**Pareto Rules for Malaria Super-spreaders and Super-spreading**

Laura Cooper1,2, Su Yun Kang3, Donal Bisanzio3, Kilama Maxwell4, Isabel Rodriguez-Barraquer5, Bryan Greenhouse6, Chris Drakeley7, Emmanuel Arinaitwe4,7, Sarah Staedke7, Peter W. Gething3, Philip Eckhoff8, Robert C. Reiner Jr. 9, Simon I. Hay3,9, Grant Dorsey6, Moses R. Kamya10, Steven W. Lindsay11, Bryan T. Grenfell1,12,13, \*David L. Smith9

1. Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA
2. Department of Veterinary Medicine, Cambridge University, Cambridge, UK
3. Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK
4. Infectious Diseases Research Collaboration, Kampala, Uganda
5. Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA
6. Department of Medicine, University of California, San Francisco, CA, USA
7. London School of Hygiene & Tropical Medicine, London, UK
8. Institute for Disease Modeling, Bellevue, WA, USA
9. Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA
10. School of Medicine, Makerere University College of Health Sciences, Kampala, Uganda
11. School of Biological and Biomedical Sciences, Durham University, Durham, UK
12. Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA
13. Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

\*Materials and correspondence should be sent to: David L. Smith, Institute for Health Metrics and Evaluation, 2301 5th Ave, Seattle, WA 98121

# Abstract

Heterogeneity in transmission is a challenge for infectious disease dynamics and control. A 20-80 “Pareto” rule has been proposed to describe this heterogeneity whereby 20% of individuals (super-spreaders) account for 80% of transmission. However, it is unclear whether certain individuals are intrinsically more important for transmission; do super-spreaders exist, or is super-spreading an unpredictable but unavoidable feature of epidemic spread? We investigated super-spreading in malaria transmission at three sites in Uganda. We found transmission became less heterogeneous as transmission intensity increased; biting overall followed a 20-90 rule at the lowest transmission intensities, declining to 20-50 at the highest intensities. Household-level heterogeneity and seasonality explained only half of the variance in exposure. Our analysis suggests super-spreading is an important feature of malaria transmission but is more prominent at low-intensity transmission. Malaria super-spreaders exist, but seasonality and environmental stochasticity are of greater importance for super-spreading. Modest gains are expected from targeting super-spreaders.

# Main

Heterogeneity shapes infectious disease epidemiology and transmission. Understanding the causes and consequences of heterogeneity is important for analysis of infectious disease data and for determining target intervention coverage levels for control. A “Pareto rule” has been proposed for many infectious disease systems: 20% of hosts account for 80% of all infectious disease transmission 1. Such heterogeneity, often called “super-spreading,” has drawn interest because the efficiency of disease control could be dramatically improved if it were possible to identify and target individuals who account for most of transmission, who are sometimes called “super-spreaders” 1-6, a topic on which mathematics and mechanistic models have provided some of the most important insights7-9. Here, we test the proposed Pareto rule and provide quantitative entomological estimates of malaria super-spreaders and super-spreading using mosquito count data from a large study of malaria in Uganda.

Epidemiological theory has identified several quantitative effects of biting heterogeneity on malaria transmission. Heterogeneity in biting rates among individuals could amplify the transmission of pathogens as they invade a population, facilitating establishment of stable endemic transmission3. Heterogeneity could affect the functional relationships between the metrics used to measure malaria transmission intensity10-12. Heterogeneous exposure could be an important confounder in epidemiological studies, as individuals with higher rates of infection and clinical disease may also develop immunity more rapidly13. Since malaria immunity is only partially protective, a correlation between exposure and immunity could produce a spurious correlation between immunity and disease14-16. Finally, if it were possible to target those who are bitten most, heterogeneous biting could allow for more efficient malaria control and elimination1. These notions are motivated by mathematical theory, supported by the observation that mosquito counts tend to follow a negative binomial distribution, and are consistent with both a mechanistic understanding of mosquito behavior and studies describing heterogeneous transmission.

Hypotheses about heterogeneity and transmission – scaling relationships, confounding, and targeting – relate to the fraction of heterogeneous biting that is consistently and predictably associated with households or individuals, as opposed to other sources of variability, primarily seasonality, and environmental heterogeneity. Mosquito biting patterns, human exposure to malaria parasites, and malaria infection risk vary seasonally and are demonstrably heterogeneous when examined using molecular 17-20 or statistical techniques21, and have been associated with several factors3. Heterogeneity in biting over time can also arise from factors such as wind and human activity patterns, hereafter referred to as environmental stochasticity. Notably, heterogeneity in transmission can also be affected by differences in the infectiousness of individuals 22, which may be unrelated to biting rates.

