CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Simplified binomial estimation of human malaria transmission exposure distributions based on hard classification of where and when mosquitoes are caught: statistical applications with off-the-shelf tools
Authors	Killeen, Gerry F.;Monroe, April;Govella, Nicodem J.
Publication date	2021-08-01
Original Citation	Killeen, G. F., Monroe, A. and Govella, N. J. (2021) 'Simplified binomial estimation of human malaria transmission exposure distributions based on hard classification of where and when mosquitoes are caught: statistical applications with off-the-shelf tools', Parasites & Vectors, 14 (1), 314, (8pp). doi: 10.1186/s13071-021-04884-2
Type of publication	Article (peer-reviewed)
Link to publisher's version	10.1186/s13071-021-04884-2
Rights	© The Author(s) 2021. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data. - https://creativecommons.org/licenses/by/4.0/
Download date	2026-10-09 11:42:56

Item downloaded from	https://hdl.handle.net/10468/13301

Parasites & Vectors

REVIEW Open Access

Simplified binomial estimation of human malaria transmission exposure distributions based on hard classification of where and when mosquitoes are caught: statistical applications with off-the-shelf tools

Gerry F. Killeen[1*], April Monroe[2] and Nicodem J. Govella[3,4]

Abstract

The impacts and limitations of personal protection measures against exposure to vectors of malaria and other mosquito-borne pathogens depend on behavioural interactions between humans and mosquitoes. Therefore, understanding where and when they overlap in time and space is critical. Commonly used approaches for calculating behaviour-adjusted estimates of human exposure distribution deliberately use *soft classification* of where and when people spend their time, to yield nuanced and representative distributions of mean exposure to mosquito bites across entire human populations or population groups. However, these weighted averages rely on aggregating individual-level data to obtain mean human population distributions across the relevant behavioural classes for each time increment, so they cannot be used to test for variation between individuals. Also, these summary outcomes are quite complex functions of the disaggregated data, so they do not match the standard binomial or count distributions to which routine off-the-shelf statistical tools may be confidently applied. Fortunately, the proportions of exposure to mosquito bites that occur while indoors or asleep can also be estimated in a simple binomial fashion, based on *hard classification* of human location over a given time increment, as being either completely indoors or completely outdoors. This simplified binomial approach allows convenient analysis with standard off-the-shelf logistic regression tools, to statistically assess variations between individual humans, human population subsets or vector species. Such simplified binomial estimates of behavioural interactions between humans and mosquitoes should be more widely used for estimating confidence intervals around means of these indicators, comparing different vector populations and human population groups, and assessing the influence of individual behaviour on exposure patterns and infection risk. Also, standard sample size estimation techniques may be readily used to estimate necessary minimum experimental scales and data collection targets for field studies recording these indicators as key outcomes. Sample size calculations for field studies should allow for natural geographic variation and seasonality, taking advantage of rolling cross-sectional designs to survey and re-survey large numbers of separate study locations in a logistically feasible manner.

*Correspondence: gerard.killeen@ucc.ie
[1] School of Biological, Earth & Environmental Sciences and Environmental Research Institute, University College Cork, Cork, Republic of Ireland
Full list of author information is available at the end of the article

© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Keywords: Malaria, *Plasmodium*, Arbovirus, Lymphatic filariasis, Mosquito, *Anopheles*, *Culex*, *Aedes*, *Mansonia*, Human–vector interaction, Behaviour

Background

Assessments of malaria control measures that target human-biting adult mosquitoes require clear understanding of how effectively they protect individual end users and how gaps in personal protection arise [1–4]. Improved understanding of behavioural interactions between humans and mosquitoes, especially where and when they overlap in time and space, is critical to estimating the impact of personal protection measures such as insecticide-treated nets (ITNs) and identification of where and when supplementary vector control tools are needed [1–4]. Similar issues are relevant to personal protection against a variety of other vector-borne pathogens, especially arboviruses such as dengue, Zika and chikungunya carried by mosquito species like *Aedes aegypti* that often bite people while they are active outdoors [5].

