Estimating the size of populations at high risk of malaria in two Operational Districts in Cambodia: a household-based survey

Jerry O Jacobson, Dyna Doum, Neil F Lobo, Siv Sovannaroth, Allison Tatarsky, David J McIver

Submitted to: JMIR Public Health and Surveillance
on: March 19, 2024

Table of Contents

Estimating the size of populations at high risk of malaria in two Operational Districts in Cambodia: a household-based survey

Jerry O Jacobson[1]; Dyna Doum[2]; Neil F Lobo[1, 3]; Siv Sovannaroth[4]; Allison Tatarsky[1]; David J McIver[1]

[1]Malaria Elimination Initiative Institute for Global Health Sciences University of California, San Francisco San Francisco US
[2]Health Forefront Organization Phnom Penh KH
[3]University of Notre Dame South Bend US
[4]National Center for Parasitology, Entomology and Malaria Control Phnom Penh KH

Corresponding Author:
David J McIver
Malaria Elimination Initiative
Institute for Global Health Sciences
University of California, San Francisco
1975 4th St
San Francisco
US

Abstract

Background: Cambodia is targeting elimination of malaria by 2025. The last remaining pockets of malaria in Cambodia are concentrated among populations exposed to forested areas, but the size of these populations is not well understood. In order to plan for procurement and distribution of vector control tools, chemoprophylaxis, and other commodities for malaria prevention and surveillance, a robust estimation of the size of the most at-risk populations is required.

Objective: To estimate the size of the population at highest-risk of malaria exposure in two operational districts in Cambodia.

Methods: In April 2023, a multi-stage, in-person survey was conducted among residents in the two operational districts (OD) in Cambodia with the highest malaria burden at the time of the study: Sen Monorom in Mondulkiri province and Phnom Srouch in Kampong Speu province. In each OD, 10 villages were randomly selected for inclusion in the study, and 35 households were randomly selected for inclusion from each village. To estimate the number of individuals at high risk of malaria—defined as residing within 1km of a forest—the survey collected information on the distance of their primary residence to forested areas and their travel patterns to forested areas. The survey also collected data on whether the individuals surveyed had received a BITE forest pack as part of the larger BITE project, in order to develop smaller scale estimates of the number of individuals exposed to forests in each of four randomly selected villages in Sen Monorom OD, via the "multiplier method".

Results: In Sen Monorom, 138 households and 872 individuals were enrolled in the study, and in Phnom Srouch 163 households and 844 individuals were enrolled. In both ODs, the ratio of males to females was approximately equal, with median age of 22.0 years in Sen Monorom, 24.5 years in Phnom Srouch (total age range 3-86)). The first-stage survey indicated that approximately 79% of residents in Sen Monorom, representing an estimated 17,315 individuals, are at high risk for malaria exposure. In Phnom Srouch, just 25% of the population fit the high-risk criteria, representing an estimated 1,945 individuals. Between 125 and 186 individuals were estimated to be at high risk of malaria in each of the four villages where the multiplier method could be applied.

Conclusions: This study provides an estimate of the number of individuals in two operational districts in Cambodia who are considered to be at high risk for malaria infection. These estimates can be used to help plan malaria control and elimination efforts in the future. This paper provides a relatively simple method for undertaking additional at-risk population size estimate studies in other areas in Cambodia and beyond.

(JMIR Preprints 19/03/2024:58584)
DOI: https://doi.org/10.2196/preprints.58584

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Title: Estimating the size of populations at high risk of malaria in two Operational Districts in Cambodia: a household-based survey

Authors: Jerry O. Jacobson[1], Dyna Doum[2], Neil F Lobo[1,3], Siv Sovannaroth[4], Allison Tatarsky[1], David J McIver[1*]

1 Malaria Elimination Initiative, Institute for Global Health Sciences, University of California San Francisco, San Francisco, United States
2 Health Forefront Organization, Phnom Penh, Cambodia
3 University of Notre Dame, South Bend, Indiana, United States
4 National Centre for Parasitology, Entomology, and Malaria Control, Phnom Penh, Cambodia

*Corresponding author:
David J McIver
Nanaimo, British Columbia, Canada
+1 778 269 2965
davidjmciver@gmail.com

Abstract (max 450)

Background: Cambodia is targeting elimination of malaria by 2025. The last remaining pockets of malaria in Cambodia are concentrated among populations exposed to forested areas, but the size of these populations is not well understood. To plan for procurement and distribution of vector control tools, chemoprophylaxis, and other commodities for malaria prevention and surveillance, robust estimates of the population at greatest risk – those frequently exposed to forested areas – is required.

Objective: To estimate the number of forest-exposed individuals residing in the highest burden Operational Districts (ODs) in two provinces in Cambodia with active malaria transmission.

Methods: In April 2023, a multi-stage, in-person survey was conducted among residents in the two ODs in Cambodia with the highest malaria burden: Sen Monorom in Mondulkiri province and Phnom Srouch in Kampong Speu province. In each OD, 10 villages were randomly selected and 35 households were randomly selected from each selected village. To estimate the number of individuals at high risk of malaria—defined as residing within 1km of a forest or traveling at least once per week to the forest—respondents were asked the distance from their household to the nearest forested area, and their travel patterns to forested areas. To account for mobility (i.e., avoid double-counting), respondents also provided information on overnight stays at other households in the selected villages in the past month. In the four selected villages in Sen Monorom OD where Project BITE forest packs (an intervention in the larger research program) had been distributed prior to the survey, respondents were also asked questions to determine if they had received such a pack, in order to develop smaller scale “multiplier method” estimates of at-risk individuals in each of those villages.

Results: In Sen Monorom, 138 households and 872 individuals were enrolled in the survey, and in Phnom Srouch 163 households and 844 individuals were enrolled. The estimated percentage of female householders was 49.7% (852/1,716) across both OD s; the median age was 22.0 years in Sen Monorom and 24.5 years in Phnom Srouch (total age range 3-86). Based on mobility-adjusted survey estimates alone, 32.0% (19.9-47.2%) of residents in Sen Monorom (an estimated 12,133-20,135 individuals) and

36.0% (24.5-45.5%) of residents in Phnom Srouch (an estimated 1,717-2,203 individuals), met risk criteria for forest exposure. Between 125 and 186 individuals were estimated to be at risk in each of the four villages where the multiplier method could be applied.

Conclusion: This study provides estimates of the number of individuals potentially at high risk for malaria infection due to forest exposure in two operational districts in Cambodia. These estimates can support planning for malaria control and elimination efforts. The straightforward methods of household surveys and multipliers should be feasible for many national malaria control programs.

Keywords: Population size estimate; malaria; forest exposure; Greater Mekong Subregion

Introduction

Accurate estimates of the number of individuals at high risk for malaria are essential for national malaria programs to plan and implement prevention activities, to forecast and procure commodities, to assess and monitor program coverage, to advocate for resources, and to set targets for burden reduction and elimination. Yet, the population subgroups that remain at high risk for malaria in countries that have made significant progress toward malaria elimination are often not assessed or monitored by existing programs or population censuses.