To understand the relationship between biting heterogeneity and biting intensity, we studied heterogeneous biting using mosquito count data collected during routine entomological surveillance in a large malaria study in Uganda, where paired entomological and epidemiological data were collected on focal individuals in households . Using these data, we quantified the patterns and estimated the proportion of variance explained by household biting weights, seasonality, and environmental stochasticity. In light of these findings, we have re-examined the role of heterogeneity in malaria transmission dynamics and control.

# Results

## Seasonality & Transmission

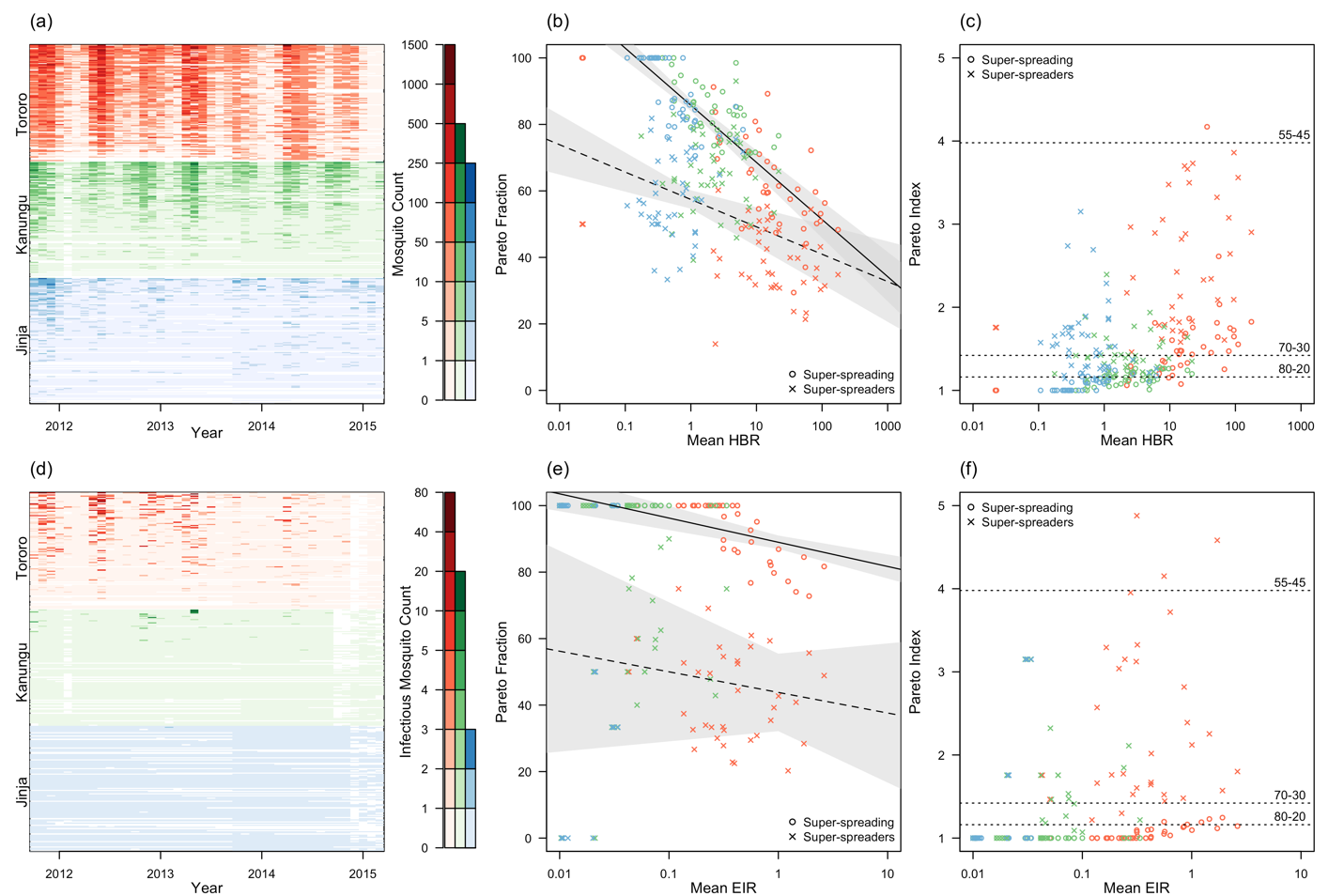
The annual entomological inoculation rate (EIR) at these three sites for the whole 3.5 years of observation was lowest in Walukuba, Jinja District (annual EIRone infectious bite, i.b.), intermediate in Kihihi, Kanungu District (annual EIR nine i.b.), and highest in Nagongera, Tororo District (annual EIR85 i.b.). The EIR has declined from what was previously reported for 201123. The analysis quantified the temporal trend in average exposure with an irregular seasonal pattern across years (denoted for day at the site, Fig. 1). This trend in Tororo also captured a sharp decline in mosquito densities following the implementation of an indoor residual spraying program by the Uganda National Malaria Control Program, with spraying in December 2014, June 2015, and December 2015, corresponding to the beginning of year three in Fig. 1, as described previously24.