While methods for weighting estimates of human exposure to bites from *Anopheles* mosquitoes according to where people spend their time have been available for decades [6], they remain underutilized [1–4, 7]. Crude indoor and outdoor biting rates are still commonly misinterpreted as being indicative of human-vector contact patterns, without any adjustment for the influence of human behaviour or the personal protection effects of ITNs [1–4, 7]. Fortunately, these deficits in current common practice may be readily addressed by relatively straightforward adjustments to data collection and analysis practice [4, 7]. More accurate representations of exposure patterns can be achieved by supplementing mosquito biting activity data with complementary surveys of a small set of human behavioural variables that capture the following over the course of the night: (i) the distribution of human populations indoors and outdoors, (ii) whether they are awake or asleep and (iii) if and when they use an ITN [4, 7]. Important examples of useful indicators that can be calculated with such behaviour-weighted approaches include the proportion of vector bites occurring indoors for an unprotected individual (π_I) and the proportion of vector bites occurring while asleep for an unprotected individual (π_S), as well as derived terms like the proportion of all vector bites directly prevented by ITN use or the proportion occurring indoors despite ITN use [4, 7]. If surveyed, estimated and interpreted consistently, these indicators can greatly improve understanding of how malaria transmission persists despite high coverage of prevention measures such as ITNs, how exposure patterns may change as supplementary vector control tools are introduced, and the potential impacts of these new tools [1–4, 7].

The most widely used approaches for calculating behaviour-adjusted estimates of human exposure distribution deliberately use *probabilistic* or *soft classification* to allow for the considerable variability between individual people in terms of where and how they spend their time [4, 7]. The probabilities that any given individual will be indoors or outdoors during a given time increment is estimated as the proportion recorded as doing so through questionnaire or observational surveys of the human population [4, 7]. These probabilities are then used to weight entomological measurements of human exposure to mosquito bites occurring indoors and outdoors, yielding nuanced and representative distributions of mean exposure to biting mosquitoes across entire human populations or population groups [4]. However, these weighted averages rely on aggregating individual-level data to obtain mean human population distributions across the relevant behavioural classes for each time increment. These summary outcomes are therefore quite complex functions of the disaggregated data, so they do not match the standard binomial or count distributions to which routine off-the-shelf statistical tools may be confidently applied. Consequently, testing for variation between individuals, much of which may be associated with epidemiologically important covariates such as age, sex, occupation and housing [7–14], requires advanced Bayesian techniques that are beyond the reach of most field entomologists and epidemiologists.

Fortunately, the proportions of exposure to mosquito bites that occur indoors or while asleep can also be estimated in a more simplified binomial fashion, based on *hard classification* of human location at a given time increment, as being either completely indoors or completely outdoors [15]. Such clear-cut assignment of humans to either location compartment then allows each mosquito caught attacking a person to be simplistically and unambiguously assigned to one of three categories, on the basis that it was either: (i) caught at a time and place when most people may practically protect themselves by using an ITN, (ii) caught at a time and place when people cannot practically use an ITN or (iii) caught at a time and place assumed to be irrelevant to normal exposure patterns because the majority of people are elsewhere (Fig. 1). This simplification is obviously cruder and less precise because it disregards

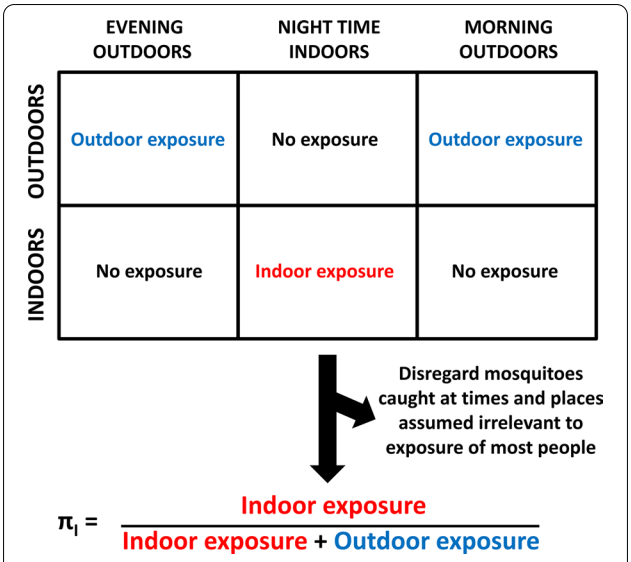

Fig. 1 A schematic illustration of how the proportions of human exposure to mosquitoes occurring indoors may be estimated as a simple binomial indicator based on hard classification of where human individuals and populations spend their time [16–21]

many valid observations of mosquito-human interactions that occur in the evenings and mornings, when some people are asleep indoors while others are awake outdoors.

However, it does offer the advantage of allowing convenient analysis with standard logistic regression methods, to statistically estimate confidence intervals around means, compare vector species and human population groups and assess the influence of individual behaviour on exposure patterns and malaria risk [16–21] (Additional file 1: Poster S1). Also, because such hard classification techniques allow these key indicators to be calculated in simple binomial form, standard sample size estimation techniques may be readily used to estimate necessary minimum experimental scales and data collection targets required for field studies. While crude hard classification understandably appears to provide less precise estimates, no consistent trend toward under- or overestimation was obvious in a multi-country study encompassing ten different mosquito populations from across Africa [9]. Furthermore, hard and soft classification approaches have different advantages and disadvantages, so both may be included for complementary purposes in a given report [9].

