially high during this period since the participants were recruited during the malaria season. A decline in the prevalence of *P. falciparum* parasite was observed between 2006 to 2013 from 30.21% to 8.78%. This was followed by a slight increase in 2014 and 2015 then another decline in 2016 to 4.32% in 2017.

*Table 1: Prevalence's of Plasmodium falciparum positivity, fever, and presumptive malaria (fever + parasitemia) in the pre-transmission season cross-sectional survey and the number of active follow-up events in the Junju Cohort. ¶ Participants were recruited for the original vaccine study in 2005 and 2006 and sampling of participants extended into the high transmission season each year. *Excluded in the main analysis due to health workers' strike.*

Year	Samples (n)	Mean age in years (min-max)	P. *falciparum* (%)	Fever (%)	Presumptive malaria (%) †	Active follow-up events (n)
2005¶	372	3.9 (1-9.0)	49.05%	0.81%	0.54%	1029
2006¶	300	4.5 (1.5-9.5)	31.21%	0.67%	0.33%	615
2007	339	4.8 (<1-11.0)	15.93%	3.24%	1.18%	1816
2008	341	5.4 (<1-12.0)	29.62%	3.52%	3.52%	884
2009	352	5.8 (<1-13.0)	20.17%	2.56%	1.14%	958
2010	377	6.5 (<1-14.0)	27.59%	3.98%	2.65%	957
2011	377	7.1 (<1-12.7)	23.08%	1.86%	1.33%	891
2012*	399	7.1 (<1-13.7)	16.79%	1.75%	0.50%	696

2013	410	7.5 (<1-14.7)	8.78%	1.46%	0.24%	1483
2014	404	8.2 (<1-15.7)	14.25%	1.49%	0.50%	1619
2015	400	8.3 (<1-16.5)	17.36%	1.25%	0.25%	1319
2016	316	7.8 (<1-15.0)	11.61%	1.27%	-	1087
2017	335	7.1 (<1-15.0)	4.32%	1.19%	-	1050
Total	**4722**	**78.2 (1-198)**	**20.69%**	**2.19%**	**1.06%**	**14404**

Relationship of fever to parasitaemia over time

The probability that a fever case was malaria attributable at a given parasite density λ changed gradually over the study period as shown in Figure 2. The MAF was estimated using the Bayesian Latent class model and logistic regression. The Bayesian latent class gave a lower MAF estimate, Bland-Altman bias =0.20(0.16-0.24), compared to the logistic model. After estimating the sensitivities and specifities of different parasite densities, the optimal parasite cut-off was selected using the logistic regression for the different years [Supplementary Table 1]. However, the number of malaria positive individuals did not vary significantly with the new thresholds compared to the previously defined 2500p/µl threshold [Supplementary Figure 1], despite the changing patterns. Notably, the Bayesian latent class approach and the logistic power models approximated a similar pattern of MAF but the estimates were lower in the former model. Comparable

patterns were also observed in the probabilities, λ_i, predicted from the individual parasite densities [Supplementary Figure 1].