Macintosh HD:Users:dave:Dropbox:Work:Manuscripts:HeteroPowerLaws:SuperspreadingFigures:Superspreading_Figure3.pdf

**Fig. 1. Temporal trends in average daily mosquito exposure (*i.e.* HBR)**, , for each site – (a) Jinja, (b) Kanungu, (c) Tororo – plotted for each one of the three study sites along with credible intervals (dashed). Seasonal signals for study sites with lower transmission have been plotted for Kanungu and Tororo to show the scale of the differences in the mosquito counts among sites.

## Pareto Rules for Human Biting Rates

The proposed Pareto rule, when applied to malaria super-spreading, predicts that 20% of the population would receive 80% of all bites1 across settings and seasons. Using monthly catch data for both daily human biting rate (HBR) and EIR, we found that the proportion of all bites received monthly by the most exposed quintile of households for that period varies from above 90% for daily HBR rates of less than 0.3 to approximately 50% for daily HBR of 175 (Figs. 2b and 2d). Even though the variance in biting increases, the dispersion of these distributions decreases dramatically as the mean biting increases. These data suggest that the proposed 20-80 rule does not hold across the spectrum of transmission. Instead, biting becomes less heterogeneous overall as biting intensity increases.

 **­­**

**Fig. 2. The counts and Pareto analysis for all Anophelines (top row) and for infected Anophelines (bottom row).** (a) Anopheline mosquito catch counts by month and household (each household is on one line), sorted within each site by the median counts for each household. Darker colors indicate higher counts. (b) Monthly Pareto fractions for super-spreading (circles) and super-spreader (x’s) by mean monthly HBR. Linear fit lines for super-spreading (solid) and super-spreaders (dashed) are shown with confidence intervals in grey. (c) The Pareto index, plotted vs. the logged mean monthly HBR for both super-spreading (circles) and super-spreaders (x’s). The dashed lines are values of the Pareto index that give the 80-20, 70-30, and 55-45 distributions. The range is restricted to 1-5 (see Table 1 for outliners). (d) Sporozoite-positive anopheline mosquito catch counts by month and household (each household is on one line), sorted within each site by the median counts for each household. Darker colors indicate higher counts. (e) Monthly Pareto fractions for super-spreading (circles) and super-spreaders (x’s) by mean monthly EIR. Linear fit line for super-spreading (line) and super-spreaders (dashed) are shown with confidence intervals in grey. (f) The Pareto index, plotted vs the logged mean monthly EIR for both super-spreading (circles) and super-spreaders (x’s). As with c), the range is restricted to 1-5 (see Table 1 for outliners).

## Heterogeneity & Transmission

At the crux of the problem of understanding the consequences of heterogeneity on transmission and for targeting and other epidemiological phenomena is that the observed negative binomial distribution of catch counts could have arisen with the contribution of super-spreaders represented along a spectrum. To assess superspreading, we conducted a study to estimate the household biting weights for each household at each site (, which are interpreted as the ratio of a household’s expectation to the overall expectation for the population. One hypothetical process, conceptually located at one extreme of this spectrum, is that after accounting for temporal heterogeneity and sampling noise, household biting weights would account for all the remaining variance in HBR; super-spreading would be entirely attributable to super-spreaders. At the other extreme of this spectrum, daily environmental stochasticity would account for all remaining variance, leaving nothing explained by super-spreaders. Points along the spectrum different fractions of heterogeneity apportioned to household biting weights (i.e., super-spreaders) versus environmental stochasticity (super-spreading without super-spreaders).

Our data and analysis allow us to evaluate where these populations are along this continuum, as our estimated daily seasonal expectation for the HBR and the household biting weights (i.e., and allow us to evaluate what fraction of the variance is explained by these two factors, leaving the rest explained by environmental stochasticity, measurement errors, and sampling noise. We estimated the proportion of variance associated with seasonality, and next with seasonality and household biting weights (Fig. 3b). Approximately half of the variance was attributable to the combined effects of seasonality and household biting weights. After accounting for seasonality, the proportion of the variance explained by household biting weights ranked in reverse order to biting intensity; household biting weights accounted for the highest fraction of the variance at the lowest-intensity site.