Here, we present a brief explanation and discussion of relatively simple methodological options for applying such hard classification techniques to address common objectives in epidemiological entomology [22, 23]. We place particular emphasis on relatively simple approaches and tools that are accessible to numerate entomologists familiar with widely available, off-the-shelf analytical tools.

Simplified binomial estimation of human exposure distributions based on hard classification of where and when mosquitoes are caught

Simplified binomial estimation of the proportion of exposure of unprotected individuals to mosquito bites that occurs indoors requires hard classification of human behaviours (Box 1), assuming the night is split into distinct periods, during which all exposure is assumed to occur either entirely indoors or entirely outdoors (Fig. 1). Here, we rely on examples relevant to malaria transmission by predominantly nocturnal vectors, so we use the term *night* to inclusively denote all times at which such *Anopheles* mosquitoes are active, even if that includes daylight hours before dusk and after dawn. For more diurnal vectors of other pathogens such as dengue, chikungunya and Zika, the term *day* may be more appropriate and can include all 24 h of the daily cycle [15] whenever relevant.

The calculations described in Box 1 have already featured in several published applications [15–21] that readers may draw on as illustrative examples when adapting these techniques to their own needs. In the first of these examples [15], it was demonstrated that most human exposure to *Anopheles funestus* and *An. quadriannulatus* in southeastern Zambia occurred indoors in the absence of a protective ITN ($\pi_{I,u} \geq 0.97$), because these mosquito species are most active at times of the night when humans are indoors ($P_{FL,I} \geq 0.90$). Note, however, that if one only accounts for the behaviour of the mosquitoes, as observed attacking participants in human landing catches who artificially spend equal amounts of time indoors and outdoors for the purpose of such experiments, these two mosquito populations had no apparent preference for feeding indoors ($P_I \approx 0.5$) [15]. Subsequent pooled analysis of similar data from distinct sites scattered across Africa revealed similar patterns [16], confirming that the most important vectors on the continent are not innately endophagic in the strict sense but rather highly nocturnal. Therefore, it was concluded that high proportions of human exposure to malaria transmission have historically occurred indoors because that is where most people sleep at night [1, 3, 23, 25]. Further applications of such simplified binary formats for behavioural interaction indicators include demonstrating statistically significant changes in human exposure distributions following scale up of ITNs [17, 18, 21].

Beyond comparing vector species and human population groups, these simplified binary behavioural interaction indicators (Box 1) can also be used to assess the

influence of variable, often idiosyncratic individual behaviours on exposure patterns and malaria risk [16–21]. Whenever possible, human behaviour data should therefore be collected in a disaggregated format that is linkable to individual human study participants, so that the epidemiological importance of differences between individuals and population subgroups can be formally assessed [4] with statistical contrasts using standard logistic regression models and off-the-shelf software [19, 20]. For ease of application to readers own data sets, a Microsoft Excel® template for calculating individual-level estimates of π_I and π_S is provided and populated with a sample of anonymized data from previous large-scale cross-sectional surveys [20] for illustrative purposes (Additional file 2: Dataset S1). The additional insight that may be obtained from disaggregated data with such individual-level calculations is exemplified by comparing the in-depth epidemiological analyses reported along with the original version of this template [20] with the much broader, population-wide mean overview obtained through the preceding entomological report [26].

Box 1 Mathematical description of how simplified binomial estimates of human exposure distributions may be calculated based on hard classification of where and when mosquitoes are caught

First, one must divide the night, or entire 24-h day if needs be, into hour-long survey time intervals ($t = 0$, 1, 2, 3…23). These hour-long intervals are then conceptually grouped into three distinct periods, during which all exposure is assumed to occur either entirely indoors or entirely outdoors (Fig. 1). Note, however, that even previous surveys of where and when people spent their time in the same context may not necessarily remain fully representative going forward, so it may be unwise to assume population-wide mean cut-off points that separate these three distinct periods *a priori*. Also note that one of the most important applications of this approach is to enable analysis of the behavioural variations between individuals in a population, so it is essential to collect both human and mosquito behaviour data in fully disaggregated form at time intervals of ≤ 1 h in duration. Once the data have been collected in appropriately disaggregated hour-by-hour format, the middle period of the night when all exposure is assumed to occur while asleep indoors (Fig. 1) is defined as beginning at the first (F, I) and ending at the last (L, I) of the surveyed hour-long time (t) intervals when either (i) a *human individual* reported being indoors ($I_t = 1$) or (ii) the *majority of a human population* were indoors ($I_t > 0.5$), because they reported that they had already entered their houses for the night and had not yet left for the day ($F, I \le t \le L, I$). The remaining periods of the night, before ($t < F, I$) and after ($t > L, I$) this interval, correspond to periods when either (i) an individual or (ii) most people in a population are outdoors ($I_t < 0.5$). Correspondingly, the proportion of human exposure for unprotected (u) non-users of ITNs that occurs indoors ($\pi_{I,u}$) may be approximately calculated by dividing the number of mosquitoes caught indoors during the period that most people are indoors (N_I) by itself plus the number of mosquitoes caught outdoors (N_o) outside of that period [15, 17, 18]:

$$\pi_{I,u} = \sum_{t=F,I}^{L,I} [N_{I,t}] / \left(\sum_{t=0}^{F,I-1} [N_{O,t}] + \sum_{t=F,I}^{L,I} [N_{I,t}] + \sum_{t=L,I+1}^{23} [N_{O,t}] \right) \quad (1)$$

The equivalent binomial calculations may also be made for the proportion of exposure to mosquito bites of unprotected individuals which occurs while asleep ($\pi_{S,u}$), using the first (F, S) and last (L, S) hour-long intervals when either (i) an individual ($S_t = 1$) or (11) most people in a population were asleep ($S_t > 0.5$):

$$\pi_{S,u} = \sum_{t=F,S}^{L,S} [N_{I,t}] / \left(\sum_{t=0}^{F,S-1} [N_{O,t}] + \sum_{t=F,S}^{L,S} [N_{I,t}] + \sum_{t=L,S+1}^{23} [N_{O,t}] \right) \quad (2)$$

To more clearly interpret the $\pi_{I,u}$ and $\pi_{S,u}$ estimates obtained [16], the two following underlying determinants of these outcomes may also be calculated [15, 17, 18]. The propensity of vectors to feed indoors is reflected in the proportion of mosquitoes captured indoors (P_I):

$$P_I = \sum_{t=0}^{23} [N_{I,t}] / \sum_{t=0}^{23} [N_{I,t} + N_{O,t}] \quad (3)$$

The propensity of vectors to feed at times when people are indoors is reflected in the proportion of all mosquitoes caught that were captured during hours when the majority of people were indoors ($P_{FL,I}$):

$$P_{FL,I} = \sum_{t=F,I}^{L,I} [N_{I,t} + N_{O,t}] / \sum_{t=0}^{23} [N_{I,t} + N_{O,t}] \quad (4)$$

Similar calculations can be made for the propensity of vectors to feed at times when people are asleep ($P_{FL,S}$), which may paint a more nuanced and accurate picture of the underlying behavioural drivers of human exposure distribution in some contexts [20, 24]:

$$P_{FL,S} = \sum_{t=F,S}^{L,S} [N_{I,t} + N_{O,t}] / \sum_{t=0}^{23} [N_{I,t} + N_{O,t}] \quad (5)$$

All these indicators of propensity of vectors to feed indoors (P_I) and during the night time hours predominantly spent indoors ($P_{FL,I}$, referred to as nocturnality [17] or nocturnal biting [18] in previous publications) or asleep ($P_{FL,S}$) may be formally tested for vector preference (P_I,$P_{FL,I}$ or $P_{FL} \neq 0.5$) in terms of the significance of the differences of these estimates from the null hypothesis (P_i,$P_{FL,I}$ or $P_{FL} = 0.5$).

Simplified sample size calculations for studies surveying the proportions of human exposure to malaria vectors occurring indoors or while asleep

Because these hard classification techniques allow the proportions of human biting exposure occurring indoors ($\pi_{I,u}$) and while asleep ($\pi_{S,u}$) to be calculated in simple binomial form, it is also possible to apply standard sample size estimation techniques [27] to plan necessary minimum experimental scales and data collection targets for field studies (Fig. 2). To enable sample size estimation with the most intuitive and readily accessible statistical tools, here an example is provided (Box 2) that adapts well-established simple deterministic statistical models, originally formulated for cluster-randomized trials with disease infection prevalence as the primary binomial outcome [27]. Of course, more advanced, simulation-based stochastic approaches may also be applied to sample size calculations for surveying these same binary indicators [28], and these more intricate techniques may be more rigorous and appropriate for teams with sufficient analytical capacity. However, relevant analytical capacity remains underdeveloped for malaria generally, and entomology in particular, in endemic tropical countries [3]. Therefore, the simple but widely accepted deterministic models [27] and calculation tools (Additional file 3: Dataset S2) used in Box 2 may represent an accessible and practical alternative for teams at an earlier stage of analytical capacity development.