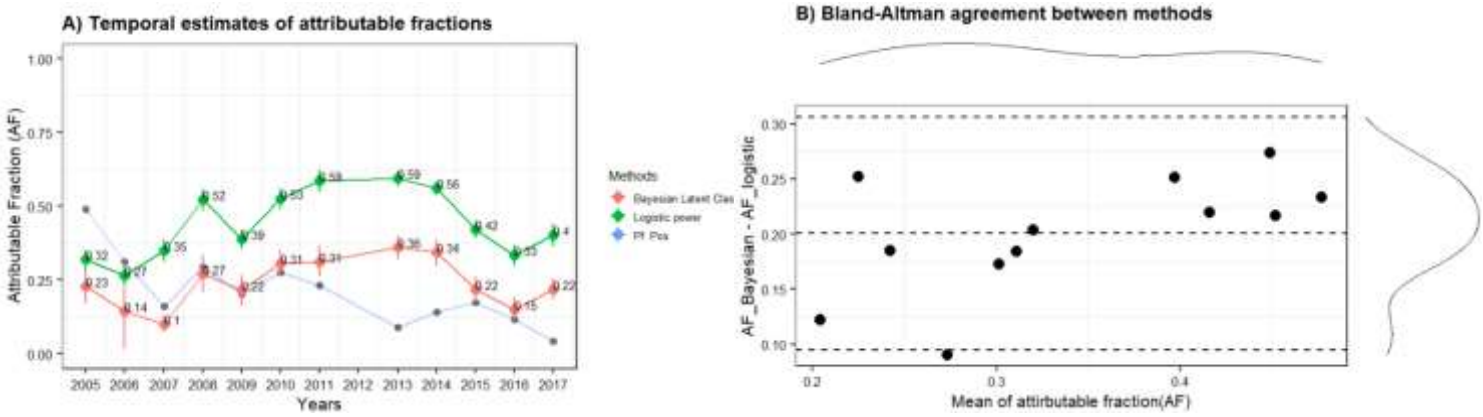

Figure 2: A) Temporal estimates of attributable fraction (AF) from 2005 to 2017 using Bayesian Latent class models and logistic power models. Pf. Pos is the prevalence of parasite positivity during the cross-sectional bleed. B) Bland-Altman plot of agreement

Non-febrile individuals.

An interval estimate for the prevalence of malaria fever was estimated using the Bayesian latent class model. The Bayesian individual probabilities for non-febrile participants with parasitemia were adjusted using the interval estimate for the prevalence of malaria fever. This is shown in Supplementary Figures 2 and 3 where non-febrile cases had a lower, λ_i ,compared to the febrile cases for the parasite positive individuals. Detailed implementation of the methodology is included in the repository as OpenBUGS and R codes.

Impact of age on MAF.

The MAF estimates were higher for older age groups than the children <1 year as shown in Figure 3A below. Additionally, the predicted individual probabilities declined with age

as shown in Figure 3B and likewise the logistic power model had higher estimates and smaller range than the Bayesian latent class predictions [Supplementary Figure 4]. This shows as expected that the age groups of 1-5 years and 5-10 years had a higher probability of having malaria compared to the other age groups. The younger age groups had a higher specificity and sensitivity intersection (Figure 3C) indicating a lower parasite density threshold for clinical episodes compared to the older age groups.

To assess whether the variation in case definitions had a significant impact on the estimates of correlates of protection. The performance of the predicted probabilities from