Consistent with this, the Pareto analysis for super-spreaders also suggests household biting weights accounted for a small fraction of the total heterogeneity. Even in the lowest-intensity site, where 20% of the population received between 80% and 100% of bites at any one point in time, environmental stochasticity reshuffled the identities of the households with the most mosquitoes each month, lowering the estimated degree of aggregation. The analysis for super-spreading thus overestimates the proportion of exposure that is targetable because it selects a different set of highly exposed households each month. Because of environmental variability, the Pareto statistics for super-spreaders necessarily indicate less aggregation than for super-spreading, but the statistics for super-spreaders are a better measure of the likely impact of targeting. For super-spreaders, the fraction of bites and infectious bites on the top 20% was substantially lower, ranging from 70% down to 40% (Figs. 2b and 2d). Because the analysis focuses on super-spreaders, it may not be true that the top 20% account for at least 20% of the bites in every month. Here, is and Table 1

The utility of the Pareto analysis is at least partly limited when the data are dominated by zero counts, as they are for HBR at the lowest intensities and for EIR at two of the sites. When there are only a handful of positive observations, the Pareto analysis reports that fact almost tautologically. With that caveat, we note that the super-spreaders analysis for EIR suggests that there is virtually no relationship between super-spreading and super-spreaders. There was, in fact, very little indication of any Pareto pattern for super-spreaders in the EIR counts data. This is perhaps due to the expectation that infectious bites, being found only in older mosquitoes, should tend to be more widely dispersed.

Theory suggests that heterogeneous biting through super-spreaders plays an important role in sustaining transmission; populations are more readily invaded by malaria parasites in models where biting is concentrated on a few individuals. Heterogeneous biting amplifies transmission between mosquitoes because those humans who are most frequently bitten are most likely to become infected, which leads to a greater expectation of infections in other mosquitoes25-29. Putting this relationship into context for understanding transmission dynamics and control, transmission is amplified by a factor that is related to the coefficient of variation of the distribution of these biting weights. When biting is heterogeneous and malaria is rare, correlations in mosquito feeding on humans who are most likely to have been infected amplify transmission so that the criterion for invasion is 3. The effect size of heterogeneity on invasion is thus equal to 1 +, where is the squared coefficient of variation of biting rates.

For heterogeneity to amplify transmission, differences in biting rates must persist, so it is not the overall shape of the biting distribution that matters, but average differences among individuals in their biting rates over the infectious period. For super-spreaders and super-spreading, the relevant quantity is the average ratio for each household of the observed count to the expected site-wide value for that day in a population, what we have called household biting weights (denoted ). In our notation, is simply the variance of the distribution of biting weights, so transmission would be amplified by a factor 3. We fit Gamma distributions to the estimated household biting weights at each site using MLE, , which were thus constrained to have a mean of one and both variance and coefficient of variation (Fig. 3a). Using the estimated distributions for biting weights, heterogeneous biting alone would amplify transmission during an invasion by a factor of 2.2 in Jinja, by 2.5 in Kanungu, and by 1.5 in Tororo. Notably, the basic reproductive number increases with human population density and declines with heterogeneity3, so the general rule for amplification may not apply to Tororo, which has rural populations with the highest transmission intensities.

**Macintosh HD:Users:dave:Dropbox:Work:Manuscripts:HeteroPowerLaws:SuperspreadingFigures:SuperspreadingFigure4.pdfFig. 3.** (a) Distribution of household biting weights (i.e., the points are the fitted values from the fitting procedure) and a gamma distribution fitted to describe the points using MLE (solid lines, plotted to the 99th quantile) for Jinja, ; Kanungu, ; Tororo, . (b) Proportion of variance explained by household biting weights (, seasonality in the HBR (), and environmental stochasticity, measurement errors, and sampling noise.

# Discussion

Though many studies have examined the properties of mosquito counts within a site over time30, this study describes, for the first time, the distribution of relative entomological risks for malaria transmission among households at multiple sites across a range of mosquito densities and transmission levels and through multiple seasons. We have shown that some of the observed heterogeneity is predictably associated with individual households over time. In this study, it was not generally true that 20% of the houses had 80% of the mosquitoes. While the variance of these distributions did increase with the mean, the shape of the distributions all suggested biting would become more evenly distributed as the intensity increased. The Pareto fraction varied from 20-90 (i.e., 20% of houses had 90% of mosquitoes) at low intensity to only 20-50 at high intensity. One ecological explanation for this finding is that when mosquito numbers are low it is likely because there are few breeding sites and thus even fewer close to settlements. An important question is how well these relationships hold across vector species and the broad range of ecological settings that characterize transmission of malaria and other vector-borne diseases31,32. It should be noted, however, that the study directly tested the infectivity of up to 50 mosquitoes from each household per night, but that we relied on models to quantify higher mosquito numbers in those samples. As a result of that modelling, our analysis may not have accurately captured outlier occurrence in areas with higher mosquito density.