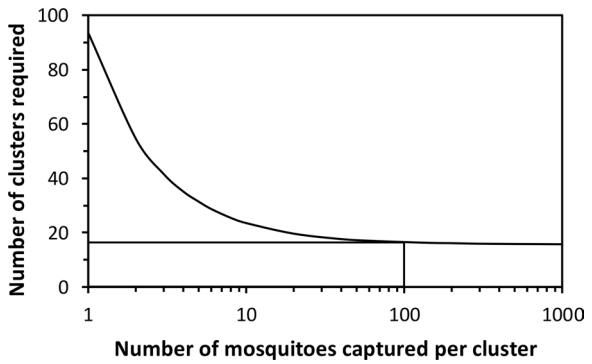

Fig. 2 The predicted relationship between the number of mosquitoes caught per cluster and the number of village-scale population clusters required to achieve 80% power if all other assumed parameters are exactly as described in Box 2. Note that in the field of epidemiology, the simple term *cluster* usually refers a geographically distinct, but often demographically defined, unit of observation that may be considered independent in the statistical sense [27]

Note that while many investigators may not need to implement surveys across such extensive scales, they should nevertheless allow for the fact that some natural variation of $\pi_{I,u}$ and $\pi_{S,u}$ is likely to occur among different sub-villages, villages or other geographic subunits/population clusters far enough apart to be considered independent units of observation. Furthermore, such natural spatial variation may be further exaggerated by natural seasonality, which may also be of interest in its own right. Therefore, sample size calculations should always allow for clustering and covariance within clusters, as in the example provided in Box 2. Note that it may not be logistically feasible to survey all different study locations simultaneously, especially on scales of entire districts or countries. Rolling cross-sectional designs that survey clusters sequentially (e.g. [14, 31]), rather than at the same time, may therefore be necessary to survey and resurvey the large numbers of distinct locations sometimes required to achieve sufficient power (Fig. 2).

Box 2 An example of a sample size calculation for a nationally representative survey of the proportion of human exposure to malaria vectors occurring indoors in the United Republic of Tanzania

A nationally representative survey of key mosquito and human behaviours is underway in Tanzania, for which one of the two key primary outcomes is the proportion of human exposure occurring indoors in the absence of an ITN ($\pi_{I,u}$). For practical logistics reasons, this very large-scale study has a rolling cross-sectional design, enabling each of the village-scale independent units of observation (referred to as *clusters* in epidemiological terminology [27]) to be feasibly surveyed in a sequential manner once per year. This is an observational study, so the groups of clusters being contrasted will be eco-epidemiological strata rather than the interventions groups allocated in an equivalent experimentally controlled trial. For example, one *a priori* hypothesis of this particular study is that high humidity may enable mosquitoes to feed earlier in the evening and later in the morning without becoming desiccated, leading to lower values of $\pi_{I,u}$ and attenuated impacts of ITNs. The simplest

sample size calculation that may be carried out therefore assumes that all the surveyed clusters will be ranked *post hoc* based on humidity measurements and then split into two quantiles, namely the most humid half of the surveyed clusters versus the least humid half. Based on the following assumed input parameters, and using Eq. 4 from Hayes and Bennett's widely cited article [27], the minimum number of village-scale survey clusters (c) required to achieve 80% power for contrasts between two such groups was estimated to be 16, assuming a minimum of 100 mosquitoes are caught in each village-scale cluster (Fig. 2, Additional file 3: Dataset S2) at times and places when most humans are exposed (Fig. 1). The calculations underpinning Fig. 2 and detailed in Additional file 3: Dataset S2 are explained as follows. For the outcome of interest, namely the proportion of human exposure occurring indoors ($\pi_{I,u}$), the following contrast between two previously surveyed settings was assumed to be representative of the level of variation that could be epidemiologically relevant, because persistent residual malaria transmission depends more directly on the relative size of gaps in *de facto* protective coverage ($1\text{-}C_p$, where $C_p = \pi_{I,u}\ C$ [29]) than upon protective coverage *per se* (C_p) [24]): One rural Tanzanian setting where 90% of exposure to *Anopheles gambiae* was estimated to occur indoors and another where 73% of exposure to the same species occurred indoors in an urban setting [26], so $\pi_{I,u,1} = 0.90$ and $\pi_{I,u,2} = 0.73$, respectively. While this may sound like a relatively minor difference, when looked at in terms of a potential gap in protective coverage ($1\text{-}C_p$) [24], it represents almost a tripling of the fraction of exposure occurring outdoors, from 10 to 27%. Based on experience and results from limited studies in Tanzania, plus those reported from several rural settings across Africa [16], only a modest coefficient of variation between clusters (k) of 0.20 was assumed. Additionally, a worst case scenario was assumed in terms of mosquito abundance and capture success, with the number of mosquitoes in each cluster (generally a village or sub-village in most epidemiological or entomological studies, and specifically the former in this particular study) collected at times and places when human exposure may occur (Fig. 1) assumed to be relatively low ($n = 100$ mosquitoes per cluster). As illustrated in Fig. 2 and explained as follows, a total of 16 clusters per stratum are expected to yield sufficient power to test for between-stratum differences in $\pi_{I,u}$ of the magnitude described above, even if far less than 100 mosquitoes are caught in some of the surveyed clusters.