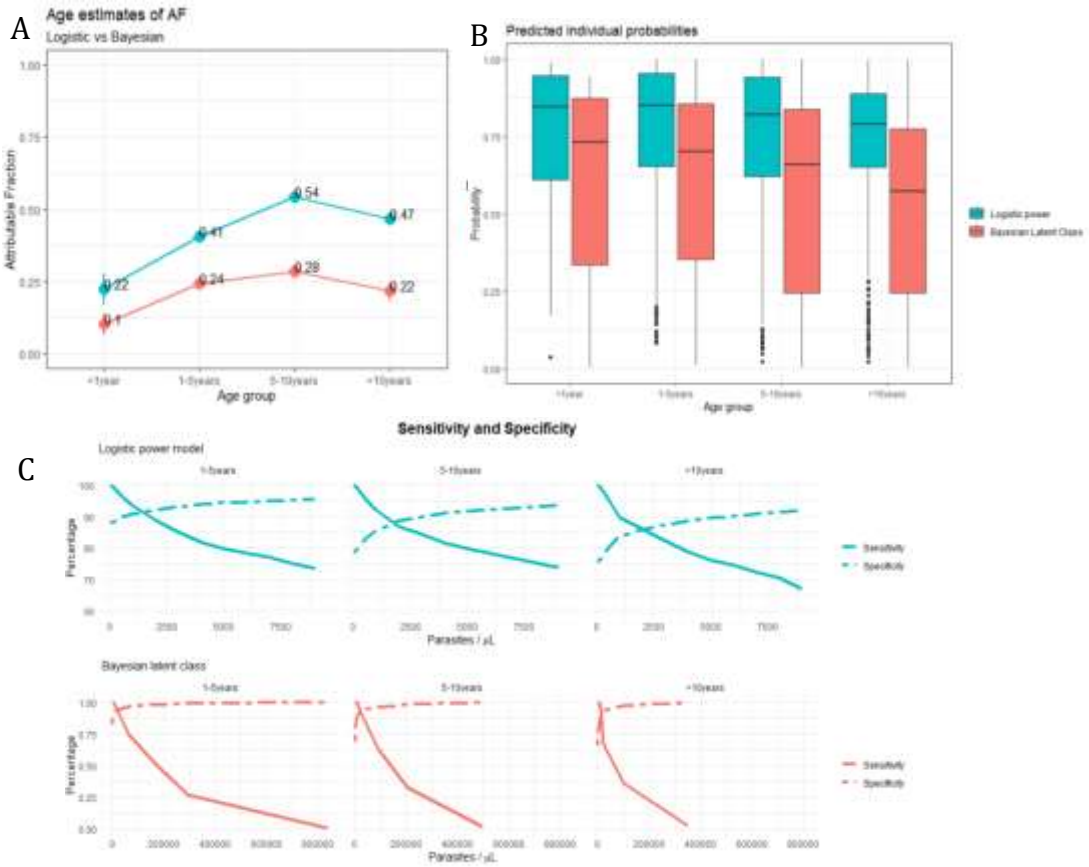

Figure 3: Malaria attributable fractions and probabilities over age group for all the study participants

the Bayesian latent class approach and the binary outcome based on the newly defined parasite density thresholds plus fever were compared. Data on antibody responses to

238 selected *Plasmodium falciparum* merozoite antigens for a study done in Kilifi was used
239 for the comparison. Specifically, the data had antibody measurements for the survey
240 conducted in Junju in the year 2008 and a subset of the Chonyi cohort in the year 2000
241 (28).

242 A cut-off of 2500 parasites/µl [Supplementary Table 1] plus fever was used to define the
243 binary outcome (malaria positive, parasites $\geq$ 2500 parasites/µL or negative otherwise).
244 The loess predicted probabilities from Bayesian latent class models were used as the
245 response variable to fit the zero-one inflated beta regressions. Table 2 shows that using
246 the probability as the outcome gave comparable point estimates with the binary outcome.
247 The binomial multilevel models however, had high standard errors and Bayesian
248 Information Criterion (BIC) values.

249 *Table 2: A comparison of a binary and probability outcome using high vs low antibody levels in Junju 2008 and*
250 *Chonyi 2000 cohort. Coef. – Regression Coefficients, SE- Standard Error and BIC – Bayesian Information Criterion,*

		Binary Outcome		**Probability Outcome**		**Probability Bayesian**
		Coef. (SE)	**BIC**	**Coef. (SE)**	**BIC**	**Coef. (SE)**
Junju Cohort	AMA1	-0.30 (0.32)	427.73	-0.22 (0.20)	416.026	-0.23 (0.19)
	MSP2	0.09 (0.34)	428.6	-0.04 (0.20)	417.3474	-0.04 (0.21)
	MSP3	-0.24 (0.35)	428.18	-0.16 (0.19)	415.947	-0.19 (0.21)
Chonyi Cohort	AMA1	0.18 (0.34)	426.8972	-0.24 (0.11)	390.0447	-0.26 (0.13)
	MSP2	-0.67 (0.54)	425.4389	-0.33 (0.14)	401.3226	-0.34 (0.15)
	MSP3	0.68 (0.35)	423.5147	-0.11 (0.12)	410.5933	-0.12 (0.14)

Discussion

In areas with high malaria transmission, differences in the prevalence of malaria fever can occur due to change in transmission intensities or differences in levels of immunity in various subsets of the population like age groups (3). The present study shows a variation of transmission intensity over time, and how this contributes to variation in the MAF. A previous study done in Kilifi showed that immunity to malaria is affected by age and transmission (3). The study compared Chonyi, a high transmission area and Ngerenya, a low transmission area. The sites had a variable age-specific clinical disease pattern with Ngerenya having a higher MAF compared to Chonyi overall and specifically for the older age group of 5-19years. Shifting MAF was also observed with changing transmission patterns in the current study. A shift in malaria transmission intensities and malaria epidemiology has been reported in different endemic areas (15,34). Therefore, it is important to review MAF and case definitions with changing transmission settings. Furthermore, transmission intensity correlates with the rate of acquisition of natural immunity (35). A decrease in malaria transmission intensity led to reduced immunity which would result in a higher tendency to acquire malaria attributable fevers at lower parasite densities as was observed in this study.

