Check for updates

RESEARCH ARTICLE

REVISED Individual-level variations in malaria susceptibility and acquisition of clinical protection [version 3; peer review: 2 approved]

John Joseph Valletta[1]*[+], John W.G. Addy[2]*, Adam J. Reid [3], Francis M. Ndungu [4], Yaw Bediako [2,5], Jedida Mwacharo[4], Khadija Said Mohammed [4], Jennifer Musyoki[4], Joyce Mwongeli Ngoi[4,5], Joshua Wambua[4], Edward Otieno [4], Matt Berriman [3], Philip Bejon [4], Kevin Marsh [6], Jean Langhorne [2], Chris I. Newbold [3,7], Mario Recker [8]

[1]School of Mathematics and Statistics, University of St. Andrews, St. Andrews, UK
[2]Malaria Immunology Laboratory, Francis Crick Institute, London, UK
[3]Parasite Genomics, Wellcome Trust Sanger Institute, Hinxton, UK
[4]KEMRI/Wellcome Trust Research Programme, Kilifi, Kenya
[5]West African Centre for Cell Biology of Infectious Pathogens, University of Ghana, Accra, Ghana
[6]Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK
[7]MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK
[8]Centre for Ecology and Conservation, University of Exeter, Penryn, UK

* Equal contributors [+] Deceased author

Abstract
After decades of research, our understanding of when and why individuals infected with *Plasmodium falciparum* develop clinical malaria is still limited. Correlates of immune protection are often sought through prospective cohort studies, where measured host factors are correlated against the incidence of clinical disease over a set period of time. However, robustly inferring individual-level protection from these population-level findings has proved difficult due to small effect sizes and high levels of variance underlying such data. In order to better understand the nature of these inter-individual variations, we analysed the long-term malaria epidemiology of children ≤12 years old growing up under seasonal exposure to the parasite in the sub-location of Junju, Kenya. Despite the cohort's limited geographic expanse (ca. 3km x 10km), our data reveal a high degree of spatial and temporal variability in malaria prevalence and incidence rates, causing individuals to experience varying levels of exposure to the parasite at different times during their life. Analysing

Open Peer Review

Approval Status ✓ ✓

	1	2
version 3 (revision) 25 Feb 2022		✓ view
version 2 (revision) 13 Sep 2021	✓ view	? view
version 1 02 Feb 2021	? view	

1. **Lucy C. Okell** , Imperial College London, London, UK

2. **Wendy Prudhomme-O'Meara**, Duke

individual-level infection histories further reveal an unexpectedly high variability in the rate at which children experience clinical malaria episodes. Besides exposure to the parasite, measured as disease prevalence in the surrounding area, we find that the birth time of year has an independent effect on the individual's risk of experiencing a clinical episode. Furthermore, our analyses reveal that those children with a history of an above average number of episodes are more likely to experience further episodes during the upcoming transmission season. These findings are indicative of phenotypic differences in the rates by which children acquire clinical protection to malaria and offer important insights into the natural variability underlying malaria epidemiology.

Keywords
Plasmodium falciparum, clinical malaria, malaria susceptibility, naturally acquired immunity, spatial heterogeneity, longitudinal cohort study

University, Durham, USA

Moi University College of Health Sciences, Eldoret, Kenya

Christine Markwalter iD, Duke University, Durham, USA

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the KEMRI | Wellcome Trust gateway.

Introduction

Individuals growing up in *P. falciparum* malaria endemic areas acquire a general state of immunity against clinical malaria through repeated exposure to the parasite. This process of naturally acquired protection is still poorly understood[1] but believed to involve the build-up of a repertoire of immune responses against the myriad of antigenic targets that the parasite displays over its lifecycle during *in vivo* infections[2–4]. Although protection against life-threatening disease may be acquired through a small number of infections only[5], individuals remain prone to experience clinical episodes throughout childhood and sometimes even into their late teens or early adulthood, depending on the intensity of transmission (acquisition of protection is generally faster in high transmission settings with year-round transmission than in settings with little and interrupted transmission[6]).

Research into the complexity of clinical protection often relies on cohort studies, where antigen-specific immune responses are correlated against the incidence of disease. These studies have been key in advancing our understanding of the role of variant surface antigens in both disease severity and natural acquired protection, for example, and are instrumental for the discovery of novel candidate targets for vaccine research[7,8]. Unfortunately, finding robust correlates of protection is made complicated by at least two factors. First, the observed effect sizes are often small, in particular with respect to the variability of the underlying data, and it is difficult to ascertain how much these reflect the true effects sizes due to the many confounding factors influencing the observations. Second, protection itself is not a dichotomous phenotype but rather related to an individual's risk, or probability of developing disease if infected. The latter is particularly problematic because in many cases we simply do not know whether someone got infected during the study period or not.

In order to capture some of the spatial and temporal variation in disease transmission, and thus some of the uncertainty underlying the risk of an infection *per se*, Olotu *et al.*[9] previously proposed an exposure index, which provides a quantitative marker for the risk of infection experienced by an individual in space and time. What this and other research[10] found was that disease prevalence, or transmission intensity, is not necessarily homogeneous across space or time. In fact, even small geographic regions can exhibit malaria *hotspots*, where individuals have a notably higher chance of getting infected than in surrounding areas[11–13]. The temporal stability of these hotspots is variable, however. Whereas some hotspots can be persistent because of environmental or geographic factors, such as proximity to standing water and so to mosquito breeding sites, others might just persist for the duration of a transmission season or two. One way or the other, this heterogeneity in exposure must be considered as an important source of variability in an individual's risk of experiencing a clinical episode in a given year, irrespective of their degree of protection.

Longitudinal cohort studies are ideally placed to offer more detailed insights into the process of naturally acquired protection (e.g. 14). That is, following individuals over time and recording their clinical episode histories can provide important information on how previous infections relate to a child's future risk of disease and how this changes as they grow up under repeated exposure to the parasite. Here we made use of data generated from a long-term birth-cohort study in Kenya. Using individual-level episode histories together with spatio-temporal disease prevalence data we demonstrate that children exhibit a high degree of phenotypic variability regarding their development of protection from clinical malaria. We further identified independent risk factors for clinical disease, which should help to guide future studies aimed at finding robust immune signatures of anti-malarial protection.

Methods
Study population
The study was conducted through the KEMRI-Wellcome Trust Research Programme (KWTRP), Kilifi, Kenya. The children investigated were part of the long-term birth cohort in Junju. Recruitment into the cohort was voluntary following sensitisation of the whole area without specific spatial or demographic selection criteria, such thatthe cohort provides an unbiased spatial representation of the village as a whole. Children were recruited into the cohort at birth and actively monitored on a weekly basis for detection of malaria episodes until 15 years of age. See 15 for further information regarding the Junju cohort.

P. falciparum (*Pf*) episodes, defined as a body temperature > 37.5°C and 2,500 parasites per microlitre of blood, are diagnosed during weekly active surveillance, where auxiliary body temperature and/or recent history of fever were recorded. Blood samples were taken from febrile children and *Pf* infection initially detected by rapid diagnostic test (RDT) and confirmed by microscopy. Apart from the children's infection status and history of clinical malaria, we also had access to their date of birth, spatial location of their homestead, and

sickle cell trait status (AA or AS; homozygous (SS) individuals were removed from this analysis).

Ethical considerations

Approval for human participation in this cohort studies was given by Kenya Medical Research Institute Ethics Research Committee, and research was conducted according to the principles of the Declaration of Helsinki, which included the administration of informed consenting in the participant's local language.

Sample selection

Our statistical analyses are based on a subset of samples taken between 2006 and 2018 and include children between the ages of 1 and 12 years (n = 544, total number of samples N = 3767). Unless stated otherwise, these samples include children with the known malaria protective sickle cell trait (heterozygous, AS; n = 92, total number of samples N = 618). Note, throughout the analysis we refer to n as the number of individuals, and refer to N as the number of samples (with $N > n$, as individuals were sampled longitudinally over the course of the study).

Prevalence index

To capture and compare the spatial distribution of malaria over time within this cohort we devised a *prevalence index* (*PI*), which provides a summary statistic of the annual prevalence at location i in year t by summing over all individuals j with recorded malaria episodes in that year in the neighbourhood of i, weighted by their spatial distances, D_{ij}. It is given as

$$PI_{i,t} = \frac{\sum_j Z_{j,t}\, e^{-aD_{ij}}}{\sum_j e^{-aD_{ij}}}$$

where $Z_{j,t}$ is a binary operator indicating whether whether the individual had a recorded episode or not. Note, locations and individuals are used somewhat interchangeably, which also accounts for the occasions where multiple children from the same homestead are simultaneously enrolled. Although clustering at the individual household-level has previously been described (e.g. 16), we here did not consider such fine spatial resolution, especially as we only focused on children and no other family / household members.

The spatial weighting factor a determines how rapidly the influence of an infection at j on the prevalence at j declines with increased spatial distance. Note, this measure does not rely on a fixed spatial radius to determine the neighbourhood of i; instead, the chosen functional form of the spatial weighting means it naturally converges towards 0, such that far away locations have little to no effect. The value of a (here a = 2) has a significant influence on the spatial heterogeneity of *PI*, with smaller values leading to smaller differences in *PI* between distant locations, and vice versa. Initial attempts to estimate a form the data were not satisfactory due to large variations between estimates based on individual years. It was therefore decided to

determine a by maximising the correlation between the resulting prevalence index and clinical episode for all individuals and all time points (*PI ~ Z*).