Even with a tenfold increase beyond the within-cluster sample size target for this study, up to 1000 mosquitoes per village is expected to yield very little improvement in overall power (Fig. 2) because doing so does not increase true replication in the strict sense [27, 30]. This is because individual mosquitoes caught within a surveyed cluster merely represent binary observations on multiple samples from within that same cluster, rather than truly independent observations from distinct cluster-level replicates *per se*. Mosquitoes sampled from the same population in the same village, experiencing the same environmental conditions at the time when it is surveyed, are obviously expected to behave more similarly to each other than to mosquitoes of the same species collected from another village in a different part of the country, where the environmental conditions may be very different, especially if it is visited at a different time of the year. Unless such intra-cluster correlation between sampled observations is allowed for, *pseudoreplication* renders invalid any subsequent analyses that erroneously treat them as independent observations [27, 30]. The deterministic predictive models of Hayes and Bennett therefore conservatively account for the expected similarities between observed individuals in each cluster in the simplest fashion possible, with a single within-cluster variance parameter [27]. Based on the envisaged protocol, the number of mosquitoes expected per surveyed cluster should therefore be considered as a single aggregate total for parameterizing n in Additional file 3: Dataset S2, regardless of how many houses are to be sampled or how many nights they are sampled over. For binary outcomes with binomial distributions, remarkably small samples can yield quite precise estimates of the mean for any given unit of observation, so even as few as 100 mosquitoes per cluster may be expected to achieve essentially the maximum possible power, thus minimizing cluster replication requirements (Fig. 2). Recognizing that excessive within-cluster sampling, to capture hundreds or even thousands of mosquitoes per village, adds negligible power to appropriate statistical tests (Fig. 2) also has encouraging implications for the robustness of the study design to seasonal or location-specific scarcities of mosquitoes: Even if < 100 mosquitoes are caught in each cluster, this is expected to have only a modest effect on the power of the study, so long as at least ten are caught (Fig. 2) in the times and places that matter (Fig. 1). Thus, even if this study falls short of this *a priori* target of 100 mosquitoes in some of the villages surveyed, this is not a major cause for concern or for alteration of the study protocol.

Conclusions

Simplified binomial estimates of the proportions of exposure to mosquito bites that occur indoors or while asleep, based on *hard classification* of human location at a given time, allow convenient statistical analysis to estimate confidence intervals around means, compare vector species and human population groups, and assess the influence of individual behaviour on exposure patterns and malaria risk. Also, because such hard classification techniques allow these key indicators to be calculated in simple binomial form, standard sample size estimation techniques may be readily used to plan *a priori* the necessary experimental scales and data collection targets required for field studies. Sample size calculations for field studies should allow for natural geographic variation and seasonality, taking advantage of rolling cross-sectional designs to survey and re-survey's large numbers of separate study locations in a logistically feasible manner.

Abbreviation

ITNs: Insecticide-treated nets.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13071-021-04884-2.

Additional file 1: Poster S1. A poster illustrating the basic principles of simple binomial *hard classification* of where and when biting mosquitoes are caught, plus examples of applications of this approach using routine off-the-shelf statistical tools to compare mosquito populations and human population groups or calculate suitable sample sizes for large-scale field surveys.

Additional file 2: Dataset S1. An Excel® spreadsheet template for calculating individual-level estimates for the proportion of exposure of bites by *Anopheles gambiae* (*s.l.*) that would occur indoors or while asleep in the absence of any protective interventions, such as window screens or bed nets, in the Tanzanian city of Dar es Salaam [20], as illustrated in Additional file 1: Poster S1. This example is populated with an anonymized sample of questionnaire data describing the times residents reported having gone indoors for the evening, gone to sleep for the night, woke up in the morning and left the house in the morning, as well as published patterns of vector biting activity as measured by human landing catch in parts of Dar es Salaam with vector densities that were high enough to measure [20].