A strong rationale for developing malaria vaccines comes from cohort studies, which show that individuals continuously exposed to malaria develop immunity that initially prevents death from severe disease, and subsequently recurrent illness (1,13). The main assumption in defining correlates of protection to inform vaccine development is that malaria case definition is non-biased. Many of the studies classify the participants into

two groups (clinical malaria case and non-case) using a defined parasitemia threshold plus fever (2,3,8–10). The optimal parasite density threshold is selected from maximum combined sensitivity and specificity after fitting the case definition models (2). Additionally, the models estimate the probabilities individual episodes of fever are malaria attributable at a given density of parasitemia (2,4,18).

The logistic power model is the widely used technique for case definition (Atieli et al. 2011; Dicko et al. 2005; Mwangi et al. 2005; Olotu et al. 2010; Tom Smith, Schellenberg, and Hayes 1994) and rarely the Bayesian latent class model (4,18). However, in this study the logistic approach was observed to give higher but comparable pattern estimates with the Bayesian latent class model. Similarly, this was observed in a study done by Vounatsou et. al comparing the logistic power and Bayesian latent class model (4). The logistic model approach, however, has been reported to have a limitation of estimating imprecise standard errors and negative probabilities sometimes (6). Comparatively, in this present study, the logistic approach also gave high probability estimates with narrow variation in the low parasite densities compared to the latent class model. In vaccine studies, the latent class was reported to help in identifying possible biases in efficacy estimates since it utilizes the whole range of possible parasite density cut-offs (20). An inverse relationship of clinical malaria and age has been shown (12,36), similarly, this was observed with the estimated probabilities which decreased with age.

Continuous variables, like the probabilities we use here have been shown to have more variation information and statistical power and are sometimes preferred over the categorization of data (23).Several articles (21–24) have also pointed out problems

associated with the categorization of data. In this present study, we compare the performance of using probability and binary outcome model the association with clinical malaria (28). Assuredly, the probability model had a good statistical fit; lower BIC estimates and standard errors and gave comparable coefficients with the binary model. Also, the point estimates were similar to what was reported by Murungi et. al in the 2008 study from the same cohort (28). In this study however, they reported risk ratios that were estimated using a modified Poisson regression (37) and here we report coefficients both of them showing a lower risk of disease for individuals with high antibody measurements.

Conclusion:

The present study shows the importance of reviewing malaria case definitions and attributable fractions with changing malaria transmissions. Additionally, we show the utilization of probabilities estimated from the case definition models in modeling the association of correlates of disease. Results from Junju and Chonyi Kilifi data verify the validity of using the probability outcome to identify correlates of disease protection while still having a better statistical fit. The computational time to fit the zero-inflated models was higher compared to the binary-based regression models and a training class is required for the latent class models which can be a limitation for some cohort designs.

Approaches to estimate individual's marginal probabilities of clinical malaria over a given follow-up time would be of importance for creating parsimonious models. Further studies to compare the probabilities estimated from models utilizing the quantitative nature of

the parasite densities without grouping the data in conjunction with changing transmission would be valuable.

Ethics Declarations

Ethics approval and consent to participate

The research was given ethical approval by the University of Witwatersrand's Human Research Ethics Committee (Medical) (Clearance Certificate No. M190121) and KEMRI-Scientific and Ethics Review Unit (SERU) (Approval numbers KEMRI SSC No. 3139). Informed consent was collected from parents/guardians before the beginning of recruitment. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The scripts used for the current study are available in the GitHub repository, https://github.com/Keniajin/case_definition.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was commissioned by the National Institute for Health Research (NIHR) Global Health Research programme (16/136/33) using UK aid from the UK Government. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

K.M, F.O, S.K are funded by the National Institute for Health Research (NIHR) Global Health Research programme (16/136/33). F.O. is supported by a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation (3.2 - 1184811 - KEN - SKP) and an EDCTP Senior Fellowship (TMA 2015 SF1001) which is part of the EDCTP2 Programme supported by the European Union. S.K and E.M are supported by the DELTAS Africa Initiative under Initiative to Develop African Research Leaders (IDeAL) Grant No. DEL-15-003 and DELTAS Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) Grant No. 107754/Z/15/Z-DELTAS Africa SSACAB respectively.

Authors' contributions

Conception and Design of the work: KM, IN, AT, JM, EM and FO. Analysis: KM, JM and AT. Funding Acquisition: EM, SK and FO. Interpretation of data: KM, IN, SK, EM and FO. Work drafting: KM and FO. Review and Editing: KM, IN, AT, JM, RK, DO, LN, JT, SK, EM and FO. Project administration: LN. Software: KM, JM and AT. Supervision: EM, SK and FO. Validation: SK, EM and FO. All authors read and approved the final manuscript.

Acknowledgements

We thank all participants who took part in the cohort studies and the field workers involved in the data generation process.

References:

1. Marsh K, Kinyanjui S. Immune effector mechanisms in malaria. Parasite Immunol. 2006;28(1–2):51–60.

2. Smith T, Schellenberg JA, Hayes R. Attributable fraction estimates and case definitions for malaria in endemic. Stat Med. 1994 Nov;13(22):2345–58.

3. Mwangi TW, Ross A, Snow RW, Marsh K. Case Definitions of Clinical Malaria under Different Transmission Conditions in Kilifi District, Kenya. 2005;

4. Vounatsou P, Smith T, Smith AFM. Bayesian analysis of two-component mixture distributions applied to estimating malaria attributable fractions. J R Stat Soc Ser C (Applied Stat. 1998;47(4):575–87.

5. Smith T, Charlwood JD, Kihonda J, Mwankusye S, Billingsley P, Meuwissen J, et al. Absence of seasonal variation in malaria parasitaemia in an area of intense seasonal transmission. Acta Trop. 1993;54(1):55–72.

6. Smith T, Vounatsou P. Logistic regression and latent class models for estimating positivities in diagnostic assays with poor resolution. Commun Stat Methods. 1997;26(7):1677–700.

7. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. Biometrics. 1993;865–72.

8. Olotu A, Fegan G, Williams TN, Sasi P, Ogada E, Bauni E, et al. Defining clinical malaria: the specificity and incidence of endpoints from active and passive surveillance of children in rural Kenya. PLoS One. 2010;5(12):e15569.

9. Dicko A, Mantel C, Kouriba B, Sagara I, Thera MA, Doumbia S, et al. Season, fever prevalence and pyrogenic threshold for malaria disease definition in an endemic area of Mali. Trop Med \& Int Heal. 2005;10(6):550–6.

10. Atieli HE, Zhou G, Lee M-C, Kweka EJ, Afrane Y, Mwanzo I, et al. Topography as a modifier of breeding habitats and concurrent vulnerability to malaria risk in the western Kenya highlands. Parasites \& vectors. 2011;4(1):241.

11. Olotu A, Fegan G, Wambua J, Nyangweso G, Ogada E, Drakeley C, et al. Estimating Individual Exposure to Malaria Using Local Prevalence of Malaria Infection in the Field. PLoS One [Internet]. 2012 [cited 2021 Jun 4];7(3):e32929. Available from: www.plosone.org

12. Mogeni P, Williams TN, Fegan G, Nyundo C, Bauni E, Mwai K, et al. Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. Grais RF, editor. PLOS Med. 2016 Jun;13(6):e1002047.

13. Osier FH, Mackinnon MJ, Crosnier C, Fegan G, Kamuyu G, Wanaguru M, et al. New antigens for a multicomponent blood-stage malaria vaccine. Sci Transl Med.

2014;6(247):247ra102--247ra102.

14. Muthui MK, Mogeni P, Mwai K, Nyundo C, Macharia A, Williams TN, et al. Gametocyte carriage in an era of changing malaria epidemiology: A 19-year analysis of a malaria longitudinal cohort. Wellcome open Res. 2019;4.

15. Noor AM, Kinyoki DK, Mundia CW, Kabaria CW, Mutua JW, Alegana VA, et al. The changing risk of Plasmodium falciparum malaria infection in Africa: 2000–10: a spatial and temporal analysis of transmission intensity. Lancet. 2014;383(9930):1739–47.

16. O'Meara WP, Bejon P, Mwangi TW, Okiro EA, Peshu N, Snow RW, et al. Effect of a fall in malaria transmission on morbidity and mortality in Kilifi, Kenya. Lancet. 2008/11/06. 2008;372(9649):1555–62.