Malaria continues to be an important cause of morbidity and mortality In the Greater Mekong Subregion (GMS). While the number of cases of malaria in the GMS has decreased by 77%, and the number of deaths decreased by 97%, between 2012 and 2022 [1], the number of indigenous cases rose from 90,082 in 2021 to 170,527 in 2022 [2]. In Cambodia, as throughout the GMS, two groups at increased risk for malaria include forest dwellers (who normally reside in the forest or on the forest fringe) and forest goers (who normally reside in non-forest residential areas and travel to the forest for a variety of reasons)[3,4]. Population size estimates for forest-exposed populations are generally unavailable in Cambodia.

This study aimed to apply a simple and replicable method—a household survey—to estimate the number of people at risk of malaria due to forest exposure in the two Health Operational Districts (ODs) with the greatest annual parasite incidence within the two highest-burden provinces in Cambodia at the time of this study: Sen Monorom OD in Mondulkiri province and Phnom Srouch OD in Kampong Speu province. In 2021, these ODs had an annual parasite incidence of 5.7 and 2.5 cases per 1,000 population, respectively, based on data from the Malaria Information System of the Cambodia National Center for Parasitology, Entomology and Malaria Control (CNM). In addition, the study aimed to develop smaller area estimates for selected villages, by combining the household survey data with Project BITE implementation research study data, using the “multiplier method”. Project BITE (Bite Interruption Toward Elimination) is a large research program that included the distribution and evaluation of “forest packs” containing vector control tools to at-risk populations in targeted ODs. Various refinements were introduced in the design of the survey to reduce potential error in the population size estimates due to the at-risk population’s frequent travel away from the household and possible under-reporting of traveling to the forest due to the illicit nature of some forest activities.

Methods

Project BITE program data

Between October 2022 and January 2023, Project BITE, CNM and Malaria Consortium - an international non-governmental organization (NGO) supporting malaria elimination activities in Cambodia - distributed “forest packs” primarily via village malaria workers to forest dwellers and forest goers in Sen Monorom and Phnom Srouch (forest rangers were also targeted, but not included in the present study). The packs contained mosquito bite prevention tools (topical repellent, volatile pyrethroid spatial repellent, and etofenprox treatment for clothing) and were distributed monthly, for four months, in villages that were active or recently active *P. falciparum* foci. In total, 14,000 packs were distributed in Sen Monorom OD and 6,731 packs in Phnom Srouch OD to a total of 5,744 individuals across both ODs. Individuals could choose to receive and use any or all of the items in the pack. Forest dwellers were defined as individuals who resided in a living structure inside the forest or within 1 kilometer of the forest edge, while forest goers were defined as those traveling at least once per week to the forest, while residing at least 1 kilometer from the forest edge. At each distribution round, field teams aimed to provide packs to the same households and individuals, however identities were not verified.

Household survey

Between April 3rd and 18th, 2023, a household survey was conducted to estimate the number of forest dwellers and goers in the two selected ODs, in all villages with any *P. falciparum* (*Pf*) cases in 2022. In Phnom Srouch, all 10 villages with active foci were selected. In Sen Monorom, 10 of 38 villages (26%) with active foci were randomly selected by probability proportional to size (PPS) without replacement, reaching the desired sample size[5], with size defined as the number of *Pf* cases per population. In each selected village, 35 households were selected by simple random sampling from a list of residential dwellings developed with the assistance of village leaders. A study team consisting of an interviewer and a village malaria worker (VMW) went to the selected households in randomized order until information had been gathered on the target sample size of 80 household members aged ≥ 3 years. In each household, the household head or other adult aged 18 years or older present, was asked to respond to a series of questions on behalf of all such individuals (henceforth, survey “participants” or “householders”), including all usual residents and any visitors who had slept at the household on the previous night, whether or not they were family members. The survey (Appendix 1) was interviewer-administered at the household by tablet and queried demographic characteristics, patterns of sleeping away from the household in the past four weeks, and receipt of BITE tools. While the primary respondent was asked to complete the survey on behalf of all householders, any householder present could respond for themselves if they provided consent. If participants indicated that the household was ≥ 1 km from the forest, additional questions were asked regarding travel to the forest to date during the year.

As confidentiality protections, GPS coordinates of households were not recorded, respondents were not asked about the nature of forest activities, and only initials—rather than names—were recorded on survey forms.

Survey sample size calculations

The survey sample size was determined based on a desired precision of +/- 5% for the estimated percentage of householders that belong to the target population in a given OD, based on a 25% baseline prevalence of belonging to the target population (based

on rough data previously collected by Project BITE), a total OD population size of 50,000 eligible residents, a non-response rate of 10%, a design effect (DEFF) of 2.5 due to the multi-stage sampling design, and a 95% confidence level. These parameters led to a minimum sample of 796 residents per OD. Results were similar when assuming a total population of 25,000 eligible residents. Assuming an average of three individuals per household would meet age and other eligibility criteria led to a minimum of 265 households per OD; however, since there were no available data to check this assumption, the study team planned, conservatively, that it might be necessary to visit up to 30% more, or 345 households. Budget considerations allowed for sampling up to 10 villages per OD. Dividing the sample evenly across villages led to sample size requirements of 80 residents and 35 households per village. However, once the sample size of 80 residents was met, no further households were enrolled in the village. If the final household had more eligible residents than required to meet the target number of residents, data on all eligible residents were collected in the survey.

Statistical analysis

Participants who indicated that their household was located “inside the forest or within one kilometer of the forest” were classified as “forest dwellers”. Where responses among householders differed (16 households), all householders were classified according to the majority response. Participants not classified as forest dwellers were classified as “forest goers” if they had travelled to the forest at least one day per week anytime during the dry season “so far this year, up to today”.

Household-based size estimates

The number of people at risk in each OD was calculated by multiplying the resident population of the OD by the survey-estimated proportion of residents who met criteria as a forest dweller or goer. To develop a 95% confidence interval (CI), the limits of the CI of the proportion were multiplied by the resident population. The population counts for Sen Monorom OD were from the Cambodia national Malaria Information System (MIS). For Phnom Srouch OD, we obtained the population counts by calling all of the village heads, since the data in the MIS for Kampong Speu province were incomplete.

Multiplier size estimates

We employed the “multiplier method”, often used in HIV epidemiology [6,7], as a second size estimation strategy. The “multiplier method” draws on two data sources. One is a representative survey—like the household survey described here—and the second is a count of the number of people in a known subgroup of the target population (the “multiplier”). The multiplier in this study was the number of people who had been given a Project BITE forest pack, based on Project BITE program data. Multiplier-based size estimates could only be calculated for villages where forest packs had been distributed. In each such village, the multiplier size estimate was calculated as M/p, where M was the number of recipients of forest packs, approximated as one fourth of the number of forest packs distributed in the village (whether complete or incomplete) over the four distribution rounds. The parameter p was the survey-estimated proportion of at-risk householders (whether forest dwellers or goers) who reported (1) receiving “any items to help prevent mosquito bites from a VMW as part of Project BITE” between October 2022 and January 2023 and (2) mentioned topical repellent, spatial repellent, or treated clothing, when asked in an open-ended question which items they received. The multiplier size estimates were calculated only for villages in Sen Monorom, due to data quality issues with the BITE forest pack distribution records for Phnom Srouch. CIs for the multiplier-based size estimates were calculated as in Fearon et al.[8] to account

for sampling error in *p*.