Statistical modelling

In order to determine potential risk factors underlying an individual's probability of experiencing a clinical malaria episode, we built a Bayesian hierarchical logistic regression model. The modelled outcome was *GotEpisode*, a binary response variable (yes/no) indicating whether an individual experienced a clinical episode in a given year or not. The explanatory variables were *PI*, *Age*, *BirthQuarter*, and *Genotype* (AA or AS, heterozygous sickle cell trait). To test the effect of an individual's previous episode history, we built an additional model using *PrevEpisodes* along-side *PI* and *Age*, and the interaction between *Age* and *PrevEpisodes*, as explanatory variables. Here, *PrevEpisodes* refers to the child's number of previously recorded clinical episodes. In both cases we included a child-specific intercept (commonly known as a random effect in the frequentist literature), to accommodate the longitudinal nature of these data, where individuals were repeatedly sampled over multiple years over the duration of the study. To assess the effect of between-child variation, we compared the model against an equivalent non-hierarchical model assuming complete independence of all data points. Model comparison was done based on the models' expected log predictive density (ELPD) using Pareto smoothed importance-sampling leave-one-out cross-validation (PSIS-LOO) from the *loo* package in R. All continuous variables were centred and scaled (i.e. set mean = 0 and standard deviation = 1) to improve model convergence and to allow for better comparison between the estimated effect sizes. A zero-mean normal prior (with standard deviation = 10) was placed on all regression coefficients, and a gamma prior (with shape and scale parameter = 1) was placed on the standard deviation for the group-level effect (i.e. the standard deviation for the child-specific intercepts). The model parameters were jointly inferred by means of MCMC sampling, using the *rstanarm* (version 2.21.1) R package. A total of 2000 iterations (including 1000 iterations for warm-up) were used for each of four chains run in parallel. Convergence was assessed via trace plots and the Rhat convergence diagnostic. Marginal effects for the explanatory variables on the response were plotted using the *marginal_effects* function in the *brms* (version 2.13.5) R package. R version 3.6.3 was used to produce all the figures and perform all of the analyses.

Results

Malaria prevalence fluctuates over time and space

Figure 1A illustrates the annual variation in the total number of clinical episodes and proportion of individuals with a recorded clinical episode between 2006 and 2018 based on children under the age of 12 years enrolled in this cohort. Although prevalence and total number of episodes follow a qualitatively similar trend, years of high prevalence, such as 2014 and 2018, are also characterised by individuals experiencing multiple episodes in the same year (Figure 1B and 1C). That is, in those years we not only see more individuals with a clinical episode but a disproportional increase in the numbers of

Figure 1. Temporal variation in malaria incidence. A. Temporal variance in the number of clinical episodes and proportion of individuals with a recorded episode in Junju, Kenya, between 2006 and 2018. **B**,**C**. Temporal variation in the distribution of children experiencing 1 (blue bars), 2 (green bars), or ≥ 3 clinical episodes (yellow bars) per year between 2006 and 2018, highlighting an *excess* of multiple episodes in years of high transmission (2014 and 2018).

episodes per individual, which could be indicative of a prolonged or more intensive transmission season or a possibly a change in the circulating parasite population.

The strong inter-annual variability is also mirrored in the spatial distribution of malaria prevalence, as shown by means of the prevalence index (PI, see Methods) for the years 2007 to 2018 in Figure 2. Not only do these maps show significant year-on-year fluctuations but they also reveal temporal trends with some regions having much higher prevalence rates in some years than others. Furthermore, we find that in high prevalence years (2014 and 2018), malaria incidence was not confined to particular areas but was observed almost across the entire cohort, which could be indicative of more intense and/or longer transmission seasons. Note, because PI is based on recorded symptomatic episodes in children only we cannot say whether the observed spatial patterns are indicative of

dynamically changing transmission hotspots or of potential heterogeneities in the immunity landscape across the cohort, or both.

High individual-level variability in clinical episode histories

The spatio-temporal analyses above imply that individuals growing up in the cohort are likely to experience different levels of exposure to the parasite at different points during their lives. Consequently, individual histories of clinical malaria episodes are expected to be equally variable. This is exemplified in Figure 3, where individual episode histories, by means of the cumulative number of episodes, are plotted over time for children born between 2006 and 2011, stratified by the individuals' year of birth and genotype (AA/AS). Besides the high variance in the rate at which individuals acquire clinical episodes, these graphs also suggest that neither the cumulative
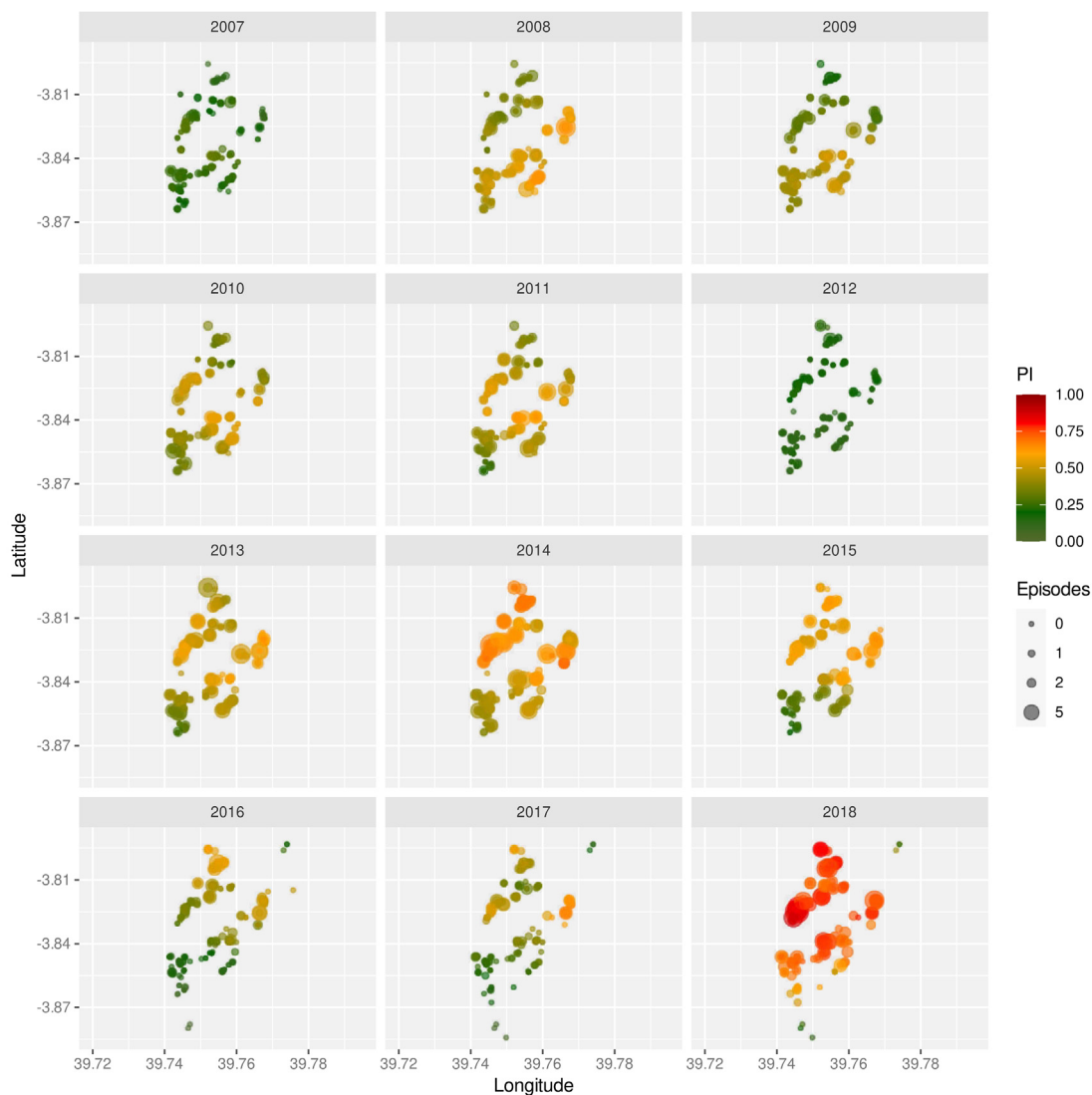
Figure 2. Spatio-temporal variation in malaria prevalence. Geographic homestead location of children enrolled in the Junju cohort, stratified by year. The size of each point correspond to the child's number of recorded episodes that year, with the colour indicating their prevalence index (PI).
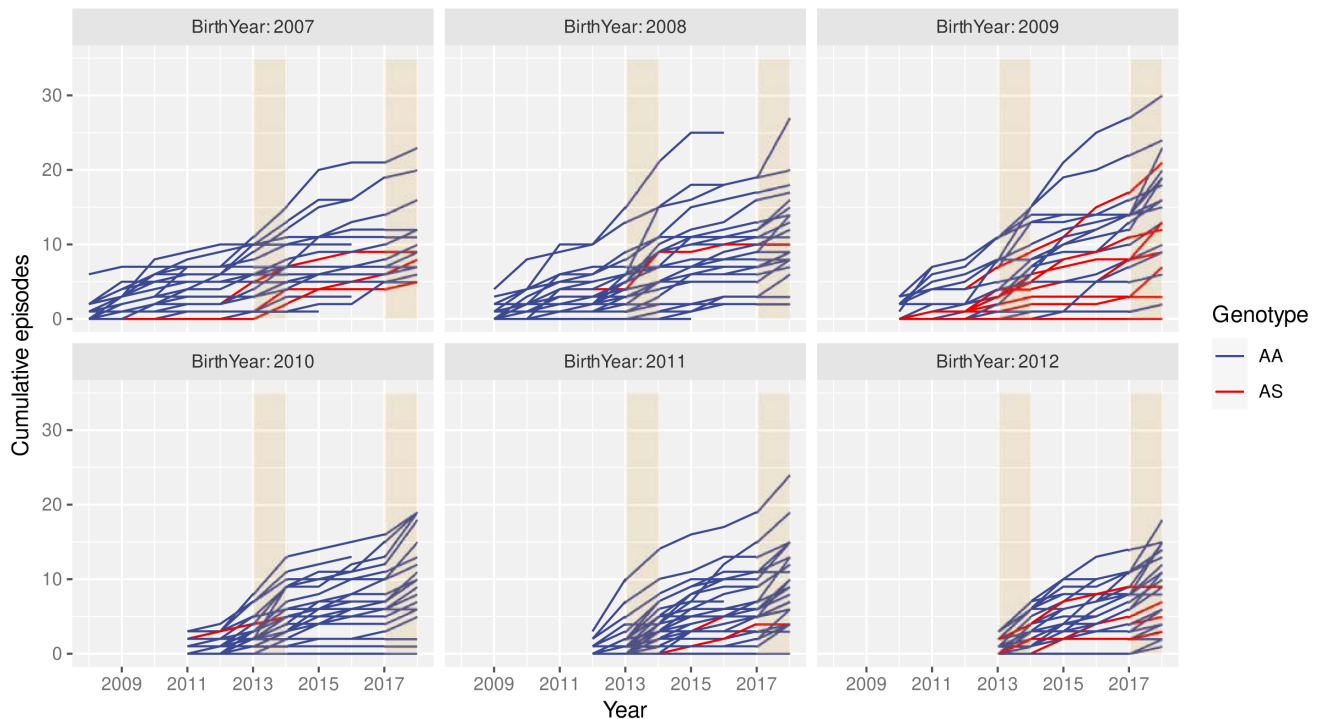
Figure 3. High variance in clinical episode histories. Graphs showing individual children's cumulative number of clinical malaria episodes over time, stratified by birth year. Red lines indicate children with the sickle cell trait (AS). Orange rectangles highlight high malaria transmission years (2014 and 2018).

number of episodes nor the rate at which episodes are acquired (i.e. number of episodes per year) are simply a function of age. That is, in years of particularly high transmission, e.g. in 2014 or 2018, most children seem susceptible to a clinical episode regardless of their age or previous episode history.