Additional file 3: Dataset S2. An example of a sample size calculation for a nationally representative survey of the proportion of human exposure to malaria vectors occurring indoors in the United Republic of Tanzania, using an Excel® spreadsheet template to apply Eq. 4 of the classic paper by Hayes and Bennett [27], as explained in Box 2 and illustrated in Fig. 2 and Additional file 1: Poster S1.

Acknowledgements

We thank Dr. Paul Johnson and Dr. Mafalda Viana for their insightful comments on the manuscript. We also thank Dr. Amy Ratcliffe and Prof. Steven Lindsay for sharing the Excel® template version of the original Hayes and Bennett sample size calculation models [27], which we adapted herein to these mosquito behaviour indicators (Additional file 3: Dataset S2).

Authors' contributions

GFK developed the analytical methodology, prepared the initial draft of the manuscript and finalized it in consultation with the other authors. GFK and NJG applied the methodology described in Fig. 2, Box 2 and Additional file 3: Dataset S2 to planning national-scale surveys of behavioural interactions between humans and mosquitoes. NJG and AM contributed to editing and finalization of the manuscript. All authors read and approved the final manuscript.

Funding

GFK is supported by an AXA Research Chair award, jointly funded by the AXA Research Fund and the College of Science, Engineering and Food Sciences at University College Cork. NJG is supported by the African Research Leaders Award (Grant Ref: MR/T008873/1), jointly funded by the UK Medical Research Council (MRC) and the UK Foreign, Commonwealth & Development Office (FCDO) under the MRC/FCDO Concordat agreement, which is part of the EDCTP2 programme supported by the European Union.

Availability of data and materials

All relevant equations, calculation tools and sample data are included as additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Biological, Earth & Environmental Sciences and Environmental Research Institute, University College Cork, Cork, Republic of Ireland. [2]Johns Hopkins Center for Communication Programs, Baltimore, MD, USA. [3]Environmental Health and Ecological Sciences Department, Ifakara Health Institute, Ifakara, Tanzania. [4]The School of Life Science and Bioengineering, The Nelson Mandela, African Institutions of Science and Technology, Tengeru, P.O. Box 447, Arusha, United Republic of Tanzania.

Received: 16 April 2021 Accepted: 21 July 2021
Published online: 03 August 2021

References

1. Killeen GF. Characterizing, controlling and eliminating residual malaria transmission. Malar J. 2014;13:330.
2. Killeen GF, Marshall JM, Kiware SS, South AB, Tusting LS, Chaki PP, et al. Measuring, manipulating and exploiting behaviours of adult mosquitoes to optimize malaria vector control impact. BMJ Global Health. 2017;2:e000212.
3. Killeen GF, Chaki PP, Reed TE, Moyes CL, Govella NJ. Entomological surveillance as a cornerstone of malaria elimination: a critical appraisal. In: Dev V, Manguin S, editors. Towards malaria elimination-a leap forward. London: InTech; 2018. p. 403–29.
4. Monroe A, Moore S, Okumu F, Kiware S, Lobo NF, Koenker H, et al. Methods and indicators for measuring patterns of human exposure to malaria vectors. Malar J. 2020;19:207.
5. Weetman D, Kamgang B, Badolo A, Moyes CL, Shearer FM, Coulibaly M, et al. *Aedes* mosquitoes and *Aedes*-borne Arboviruses in Africa: current and future threats. Int J Environ Res Public Health. 2018;15:220.
6. Garrett-Jones C. A method for estimating the man-biting rate. Geneva: World Health Organization; 1964. p. 22.
7. Monroe A, Moore S, Koenker H, Lynch M, Ricotta E. Measuring and characterizing night time human behaviour as it relates to residual malaria