Sampling weights

Survey-estimated proportions were weighted, with sampling weights calculated as the inverse of the probability the participant's household was selected:

P(participant selected) = P(village selected) x P(household selected)

In Sen Monorom, P(village selected) was the probability the village was selected by PPS, as calculated by the *Sampling* package in *R* statistical software, whereas in Phnom Srouch it was one, as all villages with active foci were surveyed. P(household selected) was the sampling fraction of households: (#households enrolled) / (#households in village). The standardized sampling weights ranged from -0.7 to 2.6 in Sen Monorom and from -1.3 to 1.8 in Phnom Srouch.

Mobility adjustment

To reduce the likelihood of double-counting individuals who frequently spend the night at multiple households in the same OD—a common occurrence in these locations—we developed mobility correction factors for the OD- and village-level size estimates. The OD-level correction reduced the household-based proportions at risk (and corresponding CIs) by 50% of the estimated percentage of at-risk householders in the OD who had stayed overnight at another household in the same OD for ≥ 2 of the past four weeks. These parameters are based on World Health Organization guidance for mapping of key populations for HIV[9]. The village-level multiplier size estimates were similarly adjusted, based on at-risk householders in the village who had stayed overnight elsewhere in the same village.

All analyses were conducted using multi-stage survey procedures in Stata® version 15.1 (StataCorp, College Station, Texas).

Results

Recruitment

In Phnom Srouch and Sen Monorom ODs, 163 and 138 households, respectively, were enrolled in the survey across the 10 villages selected in each OD. Primary respondents for the households reported data on a total of 844 household members (residents or previous-night visitors) in Phnom Srouch (mean per household, 5.2; range, 1-10) and 872 household members in Sen Monorom (mean per household, 6.3; range, 1-30). All households approached were enrolled on the first visit attempt. The number of households enrolled per village ranged from 14 to 20 in Phnom Srouch and from 6 to 18 in Sen Monorom.

Demographics

The mean age of householders was estimated at 29.1 years in Phnom Srouch and 26.0 years in Sen Monorom (Table). Most were aged 18 to 59 years (63.4% in Phnom Srouch and 56.3% in Sen Monorom), with a considerable number of children aged 3 to 17 years (from 29.5% to 38.5%). More than half of householders were female in Phnom Srouch (53.0%) while 45.7% were female in Sen Monorom.

Table 1. Estimated demographics of householders ages ≥ 3 years in the study ODs

	Phnom Srouch (N=844)		Sen Monorom (N=872)	
	n	Percent (95% CI)	n	Percent (95% CI)
Sex				
Female	450	53.0 (50.0-56.0)	402	45.7 (41.1-50.5)
Male	394	47.0 (44.0-50.0)	470	54.3 (50.0-59.0)
Age (years)				
3-17	249	29.5 (26.3-32.8)	349	38.5 (34.6-42.6)
18-29	239	63.4 (60.0-66.7)	228	56.3 (53.1-59.4)
30-59	291		252	
≥ 60	65	7.1 (5.4-9.4)	43	5.2 (3.6-7.4)
Mean	-	29.1 (28.1-30.2)	-	26.0 (24.9-27.2)
Median	-	25.0 (24.0-26.0)	-	22.0 (21.0-23.0)

Notes: CI, confidence interval; Percentages and CIs are weighted.

Household-based size estimates for population at-risk

In Sen Monorom, an estimated 79.2% of householders were found to be at risk for malaria as either a forest dweller or goer compared to 24.8% in Phnom Srouch (Table). In most villages in Sen Monorom, all participants were dwellers and no village had both dwellers and goers. In Phnom Srouch, two of the selected villages had both dweller and goer participants.

Multiplying the proportion at risk by the total OD population led to estimates of 2,371 and 20,613 individuals at risk in Phnom Srouch and Sen Monorom, respectively (Table).

Table 2. Forest exposure among householders aged ≥ 3 years

	Phnom Srouch (N=844)		Sen Monorom (N=872)	
	n	Percent (95% CI)	n	Percent (95% CI)
Forest dweller	132	17.3 (13.9-21.4)	676	77.6 (52.0-91.7)
Forest goer	58	7.5 (5.7-9.9)	30	1.6 (0.2-11.1)
Dweller or goer (“total at risk”)	190	24.8 (21.9-28.1)	706	79.2 (55.5-92.1)

Notes: CI, confidence interval; Percentages and CIs are weighted.

Table 3. District-level population size estimates, based on householder percentage at risk

District	Population[a]	Survey-estimated percentage at risk (95% CI)[b]	Estimated individuals at risk (95% CI)
Phnom Srouch	9562	24.8 (21.9-28.1)	2371 (2094-2687)
Sen Monorom	26026	79.2 (55.5-92.1)	20613 (14444-23970)

Notes: [a] residents in villages with *Pf* cases in 2021; estimates and CIs are weighted; [b] from Table .

The estimates in Table are not adjusted for mobility. Among at-risk participants, 36.0%

in Phnom Srouch and 32.0% in Sen Monorom had stayed overnight at another household in the same OD for ≥ 2 of the past four weeks (Table). Reducing the size estimates by 50% of these proportions led to mobility-adjusted estimates of 1,945 and 17,315 OD residents at risk, respectively. Notably, the reason reported for 95.8% (387/404) of these stays was “work”.

Table 4. District-level population size estimates, adjusted for mobility

District	At-risk householders meeting mobility criteria		Mobility-adjusted individuals at risk
	(n)	(%) (95% CI)	
Phnom Srouch	68 / 190	36.0 (24.5-45.5)	1945 (1717-2203)
Sen Monorom	280 / 706	32.0 (19.9-47.2)	17315 (12133-20135)

Notes: estimates and CIs are weighted.

Multiplier size estimates

Of the nine villages where BITE forest packs were distributed, four were subsequently randomly selected into the survey (Table). All participants in these villages were forest dwellers, based on their survey responses. Between 67.0% to 98.8% of the at-risk population across these villages were estimated to have received the forest packs prior to the survey.

Combining these proportions with the total populations of the villages, we estimate 160, 195, and 162 residents at risk, respectively (Table). Multiplier estimates could not be calculated for Chak Cha village because no survey participants there had received the BITE packs.

The estimated percentage of householders who met the within-village mobility criterion varied widely, from 6.0% in Kdaoy to 43.2% in Pu Char (Table). The mobility adjustment reduced the size estimates in the three villages to 125, 186, and 157 individuals at risk, respectively.

However, the 95% CI limits of the multiplier estimates (before and after the mobility adjustment) were greater than the total number of residents in K uon and Kdaoy village. Furthermore, in all three villages, the CI’s lower limit was less than our approximation of the number of people who received forest packs. We therefore report an additional “uncertainty interval” bounded by the number of BITE pack recipients and the village population (Table).

Table 5. Village-level population size estimates, based on the multiplier method

Village	Resident population	Survey participants classified as at risk	At-risk survey participants who received BITE tools		Individuals given BITE tools[a] as per distribution records (*M*)	Estimated individuals at risk (95% CI)
			n	(%) (95% CI) (*p*)		
Chak Cha	1500	82 / 82	0	-	160	-
Pu Char	329	88 / 88	59	67.0 (66.8-67.3)	107	160 (129-190)
K uon	130	83 / 83	70	84.3 (84.0-84.6)	[b]164	195 (165-224)
Kdaoy	275	83 / 83	82	98.8 (98.7-98.8)	160	162 (137-187)

Notes: CI, confidence interval; percentages and CIs are weighted estimates; [a] approximated as ¼ (BITE packs distributed). [b] This figure, gathered from distribution records, is greater than the population size of the village, which was based on the available village registry. Table includes villages included in both the survey and program implementation.

Table 6. Village-level population size estimates, adjusted for mobility

Village	Householders meeting mobility criteria		Mobility-adjusted individuals at risk (95% CI)	Uncertainty Interval[a]
	(n)	(%) (95% CI)		
Chak Cha	12 / 82	14.6 (14.4-14.8)	-	-
Pu Char	38 / 88	43.2 (42.7-43.6)	125 (101-149)	107-149
K uon	7 / 83	8.4 (8.2-8.7)	186 (158-215)	*cannot determine*[b]
Kdaoy	5 / 83	6.0 (5.9-6.2)	157 (133-181)	160-181

Notes: [a]The uncertainty interval goes from the number of individuals who received BITE tools to the population size of the village. [b] The upper limit of the uncertainty interval is meant to be the total population size of the village; however, BITE pack distribution data indicates that more packs were distributed than there are individuals in the village.

Discussion

Based on a survey of forest-exposed individuals in two high malaria transmission ODs in Cambodia, this study estimates that there are 1,945 individuals in Phnom Srouch OD and 17,315 individuals in Sen Monorom OD who may be at high risk for malaria based on their proximity or travel to forested areas. A higher proportion of people in Sen Monorom (nearly 80%) were classified as at risk compared to Phnom Srouch (approximately 25%), which is likely a result of decreasing forest stands in Kampong Speu province [10]. These estimates of the population at risk can provide essential data to plan and budget for government or NGO-led anti-malaria campaigns and support the targeting of health programs to those likely to be at greatest risk.

As countries in the GMS continue to move toward malaria elimination, the populations at risk of malaria have become smaller and are increasingly those exposed to forests[3,11,12]. Unlike in high-burden settings, where vector control centers on mass distribution or campaigns over large areas and populations, in low-burden settings it becomes increasingly important to target and tailor prevention to the diminishing population of individuals at continued risk. Intervention targeting may also require alternative delivery methods to access hard-to-reach populations[13]. To efficiently target surveillance, prevention, and control activities, national malaria programs (NMPs) require accurate estimates of the size and locations of the subgroups at risk. This study demonstrated a straightforward approach using household surveys, which is likely to be feasible in contexts where NMPs can enlist the support of local health facility staff or village or mobile malaria workers, and where most individuals at risk can be enrolled via their households. In contrast to recent applications of the multiple source capture-recapture method for population size estimation in Lao PDR [14] and Namibia [13], which have relied on serial household surveys and/or multiple sources of program data, the methods described here require fewer data sources, fewer contacts with participants, and less complicated statistical analysis.

Findings in this study indicate that, of the two ODs, which were among the highest risk districts in Cambodia, a much higher proportion of at-risk individuals, as defined in this study, was estimated in Sen Monorom OD in Mondulkiri province compared to Phnom Srouch OD in Kampong Speu province. We speculate that this difference is largely driven by two factors: the amount of intact forest remaining in Sen Monorom OD and the amount of travel to the forest from the OD. In fact, the former factor directly

influences the latter; when there is more forest available, people are more likely to travel to the forest for purposes of logging, hunting, collection of non-timber forest products, farming, and other activities. In contrast to Mondulkiri, forested areas in Kampong Speu are becoming increasingly rare[10], with fewer people making a living in forest-related occupations.

The study was conducted in April, after the end of the malaria season, which typically runs from August to December or early January. The malaria season coincides with the rainy season and when people are likely to be travelling to the forest. While people may have stopped going to the forest as frequently by the time this study was conducted, the survey used was able to capture any travel to the forest within the last three months, which should identify those individuals who typically travel to the forest during the rainy season. However, people who may have travelled to one of the ODs in the study from another location temporarily (i.e., migrant workers, migrant populations, seasonal workers, etc.) may have been missed by this survey if they returned to their homes between the end of the rainy season and when the study began.

It is helpful, for purposes of malaria elimination, to have multiple size estimation tools, because there are limitations to any single method[15]. Multiple strategies also allow countries greater flexibility to select the best approach for a given context. However, the size estimates generated here using the multiplier method—which could only be undertaken in the four villages surveyed where BITE tools had been distributed—produced, in part, unrealistic results. Although all survey participants in these villages were forest dwellers, which suggests that the entire village population was proximate to or inside the forest, the multiplier estimates of 160 (95% CI, 129-190) individuals at risk in Pu Char village and of 162 (137-187) in Kdaoy village were far below the resident populations of 329 and 275 people, respectively. This level of under-estimation seems too great to be due to children aged <3 years in the population counts, who were not included in the surveys. Instead, the large error is likely due to the fact that the number of unique individuals who received forest packs needed to be approximated by assuming the same individuals were reached during each round of distribution. Conversely, the multiplier estimate of 195 (165-224) individuals at risk in K uon village is larger than the total village population of 130 residents, which may be due to under-estimation of the parameter p (the survey percentage), potentially due to under-reporting. The K uon over-estimate persisted even after accounting for householders who frequently stayed at other households in the same village.

This study was based on a probability survey of households, so that it can be considered representative of the household population in each OD. Since respondents were successfully enrolled from all selected households, with no refusals, there was no potential for bias due to differential rates of participation in the study. The study team attributes this to close coordination with village leaders and the presence of VMWs on the field team. Furthermore, there is likely to be minimal error owing to residents' being away from the selected households at the time of the survey since data were collected on all household members and recent visitors.

While this study had the benefit of being part of the larger Project BITE research program, which included providing forest packs to some villages independent of this population size estimation study, there are other approaches to leveraging existing malaria service delivery or surveillance activities to enable more accurate estimates using the multiplier method. For example, if a population size estimate study is being conducted by interviewing individuals who arrive at a public health clinic, the proportion of interviewed individuals that received chemoprophylaxis from that clinic or local VMWs might be used as a multiplier to further refine estimates. There are

numerous such ways that a sample population can be sub-divided based on receiving or enrolling in another service in order to create more accurate estimates.

When conducting future population size estimate surveys, it is recommended to first engage in relationship building with target villages and populations, especially if the survey is enquiring about potentially elicit or illegal activities (as was this case in this study, where illegal logging is common among the target population). Creating a relationship based on mutual trust will allow for greater honesty in survey responses, and thereby more accurate estimates. Furthermore, it is important to reach those households or individuals which were randomly selected, and if necessary multiple attempts should be made. Especially in a study like this, where mobility and exposure to forests is central to the population being studied, care must be taken that selected individuals are not skipped if not immediately available, as it is likely that the person is unavailable precisely because they are part of the population being studied. The study team must be flexible with where and when they are on site to conduct interviews, and work with village leaders to communicate the needs and enrolment criteria of the study, in order to achieve maximum representation from the randomly selected group.

Limitations

Our findings are subject to important limitations. The size estimates do not capture any individuals who resided entirely at work sites (in forested areas or otherwise) and did not belong to a household in the village, such as foreign workers. Future studies may improve upon this by working with local village chiefs and other leaders to identify these groups and include them in the sampling frame of households. The mobility adjustment prevented double-counting of individuals who may live between two households, but not three or more. There may also be a small degree of error due to householders' not knowing whether their households were located within 1km of the forest; yet, in only 16 of the 301 households surveyed (5%) was there disagreement among participants on this measure and resolving it (by assuming the majority response was correct) changed our estimate of the prevalence of forest dwellers by less than 1%. The size estimates also do not include anyone living away from the forest who traveled there less than weekly, or who began doing so subsequent to the survey (i.e., after mid-April). The OD-level size estimates—which were based on percentages of householders—may have been under-estimated if the primary respondents (heads of household or other adults present at the time) were unaware of others' forest-going activity. However, in this setting of small villages, the study team believes this is unlikely as most family and community members are aware of the activities of others. A source of over-estimation was that available population counts which were used to multiply the percentages at risk estimated from the survey could not be subset by age, whereas survey data could only be collected on individuals older than age three due to concerns raised during ethical review of the study protocol. For a more aligned calculation, future surveys should collect data on the ages of all householders (regardless of whether they are considered study eligible) in order to estimate the age distribution in the general population, if unavailable from census data.

A multiplier estimate for Chak Cha village could not be calculated because none of the survey participants had received the forest pack (leading to a p of 0); this was due to low coverage of the tools in the village (about 7%, according to program data). To avoid this situation, the authors recommend using a publicly available sample size calculator (see annexes of WHO, 2017) to determine the size of the multiplier required when planning for multiplier size estimates. It is also important to note that the "services multiplier method"—commonly used in HIV epidemiology[6,7]—works best when

program data accurately reflect the number of unique individuals who have accessed a specific service in the study area during a specific time period. Recent adaptions of the capture-recapture method have been introduced that do not require program data[16,17].

Finally, we note that the precision of the size estimates for Sen Monorom (where villages were randomly sampled) was far less than in Phnom Srouch (where all villages with recent *P.f.* cases were surveyed) due to the considerable degree of village-level variation in our risk measures and the relatively small number of households selected per village, owing to budget limitations. To improve precision, future studies should aim to collect data from a larger number of households, and ensure this target is met even after the planned sample of individuals has already been achieved.

Conclusion

This study provides a robust and replicable method for estimating the number of individuals at high risk of malaria due to forest exposure in the last remaining pockets of malaria transmission in Cambodia and across the Greater Mekong Subregion. With perpetually constrained budgets for malaria elimination activities, having a reliable understanding of the estimated size of the population at risk for malaria allows national malaria programs to appropriately plan for and implement critical prevention and surveillance activities.

Acknowledgments

The authors thank the participants of the study, village leaders, and village malaria workers for their participation and assistance in this study. The authors also thank the Innovative Vector Control Consortium (IVCC) for the management of the grant which funded this work and for their technical collaboration and partnership. The project was funded by the Australian Department of Foreign Affairs and Trade (DFAT) through the Innovative Vector Control Consortium (IVCC) to the UCSF Malaria Elimination Initiative (MEI)(Grant #A134328).

Conflicts of Interest

None declared.

Declarations

This study was approved by the National Ethics Committee for Health Research of the Ministry of Health of Cambodia (ref. no. 241 NECHR) and the University of California, San Francisco (ref. no 22-38096). The purpose of the study was explained to all study participants and oral informed consent was obtained. Participants understood that they were free to remove themselves from the study at any time without repercussion. All the data and samples were deidentified and coded towards analysis following IRB guidelines.

Authors' Contributions

JJ, DD, DJM, and AT contributed to the design of the study. DD led the collection of the data. JJ conducted all analyses, with critical review from DD and DJM. All authors contributed to critical review and editing of the final manuscript.

References

1. Manzoni G, Try R, Guintran JO, Christiansen-Jucht C, Jacoby E, Sovannaroth S, Zhang Z, Banouvong V, Shortus MS, Reyburn R, Chanthavisouk C, Linn NYY, Thapa B, Khine SK, Sudathip P, Gopinath D, Thieu NQ, Ngon MS, Cong DT, Hui L, Kelley J, Valecha NNK, Bustos MD, Rasmussen C, Tuseo L. Progress towards malaria elimination in the Greater Mekong Subregion: perspectives from the World Health Organization. Malar J 2024 Mar 1;23(1):64. doi: 10.1186/s12936-024-04851-z

2. World Health Organization. World Malaria Report 2023. Geneva: World Health Organization; 2023. Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023 [accessed Feb 14, 2024]

3. Sandfort M, Vantaux A, Kim S, Obadia T, Pepey A, Gardais S, Khim N, Lek D, White M, Robinson LJ, Witkowski B, Mueller I. Forest malaria in Cambodia: the occupational and spatial clustering of Plasmodium vivax and Plasmodium falciparum infection risk in a cross-sectional survey in Mondulkiri province, Cambodia. Malar J 2020 Nov 19;19(1):413. doi: 10.1186/s12936-020-03482-4

4. Kerkhof K, Sluydts V, Heng S, Kim S, Pareyn M, Willen L, Canier L, Sovannaroth S, Ménard D, Sochantha T, Coosemans M, Durnez L. Geographical patterns of malaria transmission based on serological markers for falciparum and vivax malaria in Ratanakiri, Cambodia. Malar J 2016 Oct 19;15(1):510. doi: 10.1186/s12936-016-1558-1

5. Rosén B. On sampling with probability proportional to size. J Stat Plan Inference 1997 Aug 15;62(2):159–191. doi: 10.1016/S0378-3758(96)00186-3

6. World Health Organization. Biobehavioural survey guidelines for populations at risk for HIV. 2017. Available from: https://www.who.int/publications-detail-redirect/978-92-4-151301-2 [accessed Dec 15, 2023]

7. Johnston LG, Prybylski D, Raymond HF, Mirzazadeh A, Manopaiboon C, McFarland W. Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. Sex Transm Dis 2013 Apr;40(4):304–310. PMID:23486495

8. Fearon E, Chabata ST, Thompson JA, Cowan FM, Hargreaves JR. Sample Size Calculations for Population Size Estimation Studies Using Multiplier Methods With Respondent-Driven Sampling Surveys. JMIR Public Health Surveill 2017 Sep 14;3(3):e59. PMID:28912117

9. World Health Organization. Estimating sizes of key populations: guide for HIV programming in countries of the Middle East and North Africa. World Health Organization. Regional Office for the Eastern Mediterranean; 2016. Available from: https://iris.who.int/handle/10665/327293 [accessed Dec 15, 2023] ISBN:978-92-9022-131-9

10. Chhinh N, Rath S, Choeun K. The Alteration of Flood Peak Discharge by Land Cover Change in Prek Thnot Watershed, Kampong Speu Provine, Cambodia. Cambodia J Basic Appl Res 2023;5(2):22–33. doi: 10.61945/cjbar.2023.5.2.04

11. Nofal SD, Peto TJ, Adhikari B, Tripura R, Callery J, Bui TM, von Seidlein L, Pell C. How can interventions that target forest-goers be tailored to accelerate malaria elimination in the Greater Mekong Subregion? A systematic review of the

qualitative literature. Malar J 2019 Feb 1;18(1):32. doi: 10.1186/s12936-019-2666-5

12. Sanann N, Peto TJ, Tripura R, Callery JJ, Nguon C, Bui TM, Nofal SD, von Seidlein L, Lek D, Dondorp AM, Cheah PY, Pell C. Forest work and its implications for malaria elimination: a qualitative study. Malar J 2019 Nov 27;18(1):376. PMID:31771587

13. Smith JL, Mumbengegwi D, Haindongo E, Cueto C, Roberts KW, Gosling R, Uusiku P, Kleinschmidt I, Bennett A, Sturrock HJ. Malaria risk factors in northern Namibia: The importance of occupation, age and mobility in characterizing high-risk populations. PLOS ONE Public Library of Science; 2021 Jun 25;16(6):e0252690. doi: 10.1371/journal.pone.0252690

14. Rerolle F, Jacobson JO, Wesson P, Dantzer E, Lover AA, Hongvanthong B, Smith J, Marshall JM, Sturrock HJW, Bennett A. Population size estimation of seasonal forest-going populations in southern Lao PDR. Sci Rep Nature Publishing Group; 2021 Jul 20;11(1):14816. doi: 10.1038/s41598-021-94413-z

15. Jacobson JO, Cueto C, Smith JL, Hwang J, Gosling R, Bennett A. Surveillance and response for high-risk populations: what can malaria elimination programmes learn from the experience of HIV? Malar J 2017 Jan 18;16(1):33. PMID:28100237

16. Doshi RH, Apodaca K, Ogwal M, Bain R, Amene E, Kiyingi H, Aluzimbi G, Musinguzi G, Serwadda D, McIntyre AF, Hladik W. Estimating the Size of Key Populations in Kampala, Uganda: 3-Source Capture-Recapture Study. JMIR Public Health Surveill 2019 Aug 12;5(3):e12118. doi: 10.2196/12118

17. McIntyre AF, Mitchell A, Stafford KA, Nwafor SU, Lo J, Sebastian V, Schwitters A, Swaminathan M, Dalhatu I, Charurat M. Key Population Size Estimation to Guide HIV Epidemic Responses in Nigeria: Bayesian Analysis of 3-Source Capture-Recapture Data. JMIR Public Health Surveill 2022 Oct 26;8(10):e34555. PMID:36287587

Appendix 1: Survey questionnaire

Section 1: Household Identification

Date____/_____/_____
1. Supervisor Number [_____]
2. OD Code [_____]
3. Village Code [_____]
4. Household ID [_____]
5. Household ID [_____]
6. Participant ID [_____]
7. Does this individual meet the inclusion criteria for the study? If No, STOP

- No
- Yes

8. Has this individual read and understood the informed consent form, and agree to participate in the study? If No, STOP

- No
- Yes

9. What visit attempt is this? [_____]

INSTRUCTIONS TO INTERVIEWER

Participant should provide information for all residents and visitors aged 3 and older, whether or not they are at home at the time of the visit, and whether or not they are part of the same family.

A resident is anyone who regularly sleeps at this household.

A visitor is anyone else who slept at this household last night.

The participant should answer on behalf of all residents and visitors, to the best of his or her ability. However, the participant should feel free to confer with them in order to provide a more accurate response.

Up to 3 visit attempts will be made for selected households, before selecting a replacement household.

Please tell me the names of all people who usually live in this household (in any structure), and anyone else who slept here last night. But only those who are aged 3 or older. To protect confidentiality, I will write down only the first TWO letters of each person's given name. If two people here have the same first two letters, I will write down the first THREE letters. For example, instead of "Chea", I would write "CH". This will be the person's nickname for all of the questions.

Section 2: Household Listing

Now we would like to ask you some information about time spent at this household and at other places in the past 4 weeks.

We would first like to ask you some information about the members of this household, including anyone who usually sleeps

here and any recent visitors.

Q1. Individual # [_____]
Q2. Nickname (first 2/3 letters of given name) [_____]
Q3Y. Age (years) [_____]
Q3M. Age (months) [_____]

Q4. What is the relationship of (NICKNAME) to the head of household?

- Head of household
- Spouse (wife/husband/partner)
- Son or daughter
- Son-in-law or daughter-in-law
- Adopted/foster/stepchild
- Grandchild
- Parent
- Aunt or uncle
- Grandparent
- Other relative
- Not related (friend, vistor)

Q5. Is (NICKNAME) male or female

- Male
- Female

Q6. In the past 4 weeks, did (NICKNAME) stay overnight any place else away from this household?

- No
- Yes

Q7. If (NICKNAME) stayed overnight outside this household in the past 4 weeks, what was the primary reason?

- Work
- Away at school / studying
- Staying with relatives or friends
- Holidays / travel for leisure
- Hospitalized or taking care of patient
- Don't know
- Other

Describe other
[_____]

Q8. Is the other place where (NICKNAME) stayed overnight... in this village, in another village in this district, or outside this district?

- This village
- Another village in this district
- Outside of district
- Don't know

Q9. In the past 4 weeks, approximately how many weeks did (NICKNAME) spend at this other place?

- Less than one week
- 1 week
- 2 weeks
- 3 weeks
- 4 weeks
- Don't know

Q10. In the past 4 weeks, approximately how many weeks did (NICKNAME) spend the night here at this household?

- Less than one week
- 1 week
- 2 weeks
- 3 weeks
- 4 weeks
- Don't know

Is this household inside the forest or within 1km of the forest?

- No
- Yes

Q11. During the dry season so far this year up to today, did (NICKNAME) spend any time in the forest?

- No
- Yes
- Don't know

Q12. During the dry season so far this year up to today, did (NICKNAME) ever go to the forest at least 1 day per week?

- No
- Yes
- Don't know

Q13. In which months of this dry season did (NICKNAME) go to the forest at least 1 day per week? Select all that apply

- January
- February
- March
- April
- Don't know

Section 3: Participation in BITE Activities

Did this village receive products from BITE?

- ○ No
- ○ Yes

We would now like to ask you about some recent mosquito bite prevention activities in this district.

Q14. Between October and January, did (NICKNAME) provide a blood sample to Project BITE?

- ○ No
- ○ Yes
- ○ Don't know

Q15. Between October and January, did (NICKNAME) receive any items to help prevent mosquito bites from a VMW as part of Project BITE?

- ○ No
- ○ Yes
- ○ Don't know

Q16. What kinds of items for mosquito bite prevention did (NICKNAME) receive from a VMW?
Wait for participant respond spontaneously. Do not read the options.

- ☐ Topical Repellent
- ☐ Spatial Repellent
- ☐ Etofenprox treatment for clothing
- ☐ None of the above
- ☐ Don't know
- ☐ Other

Describe other
[_____]

Appendix 2: Data Analysis Stata Code

```
* Cambodia PSE for forest workers using HH survey and multipliers

#delim ;
set more off;

* folders;
local       dir       "~/Documents/Projects/Projects-working/UCSF/Cambodia
PSE/Analysis";
local datadir = "`dir'/../Data";
local data_prelim = "`datadir'/PSE Full Dataset 18apr23.xlsx";
local   data_final   =   "`datadir'/PSE   Full   Dataset   cleaned   updated
01jul23.xlsx";
local sheet = "data_fmtd";
local dataout = "pse cambodia";

cd "`dir'";

* import final excel data;
import excel using "`data_final'", sheet("`sheet'") cellrange(A2)
     firstrow case(lower) clear;

compress;
save "`dataout'", replace;
*/;

* reload saved data;
capture log close;
log using "pse_cambodia.log", replace;
use "`dataout'", clear;

do labels;

* dates;

* "date" reflects end date, which differs from start date;
* some dates are missing;
* tabulate start and end dates;

gen start_mo = substr(start, 6, 2);
gen start_day = substr(start, 9, 2);
destring start_mo start_day, replace force;

gen end_mo = substr(end, 6, 2);
gen end_day = substr(end, 9, 2);
destring end_mo end_day, replace force;

ren start start_old;
ren end end_old;

gen start = mdy(start_mo, start_day, 2023);
gen end = mdy(end_mo, end_day, 2023);
format start end %td;
lab var start "start date of survey, from machine recorded start";
```

```
lab var end "end date of survey, from machine recorded end";
drop start_mo start_day end_mo end_day;
tab start;
tab end;

ren date date_old;
lab var date_old "end date as reported, but some missing";

* svyset for Phnom Sruouch (od==1);
*svyset     hhuic     [pw=weight],     strata(vcode)     fpc(fpc_village)
vce(linearized);

* svyset for Senmonorom (od==2);
*svyset vcode [pw=weight], fpc(fpc_od) vce(linearized) || hhuic;



* IDENTIFIERS;

lab def odl 1 "Phnom Srouch" 2 "Senmonorom";
lab val od odl;

* hh UIC;
* there should be 301 HHs, as in the HH-level spreadsheet;
quietly duplicates report od vcode hhid2;
return list;
gen hhuic = string(od) + string(vcode) + string(hhid2);
lab var hhuic "Household UIC";


/* Eligibility;
* Dave confirmed that all participants were eligible and consented,
* that "no" values in data reflect data issue
*/;

tab elig, m;
tab consent, m;
summ ageyr,d;

* TRIM STRINGS;

foreach x of var village_name hhinforest goforest_any goforest_1daywk
     bite_bloodsample   bite_self   bite_gotitems   bite_top   bite_spat
bite_etof
     eligible consented nickname
     relation sex overnight overnight_rsn overnight_loc overnight_time*
{;
     replace `x' = trim(`x');
};

* SAMPLING WEIGHTS;
*  weights  are  calculated  in  "Sampling  plan  &  weights  PSE  Cambodia
15jun23.xlsx";

gen weight=.;
```

```
* Phnom Srouch;
replace weight=1.71 if vcode==1 & od==1;
replace weight=2.24 if vcode==2 & od==1;
replace weight=1.51 if vcode==3 & od==1;
replace weight=2.01 if vcode==4 & od==1;
replace weight=1.71 if vcode==5 & od==1;
replace weight=1 if vcode==6 & od==1;
replace weight=1.01 if vcode==7 & od==1;
replace weight=1.55 if vcode==8 & od==1;
replace weight=1.5 if vcode==9 & od==1;
replace weight=1 if vcode==10 & od==1;


* Mondulkiri / Senmonorom;
replace weight=2.577 if vcode==5 & od==2;
replace weight=37.413 if vcode==11 & od==2;
replace weight=13.402 if vcode==17 & od==2;
replace weight=7.211 if vcode==24 & od==2;
replace weight=3.301 if vcode==28 & od==2;
replace weight=12.541 if vcode==30 & od==2;
replace weight=2.49 if vcode==33 & od==2;
replace weight=1.868 if vcode==35 & od==2;
replace weight=2.491 if vcode==37 & od==2;
replace weight=1 if vcode==38 & od==2;
assert weight!=.;
by od, sort: table vcode, c(min weight max weight);

* un-normalize weights, to see if it improves CIs;
replace weight = weight * 9.4 if od==1;
replace weight = weight * 4.3 if od==2;

* Finite population corrections (FPCs);


* OD-level FPCs;
gen fpc_od=.;
replace fpc_od=9562 if od==1;
replace fpc_od=26026 if od==2;
lab var fpc_od "#residents in OD in villages with pf cases in 2022";
assert fpc_od!=.;

* village-level FPCs;

* list participants per village, to calculate village-level FPC in Excel;
by od, sort: tab vcode od;

gen fpc_village=.;

replace fpc_village=1214 if vcode==1 & od==1;
replace fpc_village=1266 if vcode==2 & od==1;
replace fpc_village=1041 if vcode==3 & od==1;
replace fpc_village=1277 if vcode==4 & od==1;
replace fpc_village=1107 if vcode==5 & od==1;
replace fpc_village=670 if vcode==6 & od==1;
```

```
replace fpc_village=608 if vcode==7 & od==1;
replace fpc_village=838 if vcode==8 & od==1;
replace fpc_village=899 if vcode==9 & od==1;
replace fpc_village=642 if vcode==10 & od==1;

replace fpc_village=669 if vcode==5 & od==2;
replace fpc_village=981 if vcode==11 & od==2;
replace fpc_village=553 if vcode==17 & od==2;
replace fpc_village=321 if vcode==24 & od==2;
replace fpc_village=264 if vcode==28 & od==2;
replace fpc_village=1240 if vcode==30 & od==2;
replace fpc_village=210 if vcode==33 & od==2;
replace fpc_village=329 if vcode==35 & od==2;
replace fpc_village=130 if vcode==37 & od==2;
replace fpc_village=275 if vcode==38 & od==2;

lab var fpc_village "#residents in village";
assert fpc_village!=.;

* DEMOGRAPHICS;

gen male=sex=="male";
gen agecat = cond(ageyr<=17, 1,
                cond(ageyr<=59, 2,
                cond(ageyr>=60 & ageyr!=., 3, .)));
lab def agel 1 "3-17" 2 "18-59" 3 ">=60";
lab values agecat agel;

tab od sex, row m;
tab od agecat, row m;


preserve;
     keep if od==1;
     svyset    hhuic    [pw=weight],    strata(vcode)    fpc(fpc_village)
vce(linearized);
     svy: tab sex, ci;
     svy: tab agecat, ci;
     svy: mean ageyr;
     epctile ageyr, p(50) svy;
restore;

preserve;
     keep if od==2;
     svyset vcode [pw=weight], fpc(fpc_od) vce(linearized) || hhuic;
     svy: tab sex, ci;
     svy: tab agecat, ci;
     svy: mean ageyr;
     epctile ageyr, p(50) svy;
restore;



* RISK CRITERIA
```

```
* forest dweller;
replace hhinforest = trim(hhinforest);
assert inlist(hhinforest, "yes", "no");
gen dweller = hhinforest=="yes";
tab od dweller, row m;
lab var dweller "forest dweller";

* check that hhinforest is same for all participants in hh;
sort hhuic participantid;
l od vcode hhuic hhinforest participantid dweller if dweller[_n] != dweller[_n-1] & hhuic[_n]==hhuic[_n-1], sepby(hhuic) noobs;
 assert dweller[_n] == dweller[_n-1] if hhuic[_n]==hhuic[_n-1];


* forest goer;
tab goforest_any goforest_1day if !dweller,m;
gen goer = goforest_1day=="yes";