The effect of age on the risk of clinical malaria

Under the assumption of naturally acquired immunity, whereby individuals acquire protection against clinical malaria through repeated infections, we would expect age and/or the number of previously experienced malaria episodes to be correlated with a reduced risk of clinical malaria. Having access to individual episode histories we can thus analyse how both age and previous number of episodes alter the risk of future episodes. As illustrated in Figure 4, and against expectation, individuals who experience an episode in a given year are not different in age (Figure 4A) but appear to have acquired more previous episodes on average than those who did not (Figure 4B). Importantly, the fact that individuals with more episodes seem more at risk of a further episode does not seem to be an age effect but persists as individual grow older, which is demonstrated in Figure 5 by means of age-stratified distribution of previous number of episodes in individuals with and without a clinical episode.

Birth time of year affects the risk of clinical malaria

Previous studies have shown that *in utero* exposure to *P. falciparum can affect a child's future susceptibility to malaria (e.g.* 16–19). Equally, poor nutrition during pregnancy has also been linked with long-term negative consequences for disease susceptibility (e.g. 20–23. We therefore asked whether the birth time of year has an effect on a child's experience of clinical malaria episodes. For this we looked at the total number of clinical episodes acquired by a certain age (here we set age = 6 years), stratified by birth quarter (Q1-Q4). As clearly indicated in Figure 6A, children born towards the end of the year (Q4) appear to be at a lower risk of a clinical episode on average than those who are born earlier in the year.

Next we investigated the temporal stability of this protective effect, for which we calculated the age-dependent risk ratio of children born in Q4 relative to those born in Q1–Q3. As can be seen in Figure 6B, Q4 children under the age of 6 years have a significantly reduced risk of experiencing a clinical malaria episode than those born earlier. This protective effect wanes as children get older, however, possibly under the influence of other environmental factors, including continuous exposure to the parasite.
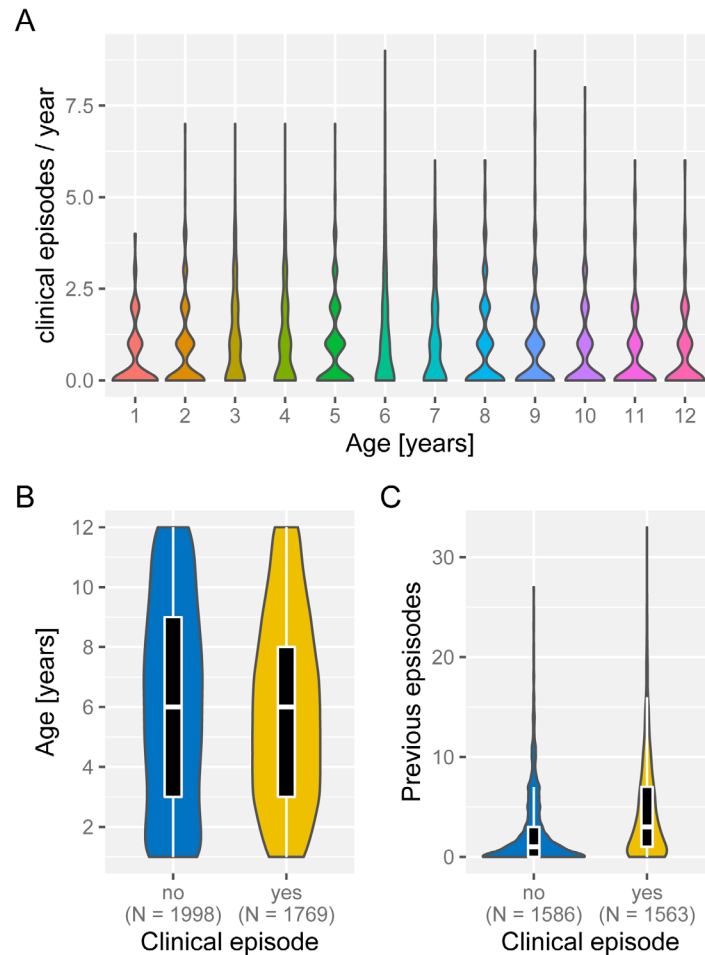
Figure 4. Effect of age and exposure history on risk of clinical malaria. Box and whisker plots indicate the median and inter quartile ranges of age and previous number of episodes of individuals with or without a recorded episode in a given year. Due to confounding effect of the sickle cell trait, AS individuals were removed in (**C**).

Risk factors of clinical malaria

The exploratory analyses presented above have revealed a high degree of heterogeneity in both malaria prevalence and in individual episode histories. In order to determine if and by how much the latter is simply a function of the former, and to what degree other factors have an independent effect on an individual's risk of experiencing a clinical malaria episode we built a Bayesian hierarchical logistic regression model (see Methods). Specifically we investigated the effect of *PI*, *genotype* (AA/AS), age and birth quarter on the probability of a child experiencing a clinical episode in a given year.

Figure 7A shows the credible intervals (CI) of the regression coefficients (median plus 80% and 95% credible intervals). As expected, malaria prevalence (*PI*) has the strongest effect on the risk of a clinical episode (median effect size estimate: 0.96, 95% CI [0.87, 1.06]). Equally expected is the protective effect of the sickle cell trait (-0.89, 95% CI [-1.23, -0.55]).

In line with the above exploratory analysis (Figure 6), birth quarter Q4 also has a protective effect (-0.42, 95% CI [-0.8, -0.03]). In contrast to the similarity in the age distribution between individuals who experience a clinical episode or not (Figure 4A), accounting for inter-individual level variation we now find that older children have a slightly reduced risk of experiencing an episode compared to younger ones (-0.15, 95% CI [-0.24, -0.07]). Figure 7B–D illustrate the marginal effects of the four covariates on the probability of a clinical episode, again showing the protective effect of birth quarter Q4, the sickle cell trait (AS), and, to a smaller and more uncertain degree, age.

Our results suggest that beside exposure, by means of disease prevalence in the surrounding area (*PI*), individual-level differences in susceptibility could have a significant effect on the risk of experiencing a clinical episode in a given year. Apart from the well-known protective sickle cell trait, these differences may
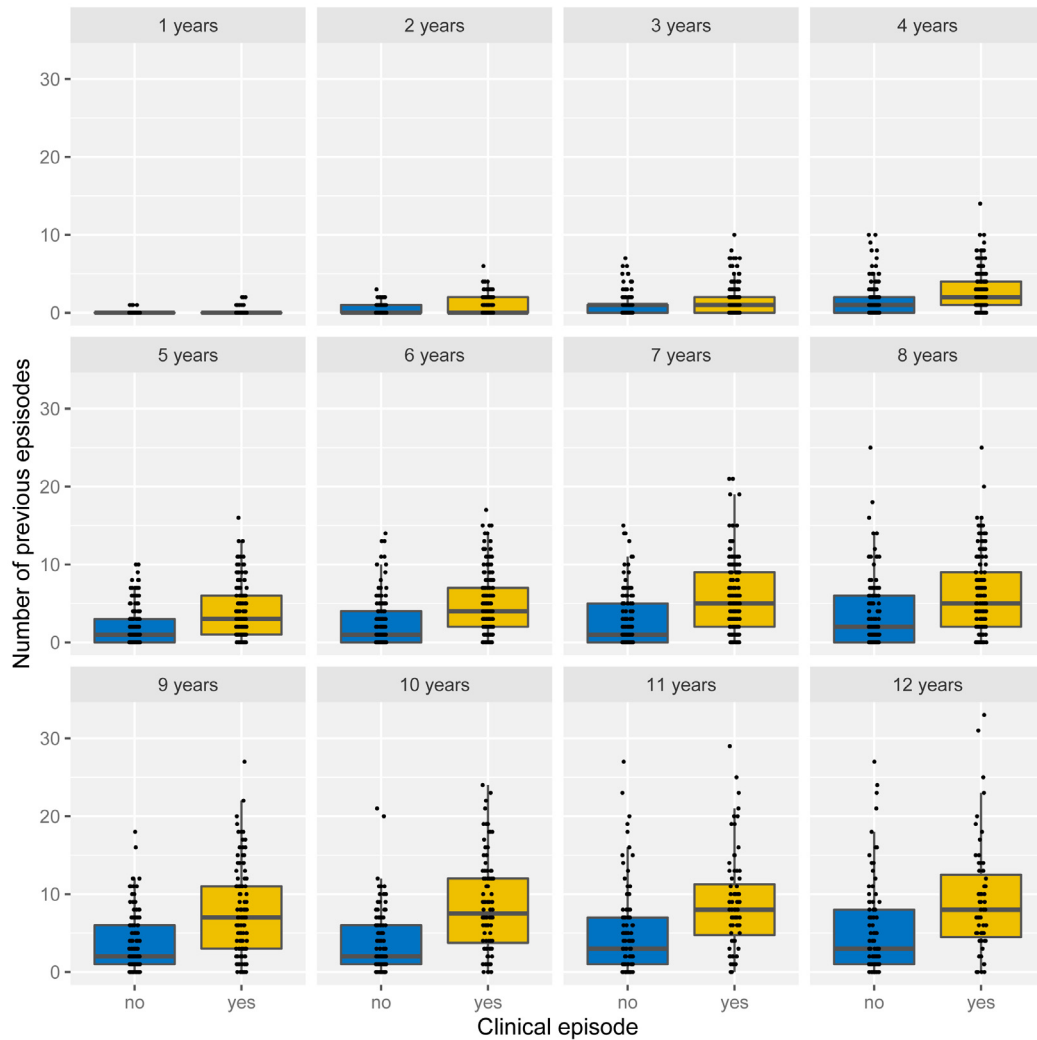
Figure 5. Age-stratified distribution of episode history in relation to risk of clinical malaria. Box and whisker plots indicate the median and inter quartile ranges of age and previous number of episodes of individuals with or without a recorded episode in a given year. Individuals with sickle cell trait removed.

stem from the birth time of year, for example, as demonstrated here, as well as other inter-individual differences accounted for by the hierarchical nature of our model (confirmed by means of model comparison against a non-hierarchical model based on the estimated log probability density, ELPD: -2136, hierarchical model, vs. -2326, non-hierarchical model; see Methods).