17. Snow RW, Kibuchi E, Karuri SW, Sang G, Gitonga CW, Mwandawiro C, et al. Changing malaria prevalence on the Kenyan coast since 1974: climate, drugs and vector control. PLoS One. 2015;10(6):e0128792.

18. Vounatsou P, Smith T, Kitua AY, Alonso PL, Tanner M. Apparent tolerance of Plasmodium falciparum in infants in a highly endemic area. Parasitology. 2000;120(1):1–9.

19. Small DS, Cheng J, Ten Have TR. Evaluating the efficacy of a malaria vaccine. Int J Biostat. 2010;6(2).

20. Smith TA. Measures of clinical malaria in field trials of interventions against

413 Plasmodium falciparum. Malar J. 2007;6(1):1–7.

414 21. Thoresen M. Spurious interaction as a result of categorization. BMC Med Res
415 Methodol. 2019;19(1):1–8.

416 22. Collins GS, Ogundimu EO, Cook JA, Manach Y Le, Altman DG. Quantifying the
417 impact of different approaches for handling continuous predictors on the
418 performance of a prognostic model. Stat Med. 2016;35(23):4124–35.

419 23. Altman DG, Royston P. The cost of dichotomising continuous variables. Bmj.
420 2006;332(7549):1080.

421 24. Frøslie KF, Røislien J, Laake P, Henriksen T, Qvigstad E, Veierød MB.
422 Categorisation of continuous exposure variables revisited. A response to the
423 Hyperglycaemia and Adverse Pregnancy Outcome (HAPO) Study. BMC Med Res
424 Methodol. 2010;10(1):1–5.

425 25. Weinberg CR. How bad is categorization? Epidemiology. 1995;345–7.

426 26. Scott JAG, Bauni E, Moisi JC, Ojal J, Gatakaa H, Nyundo C, et al. Profile: The Kilifi
427 health and demographic surveillance system (KHDSS). Int J Epidemiol.
428 2012;41(April):650–7.

429 27. Bejon P, Mwacharo J, Kai O, Mwangi T, Milligan P, Todryk S, et al. A phase 2b
430 randomised trial of the candidate malaria vaccines FP9 ME-TRAP and MVA ME-
431 TRAP among children in Kenya. PLOS Clin Trial. 2006;1(6):e29.

432 28. Murungi LM, Kamuyu G, Lowe B, Bejon P, Theisen M, Kinyanjui SM, et al. A

threshold concentration of anti-merozoite antibodies is required for protection from clinical episodes of malaria. Vaccine. 2013;31(37):3936–42.

29. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria; 2021. Available from: https://www.r-project.org/

30. Spiegelhalter D, Thomas A, Best N, Lunn D. OpenBUGS user manual. Version. 2007;3(2):2007.

31. Buis M. ZOIB: Stata module to fit a zero-one inflated beta distribution by maximum likelihood. 2012;

32. Liu F, Kong Y. zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression. R J. 2015;7(2):34.

33. Bejon P, Warimwe G, Mackintosh CL, Mackinnon MJ, Kinyanjui SM, Musyoki JN, et al. Analysis of immunity to febrile malaria in children that distinguishes immunity from lack of exposure. Infect Immun. 2009;77(5):1917–23.

34. Nkumama IN, O'Meara WP, Osier FHA. Changes in malaria epidemiology in Africa and new challenges for elimination. Trends Parasitol. 2017;33(2):128–40.

35. Polley SD, Mwangi T, Kocken CHM, Thomas AW, Dutta S, Lanar DE, et al. Human antibodies to recombinant protein constructs of Plasmodium falciparum Apical Membrane Antigen 1 (AMA1) and their associations with protection from malaria. Vaccine. 2004;23(5):718–28.

36. Snow RW, Omumbo JA, Lowe B, Molyneux CS, Obiero J-O, Palmer A, et al.

Relation between severe malaria morbidity in children and level of Plasmodium falciparum transmission in Africa. Lancet. 1997;349(9066):1650–4.

37. Zou G. A Modified Poisson Regression Approach to Prospective Studies with Binary Data. Am J Epidemiol. 2004;159(7):702–6.

Supplementary Files

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

- SupplementaryFigures.pptx