replace goer=. if goforest_any=="";
replace goer=0 if dweller==1;
assert goer!=1 if dweller==1;
tab goer if !dweller,m;
lab var goer "forest goer";

* at risk:  either dweller or goer;
* (dweller has no missings);
gen atrisk = dweller==1 | goer==1 if goer!=.;
lab var atrisk "forest dweller or goer";
tabstat dweller goer atrisk, s(n sum mean) by(od) notot;

* Unweighted risk proportions;
by od, sort: ci proportion dweller goer atrisk;

*** Weighted risk proportions - weighting scheme differs by OD ***;

* (svyset interprets fpc as N (pop size) if fpc>1);


* Phnom Srouch;
* villages are strata since all were included;
preserve;
     keep if od==1;

     * ESTIMATES USING UNSTANDARDIZED WEIGHTS - PHNOM SROUCH;
     svyset    hhuic    [pw=weight],    strata(vcode)    fpc(fpc_village) vce(linearized);
     svy: proportion dweller goer atrisk;

     * ESTIMATES USING STANDARDIZED WEIGHTS - PHNOM SROUCH;
     qui summ weight;
     gen sweight = (weight - r(mean)) / r(sd);
     *svyset    hhuic    [pw=sweight],    strata(vcode)    fpc(fpc_village) vce(linearized);
     *svy: proportion dweller goer atrisk;

     di "STANDARDIZED SAMPLING WEIGHTS for Phnom Srouch";
```

```
      summ sweight, d;
restore;

* Senmonorom;
* villages are stage 1 since they were sampled;
preserve;
      keep if od==2;

      * ESTIMATES USING UNSTANDARDIZED WEIGHTS - SENMONOROM;
      svyset vcode [pw=weight], fpc(fpc_od) vce(linearized) || hhuic;
      svy: proportion dweller goer atrisk;

      * ESTIMATES USING STANDARDIZED WEIGHTS - SENMONOROM;
      qui summ weight;
      gen sweight = (weight - r(mean)) / r(sd);
      *svyset vcode [pw=sweight], fpc(fpc_od) vce(linearized) || hhuic;
      *svy: proportion dweller goer atrisk;

      di "SAMPLING WEIGHTS for Senmonorom";
      summ sweight, d;
restore;

* Village-level variation in risk estimates (unweighted);
* the cause of the wide CIs;
by od, sort: table vcode, c(mean dweller mean goer);



/*** MOBILITY CORRECTION FACTORS FOR DISTRICTS ****
*
* survey participants were:
*   residents, anyone who regularly sleeps at this HH
*    visitors, anyone else who slept at this HH last night
*/;


* In  the past 4 weeks, did any at-risk participants stay overnight ;
* in other villages included in the study?;
tab od overnight if atrisk, m;

* a lot of mobility within village and district;
by od, sort: tab overnight overnight_location if atrisk,m;

* were they gone long enough to risk getting counted ;
* as a "usual resident" in another HH in the study area?;
*    18% of those who traveled spent >=2 weeks in the past 4 weeks there;
*    48% spent >=1 week;
tab overnight_time_there if atrisk & overnight=="yes";
tab    overnight_time_here    overnight_time_there    if    atrisk    &
overnight=="yes";

* Mobility identifier for districts;
assert inlist(overnight, "yes", "no");
gen mobile_district = inlist(overnight_location,
                                      "another_village_in_this_district",
```

```
                                        "this_village");
tab od mobile_district if atrisk, row m;


* Phnom Srouch;
preserve;
     keep if od==1;
     svyset    hhuic    [pw=weight],    strata(vcode)    fpc(fpc_village)
vce(linearized);
     svy, subpop(if atrisk): proportion mobile_district;

restore;

* Senmonorom;
preserve;
     keep if od==2;
     svyset vcode [pw=weight], fpc(fpc_od) vce(linearized) || hhuic;
     svy, subpop(if atrisk): proportion mobile_district;
restore;

* Most stayed overnight elsewhere for WORK;
tab overnight_rsn if mobile_district==1;


*** Mobility correction factor for villages ***;

* Mobility identifier for districts;
gen mobile_village = overnight_location=="this_village";
tab od mobile_district if atrisk, row m;



* MULTIPLIER ESTIMATES;

* whether village got BITE - version corrected by Dyna;
* 0/1/99.  99s represent villages where BITE was distributed just once
(instead of 4 times);
* exclude 99s from the multiplier estimates;
*  exclude  Phnom  Srouch  (Kampong  Speu  province)  from  the  multiplier
estimates;

assert inlist(bite_corrected, 0, 1, 99);
recode bite_corrected (99=0), gen(bite_village);
replace bite_village=0 if od==1;
lab var bite_village "Village in Senmonorom OD and received 4 rounds of
BITE distribution";
assert inlist(bite_village, 0, 1);

* check whether same within village (confirmed);
tab vcode bite_village,m;
table vcode, c(min bite_village max bite_village);
tab village_name if bite_village==1;

* area unique identifier (sub-annex or main village);
gen area = string(vcode) + " " + village_name;
```

```
* Total N and #at-risk in areas;
table area if bite_village==1, c(count vcode sum dweller sum goer sum
atrisk);


* received any BITE items?;

* check consistency of responses;
tab bite_got bite_topical,m;
tab bite_got bite_spatial,m;
tab bite_got bite_etof,m;
tab bite_ot,m;
tab bite_oth_specify,m;

assert inlist(bite_got, "yes", "no", "n/a");

foreach x in "topical" "spatial" "etof"
{;
     assert inlist(bite_`x', "0", "1", "n/a");
     gen `x' = bite_`x'=="1" if inlist(bite_`x', "0", "1");
     replace `x'=0 if bite_got=="no";

     * respondent said their village didn't get BITE, so they weren't
asked
     * if they personally got BITE items (recorded as 'n/a');
     * recode as 0 so they are included in the denominator;
     replace   `x'=0   if   bite_village==1   &   bite_selfreport=="no"   &
bite_got=="n/a";
     lab var `x' "received `x' during period";
};
gen bite_any = (topical==1 | spatial==1 | etof==1) if topical!=.;
lab var bite_any "received a BITE tool during period";

* got any BITE items, all areas in survey;
by od, sort: summ topical spatial etof bite_any;

* got any BITE items, multiplier areas only;
summ topical spatial etof bite_any if bite_village;

* # participants who received BITE tools;
tab area bite_any if bite_village==1 & atrisk==1;

* calculate p(received BITE) w/CI by village, Senmonorom OD only;
keep if od==2;
     svyset vcode [pw=weight], fpc(fpc_od) vce(linearized) || hhuic;
     svy, subpop(if bite_village==1 & atrisk==1): proportion bite_any;

foreach a in    "30 Chak Cha Village"
                     "35 Pu Char"
                     "37 K uon"
                     "38 Kdaoy"
{;
     di " ";
     di "AREA: `a'";
```

```
        di " ";
        svy, subpop(if bite_village==1 & area=="`a'"): proportion bite_any;

        * mobility correction factor for `a';
        tab mobile_village if area=="`a'";
        svy, subpop(if area=="`a'"): proportion mobile_village;

        * unadjusted estimate for `a' to calculate DEFF;
        ci prop bite_any if bite_village==1 & area=="`a'";
  };
```