Our study found strong support for super-spreading, but weaker support for the role of super-spreader households in malaria transmission. Some of the variability in biting was associated with particular households over the study, but our analysis also shows that environmental stochasticity accounts for approximately half of the variance; despite marked differences among households in their propensity to have mosquitoes, the identities of the houses receiving the most bites changed over time due to other factors, broadly lumped together under the category of environmental stochasticity, though some of the variance is clearly associated with the sampling process and some of it could be due to measurement errors. The main sources of environmental stochasticity have not been quantified but could include heavy rain and high winds at night, changes in wind direction and strength when mosquitoes are searching for a blood meal, breeding sites appearing and disappearing around the study sites, and pulses of mosquitoes emerging from water bodies. The quantity relevant for malaria transmission is the fraction of bites received by houses in the top quintile of biting weights. Our study suggests the Pareto fraction for super-spreaders is closer to 20-50, varying from 20-60 to 20-40. Though these are less heterogeneous than the proposed 20-80 rule, household heterogeneity would still have quantitative effects on transmission, invasion, exposure and immunity, the design of malaria studies, and the prospects for targeting. Our analysis suggests that all these effects would be strongest at low transmission intensity.

Translating general information about heterogeneity into a method for targeting would require finding covariates that would identify households with high biting weights and extrapolating these causes to establish relationships that could be used to target other households in the region. Some of this analysis has already been done for these sites; studies have found significant associations at some sites between the degree of urbanicity33, elevation, household quality or ease of household entering34, the number and age distribution of people living in a household, distance from water, and the enhanced vegetation index33. Collectively, these studies have established an evidence base and reasonable expectations about the gains in efficiency for targeting.

Since malaria immunity develops poorly and is only partially protective, and since most immune markers tend to be correlated with both exposure and immunity, immunity to malaria may be spuriously associated with disease14-16. These Pareto rules, though not directly relevant for epidemiological studies of malaria or for disentangling cause and effect because they describe overall patterns without identifying who in a population is most exposed, can be used to inform and improve the design of epidemiological studies across the spectrum of transmission. Household biting weights, which do convey information about relative rates of exposure, can be estimated with some effort and when considered along with environmental and sampling noise, provide a more accurate measure of individual exposure that can be used to validate other metrics of exposure or to disentangle cause and effect in studies of the protective effects of malaria immunity. Household biting weights are thus potentially useful for the proper design of epidemiological studies of malaria transmission, though our study suggests they are .

These ideas could be applied to other infectious disease systems. Observed heterogeneity is probably caused by multiple factors, and while we found important effects of seasonality and environmental stochasticity, other concerns may be important for pathogens with other modes of transmission. Heterogeneity is a basic feature of infectious disease transmission, but without some quantitative understanding of heterogeneity across the spectrum of transmission intensity, there is only a weak basis for understanding its consequences on pathogen transmission dynamics or control. Our analysis provides some basis for including heterogeneous biting in mechanistic models of mosquito-borne pathogens; for proper study designs and analysis of epidemiological data; and for evaluating the prospects of improving disease control through targeting. These other factors must be considered in understanding the consequences of heterogeneous biting for malaria and heterogeneous transmission, and more generally for infectious disease epidemiology, transmission dynamics, and control.

# Methods

We examined heterogeneous transmission of malaria in a recent study at three sites with markedly different levels of malaria transmission in Uganda (Figs 1, 2); *Anopheles gambiae* s.l. and *An. funestus* s.l. are the dominant vectors, and a substantial fraction of exposure occurs at night in or around the home23,33,35. The study followed 330 houses in three locations monthly for 3.5 years. All children 0.5-10 years of age from each household were enrolled and followed longitudinally over the whole course of the study or until reaching the age of 11. All household members were given a long-lasting insecticidal net at the time of enrolment; bed net usage was high (98%), long-lasting, and consistent across study sites. Mosquitoes were caught in a Centers for Disease Control (CDC) light trap placed next to the bed in a room where at least one child participant was sleeping. The trap was placed indoors in the evening and retrieved the next morning, and mosquitoes were sorted, counted, and up to 50 per household were tested for sporozoites using an enzyme-linked immunosorbent assay (ELISA)23. Counts by CDC light traps were compared with other trapping methods and were highly correlated, albeit with different totals23. Details of this study, including study sites, detailed methods, the overall study design, and ethical approval have been described elsewhere23,35.

## Analysis

The HBR and the EIR are entomological measures of malaria exposure: the number of bites by vector mosquitoes and by infectious vector mosquitoes, per person, per day. Here, we take the CDC light trap counts of mosquitoes and infectious mosquitoes as our estimated HBR and EIR.

To compute the EIR, the HBR is multiplied by the sporozoite rate (SR). For households, the SR was a complicated statistic because a maximum of 50 mosquitoes were examined for sporozoites from houses where more than 50 mosquitoes were counted in a night. For such households, EIR data were thus computed using the observed HBR multiplied by the SR from the sample of 50. To average over the sampling stochasticity, the SR was also calculated as a site-wide average for each month. For the period of October 2013 to March 2014, for which there were low overall mosquito counts and no sporozoite-positive mosquitoes, SR is estimated as the average of the two previous years: for example, October 2011 and 2012 were averaged to give an estimate for the October 2013 SR.

As a first measure of heterogeneous biting, we use the Pareto fraction for super-spreading, the proportion of all biting received by the most exposed quintile of the population. This fraction is an intuitive measure of dispersion and a simple way of describing the potential gains in efficiency if malaria control were targeted. The Pareto fraction for super-spreading for a given month is defined as the proportion of the total HBR or EIR received by the 20% of households with the highest HBR or EIR in that month. The Pareto fraction for super-spreaders is defined as the proportion of the total HBR or EIR received by the 20% of households with the highest biting weights (, see model description). The key difference between the Pareto fractions for super-spreading and for super-spreaders is that the latter measures biting by households that were consistently bitten most across the whole study, not just for the month.

As a second measure of heterogeneous biting, we compute the Pareto index, ….

To identify super-spreaders, we also conducted a statistical analysis of the data to quantify seasonality, household biting, and the residual errors. Mosquito counts for the household on sampling day at site are assumed to arise from a zero-inflated negative binomial (ZINB) distribution,

where is a shape parameter quantifying the amount of overdispersion. Note that is a function of , , and . The parameters and can be modeled as a function of a set of explanatory variables. For it is common to use a logistic regression with a logit link function, as it describes a binomial process. A log link function is used to model the dependence of on a different (or same) set of covariates. The log link function ensures that the estimated will not be negative, regardless of parameter values. In our case, we model the dependence of and on the same set of explanatory variables.

,

Here denotes the household at site ; quantifies the effects of household biting of the household at site ; denotes sampling day at site which is modeled as random effects; and the random effects are environmental stochasticity and sampling errors (a random variable that is Gaussian independently and identically distributed). The additional variation among mosquito counts that is not accounted for by the covariates is modeled via , which in this model was gamma distributed. The available time points (sampling days, ) are modeled as temporally structured random effects, ensuring that contiguous periods are likely to be similar, but allowing for flexible shapes in the evolution curve. First- and second-order random walks, and autoregressive processes of order 1 and order 2 are considered for modeling 36. The regression models are implemented in an integrated nested Laplace approximation framework via the R-INLA package37. For convenience and consistency, default prior specifications in R-INLA have been chosen for each of the prior distributions. A related simulation study was conducted to identify methods that could accurately estimate heterogeneity 38. This method – analysis using a ZINB model39 – was shown to give accurate estimates of household biting weights, seasonal patterns, and environmental stochasticity. The output of this analysis was a set of estimated quantities describing a household biting weight for each house at each site, , and a daily, site-specific expected number of bites by mosquitoes (i.e. the HBR), . In Figure 2a, the reported values are the result of an MLE analysis for a gamma distribution constrained to have a mean of 1. A table for each site containing a row for each observation: the count, the computed expected number of bites for that day, the biting weight for the individual is available for download (<https://figshare.com/articles/Uganda_environmental_covariates_csv/6797408/3>) .

# Data availability

The primary data generated and analysed during the current study are available at figshare (<https://doi.org/10.6084/m9.figshare.6797408.v3>). Other data from the same study can be found in the Clinical Epidemiology Database Resources repository (<https://clinepidb.org/ce/app>).

# Code availability

All code used for these analyses was also used for a previous paper. It is publicly available online at <https://gatesopenresearch.s3.amazonaws.com/supplementary/12838/f7b1ea69-1fb5-46e5-98d8-5601c4d6f731.R>.

# References

1. Woolhouse, M. E. J. *et al.* Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *PNAS* **94,** 338–342 (1997).

2. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438,** 355–359 (2005).

3. Smith, D. L., McKenzie, F. E., Snow, R. W. & Hay, S. I. Revisiting the basic reproductive number for malaria and its implications for malaria control. *PloS Biol.* **5,** e42 (2007).

4. Bousema, T. J., Griffin, J. T., Sauerwein, R. W. & Smith, D. L. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med* (2012). doi:10.1371/journal.pmed.1001165.t001

5. Bousema, T. J. *et al.* Identification of hot spots of malaria transmission for targeted malaria control. *J. Infect Dis* **201,** 1764–1774 (2010).

6. Bejon, P. *et al.* Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med* **7,** e1000304 (2010).

7. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control*. (Oxford University Press, 1991).

8. Perkins, T. A., Scott, T. W., Le Menach, A. & Smith, D. L. Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission. *PLoS Comput Biol* **9,** e1003327 (2013).

9. Keeling, M. J. & Grenfell, B. T. Individual-based perspectives on R(0). *J. Theor. Biol.* **203,** 51–61 (2000).

10. Smith, D. L., Drakeley, C. J., Chiyaka, C. & Hay, S. I. A quantitative analysis of transmission efficiency versus intensity for malaria. *Nature Comm* **1,** 108 (2010).

11. Smith, D. L., Dushoff, J., Snow, R. W. & Hay, S. I. The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* **438,** 492–495 (2005).

12. Tusting, L. S., Bousema, T. J., Smith, D. L. & Drakeley, C. J. Measuring changes in *Plasmodium falciparum* transmission: precision, accuracy and costs of metrics. *Adv. Parasitol.* **84,** 151–208 (2014).

13. Carneiro, I. *et al.* Age-patterns of malaria vary with severity, transmission intensity and seasonality in sub-Saharan Africa: a systematic review and pooled analysis. *PLoS ONE* **5,** e8988 (2010).

14. Bejon, P. *et al.* Effect of the pre-erythrocytic candidate malaria vaccine RTS,S/AS01E on blood stage immunity in young children. *J. Infect Dis* **204,** 9–18 (2011).

15. Bousema, T. J., Kreuels, B. & Gosling, R. Adjusting for heterogeneity of malaria transmission in longitudinal studies. *J. Infect Dis* **204,** 1–3 (2011).

16. Greenhouse, B. *et al.* Antibodies to *Plasmodium falciparum* antigens predict a higher risk of malaria but protection from symptoms once parasitemic. *J. Infect Dis* **204,** 19–26 (2011).

17. Guelbeogo, W. M. *et al.* Variation in natural exposure to anopheles mosquitoes and its effects on malaria transmission. *eLife* **7,** (2018).

18. Michael, E. *et al.* Quantifying mosquito biting patterns on humans by DNA fingerprinting of bloodmeals. *Am. J. Trop. Med. Hyg.* **65,** 722–728 (2001).

19. De Benedictis, J. *et al.* Identification of the people from whom engorged *Aedes aegypti* took blood meals in Florida, Puerto Rico, using polymerase chain reaction-based DNA profiling. *Am. J. Trop. Med. Hyg.* **68,** 437–446 (2003).

20. Scott, T. W., Githeko, A. K., Fleisher, A., Harrington, L. C. & Yan, G. DNA profiling of human blood in anophelines from lowland and highland sites in western Kenya. *Am. J. Trop. Med. Hyg.* **75,** 231–237 (2006).

21. Smith, T. A. Estimation of heterogeneity in malaria transmission by stochastic modelling of apparent deviations from mass action kinetics. *Malaria J* **7,** 12 (2008).

22. Gonçalves, B. P. *et al.* Examining the human infectious reservoir for Plasmodium falciparum malaria in areas of differing transmission intensity. *Nature Comm* 1–11 (2017). doi:10.1038/s41467-017-01270-4

23. Kilama, M. *et al.* Estimating the annual entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles gambiae s.l.* using three sampling methods in three sites in Uganda. *Malaria J* **13,** 111 (2014).

24. Katureebe, A. *et al.* Measures of malaria burden after long-lasting insecticidal net distribution and indoor residual spraying at three sites in Uganda: a prospective observational study. *PLoS Med* **13,** e1002167 (2016).

25. Dietz, K. in *Vito Volterra Symposium on Mathematical Models in biology* (eds. Barigozzi, C. & Levin, S. A.) 264–277 (Springer-Verlag, 1980).

26. Dietz, K. in *Malaria* (eds. Wernsdorfer, W. & McGregor, I. A.) 1091–1133 (Churchill Livingstone, 1988).

27. Dye, C. & Hasibeder, G. Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others. *Trans R Soc Trop Med Hyg* **80,** 69–77 (1986).

28. Lloyd, A. L., Zhang, J. & Root, A. M. Stochasticity and heterogeneity in host-vector models. *J. R. Soc. Interface* **4,** 851–863 (2007).

29. Smith, D. L., Dushoff, J. & McKenzie, F. E. The risk of a mosquito-borne infection in a heterogeneous environment. *PloS Biol.* **2,** e368 (2004).

30. Lindsay, S. W. *et al.* Exposure of Gambian children to *Anopheles gambiae* malaria vectors in an irrigated rice production area. *Med Vet Entomol* **9,** 50–58 (1995).

31. Bisanzio, D. *et al.* Modeling the spread of vector-borne diseases on bipartite networks. *PLoS ONE* **5,** e13796 (2010).

32. Ferreri, L. *et al.* Pattern of tick aggregation on mice: larger than expected distribution tail enhances the spread of tick-borne pathogens. *PLoS Comput Biol* **10,** e1003931 (2014).

33. Kigozi, S. P. *et al.* Associations between urbanicity and malaria at local scales in Uganda. *Malaria J* **14,** 374 (2015).

34. Wanzirah, H. *et al.* Mind the gap: house structure and the risk of malaria in Uganda. *PLoS ONE* **10,** e0117396 (2015).

35. Kamya, M. R. *et al.* Malaria transmission, infection, and disease at three sites with varied transmission intensity in Uganda: implications for malaria control. *Am. J. Trop. Med. Hyg.* **92,** 903–912 (2015).

36. Sørbye, S. H. & Rue, H. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics Miami* **8,** 39–51 (2014).

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

38. Kang, S. Y. *et al.* Heterogeneous exposure and hotspots for malaria vectors at three study sites in Uganda. *Gates Open Res* **2,** 32 (2018).

39. Yau, K. K. W., Wang, K. & Lee, A. H. Zero-Inflated Negative Binomial Mixed Regression Modeling of Over-Dispersed Count Data with Extra Zeros. *Biom. J.* **45,** 437–452 (2003).

# Acknowledgments

We would like to thank those in the study communities who participated in this study or helped collect mosquitoes. Funding for the research and support for KM, SWL, GD, MRK, CD, and DLS came from National Institutes of Health as part of the International Centers of Excellence in Malaria Research (ICMER) program (U19AI089674). The authors wish to acknowledge the Infectious Diseases Research Collaboration (IDRC) for administrative and technical support. BTG, DLS, PWG, SYK, SWL, DB, and SIH receive support from the Bill and Melinda Gates Foundation (105338, OPP1068048, OPP110495, OPP1106023, OPP1091919). PWG is a Career Development Fellow (K00669X) jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, also part of the EDCTP2 programme supported by the European Union. SIH is funded by a Senior Research Fellowship from the Wellcome Trust (095066).

# Author contributions

BG, CD, DLS, EA, GD, KM, MRK, SGS, and SWL designed and conducted the linked epidemiological and entomological PRISM studies. BTG, DLS, LVC, and SYK conceived and designed the analysis. DLS, LVC, and SYK participated in the management and analysis of the data. DLS, LVC, and SYK wrote the original draft of the manuscript. All authors participated in the writing of the final manuscript. All authors read and approved the final manuscript.

# Competing interests

The authors declare no competing interests.

# Figure legends and tables

**Fig. 1. Temporal trends in average daily mosquito exposure**, , for each site – (a) Jinja, (b) Kanungu, (c) Tororo – plotted for each one of the three study sites along with credible intervals (dashed). Seasonal signals for study sites with lower transmission have been plotted for Kanungu and Tororo to show the scale of the differences in the mosquito counts among sites.

**Fig. 2. The counts and Pareto analysis for all Anophelines (top row) and for infected Anophelines (bottom row).** (a) Anopheline mosquito catch counts by month and household (each household is on one line), sorted within each site by the median counts for each household. Darker colors indicate higher counts. (b) Monthly Pareto (solid line, circles) and modified Pareto (dashed, x’s) indices by mean monthly HBR. Linear fit line for Pareto and modified Pareto are shown with confidence intervals in grey. (c) Sporozoite-positive anopheline mosquito catch counts by month and household (each household is on one line), sorted within each site by the median counts for each household. Darker colors indicate higher counts. (d) Monthly Pareto (solid line, circles) and modified Pareto (dashed line, x’s) indices by mean monthly EIR. Linear fit line for Pareto and modified Pareto are shown with confidence intervals in grey.

**Fig. 3. (a)** Distribution of household biting weights (i.e., the points are the fitted values from the fitting procedure) and a gamma distribution fitted to describe the points (solid lines, plotted to the 99th quantile) for Jinja, ; Kanungu, ; Tororo, . **(b)** Proportion of variance explained by household biting weights (, seasonality (), and environmental and sampling noise.