These individual-level differences, on the other hand, also imply that a child's episode history could in fact be indicative of their risk of experiencing a future episode, which has already been alluded to in Figure 5. To quantify this further, we used a child's previous number of clinical episodes as an explanatory variable in our model, alongside PI and Age. As the number of previous episodes is strongly influenced by the sickle cell trait, these children were removed from this analysis. As a reminder,

all variables are centred (zero mean), which allows us to make inferences on the predicted changes in risk with respect to deviations from the variable's mean value. Figure 8 demonstrates that children with an above average number of previous episodes are at a much higher risk of an additional episode than similar aged children with fewer episodes (median effect size estimate for 1–4 year old children: 1.14, 95%CI [0.79, 1.54]). Equally, older children appear to be more protected than younger children who have a similar number of previous episodes. In fact, accounting for the interaction between age and episode history we now also find that the effect of age is much more pronounced (Figure 8C,D). The overall negative effect of the interaction between age and number of previous episodes therefore strongly suggests an independent protective effect of age, regardless of the previous number of episodes.
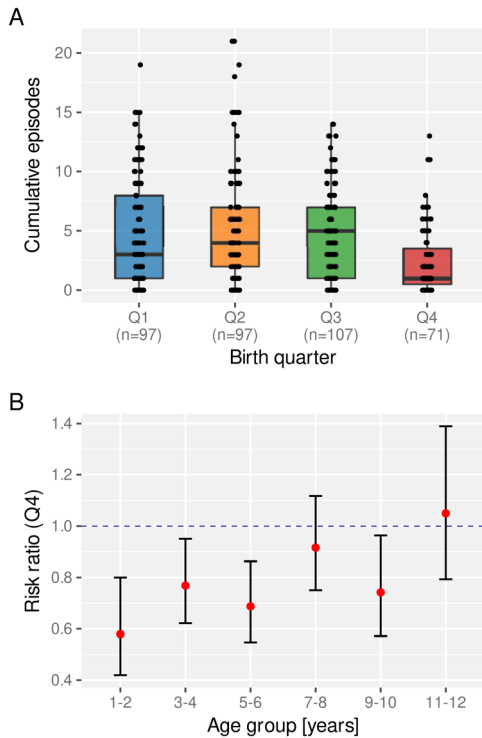
Figure 6. The effect of birth quarter on the risk of clinical malaria. A. Distribution of the number of clinical episodes experienced by an individual child by the age of 6 years stratified by birth quarter. Box and whisker plots indicate the median and inter quartile ranges. **B**. Risk ratio of experiencing a clinical episode in a given year between children born in Q4 compared to those born between Q1 and Q3.

Taken together, our results are indicative of significant (inherent or induced) differences in the qualitative and quantitative nature by which children experience clinical malaria episodes and acquire clinical protection under repeated exposure to the parasite.

Discussion

Here we used long-term epidemiological cohort data to investigate individual-level differences in children's development of clinical protection against *P. falciparum* malaria. Our analyses reveal high levels of spatial and temporal heterogeneity in malaria incidence, which lead to significant differences in exposure levels to the parasite as children grow up in this cohort. However, we found that the variability in the rate at which individuals experience clinical episodes is not solely explained by exposure alone but points towards child-specific differences in disease susceptibility. We believe that these are important yet often neglected sources of variation in the data collected from prospective cohort studies aimed at identifying correlates of protection.

It has previously been reported that some children growing up in a malaria endemic region appear to suffer from excess malaria episodes compared to other children before reaching a similar state of clinical protection, even when accounting for exposure and age[15]. These findings, which are broadly in line with the results presented here, add further evidence that children seem to exhibit inherent (phenotypic) differences in their susceptibility to clinical malaria episodes. Here we explicitly accounted for the inter-individual variability and concentrated primarily on an individual's risk of experiencing a clinical episode in a given year or not. What was interesting
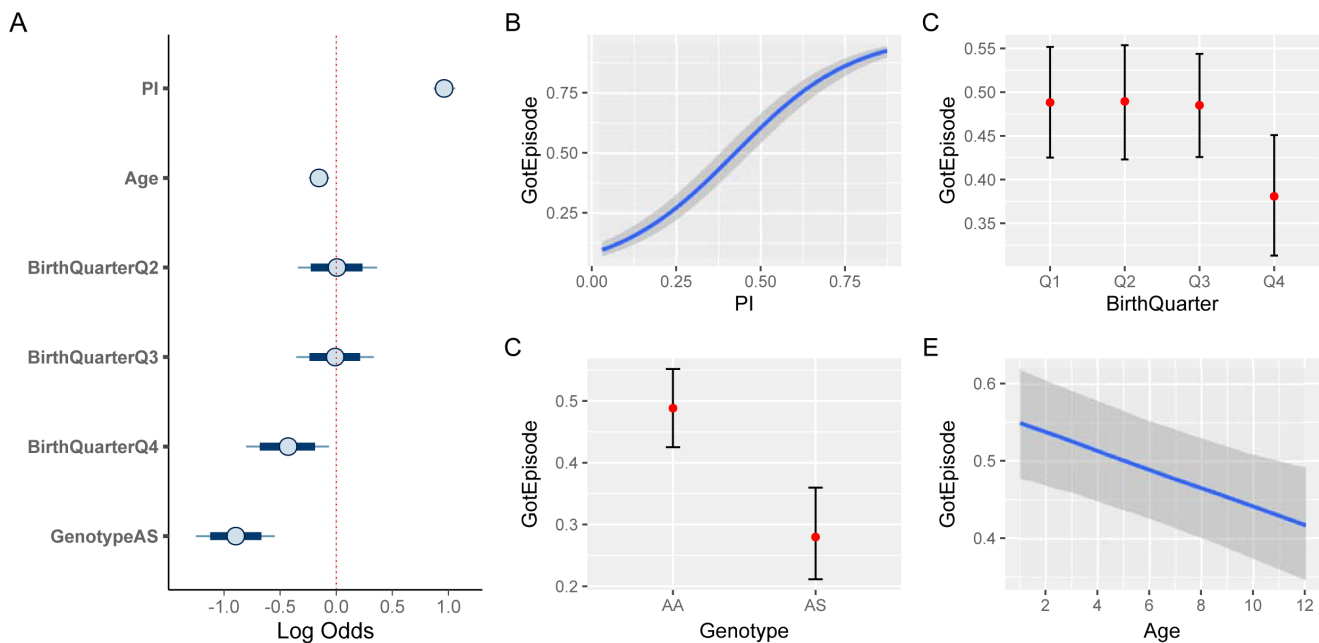
Figure 7. Risk factor analysis of clinical malaria. A. Estimated effect sizes (plus 80% and 95% CI; boxes and bars, respectively) for individual factors underlying risk of clinical malaria. **B–E**. Conditional effect plots showing the the probability of a clinical episode against prevalence index (**B**), sickle cell genotype (**C**), birth time of year (**D**), and age (**E**). The solid lines (**B**, **E**) and circles (**C**, **D**) represent the median and the shaded areas (**B**, **E**) and whiskers (**C**, **D**) the 95% prediction intervals.
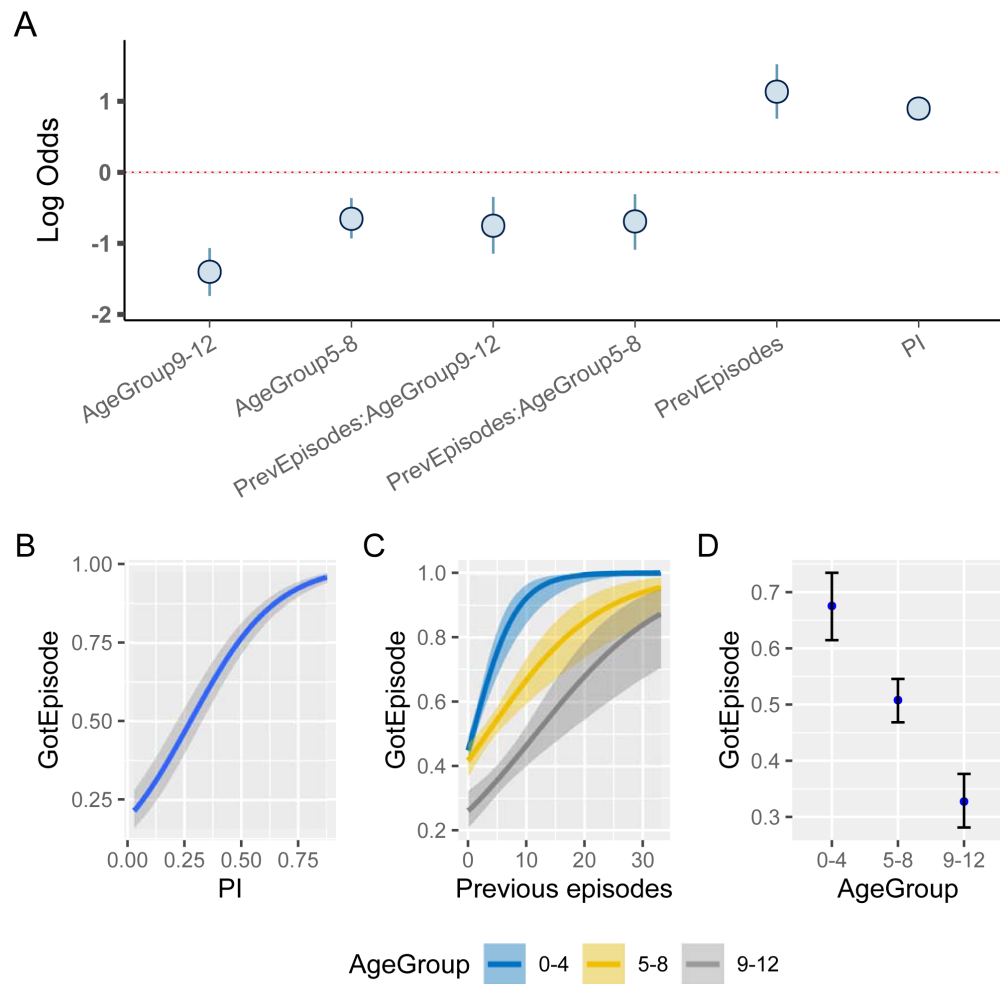
Figure 8. The effect of previous episode histories on the susceptibility to clinical malaria. A. Estimated effect sizes (plus 80% and 95% CI; boxes and bars, respectively) for individual factors underlying risk of clinical malaria. **B–D**. Conditional effect plots showing the the probability of a clinical episode against prevalence index (**B**), previous episodes (**C**), and age (**D**). The solid lines and shaded regions represent the median and the 95% prediction intervals.

to observe is that age itself only appeared to have a significant effect when accounting for this inter-individual variability. That is, although age usually needs to be factored in as a confounding factor when looking for immune correlates of protection, as older children are more likely to have experienced more episodes and have thus developed a higher degree of protection, we find that in this relatively low transmission setting and below the age of 12 years it only has a marginal effect on someone's risk of experiencing a clinical episode.

Unsurprisingly, the most significant risk factor was simply exposure. Here we devised a new prevalence index, which

provided a measure of infection prevalence in the surrounding area in a given year. It is related to the exposure index as previously introduced by Olotu *et al.*[9] but yields more spatially smoothed estimates and in this case revealed clear temporal trends in the spatial distribution of infection prevalence (Figure 2). Although our calculation of the prevalence index was based only on children with a confirmed episode, it is unlikely that these heterogeneities disappear when accounting for all individuals. Another important point to make here is that this measure is semi-quantitative in that it is based on the number of infected children, and not the total number of infections in the surrounding area. However, and as evidenced by our analyses, it is strongly correlated with the probability of a

child experiencing a clinical episode in a given year and thus captures the level of malaria transmission in the surrounding area. Our findings are also in line with previous studies that have revealed that malaria infections are not necessarily homogeneous across space but can exhibit pronounced hotspots of variable stability (e.g. 12,13,24). Unfortunately, the data analysed here did not allow us to investigate whether these spatio-temporal patterns are also reflective of locally changing transmission intensities or are purely driven by the changing immunity landscape among the study participants, or possibly both.

Apart from age and exposure, our analysis suggests that the time of year when an individual is born has a protective effect against clinical episode and that this effect can last for many years of early childhood. In this case we found that individuals born during the last quarter of the year (October - December) had a reduced risk of experiencing clinical episodes, which also manifested in a lower number of total episodes compared to other age-matched children under similar exposure conditions. In terms of timing, being born during Q4 in this cohort means that infants may acquire their first infection at the age of 4–6 months towards the end of maternal protection, assuming that the main transmission season in Junju starts around April. Alternatively, it might open up the possibility that their mothers got infected during the second trimester. It has previously been reported that malaria infections during pregnancy can alter an infant's susceptibility to clinical malaria (e.g. 16–19). It is also known that nutritional intake during pregnancy, which in many parts of the world still changes significantly between seasons, can have a lasting effect on the child's susceptibility to disease (e.g. 21). Unfortunately, we cannot say whether the effect that we observe is due to *in utero* exposure during pregnancy or first exposure under waning maternal protection. Follow-up studies will be required to answer both this question and also if similar effects can be found in other malaria endemic settings.

Another important finding is that the risk of a clinical malaria episode is not simply a decreasing function of the number of episodes experienced in the past. In fact, we found that those children who have experienced many more episodes by a certain age have a much higher baseline risk than other, similar aged children. What we cannot conclude from this analysis is whether this might be due to inherent, i.e. genetic differences in susceptibility to clinical malaria, or whether these differences are somehow induced whereby an exposure to the parasite early in life can lead to impaired immunity to malaria, which in turn leads to further episodes of malaria in the future. A recent study has also found that individuals who carry asymptomatic infections are at an increased risk of experiencing a symptomatic episode in the near future[25]. However, as the observed effect declined rapidly over time and was also seen across all age groups it is questionable whether it describes the same phenomenon as reported here. Using a systems immunological approach, we have previously described how repeated malaria episodes can lead to measurable modifications of the immune system, based on the comparison between children with few versus a large number of episodes[26]. Using the insights gained from this analysis it would thus be interesting to follow a

similar approach and compare children based on the risk factors identified herein.

There are a number of caveats in this study. The first is that we only monitored children under the age of 12 years. What is clear from our analyses is that by this age, and in this setting, many of the children are still experiencing clinical episodes at a rate similar to younger children, even though their risk of severe and life-threatening malaria had declined long before that. Due to limited sample sizes and high variability in malaria transmission over space and time, identifying children who have acquired a level of natural protection is almost impossible simply by comparing incidence of clinical malaria. In fact, even children who appeared to have *plateaued* in experiencing new episodes[24] and are thus believed to have acquired a state of clinical protection can succumb to a malaria episode once more in years of high transmission (e.g. 2018). This, on the other hand, is an important point to re-emphasize: protection itself is not a dichotomous state but can best be understood as the probability of developing clinical symptoms if infected. Monitoring individual children and their level of exposure to malaria over long periods of time is therefore required to robustly infer immunological changes underlying their transition to clinical protection. The other caveat is that our measure of prevalence is retrospective, i.e. it is a summary measure based on all recorded infections over the entire year. One could thus argue that its high correlation with clinical episodes follows a circular argument. However, as the index case was not included in its calculation and because its computation is done over the same period of time as the response (clinical episode in that year, yes or no), this should not be a reason for concern.

Conclusions

Cohort studies and results based on them are often affected by the high variability in the underlying data. This may not only drown out real effects, which can be small even when comparing clearly define phenotypes at the gene expression level[25], but can also make comparisons between studies challenging. What we have demonstrated here is that individual-level variations in observed outcomes, in this case the incidence of clinical episodes in a given year, are more than noise in the collected data and have to be taken into consideration when trying to make inferences from population-level measures. The highlighted heterogeneity in exposure to the parasite implies that children experience different transmission intensities at different times during their life. So far we do not know whether there is a (long-term) difference in children who experience the majority of their clinical episodes before the age of 6 years or after, for example. What is clear, though, is that there are pronounced differences in the children's pathway to becoming clinically protected. Identifying those who get there much faster than others might be key for broadening our understanding of natural acquired protection against malaria.

Data availability
Underlying data
Kemri Wellcome Trust Research Programme (KWTRP) Research Data Repository: Replication Data for: Individual-level

variations in malaria susceptibility and acquisition of clinical protection, (https://doi.org/10.7910/DVN/WQCKJJ)[27].

Note, spatial location data of individual homesteads has been removed due to confidentiality reasons. Data sharing requests will be reviewed by the Data Governance Committee at KEMRI CGMRC Kilifi (dgc@kemri-wellcome.org) who will manage the process and ensure that appropriate ethical approval is in place (if applicable), consent obtained caters for uses outlined in the data request, and the request is in line with the institution's policies on data sharing.

Extended data
Kemri Wellcome Trust Research Programme (KWTRP) Research Data Repository: Replication Data for: Individual-level variations

in malaria susceptibility and acquisition of clinical protection, (https://doi.org/10.7910/DVN/WQCKJJ)[27].

This project contains the following extended data:
- Data overview
- Statistical analysis

References

1. Langhorne J, Ndungu FM, Sponaas AM, *et al.*: **Immunity to malaria: more questions than answers.** *Nat Immunol.* 2008; **9**(7): 725–32.
   PubMed Abstract | **Publisher Full Text**

2. Bull PC, Lowe BS, Kortok M, *et al.*: **Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria.** *Nat Med.* 1998; **4**(3): 358–360.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Ofori MF, Dodoo D, Staalsoe T, *et al.*: **Malaria-induced acquisition of antibodies to *Plasmodium falciparum* variant surface antigens.** *Infect Immun.* 2002; **70**(6): 2982–2988.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Chan JA, Howell KB, Reiling L, *et al.*: **Targets of antibodies against *Plasmodium falciparum*-infected erythrocytes in malaria immunity.** *J Clin Invest.* 2012; **122**(9): 3227–38.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. Gupta S, Snow RW, Donnelly CA, *et al.*: **Immunity to non-cerebral severe malaria is acquired after one or two infections.** *Nat Med.* 1999; **5**(3): 340–3.
   PubMed Abstract | **Publisher Full Text**

6. Snow RW, Omumbo JA, Lowe B, *et al.*: **Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa.** *Lancet.* 1997; **349**(9066): 1650–1654.
   PubMed Abstract | **Publisher Full Text**

7. Richards JS, Arumugam TU, Reiling L, *et al.*: **Identification and Prioritization of Merozoite Antigens as Targets of Protective Human Immunity to *Plasmodium falciparum* Malaria for Vaccine and Biomarker Development.** *J Immunol.* 2013; **191**(2): 795–809.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Osier FH, Mackinnon MJ, Crosnier C, *et al.*: **New antigens for a multi-component blood-stage malaria vaccine.** *Sci Transl Med.* 2014; **6**(247): 247ra102.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. Olotu A, Fegan G, Wambua J, *et al.*: **Estimating Individual Exposure to Malaria Using Local Prevalence of Malaria Infection in the Field.** *PLoS One.* 2012; **7**(3): e32929.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

10. Kreuels B, Kobbe R, Adjei S, *et al.*: **Spatial variation of malaria incidence in young children from a geographically homogeneous area with high endemicity.** *J Infect Dis.* 2008; **197**(1): 85–93.
    PubMed Abstract | **Publisher Full Text**

11. Bousema T, Drakeley C, Gesase S, *et al.*: **Identification of hot spots of malaria transmission for targeted malaria control.** *J Infect Dis.* 2010; **201**(11): 1764–74.
    PubMed Abstract | **Publisher Full Text**

12. Bejon P, Williams TN, Liljander A, *et al.*: **Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya.** *PLoS Med.* 2010; **7**(7): e1000304.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Bejon P, Williams TN, Nyundo C, *et al.*: **A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots.** *eLife.* 2014; **3**(3): e02130.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. O'Meara WP, Mwangi TW, Williams TN, *et al.*: **Relationship between exposure, clinical malaria, and age in an area of changing transmission intensity [version 1; peer review: 2 approved with reservations].** *Am J Trop Med Hyg.* 2008; **79**(2): 185–91.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Ndungu FM, Marsh K, Fegan G, *et al.*: **Identifying children with excess malaria episodes after adjusting for variation in exposure: identification from a longitudinal study using statistical count models.** *BMC Med.* 2015; **13**(1): 183.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Mutabingwa TK, Bolla MC, Li JL, *et al.*: **Maternal malaria and gravidity interact to modify infant susceptibility to malaria.** *PLoS Med.* 2005; **2**(12): e407.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

17. Le Hesran JY, Cot M, Personne P, *et al.*: **Maternal placental infection with *Plasmodium falciparum* and malaria morbidity during the first 2 years of life.** *Am J Epidemiol.* 1997; **146**(10): 826–31.
    PubMed Abstract | **Publisher Full Text**

18. Sylvester B, Gasarasi DB, Aboud S, *et al.*: **Prenatal exposure to *Plasmodium falciparum* increases frequency and shortens time from birth to first clinical malaria episodes during the first two years of life: prospective birth cohort study.** *Malar J.* 2016; **15**(1): 379.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Kakuru A, Staedke SG, Dorsey G, *et al.*: **Impact of *Plasmodium falciparum* malaria and intermittent preventive treatment of malaria in pregnancy on the risk of malaria in infants: a systematic review.** *Malar J.* 2019; **18**(1): 304.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

20. Imdad A, Lassi Z, Salaam R, *et al.*: **Chapter 1 - Prenatal Nutrition and Nutrition in Pregnancy: Effects on Long-Term Growth and Development.** In *Early Nutrition and Long-Term Health: Mechanisms, Consequences, and Opportunities.* Elsevier Inc., 2017; 3–24.
    Publisher Full Text

21. Barker DJ, Clark PM: **Fetal undernutrition and disease in later life.** *Rev Reprod.* 1997; **2**(2): 105–12.
    PubMed Abstract

22. Leon DA, Lithell HO, Vâgerö D, *et al.*: **Reduced fetal growth rate and increased risk of death from ischaemic heart disease: cohort study of 15 000 Swedish men and women born 1915-29.** *BMJ.* 1998; **317**(7153): 241–5.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Lumey LH: **Reproductive outcomes in women prenatally exposed to undernutrition: a review of findings from the Dutch famine birth cohort.** *Proc Nutr Soc.* 1998; **57**(1): 129–35.
    PubMed Abstract | **Publisher Full Text**

24. Addy JWG, Bediako Y, Ndungu FM, *et al.*: **10-year longitudinal study of malaria in children: Insights into acquisition and maintenance of naturally acquired immunity [version 1; peer review: 2 approved with reservations].** *Wellcome Open Res.* 2021; **6**: 79. submitted.
    Publisher Full Text

25. Tran TM, Guha R, Portugal S, *et al.*: **A Molecular Signature in Blood Reveals a Role for p53 in Regulating Malaria-Induced Inflammation.** *Immunity.* 2019; **51**(4): 750–765.e10.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Bediako Y, Adams R, Reid AJ, *et al.*: **Repeated clinical malaria episodes are associated with modification of the immune system in children.** *BMC Med.* 2019; **17**(1): 60.
    PubMed Abstract | **Publisher Full Text** | **Free Full Text**

27. Recker M: **Replication Data for: Individual-level variations in malaria susceptibility and acquisition of clinical protection.** Harvard Dataverse, V1. 2020;
    http://www.doi.org/10.7910/DVN/WQCKJJ

Open Peer Review

Current Peer Review Status: ✓ ✓

Version 3

Reviewer Report 10 March 2022

https://doi.org/10.21956/wellcomeopenres.19562.r48907

✓   **Wendy Prudhomme-O'Meara**
[1] Duke Global Health Institute, Duke University, Durham, NC, USA
[2] Moi University College of Health Sciences, Eldoret, Kenya

Christine Markwalter (iD)

Duke Global Health Institute, Duke University, Durham, NC, USA

The authors considered our comments and made useful additions to the methods in particular. It is disappointing that most of their responses to our more technical comments will remain hidden in the reviewer dialogue. Some of these limitations seem to be related to the very lengthy review processes and restrictions of the journal. I still feel that fitting the parameter 'a' to maximize the correlation of PI is a bit circular. They are basically fitting PI to the outcome and then finding that it's a risk factor.

Also, please consider the following to check for a typo:
   ○ Pg 4, when discussing the spatial weighting factor a: "The spatial weighting factor a determines how rapidly the influence of an infection at j on the prevalence at j declines with increased spatial distance." I'm pretty sure they mean: "...influence of an infection j on the **_PI_** at **_i_**..."

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Malaria, epidemiology, implementation science/trials

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 2

Reviewer Report 26 January 2022

https://doi.org/10.21956/wellcomeopenres.19005.r47493

❓   **Wendy Prudhomme-O'Meara**
¹ Duke Global Health Institute, Duke University, Durham, NC, USA
² Moi University College of Health Sciences, Eldoret, Kenya

Christine Markwalter 🆔
Duke Global Health Institute, Duke University, Durham, NC, USA

This manuscript presents an analysis of how individual differences in exposure affect children's risk of experiencing a clinical malaria episode. Based on weekly surveillance in a 12-year longitudinal birth cohort in Kenya, the authors describe spatio-temporal variability in malaria incidence and heterogeneity in clinical episode histories for children living in a relatively small geographic area. They found that children with a high number of previous clinical episodes are at a higher risk of an additional episode, challenging the expectation that protective immunity, and therefore decreased risk of clinical episode, is acquired with increasing symptomatic exposure. While the manuscript is well-written and of interest to the malaria community, we have some concerns regarding the metrics used to define local exposure and potential sampling bias, which are detailed below.

Major concerns:

The Prevalence Index (PI) was described in the manuscript as capturing the spatial distribution of malaria over time within the cohort. It was also found to have the strongest effect on risk of clinical malaria compared to other covariates. However, we are concerned:

○ PI is used as a metric of local risk, but it is not clear how spatially representative the cohort is. There is no information about intervening households not enrolled in the cohort. PI assigns a number to only the local malaria cases only among cohort members which may not reflect something fundamental about the underlying epidemiology. It would be helpful to comment on how spatially representative cohort households are in Junju or how children were selected for enrollment. We went back to some previous papers and still could not quite get this clearly.

○ Some clarification of the variable definitions for PI is needed. The introduction to PI in the methods describes summing over all recorded malaria episodes in a year; however, the sums are defined over j, which is described as a location in the text. If the latter, and since $Z_{j,t}$ is binary for at least one clinical episode at location j, locations/households with vastly different rates of clinical malaria in a given year would be weighted equally, diminishing some heterogeneity in the data.

○ Given the previous two points, and particularly since this number is based on symptomatic cases, we feel that the term "prevalence" index could be misleading. "Cohort incidence index" or similar may be more appropriate.

○ It would be useful to see more information about 'a'. It is given as 2 with an indication that it was estimated from the data, but more information about a, a description of how well a fit

the data, a sensitivity analysis of the effect of *a* on the relationship between PI & episodes all would be useful in understanding this central parameter.

○ It is interesting that the zones of highest 'PI' seem to shift from north to central to south over the course of the study period. This is consistent with our finding of dynamic hotspots of fever (Platt Sci Rep). Could this be related to local levels of population immunity? Or is it influenced by the open cohort model where you have more cohort members in different zones over time?

Regarding the birth exposure analysis, is it possible that Q4 births are at lower risk simply because they have less accumulated exposure/episodes before entering the cohort? Given the incremental differences between Q1--> Q2 -->Q3-->Q4, with the largest difference seen between Q1 and Q4 births, and the effect waning with age, this seems like one possible explanation. Based on your annualized outcome (Gotepisode = annual have episode or not), I assume that a child enters the analysis only for the first calendar year AFTER their first birthday which could introduce the above bias. With the detailed data available, could a more precise metric of pregnancy exposure be estimated, based on rainfall, total cohort cases in the 9 months before their birth, or even PI associated to a birth (or pregnancy). This would be a more compelling comparison.

Other concerns:
○ Did you observe any household clustering of children with high numbers of cumulative episodes? Our work has demonstrated that infections are primarily clustered (Obala PLOS One) and also related (Nelson Nat Comm) at the household level (as measured by haplotype sharing) indicating that spatial structure of local 'prevalence' beyond the household is probably measuring a risky environment (ie mosquito density) rather than the actual risk of a nearby infection on another susceptible individual. Additional consideration/discussion of household-level factors would strengthen the overall conclusions of the manuscript.

○ Figure 5: The y axis values (0 - 3.5 previous episodes) in Figure 5 do not seem to agree with those in Figures 3 and 4B, where it appears some children experienced up to 30 previous episodes. How do you reconcile these numbers?

○ Given the large differences between years in overall episodes, shouldn't year be a covariate in the model given that children will have episode data spanning years with much different overall morbidity? In a previous analysis that included this cohort, calendar year was identified as a significant independent predictor of clinical malaria (Ref 15; Ndungu *et al*. BMC Med 2015).

○ Why was *PrevEpisodes* excluded from the risk factor analysis in Figure 7?

○ Figure 8: It would be helpful to have a graphical representation of the interaction between previous episodes and age to aid the reader in understanding your interpretation of the interaction as "older individuals will by nature have experienced more episodes and also acquired a higher degree of protection"; for example, plotting Fig 8C for multiple ages/age strata would more clearly communicate your point.

○ Reading your responses to Dr. Okell's thoughtful comments significantly aided in our understanding of different aspects of the manuscript; however, some of the indicated

modifications still seem to be missing from version 2 of the manuscript, including:
  - ○ Revised description of spatial homo/heterogeneity in results section text for Fig 2.
  - ○ Additional panel in Fig 4 showing the distribution of incidence (clinical episodes per year) for all children against age.
  - ○ Clarifying the units to effect sizes/regression coefficients for Fig 7-8
  - ○ Additional interpretation in the text for Figure 8C and 8D

- ○ Other studies have also found that previous exposure increases risk of future exposure (Sumner eLIFE, asympt exposure increases risk of future symptoms).

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Malaria, epidemiology, implementation science/trials

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 31 Jan 2022
Mario Recker, University of Exeter, Penryn, UK

We would like to thank the reviewers for their careful evaluation of our manuscript, which we hope has helped to improve clarity regarding the methodology, incl. sampling, and presentation of our results. Please see below for a point-by-point response to all concerns raised by the reviewers. Note, we have included here some additional figures to address specific points; these are not part of the manuscript as we felt they would distract from the main findings.

Major concerns:

The Prevalence Index (PI) was described in the manuscript as capturing the spatial distribution of malaria over time within the cohort. It was also found to have the strongest effect on risk of clinical malaria compared to other covariates. However, we are concerned:

- PI is used as a metric of local risk, but it is not clear how spatially representative the cohort is. There is no information about intervening households not enrolled in the cohort. PI assigns a number to only the local malaria cases only among cohort members which may not reflect something fundamental about the underlying epidemiology. It would be helpful to comment on how spatially representative cohort households are in Junju or how children were selected for enrollment. We went back to some previous papers and still could not quite get this clearly.

Response: This is an important point. Recruitment into the cohort was based on which homesteads volunteered following sensitisation of the whole area and was not subject to specific spatial or demographic selection criteria. We therefore believe that the cohort analysed here does provide an unbiased spatial representation of the village as a whole and have added this information in our revised manuscript (Methods section). Furthermore, although the prevalence index (PI) is based on recorded malaria episodes in only a small subset of the population, it strongly correlates with an individual's risk of malaria in a given year. And whilst we agree that this simple index does not capture all of the underlying factors contributing to risk, we cannot think of any reasons why it would not be reflective of the general spatio-temporal epidemiology in this area.

- Some clarification of the variable definitions for PI is needed. The introduction to PI in the methods describes summing over all recorded malaria episodes in a year; however, the sums are defined over j, which is described as a location in the text. If the latter, and since $Z_{j,t}$ is binary for at least one clinical episode at location j, locations/households with vastly different rates of clinical malaria in a given year would be weighted equally, diminishing some heterogeneity in the data.

Response: We agree that the description of how the prevalence index is derived lacked clarity, which we hope is now easier to follow in the revised version. In brief, for each individual we sum over all individuals (not homesteads) within a certain spatial neighbourhood (radius around the index location). For this we used a binary classification: 1 if the neighbouring individual had a recorded episode that year, and 0 otherwise. This way, locations or households with vastly different rates of clinical malaria, that the reviewers were alluding to, will be weighted differently; that is, a location / household with many individuals who had a recorded infection contribute more than an equally distant location with only a single individual. Spatial heterogeneity can therefore be captured whilst also achieving some spatial smoothing.

- Given the previous two points, and particularly since this number is based on symptomatic cases, we feel that the term "prevalence" index could be misleading. "Cohort incidence index" or similar may be more appropriate.

Response: This is a fair point; however, at this stage of the reviewing process, with the manuscript having been online for over a year, we feel it might be better to stick with prevalence index but to clarify this point in the revised manuscript to avoid potential misinterpretation.

- It would be useful to see more information about 'a'. It is given as 2 with an indication that it was estimated from the data, but more information about a, a description of

how well *a* fit the data, a sensitivity analysis of the effect of *a* on the relationship between PI & episodes all would be useful in understanding this central parameter.

Response: *This is another important point, which also lacked detail in the original submission. The parameter a acts like spatial smoothing, with larger values corresponding to smaller influences of spatially distant locations contributing to the prevalence index (PI), leading to more spatial heterogeneity, as shown in the graph below plotting the standard deviation of PI against a.*

Graph 1

We initially tried to use spatial statistics to infer the value from the data but found that it varied too much between individual years. In the end we therefore decided to use a value that maximised the correlation between PI and number of episodes for all individuals and all years, as shown in the graph below.

Graph 2

This value was also found to be a good 'visual' compromise between spatial smoothing and maintaining sufficient heterogeneity to demonstrate spatio-temporal trends (Fig. 2). We would usually add these graphs as supplementary material but here, given the WT OpenResearch publishing format, felt that they would distract from the main findings. We hope that interested readers will still be able to find this information in the comments section.

- ○ It is interesting that the zones of highest 'PI' seem to shift from north to central to south over the course of the study period. This is consistent with our finding of dynamic hotspots of fever (Platt Sci Rep). Could this be related to local levels of population immunity? Or is it influenced by the open cohort model where you have more cohort members in different zones over time?

Response: *We agree that this was an interesting finding and partly motivated some of the subsequent analyses. We speculate that this might indeed be driven by a dynamically changing immunity landscape, as previously indicated in Bejon et al. PloS Med (2010) and also in line with the paper referred to by the reviewers, rather than spatially biased enrolment into the cohort (for reasons mentioned above). We have added a brief discussion on this point in our revised manuscript.*

Regarding the birth exposure analysis, is it possible that Q4 births are at lower risk simply because they have less accumulated exposure/episodes before entering the cohort? Given the incremental differences between Q1--> Q2 -->Q3-->Q4, with the largest difference seen between Q1 and Q4 births, and the effect waning with age, this seems like one possible explanation. Based on your annualized outcome (Gotepisode = annual have episode or not), I assume that a child enters the analysis only for the first calendar year AFTER their first birthday which could introduce the above bias. With the detailed data available, could a more precise metric of pregnancy exposure be estimated, based on rainfall, total cohort cases in the 9 months before their birth, or even PI associated to a birth (or pregnancy). This would be a more compelling comparison.

Response: *This is an interesting hypothesis but there are reasons to believe why this cannot fully*

explain our observations. Children are enrolled at birth but for our analyses we only selected children 1 year old or above. It is true that individuals born early January have nearly an extra year of exposure compared to those born late in December and therefore could theoretically have more cumulative episodes. However, we did not find an incremental differences between Q1->Q2->Q3->Q4 that the reviewers are referring to (see Figs. 6 and 7). This becomes more apparent when we stratified by birth month, as shown below (due to low sample sizes we a birth quarter stratification in the manuscript).

Graph 3

Furthermore, although it is possible that the discrete nature of age and year can introduce the referred bias in the number of cumulative episodes, this would not affect their risk of getting a clinical episode in subsequent years, which our analysis showed was reduced for those born in Q4. We fully agree that these observations warrant more detailed analyses into the likely causes (e.g. exposure during pregnancy or exposure during first year of life), and in fact is something that we are currently working on, but we feel that this was outside the scope of this manuscript.

Other concerns:
- ○ Did you observe any household clustering of children with high numbers of cumulative episodes? Our work has demonstrated that infections are primarily clustered (Obala PLOS One) and also related (Nelson Nat Comm) at the household level (as measured by haplotype sharing) indicating that spatial structure of local 'prevalence' beyond the household is probably measuring a risky environment (ie mosquito density) rather than the actual risk of a nearby infection on another susceptible individual. Additional consideration/discussion of household-level factors would strengthen the overall conclusions of the manuscript.

Response: We did not look specifically at household clustering but agree that this might be a potentially important factor contributing to an individual's risk of infection / clinical episodes. Of a total of over 500 individuals analysed, only around 50 shared homesteads, and because we only considered children aged 1-12 years, rather than all members in a household, we do not believe that this would make a significant difference to the results presented here.

- ○ Figure 5: The y axis values (0 - 3.5 previous episodes) in Figure 5 do not seem to agree with those in Figures 3 and 4B, where it appears some children experienced up to 30 previous episodes. How do you reconcile these numbers?

Response: Thank you for pointing this out, this was due to an unfortunate mislabelling of the y-axes; what was plotted in Fig. 5 was the average number of episodes per year, i.e. cumulative episodes / age. We have now revised this figure to show total number of previous infections (going up to >30 in the 12 year old age class).

- ○ Given the large differences between years in overall episodes, shouldn't year be a covariate in the model given that children will have episode data spanning years with much different overall morbidity? In a previous analysis that included this cohort, calendar year was identified as a significant independent predictor of clinical malaria (Ref 15; Ndungu *et al*. BMC Med 2015).

Response: This is a good point; however, the effect of year is implicitly taken care of by the prevalence index, which strongly correlates with the total number of episodes and hence year, as

shown in the below graph.

Graph 4

- Why was *PrevEpisodes* excluded from the risk factor analysis in Figure 7?

Response: The reason why PrevEpisodes was excluded is because we argued that individual-level factors such as genotype or birth quarter will affect the number of previous episodes experienced by a child, i.e. PrevEpisodes itself will be a function of those factors. This is why we then carried out the subsequent analysis using only PrevEpisodes and PI as covariates, which is shown in Fig. 8.

- Figure 8: It would be helpful to have a graphical representation of the interaction between previous episodes and age to aid the reader in understanding your interpretation of the interaction as "older individuals will by nature have experienced more episodes and also acquired a higher degree of protection"; for example, plotting Fig 8C for multiple ages/age strata would more clearly communicate your point.

Response: We fully agree and have now revised the analysis to use age categories instead of using age as a continuous variable. The new figure also more clearly demonstrates the interaction effect between the number of previous episodes and age (i.e. that the effect size of PrevEpisodes decreases with increasing age).

Graph 5

- Reading your responses to Dr. Okell's thoughtful comments significantly aided in our understanding of different aspects of the manuscript; however, some of the indicated modifications still seem to be missing from version 2 of the manuscript, including:
    - Revised description of spatial homo/heterogeneity in results section text for Fig 2.
    - Additional panel in Fig 4 showing the distribution of incidence (clinical episodes per year) for all children against age.
    - Clarifying the units to effect sizes/regression coefficients for Fig 7-8
    - Additional interpretation in the text for Figure 8C and 8D

Response: We apologise for this but are equally puzzled as to why none of our resubmitted and revised material (manuscript and figures) following Dr. Okell's comments had been available online. We will try to navigate this system more carefully to make sure that all updated text and figures are available for review.

- Other studies have also found that previous exposure increases risk of future exposure (Sumner eLIFE, asympt exposure increases risk of future symptoms).

Response: Thank you for pointing us towards this interesting study, which incidentally was published after our manuscript went online. We would like to point out that the effect described therein declined rapidly over time and was observed across all age groups. It is therefore questionable whether this describes the same phenomenon underlying our observations. This is now briefly discussed in the Discussion.

Competing Interests: No competing interests were disclosed.

Reviewer Report 14 September 2021

✔

Lucy C. Okell 🆔

MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

The authors have addressed all my comments very thoroughly - thank you for the insightful responses. I have no further comments.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology, mathematical modelling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 13 July 2021

https://doi.org/10.21956/wellcomeopenres.18203.r44305

? **Lucy C. Okell** [iD]

MRC Centre for Global Infectious Disease Analysis, Department of Infectious Disease Epidemiology, Imperial College London, London, UK

This analysis harnesses a rich dataset following children over many years within one area and assessing the incidence of clinical malaria episodes. The findings are very interesting and challenge the idea that immunity to disease could be acquired after a specific number of previous disease episodes. Disease appears to be more concentrated within particular individuals than might be expected within a given area. The findings have important implications for studies of malaria susceptibility, risk factors, immune protection, and mathematical models. The analysis is detailed, thoughtful and well written.

I have some comments and clarifications:
- Could you clarify in the methods whether children were continuously recruited during the study period, or were all recruited at birth at the beginning of the study and followed up? This becomes clearer later on, but it would help interpretation e.g. of Figure 1, where it's not quite clear if year also indicates increasing average age of the cohort.

- Can you briefly state whether any interventions were provided as part of the cohort study (treatment, nets...)

- Are all episodes uncomplicated malaria, or do you count severe malaria also?

- Statistical modelling – can you explain a bit more the 'got episode' outcome versus the 'previous episodes' explanatory variable. Does 'got episode' depend on the most recent year of data for a given child, while the previous episodes are those occurring in the years before that?

- In this part of the results: "in those years we not only see more individuals with a clinical episode but a disproportional increase in the numbers of episodes per individual" – how is the disproportional part assessed – maybe based on a Poisson/negative binomial distribution or similar?

- Figure 2 is a nice depiction. Are the years with high prevalence definitely more spatially homogeneous? Visually the lower prevalence years, e.g. 2007 and 2012 also look quite homogeneous.

○ Fig 4 is very interesting. Why are sickle cells carriers removed from panel B but not A?

○ It could be helpful to have a standard plot of incidence of episodes by age near the beginning, to help the reader compare this cohort to those from other endemic settings.

○ P6 Can you clarify the units of the effect size estimates - are they the regression coefficients on the log odds scale? i.e. 0 = no effect? Same in the fig 7A legend.

○ Fig 7 – I like these plots, but is the blue just 1-yellow? It may be clearer with just one or the other. It would be good to also define 'GotEpisode' again in the legend to save the reader referring back – i.e. probability of a current clinical episode – per year, I think?

○ Fig 8. Given the interaction, more explanation is needed for panels A, C and D. What is the baseline group for age and previous episodes in the regression, and what units are used? For what age does C show the effect of previous episodes? (the youngest age perhaps, or the median?). It could be helpful to plot panel C twice, once the effect of previous episodes for a younger age and once for an older age, so the reader has more intuition how the interaction works (there should be less effect of previous episodes at older ages given the direction of interaction?). Does the effect of previous episodes ever become protective at older ages in the multivariable model? (despite not being protective in Fig 5).

○ Same query for panel D – for what number of previous episodes does this age effect apply?

○ Can you comment on the strength of correlation between previous episodes and the prevalence index? Is it possible to distinguish the two effects from each other during the regression? E.g. can a child with say 20 previous episodes come from a lower PI area?

○ A very interesting discussion. Could you explain this sentence in the last paragraph some more? "children experience different transmission intensities at different times during their life" – I was not sure how this was inferred.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Epidemiology, mathematical modelling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 03 Sep 2021

Mario Recker, University of Exeter, Penryn, UK

We would like thank the reviewer for their positive evaluation of our work and critical comments, which we believe has helped to improve the quality and readability of our manuscript. Please see below for a point-by-point response to issues raised.

- *Could you clarify in the methods whether children were continuously recruited during the study period, or were all recruited at birth at the beginning of the study and followed up? This becomes clearer later on, but it would help interpretation e.g. of Figure 1, where it's not quite clear if year also indicates increasing average age of the cohort.*
  Response: Since the initial setup of this cohort in 2005, children have been continuously recruited at or shortly after birth and actively monitored on a weekly basis for detection of malaria episodes. Furthermore, we aimed to maintain a similar average age over time by only selecting children that were between 1 and 12 years old in any given year. We have clarified this in the Method section.

- *Can you briefly state whether any interventions were provided as part of the cohort study (treatment, nets...)*
  Response: This was unfortunately omitted in the description of the data / cohort. All malaria positive children are treated with co-artemether. Bednet coverage is high in this cohort (doi: 10.1186/s12879-017-2822-x) and hence we did not consider it necessary to distribute nets specifically to study participants. We found no correlation between bednet use and outcome; hence bednet use was not considered for further analyses. This has now been added to the Methods section.

- *Are all episodes uncomplicated malaria, or do you count severe malaria also?*
  Response: This community based active surveillance of malaria with prompt provision of anti-malarials did not identify episodes of severe malaria. A malaria episode (which relates to our response variable *GotEpisode*) is defined as axillary temperature $\geq 37.5°C$ associated with >2,500 *P. falciparum* parasites per microliter of blood. This has been clarified.

- *Statistical modelling – can you explain a bit more the 'got episode' outcome versus the 'previous episodes' explanatory variable. Does 'got episode' depend on the most recent year of data for a given child, while the previous episodes are those occurring in the years before that?*

Response: As outlined in the Methods section, *GotEpisode* is a binary response variable (yes/no) indicating whether an individual experienced a clinical episode (defined as above) in a given year or not. The variable *PrevEpisodes*, on the other hand, is the recorded total number of episodes a child has experienced up until but not including the year in question. We have tried to make this clearer in the Method section.

○ *In this part of the results: "in those years we not only see more individuals with a clinical episode but a disproportional increase in the numbers of episodes per individual" – how is the disproportional part assessed – maybe based on a Poisson/negative binomial distribution or similar?*
Response: This was originally assessed by eye only as it was clear from Figure 1B that in years with low-to-medium levels of infections, less than a third of children experience more than 2 episodes per year, whereas this increases to around two thirds in years of high prevalence. We have now added another panel showing the frequency distribution of 1, 2, and 3+ episodes to the graph to make this point clearer.

○ *Figure 2 is a nice depiction. Are the years with high prevalence definitely more spatially homogeneous? Visually the lower prevalence years, e.g. 2007 and 2012 also look quite homogeneous.*
Response: This is a good point and unfortunate wording on our part. We have replaced this sentence with "*Furthermore, we find that in high prevalence years (2014 and 2018), malaria incidence was not confined to particular areas but was observed almost across the entire cohort, which could be indicative of more intense and/or longer transmission seasons.*"

○ *Fig 4 is very interesting. Why are sickle cells carriers removed from panel B but not A?*
Response: The reason for removing sickle cell carriers in Figure 4B is because it is a known protective trait. AS children will therefore have a lower total number of previous episodes and are also less likely to experience an episode in a given year, such that leaving them in would confound the results. On the other hand, they would not influence the relationship between age and the risk of an episode. We have added a note on this in the main text.

○ *It could be helpful to have a standard plot of incidence of episodes by age near the beginning, to help the reader compare this cohort to those from other endemic settings.*
Response: Thank you for the suggestion. We have now added an additional panel to Fig 4 showing the distribution of incidence (clinical episodes per year) for all children against age. This graph demonstrates that age on its own, at least in this cohort within the age range considered, is not a good predictor of clinical protection.

○ *P6 Can you clarify the units of the effect size estimates - are they the regression coefficients on the log odds scale? i.e. 0 = no effect? Same in the fig 7A legend.*
Response: All continuous variables were centred and scaled for the statistical analyses to permit better comparison between effect sizes. For those variables a unit represents one standard deviations from the mean. And yes, the regression

coefficient are log odds, with 0 indicating no effect, negative values a protective effect and positive values an increase in risk. We have added a corresponding note to the Methods section and the main text and changed the x-axis labels in Figs 7A and 8A.

○ *Fig 7 – I like these plots, but is the blue just 1-yellow? It may be clearer with just one or the other. It would be good to also define 'GotEpisode' again in the legend to save the reader referring back – i.e. probability of a current clinical episode – per year, I think?*
Response: Thank you for the suggestion, we have amended the figure accordingly (as blue was indeed 1-yellow). We have also added a reminder to the main text and figure legends that *GotEpisode* refers to a child experiencing a clinical malaria episode in a given year or not.

○ *Fig 8. Given the interaction, more explanation is needed for panels A, C and D. What is the baseline group for age and previous episodes in the regression, and what units are used? For what age does C show the effect of previous episodes? (the youngest age perhaps, or the median?). It could be helpful to plot panel C twice, once the effect of previous episodes for a younger age and once for an older age, so the reader has more intuition how the interaction works (there should be less effect of previous episodes at older ages given the direction of interaction?). Does the effect of previous episodes ever become protective at older ages in the multivariable model? (despite not being protective in Fig 5).*
Response: We fully agree that Fig 8 was hard to interpret without further explanation. As mentioned above, all numerical variables are centred and scaled, so the effects shown in panel A are with respect to a standard deviation from the mean. Panel B is probably the easiest to interpret and simply shows that an individual's risk of an episodes increases with increasing malaria prevalence in their surrounding area. What panel C shows is that of similarly aged individuals, those with a higher number of previous episodes are at a higher risk of subsequent episodes. Panel D shows this the other way around, i.e. that of children with similar numbers of previous infections, those that are older are more protected. And this interaction between age and previous number of episodes is what explains the overall small effect size of age when not accounting for it (shown in Fig 7). We have now added a more detailed interpretation of these results to the main text. With regards to age itself, and plotting panel C twice, as shown in Fig 5, the effect of the previous number of episodes does not appear to decline within the age groups considered in this analysis. Furthermore, as the data is centred, the results implicitly compare young and old children. We do agree, however, that at some point in their lives this effect should diminish; unfortunately, though, we do not have the data to explore this further.

○ *Same query for panel D – for what number of previous episodes does this age effect apply?*
Response: Please see response to previous comment.

○ *Can you comment on the strength of correlation between previous episodes and the prevalence index? Is it possible to distinguish the two effects from each other during the regression? E.g. can a child with say 20 previous episodes come from a lower PI area?*
Response: This is unfortunately not possible for this regression analysis simply because the prevalence index varies so much both over time and space, meaning it is

not possible to categorise children based on PI in a meaningful way for more than a year or two. Furthermore, we have mapped the cohort by number of previous episodes (results not shown) and could not find any clear spatial clustering of individuals with high a low number of episodes. And although we would expect that children with many episodes must come from areas where there was a lot of transmission, it is possible to find other children living in the same area with much fewer recorded episodes, which to us suggests substantial individual-level variation in the risk of experiencing a malaria episode.

○ *A very interesting discussion. Could you explain this sentence in the last paragraph some more? "children experience different transmission intensities at different times during their life" – I was not sure how this was inferred.*
Response: This statement is based on the fact that malaria prevalence in this cohort varies significantly across space and time, as illustrated in Fig. 2. So for example, a child living in the South-West might have experienced high malaria transmission during early childhood and much less so in later years, whereas a child growing up in the North-East would have been exposed to less malaria during the first few years of their life but more so at a later stage during their childhood. We currently do not know if or how this affects an individual's ability to acquire protective immunity, but maybe this is a first step in acknowledging that these differences exist and can arise even within a small geographic area.

Competing Interests: No competing interests were disclosed.