transmission in sub-Saharan Africa: a review of the published literature. Malar J. 2019;18:6.
8. Guglielmo F, Sanou A, Churcher T, Ferguson HM, Ranson H, Sherrard-Smith E. Quantifying individual variability in exposure risk to mosquito bites in the Cascades region, Burkina Faso. Malar J. 2021;20:44.
9. Hii J, Hustedt J, Bangs MJ. Residual malaria transmission in select countries of the Asia-Pacific region: old wine in a new barrel. J Infect Dis. 2021;223:S111–42.
10. Rodriguez-Rodriguez D, Katusele M, Auwun A, Marem M, Robinson LJ, Laman M, et al. Human behavior, livelihoods, and malaria transmission in two sites of Papua New Guinea. J Infect Dis. 2021;223:S171–86.
11. Finda MF, Moshi IR, Monroe A, Limwagu AJ, Nyoni AP, Swai JK, et al. Linking human behaviours and malaria vector biting risk in south-eastern Tanzania. PLoS One. 2019;14:e0217414.
12. Monroe A, Mihayo K, Okumu F, Finda M, Moore S, Koenker H, et al. Human behaviour and residual malaria transmission in Zanzibar: findings from in-depth interviews and direct observation of community events. Malar J. 2019;18:220.
13. Monroe A, Msaky D, Kiware S, Tarimo BB, Moore S, Haji K, et al. Patterns of human exposure to malaria vectors in Zanzibar and implications for malaria elimination efforts. Malar J. 2020;19:212.
14. Geissbuhler Y, Chaki P, Emidi B, Govella NJ, Shirima R, Mayagaya V, et al. Interdependence of domestic malaria prevention measures and mosquito-human interactions in urban Dar es Salaam, Tanzania. Malar J. 2007;6:126.
15. Seyoum A, Sikaala CH, Chanda J, Chinula D, Ntamatungiro AJ, Hawela M, et al. Human exposure to anopheline mosquitoes occurs primarily indoors, even for users of insecticide-treated nets, in Luangwa Valley, South-east Zambia. Parasit Vectors. 2012;5:101.
16. Huho BJ, Briët O, Seyoum A, Sikaala CH, Bayoh N, Gimnig JE, et al. Consistently high estimates for the proportion of human exposure to malaria vector populations occurring indoors in rural Africa. Int J Epidemiol. 2013;42:235–47.
17. Russell TL, Govella NJ, Azizi S, Drakeley CJ, Kachur SP, Killeen GF. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. Malar J. 2011;10:80.
18. Bugoro H, Cooper RD, Butafa C, Iro'ofa C, Mackenzie DO, Chen CC, et al. Bionomics of the malaria vector *Anopheles farauti* in Temotu Province, Solomon Islands: issues for malaria elimination. Malar J. 2011;10:133.
19. Bradley J, Lines J, Fuseini G, Schwabe C, Monti F, Slotman MA, et al. Outdoor biting by *Anopheles* mosquitoes on Bioko Island does not currently impact on malaria control. Malar J. 2015;14:170.
20. Msellemu D, Namango HI, Mwakalinga VM, Ntamatungiro AJ, Mlacha Y, Mtema ZJ, et al. The epidemiology of residual *Plasmodium falciparum* malaria transmission and infection burden in an African city with high coverage of multiple vector control measures. Malar J. 2016;15:288.
21. Sougoufara S, Thiaw O, Cailleau A, Diagne N, Harry M, Bouganali C, et al. The impact of periodic distribution campaigns of long-lasting insecticidal-treated bed nets on malaria vector dynamics and human exposure in Dielmo. Senegal Am J Trop Med Hyg. 2018;98:1343–52.
22. Garrett-Jones C, Shidrawi GR. Malaria vectorial capacity of a population of *Anopheles gambiae*: an exercise in epidemiological entomology. Bull World Health Organ. 1969;40:531–45.
23. Killeen GF. A revival of epidemiological entomology in Senegal. Am J Trop Med Hyg. 2018;98:1216–7.
24. Killeen GF, Seyoum A, Sikaala CH, Zomboko AS, Gimnig JE, Govella NJ, et al. Eliminating malaria vectors. Parasit Vectors. 2013;6:172.
25. Lindblade KA. Does a mosquito bite when no-one is around to hear it? Int J Epidemiol. 2013;42:247–9.
26. Govella NJ, Okumu FO, Killeen GF. Insecticide-treated nets can reduce malaria transmission by mosquitoes which feed outdoors. Am J Trop Med Hyg. 2010;82:415–9.
27. Hayes RJ, Bennett S. Simple sample size calculation for cluster-randomized trials. Int J Epidemiol. 1999;28:319–26.
28. Johnson PC, Barry SJ, Ferguson HM, Muller P. Power analysis for generalized linear mixed models in ecology and evolution. Methods Ecol Evol. 2015;6:133–42.
29. Kiware SS, Chitnis N, Devine GJ, Moore SJ, Majambere S, Killeen GF. Biologically meaningful coverage indicators for eliminating malaria transmission. Biol Lett. 2012;8:874–7.
30. Hurlbert SH. Pseudoreplication and the design of ecological field experiments. Ecol Monogr. 1984;54:187–211.
31. Killeen GF, Tami A, Kihonda J, Okumu FO, Kotas ME, Grundmann H, et al. Cost-sharing strategies combining targeted public subsidies with private-sector delivery achieve high bednet coverage and reduced malaria transmission in Kilombero Valley, southern Tanzania. BMC Infect Dis. 2007;7:121.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions