

# UC Irvine

## UC Irvine Previously Published Works

### Title

Human movement patterns of farmers and forest workers from the Thailand-Myanmar border.

### Permalink

<https://escholarship.org/uc/item/6190d491>

### Authors

Tun, Sai

Min, Myo

Aguas, Ricardo

et al.

### Publication Date

2021

### DOI






10.12688/wellcomeopenres.16784.2

Peer reviewed



RESEARCH ARTICLE

**REVISED** **Human movement patterns of farmers and forest workers from the Thailand-Myanmar border [version 2; peer review: 2 approved, 2 approved with reservations]**

Sai Thein Than Tun <sup>1,2</sup>, Myo Chit Min<sup>3</sup>, Ricardo Aguas <sup>1,2</sup>,  
Kimberly Fornace <sup>4,5</sup>, Gay Nay Htoo<sup>3</sup>, Lisa J. White <sup>2,6</sup>, Daniel M. Parker <sup>7,8</sup>

<sup>1</sup>Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

<sup>2</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>3</sup>Shoklo Malaria Research Unit, Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Mae Sot, Thailand

<sup>4</sup>Centre for Climate Change and Planetary Health, London School of Hygiene & Tropical Medicine, London, UK

<sup>5</sup>Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

<sup>6</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK

<sup>7</sup>Department of Population Health and Disease Prevention, University of California, Irvine, CA, 92697, USA

<sup>8</sup>Epidemiology and Biostatistics, University of California, Irvine, CA, 92697, USA

**V2** First published: 14 Jun 2021, 6:148  
<https://doi.org/10.12688/wellcomeopenres.16784.1>

Latest published: 02 Oct 2023, 6:148  
<https://doi.org/10.12688/wellcomeopenres.16784.2>

### Abstract


**Background:** Human travel patterns play an important role in infectious disease epidemiology and ecology. Movement into geographic spaces with high transmission can lead to increased risk of acquiring infections. Pathogens can also be distributed across the landscape via human travel. Most fine scale studies of human travel patterns have been done in urban settings in wealthy nations. Research into human travel patterns in rural areas of low- and middle-income nations are useful for understanding the human components of epidemiological systems for malaria or other diseases of the rural poor. The goal of this research was to assess the feasibility of using GPS loggers to empirically measure human travel patterns in this setting, as well as to quantify differing travel patterns by age, gender, and seasonality among study participants.

**Methods:** In this pilot study we recruited 50 rural villagers from along the Myanmar-Thailand border to carry GPS loggers for the duration of a year. The GPS loggers were programmed to take a time-stamped reading every 30 minutes. We calculated daily movement ranges and multi-day trips by age and gender. We incorporated remote sensing data to assess patterns of days and nights spent in forested or farm areas, also by age and gender.

### Open Peer Review

Approval Status

	1	2	3	4
<b>version 2</b> (revision) 02 Oct 2023		 view	 view	 view
<b>version 1</b> 14 Jun 2021	 view	 view		

1. **Steven Stoddard**, San Diego State University, San Diego, USA
2. **Lisa Sattenspiel** , University of Missouri, Columbia, USA
3. **Nick Ruktanonchai**, Virginia Tech, Blacksburg, USA
4. **Yao Li**, University of Maryland at College Park, College Park, USA  
Johns Hopkins Bloomberg School of Public Health

**Results:** Our study showed that it is feasible to use GPS devices to measure travel patterns, though we had difficulty recruiting women and management of the project was relatively intensive. We found that older adults traveled farther distances than younger adults and adult males spent more nights in farms or forests.

**Conclusion:** The results of this study suggest that further work along these lines would be feasible in this region. Furthermore, the results from this study are useful for individual-based models of disease transmission and land use.

### Keywords

human movement, human ecology, disease ecology, infectious disease epidemiology, Thailand-Myanmar border, forests, farms

Health, Baltimore, USA

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

**Corresponding author:** Daniel M. Parker ([dparker1@hs.uci.edu](mailto:dparker1@hs.uci.edu))

**Author roles:** **Tun STT:** Data Curation, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Min MC:** Data Curation, Investigation, Methodology, Writing – Review & Editing; **Aguas R:** Formal Analysis, Investigation, Methodology, Visualization, Writing – Review & Editing; **Fornace K:** Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Htoo GN:** Project Administration, Supervision, Writing – Review & Editing; **White LJ:** Formal Analysis, Methodology, Visualization, Writing – Review & Editing; **Parker DM:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This research was funded through a subaward from the Regional Artemisinin Initiative from the Global Fund to Fight AIDS, Tuberculosis and Malaria (COPRDDC-QMU-M-N-02/RAI). STTT was supported by the Wellcome Trust (205240) and by the Bill and Melinda Gates Foundation (OPP1193472). This study was also a part of the Wellcome-Trust Major Overseas Programme in SE Asia (220211).

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2023 Tun STT *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Tun STT, Min MC, Aguas R *et al.* **Human movement patterns of farmers and forest workers from the Thailand-Myanmar border [version 2; peer review: 2 approved, 2 approved with reservations]** Wellcome Open Research 2023, **6**:148 <https://doi.org/10.12688/wellcomeopenres.16784.2>

**First published:** 14 Jun 2021, **6**:148 <https://doi.org/10.12688/wellcomeopenres.16784.1>

**REVISED Amendments from Version 1**

We've made several revisions based on reviewer comments and critiques. In particular, we've updated the introduction/background section to include reference to Prothero, who was doing work on mobility and infectious diseases over half a century ago. We've changed the introduction so that the impetus more directly aligns with our goals in this research rather than focusing on our perception that there is a lack of empirical work on human mobility in this type of setting (rural, remote, conflict setting). We've added more details about the context of this setting in the methods section and have added some notes about illnesses among study participants. We've also attempted to make our participant recruitment approach more clear and have softened some of our statements about our results, given that this was a cohort study and the participants were not randomly selected. Finally, before we began this study we did some tests of locational accuracy on the GPS devices and we've attempted to make these results more accessible as they are important for understanding our analytic approaches and our results.

**Any further responses from the reviewers can be found at the end of the article**

**Introduction**

Human movement or travel is important with regard to infectious disease epidemiology and ecology<sup>1,2</sup>. Infectious diseases are heterogeneously distributed across landscapes. Individuals may be exposed to greater risk of acquiring infection if they move through transmission hotspots. Infectious individuals who travel may disperse pathogens across the landscape. Healthcare facilities are also heterogeneously distributed across landscapes, with ramifications for individual, household, and community access to diagnosis and treatment. Generally speaking, individuals who must travel long distances or through difficult terrain in order to seek diagnosis or treatment are less likely to receive adequate treatment<sup>3-5</sup>.

As early as the 1950s, the human movement was recognized as one of the most important factors for disease elimination and eradication<sup>6</sup>. A growing number of research projects, some focused on health, are recording human movement patterns<sup>7-12</sup>. There have been attempts to map human movement in the rural Thailand border areas to delineate and intervene the risks of malaria<sup>13-15</sup>. These projects can be broadly divided into those that are based on questionnaires/interviews and those that are based on empirical measurements (GPS devices, mobile phones, tweets, etc.) All approaches have strengths and weaknesses<sup>16</sup>. Interview/questionnaire-based approaches are prone to recollection bias and some movements may be unreported because of their nature (for example, if movements are made for illegal purposes or to places that participants don't want to discuss/report).

Mobile phone records provide a source of movement information across broad swaths of many populations<sup>17</sup>. However, the movement data are limited to the resolution of mobile tower density, and mobile phone towers are not evenly distributed across landscapes (they tend to be clustered in urban settings). There is bias in who owns and uses mobile phones as well<sup>18</sup> and mobile phone records will not allow for fine-scale mapping of the routes travelled in between locations<sup>19</sup>.

Wearable GPS devices offer extremely detailed data, but are labor intensive and dependent on volunteer cohort members. However, as the devices have become more compact (increasing wearability) and have become more affordable, their use is increasingly common<sup>10,20-26</sup>.

The main goals of this pilot project were to: I.) assess the feasibility of using GPS loggers to track human movement patterns among people living on the Thailand-Myanmar border, and II.) measure human movement patterns, including how they vary seasonally, among a cohort of participants. The results of this work have implications for further research in this region with regard to targeted public health interventions, normal travel patterns and related exposure to different environments, for individual risk of infection by various diseases (e.g. SARS-CoV-2, malaria, melioidosis), and with regard to human disease ecology. The resulting data can also be useful for calibrating human movement patterns of individuals in an individual based modelling system.

**Methods****Context of the study area**

The study area is on the Thailand-Myanmar border. Participants were recruited from clinics that serve rural, mostly underdeveloped, and low population density communities. Most participants were of the Karen ethnic group. Villages were made up of a few dozen of mostly multigeneration families living in stilt houses made of wood and thatched roof. Villages didn't normally have schools, clinics, or sanitary toilets. The houses are normally located along the main dirt road of the village. The dirt roads then continued to connect to other villages and small towns through a hilly and rugged terrain with occasional watersheds and river basins which made traversing difficult, especially during the rainy season.

Villagers made their living mostly through agriculture, but they have to undertake various types of jobs throughout the year for their subsistence. They developed land in and around their villages into farms to cultivate rice and vegetables. They farm poultry and pigs under their stilt houses. Some villagers go into the forests for hunting, or for foraging wood, and to collect wild edibles. They would go to the farms and forests overnight occasionally, and sleep without much protection from mosquitos and insects. We focused on farms and forests as places of interest in this study since apart from their homes, farms and forests might be the places the villagers spend significant amount of their time while being vulnerable to infectious diseases such as malaria.

**Data**

The study period began in March 2017 and ended in February 2018 and aimed to recruit 50 participants for a one-year duration of time (ClinicalTrials.gov Identifier: [NCT03087214](https://clinicaltrials.gov/ct2/show/study/NCT03087214), March 22 2017). The study size was purposive as this was an exploratory pilot study and no power analysis was used. Prior to the study beginning we held community engagement meetings with community elders in the Tak Province Community Ethics Advisory board (T-CAB) to explain the project. The study locations were selected because of community enthusiasm to

participate and operational feasibility. Participants were recruited from 10 villages (the lowest level of administrative division in Myanmar) near two clinics on the Thailand-Myanmar border: Wang Pah and Maw Ker Tai Clinics (*Extended data*: Figure S1<sup>27</sup>). These clinics primarily serve migrant and cross border populations and have connections to village health workers in nearby villages. We reached out to village health workers in the nearby villages to explain the project and to ask if they could help us recruit participants from their respective villages. Participants were recruited and interviewed at the respective clinics. There were no house visits and all data were collected at the clinics.

The study targeted individuals who were 18 years of age and above from the Karen or Burmese ethnic groups, who stated that they would be able to keep track of the GPS device, who were capable of walking outside of village boundaries at recruitment, and who were willing to provide written consent to the study. As incentives, participants were provided with a waterproof handbag at the beginning of the study, a headlamp in the middle of the study, and a jacket at the end of the study. The total cost of incentives per person was less than 10 GBP.

Upon recruitment, the age and gender of each participant was recorded following receipt of written informed consent (in Karen language). Participants were asked to carry i-gotU GT-600 (46x41.5x14mm) mobile GPS devices for the study. They have a reported average location error of less than 10 meters<sup>28</sup>. They were programmed to take a reading every 30 minutes. The devices are equipped with motion sensors and were set to go into a dormant mode if they sat still for longer than one hour, and to resume taking GPS readings upon detection of movement. Devices were also set to take readings at one-minute intervals if the device was moving  $\geq 15\text{km/hour}$  (travelling by vehicle rather than walking).

Field managers (one per clinic, MCM and GNH) met with study participants each month. During these meetings participants were questioned about their continued willingness to participate in the study by the field managers, their general movement patterns during the previous month and with regard to any illnesses. A newly charged GPS device was given to each participant (two GPS devices were devoted to each study participant, total of 100 devices used) during each of these monthly meetings and the GPS logger that had been carried during the previous month was collected for re-charging. The GPS device batteries last roughly 1 to 1 ½ months.

The data were transferred to a computer and stored in an encrypted folder with a unique code for each person to maintain security and anonymity. Separate data files were combined to obtain aggregated, longitudinal data for each participant. QGIS version 3.4.9 was used to generate study location maps and to visually explore the raw data. R statistical software version 4.0.3 was used for the data processing and analysis<sup>29,30</sup>, using the “sp”, “rgdal”, “raster”, “proj4”, “reshape” and “ggplot2” R packages<sup>31–36</sup>. GPS coordinates, which were originally recorded in 1984 World Geographical Coordinate System (WGS 84), were projected to UTM zone 47N to perform geographical calculations.

Land cover types (farms and forests) were classified manually (by hand) using satellite imagery from Google Earth (version 7.3.3.7786). Farms could be differentiated from forests by presence of human intervention on the vegetation cover e.g., vegetation cover in the farms were in more or less neatly arranged rows/columns. While formal ground truthing was not done after categorization, the locations of farms and forests do correspond to our experiences on the ground in these villages.

## Analysis

Our analyses focused on quantifying daily movement ranges, multi-day trips, and time spent in farm or forest areas across population strata.

The last GPS point of the day between 6pm to 12 midnight was considered to be the location where an individual spent the night. The median center of these points was assumed to be the individual’s home location. A buffer with 266 meters radius (which is the standard distance deviation of accuracy of the GPS device placed inside a bag inside a house; details in *Extended data*: Figure S2<sup>27</sup>) was created around each home to create a polygon (a GIS object with a series of x and y coordinate pairs that represents an enclosed area on a map) for home area. Polygons for the farms and forests were manually classified using satellite imagery from Google Earth.

As a proxy for how far people move each day, we calculated the maximum daily Euclidian distance, which is the furthest Euclidian distance a person was away from the location he or she slept the previous night. Multiday trips away from home were identified when the minimal daily Euclidian distances were more than 266 meters from the individual’s home location consecutively for two or more days.

The Wilcoxon rank-sum test was used to compare the distributions of maximum daily Euclidian distances. A negative-binomial generalized linear mixed-effects model was used to investigate potential associations between the total number of nights spent in the farms or forests (response variables) and other characteristics such as age group, gender, and season (exploratory variables). As there were multiple observations per individual (for each time step), a random intercept was used for individuals.

Utilization of places (home, farms or forests) for each person was estimated using two different approaches. The first method was by checking whether more than two temporally-consecutive GPS points of a person fall within a polygon designated for the person’s home (for this particular measurement, the polygon is a circle of radius 266 meters around the person’s home location), farms, or forests on each day. This is equivalent to checking if a person spent at least an hour within the same polygon. For each participant, the number of days spending in each category of place (home, farm, forest) was divided by the total number of days participated during the study period to obtain the proportion of being at the respective places.

The second method estimated the utilization of places by a biased random bridge (BRB) technique<sup>37,38</sup>. Unlike prior



methods for estimation of utilization of places such as location-based kernel density estimations (LKDE), BRB takes the activity time between successive relocations into account and models space utilization as a time-ordered series of points to improve accuracy and biological relevance while adjusting for missing values. BRB estimates the probability of an individual being in a specific location during the study time period and can be used to estimate home range (the area where individuals spend a defined percentage of their time).

To parameterize BRB models for each individual, we considered points collected more than three hours apart to be uncorrelated. However, the two temporally-consecutive points that are deemed uncorrelated by the prior cutoff, may in fact be correlated (e.g., when individuals go to sleep for more than three hours in a single location). Without manually adding points between them, this method will underestimate the usage of homes. An individual is considered stationary when the distance between two consecutive points is less than 10 meters. The minimum standard deviation in relocation uncertainty is set at 30 meters. For each individual, estimation for the usage of different places was done for the whole study period (i.e. for the duration of his/her contribution) and for each season as described below.

In Central and Southern Myanmar, the monsoon rain starts in mid-May and ends in mid-October<sup>39,40</sup>. Therefore, we split the data on 15<sup>th</sup> May 2017 and 15<sup>th</sup> October 2017, and the period between the two dates was regarded as the “rainy season”. Mid-October to mid-March is the “cool and dry season”, mid-March to mid-May is the “hot and dry season”. Combinations of the two dry seasons had been used simply as the “dry season” in some of the analyses.

### Ethics statement

Approval for this research project was obtained from the Faculty of Tropical Medicine Ethics Committee, Mahidol University (TMEC 17-007); by the Oxford Tropical Research Ethics Committee (OxTREC reference: 503-17); and by the Tak Province Community Ethics Advisory Board (T-CAB reference: TCAB-04/REV/2016). All participants provided written informed consent in the Karen language.

### Results

A total of 50 persons participated for at least two seasons during the one-year study period. The age and gender distribution of the participants can be found in [Table 1](#). Female participation was low (n= 10). Efforts were made to increase female recruitment but many women declined, stating that they

did not normally leave their homes or villages and therefore thought they would not be interesting for the study. Most participants (29 out of 50) were in the 20–40 age group. Individual duration of participation differs between participants (*Extended data*: Figure S3<sup>27</sup>). The mean percentage of days GPS points were actually observed for each participant over the study is 86.17% (median: 91.21, IQR: 79.54 to 96.93).

During the study period, a 39-year old male was diagnosed with *Plasmodium vivax* malaria. A 45-year old male sustained a non-fatal gunshot injury while going to his farm one early morning. A 30-year old male passed away from the “cancer inside stomach”, the illness that was made known to the field supervisor seven months after being in the study when his mobility became restricted.

### Daily movement ranges

The violin plot of the maximum daily Euclidian distances traveled in kilometers in log<sub>10</sub> scale ([Figure 1](#)) shows that there is a bimodal distribution for all three age groups. The violin plot is a hybrid of kernel density plot and box-plot with the axes flipped that is particularly used to describe data with multimodal distribution. In the figure the vertical axis is the distance value in kilometers with the smallest value at the bottom, and the horizontal axis shows the density value. The heights and peaks in the following results refer to the width/broadness of the violins in the horizontal axis. The first peak was between 0.01 to 0.1 kilometers (10 and 100 meters) and the second peak was between 1 and 10 kilometers. The relative heights of the two peaks differ in different age groups. For under 20s, the first peak is over 20% higher (i.e. they have higher proportion of daily maximum distance close to where they were the previous night) compared to the second peak. The difference between the two peaks in the other two age groups is less than 10%.

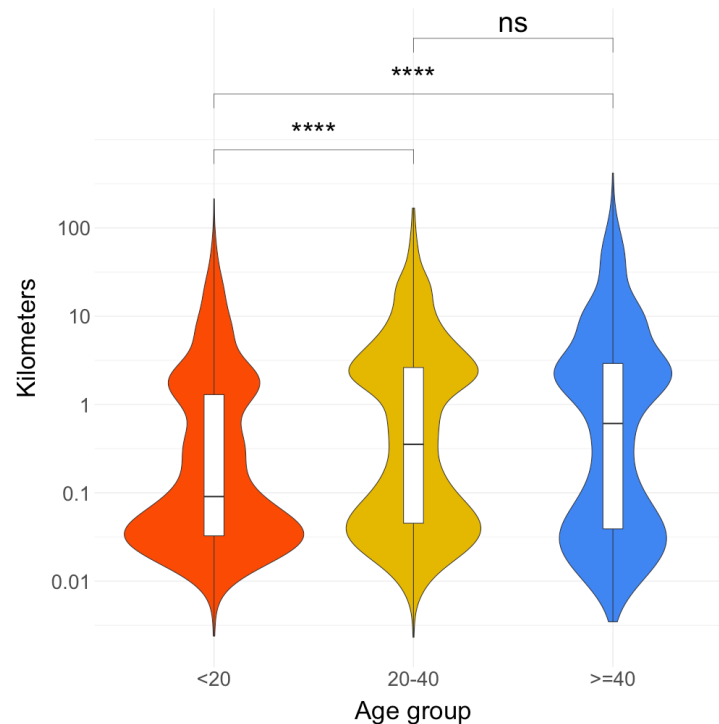
The Wilcoxon rank-sum tests provided evidence that 20–40 and over-40 age groups have greater maximum daily Euclidian distances away from home compared to under-20 age group on average. Further disaggregation of this data by gender, and age group can be found in the *Extended data*: Figure S4<sup>27</sup>.

### Multiday trips

Participants may make trips that would last several days, either because their destination could not be reached within a single day or because they stayed at their destination for several days (e.g. staying at a farm hut). Using a buffer radius of 266 meters around their home GPS points as their home locations, we calculated the number of consecutive days they spent away from home. Aside from two participants (an over-40 male and an under-20 female), all other participants had at least one trip with more than two consecutive days away from home during their participation period. Trips of less than 10 consecutive days are the most frequent among the participants. There are male outliers of over 20-years old (n=6) who took shorter consecutive day trips (2–5 days) over 10 times. Making trips of over 10 consecutive days was relatively uncommon, but 21 participants still made at least one trip of over 20 consecutive days away from home. Details are available in the *Extended data*: Figure S5<sup>27</sup>.

**Table 1. Age and gender distribution of the participants.**

Age group	Less than 20	20–40	40 and above	Total
Male	7	24	9	40
Female	2	5	3	10
Total	9	29	12	50



**Figure 1. Maximum daily Euclidian distances traveled by participants in kilometers.** Distance was calculated from the location a person was at the end of the prior night (most often, this location is their home location). Wilcoxon rank-sum test results are shown on the top of the lines connecting the age groups chosen for the tests. “ns” represents a p-value of  $> 0.05$ . \*\*\*\* represents a p-value of  $\leq 0.0001$ .

### Days spent at the forest or farm

For each participant, we identified the number of days spent at farms, forests, or at one’s home, and looked for an association between farm visits and forest visits. Here we assumed that having at least two GPS points in the polygon of a particular place constitutes using the respective place for that day, and that a person can be at various types of places in a single day. We found that if a person spent a higher proportion of days at the farms, she or he will likely spend a lower proportion of days at the forests, and vice versa, even though both being at the farms and being in the forests are possible on the same day.

Figure 2 shows the distribution of the proportion of the number of days being at the farms, forests or home for different age groups. All participants were found to be at their respective home for the majority of days. Compared to other age groups, the 20–40 age group had a higher proportion of time spent in the forests. The under-20 group had the highest proportion of time spent in the farms on average, followed by the 20–40 age group.

### Time spent at the forest or farm

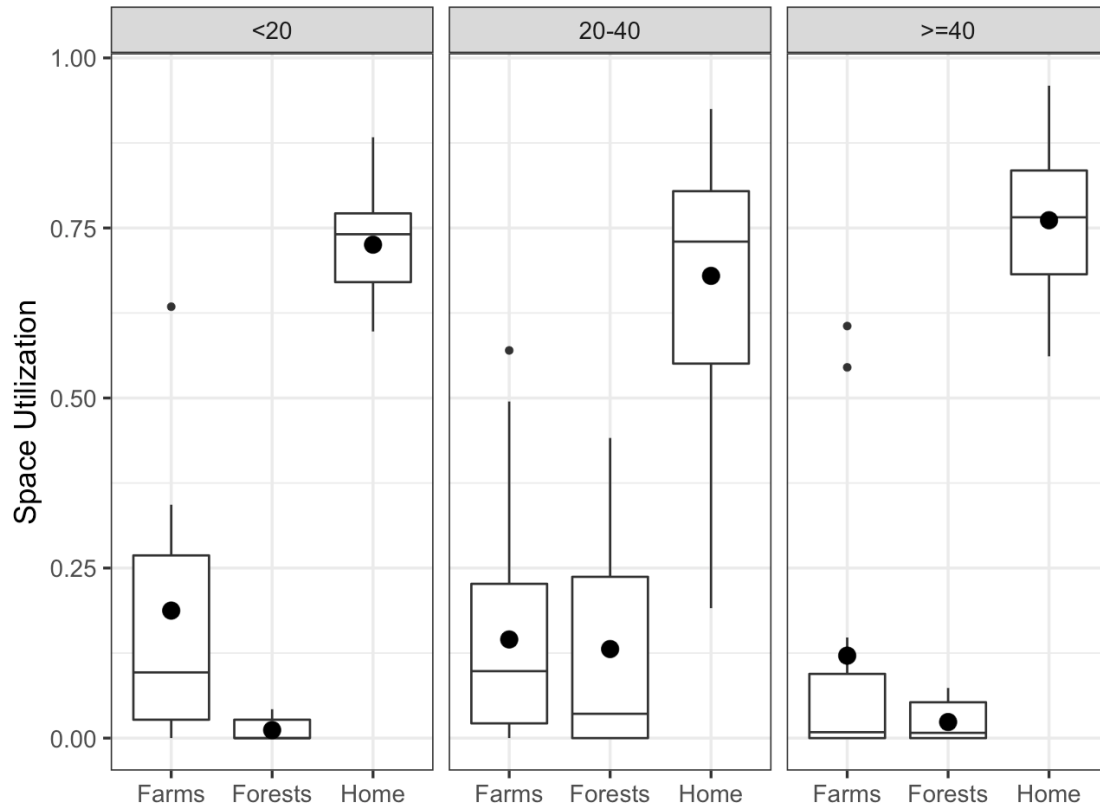
We also combined the geographic information of farms and forests with the place utilization estimated from a biased-random bridge (BRB) algorithm, and calculated the utilization of

each specific place over the study period (*Extended data: Figure S6<sup>27</sup>*). An example of the place utilization of a person can be seen in Figure 3. On average, participants in the under-20 age group spent 20.0% and 2.2% of their time in farms and forests, respectively. For the participants from the 20–40 age group the percentages are 7.6% and 7.4%, and for those in the over-40 age group, the percentages are 7.2% and 3.8%, respectively.

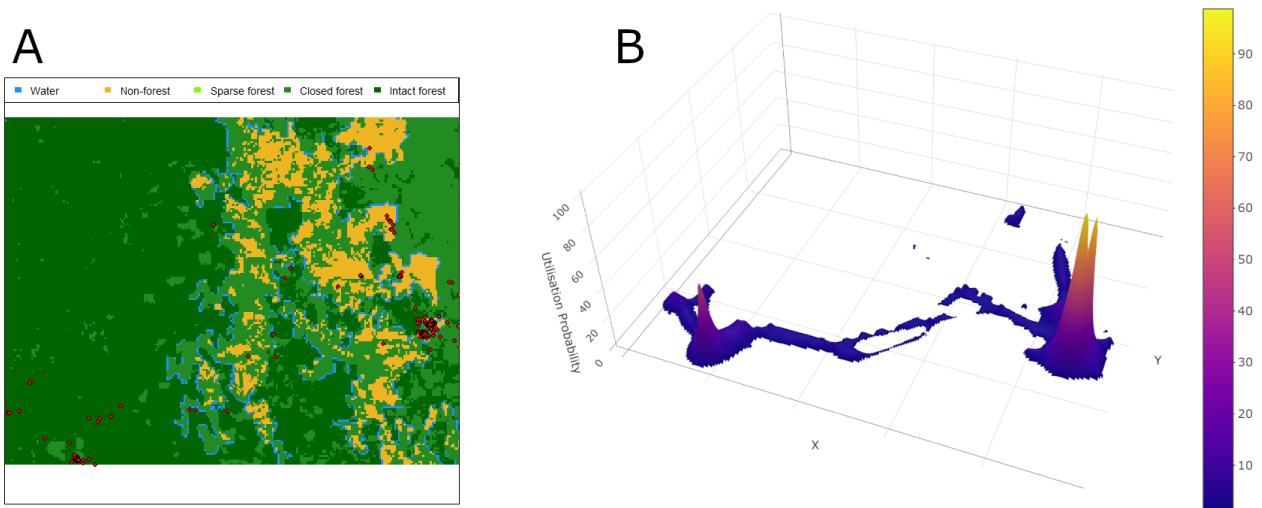
### Nights spent in the forest or farm

Being in the farms and forests at night might impose increased risks of diseases such as malaria because of potential exposure to important mosquito vector species (i.e. *Anopheles dirus*). As seen in Figure 4, we looked at the total number of nights participants spent in the farms or in the forests. Two female participants (20% of females) spent at least a night in the farm compared to 22 male participants (55% of males). As for spending at least a night in the forest, there were 21 males and only one female. Most participants in the 20–40 age group spent at least one night in the farm (18 out of 29, 62%) and in the forest (16 out of 29, 55%) whereas fewer than 35% of participants from under-20 and over-40 age groups spent a night in such places.

The negative binomial regression provided strong evidence that males in this cohort were more likely to spend nights in



**Figure 2.** Utilization of the farm, forest, and home (calculated as a proportion of number of days being at the respective places) over the participation period for different age groups. The bigger dots represent the mean values, while the smaller dots represent the outliers.

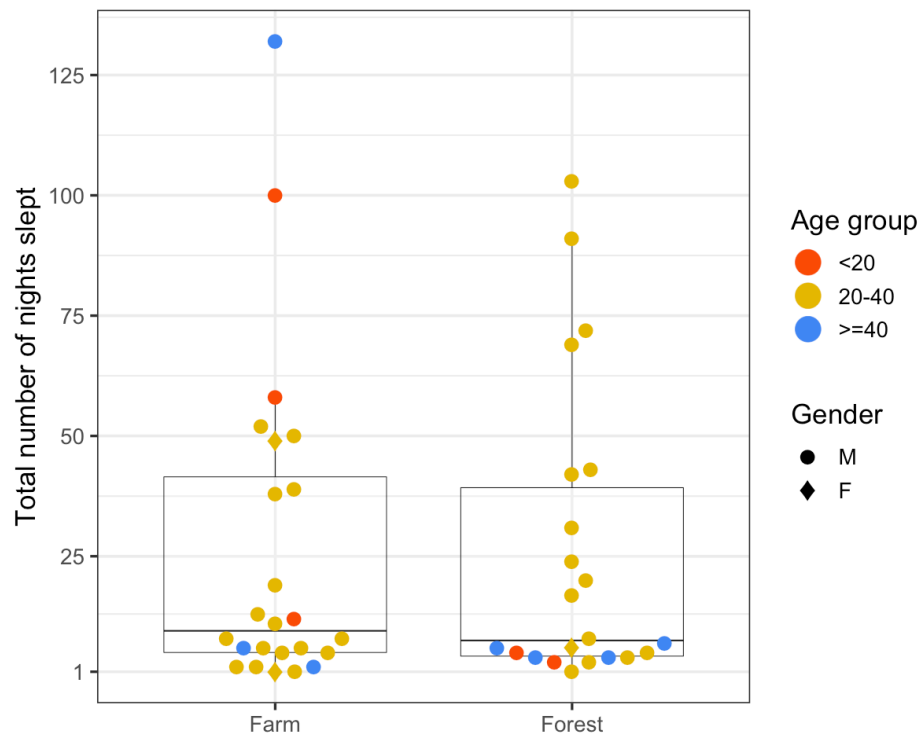


**Figure 3.** **A.** Example of GPS points (red points) recorded for a person. **B.** Corresponding utilization probability calculated from the GPS points. Its 3D version can be found [here](#).

farms ( $p=0.045$ ) and in forests ( $p=0.01$ ) compared to females, and that young adults (the 20–40 age group) were more likely to spend nights in the forest compared to the under-20 age group ( $p=0.043$ ), after controlling for the remaining variables (Table 2).

Participants may spend consecutive nights in the farms or the forests without going back home. The number of consecutive nights spent in the farms or the forests is the subset of the multi-day trips mentioned in the previous section. Figure 5 quantifies this metric for different age groups and gender. Persons of all age

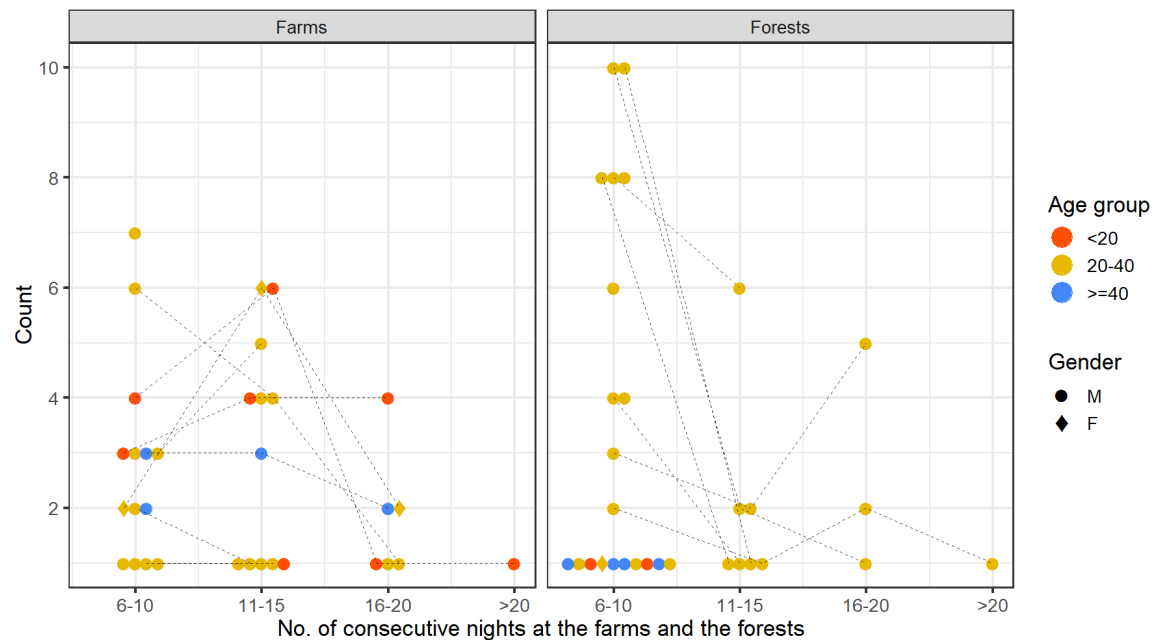




**Figure 4.** Total number of nights spent in the farms and the forests by each person over the participation period.

**Table 2.** Association between number of nights slept in farms or forests and age, gender, and season.

Predictors	Nights slept in farms		Nights slept in forests	
	Incidence rate ratios	p-value	Incidence rate ratios	p-value
(Intercept)	0.06 [0.00 – 2.01]	0.116	0.00 [0.00 – 0.17]	<b>0.005</b>
Age [<20]: comparator				
Age [>=40]	0.15 [0.01 – 3.24]	0.228	2.03 [0.09 – 46.25]	0.658
Age 20–40	1.43 [0.11 – 18.52]	0.784	16.80 [1.09 – 259.75]	<b>0.043</b>
Gender [F]: comparator				
Gender [M]	14.80 [1.07 – 205.70]	<b>0.045</b>	46.34 [2.47 – 869.07]	<b>0.010</b>
Season [Dry]: comparator				
Season [Rainy]	1.20 [0.66 – 2.19]	0.554	0.74 [0.35 – 1.58]	0.435
<b>Random effects</b>				
$\sigma^2$	1.35		1.66	
$\tau_{00}$	6.62 <sub>pid</sub>		5.66 <sub>pid</sub>	
ICC	0.83		0.77	
N	47 <sub>pid</sub>		47 <sub>pid</sub>	
Observations	89		89	
Marginal R <sup>2</sup> / Conditional R <sup>2</sup>	0.210 / 0.866		0.341 / 0.850	



**Figure 5. Number of consecutive nights spent in the farms and forests.** In each panel, each of the points in each column represents a person of a specific age group and gender, defined in the legend. A single person may contribute a point in each of the columns (e.g., in the panel named Farms, a single person may contribute one point for each of the ranges of consecutive nights). Dotted lines connect the points contributed by the same person across different columns.

groups and gender spent varying numbers of consecutive nights in the farms. An under-20 male spent the most consecutive nights (16–20 nights) in the farm. A female of 20–40 age-group and a male of over-40 age-group spent two episodes of 11–15 consecutive nights in the farm. In contrast, there was little demographic heterogeneity among those who spent consecutive nights in the forests. A few males of the 20–40 age group not only spent long periods of consecutive nights (more than six consecutive nights), but also frequently spent many short periods of consecutive nights (two to five nights) in the forests.

## Discussion

Many detailed human movement studies have been done, mainly in the regions of high socio-economic status. Our study presents an analysis of human movement in a remote rural area that has been under-studied with regard to human ecology (though do see 41,42). Compared to other studies where GPS loggers were used for a very short period of time, there is a relatively long duration of participation in our study. This makes it possible to examine potential seasonal variation.

Our data suggest a bimodal pattern of movement away from participant homes, with one peak nearby ( $\leq 100\text{m}$ ) and another one to three kilometers away from their homes (Figure 1). There were differences in these movement patterns by demography, with under-20s staying close to home on the majority of the days and both 20–40 and over-40 age groups tending to move farther away each day. We hypothesize that the reason for this difference is that over-20 age groups are more heavily

involved in subsistence activities (e.g., farming and foraging which are conducted further away from home) than the under-20 age group.

Multiday trips of less than 10 days are common among the participants. The metrics of multiday trips do not signify anything unless they are associated with the activities done during the trip which vary from visits to friends/family, getting supplies at the nearby town, farming, foraging, and other economic or subsistence activities.

All age groups in this study visited farm areas and spent the night in the farms, with no statistically significant difference found between age groups. When they spent their nights in the farms, they did it consecutively and on several occasions during the study period. Farming is one of the major forms of subsistence for rural families and it must be regarded as relatively safe compared to subsistence activities in the forests that all age groups partake in it. There was no seasonal variation in the number of nights spent at the farms in these data. Different types of crops are normally rotated over the year for cultivation in this region.

In contrast, going to and sleeping in the forests, which may involve foraging, logging, mining etc., is found to be the task for males of the 20–40 age group. The median number of nights slept in the forest among those who ever spent the night in the forest was 7.5. Only males of the 20–40 age group spent a higher number of nights in the forest than the median value. The same males (20–40 age group) were found to take

frequent and successive overnight trips to the forests. We surmise that the males in the 20–40 age group, most likely being the breadwinners of the family, are subject to any possible subsistence activities and are regarded as the most suitable persons to venture into the forests overnight despite dangers from wildlife and harsh living conditions. No seasonal variation was found in the number of nights of sleeping in the forest. In comparison, a questionnaire based movement survey conducted in similar Thai-Myanmar border area found seasonal movement patterns<sup>15</sup>.

Compared to home, sleeping places in the farms and forests may be more rudimentary, leaving people more vulnerable to medically important arthropods or other environmental risks (i.e. potentially more contact with venomous snakes, etc.). Spending several consecutive nights in the farms and forests may increase the chances of vector-borne diseases such as malaria since major malaria vectors in the area such as *Anopheles dirus*, and *Anopheles minimus* are found in the deep forests, forest edges, plantations and even in the rice fields<sup>43</sup>. Studies have found that the increased risk of malaria in forest-goers is contributed by inconsistent bed net usage, misconception that alcohol consumption or blankets provides protection against mosquito bites, non-participation in the malaria prevention activities held at the villages<sup>44</sup>.

Results from this study, particularly the space utilization data, would be useful in spatially explicit individual-based infectious disease model such as [45](#) which models the malaria elimination in the rural South East Asian region. Human mobility is a crucial part of many disease transmission dynamics, yet it has been ignored in many infectious disease models because of constraints on data and computational capacity. Compartmental models assume homogeneous mixing of individuals in their respective compartments. While they are quick to set up, they are not suitable for the disease elimination settings. Their homogeneous nature limits the modelers from exploring the impact of multiple interventions tailored towards different risk groups such as forest-goers in malaria intervention. Individual-based models could have individual specific properties and their related movement patterns thus achieving a heterogeneous population. Calibrating on the space utilization data of this study, such models could become more realistic in terms of transmission dynamics. They could provide more accurate and precise estimations to tackle infectious diseases cost-effectively.

The study has several limitations. It was a pilot study and had a limited sample size. Most participants were adult males. Potential female participants said they rarely go beyond village boundaries and thus were not eligible to be included in the study. It may have introduced a selection bias, but it points to the fact that the mobility preferences between the two genders were too different that they were essentially two different populations requiring separate analyses. The most commonly reported occupation was farming and most people in this study area, indeed, farm for at least part of the year. However, people in the study area usually perform different types of

work according to the season and assigning a single occupation to a person may not be appropriate. Employment in this region is almost entirely informal, and most working-age men will work in agriculture for part of the year and in other types of labor during other parts of the year. Responses to surveys about employment will therefore vary by the time of year, even within a single research participant.

While we believe that this cohort is representative of adult males in this setting, more studies that are demographically representative of rural villages in this setting could be useful for understanding differences in travel patterns by age and gender. Mobile GPS devices have their own limitations. As explored in the Extended data: Figure S2, their readings could be inaccurate. Because of their small size, their battery capacity was limited. During the study period, participants may have failed to carry the GPS device (intentionally if they engaged in activities that other people might think were illegal or sensitive in nature – or not). Mechanical failures may also cause problems in data collection. Even though the utmost care was taken to preserve data integrity, there could be errors and bias from data collection (due to device inaccuracies) or data manipulation. (described in the methods section under analysis and in *Extended data* Figure 2<sup>27</sup>). Categorization of land types such as farms and forests was done manually using satellite imagery. While the categories do match our authors' understanding of the area, no validation was done on the ground after categorization for this analysis. Our estimation of home location as the median center of all the GPS points where the participant spent the night, each of which in turn is derived from the last GPS point of the day between 6pm to 12 midnight, may not be robust enough to capture the actual home location. This could be overcome by having the field supervisors record each participant's home location with a GPS device in the future studies. Categorization of home area (266 meters around home location) may be too wide to discern land use that is very close to home.

Finally, the estimation of land utilization regardless of the method used is imperfect. Having two consecutive GPS points to constitute usage of the land area provide too crude a result (Figure 2). While the BRB method provide more accurate and precise estimates (Extended data: Figure S6), it is not without its caveats. The BRB approach assumes that consecutive points that were more than three hours apart were uncorrelated. Since the GPS logger went into sleep mode while stationary, the current land utilization estimation under-estimates the time spent motionless (e.g., sleeping) and hence resulting in lower usage of home in Extended data: Figure S6 compared to that in Figure 2.

## Conclusion

This study shows that it is feasible to use GPS loggers to document and quantify human movement patterns in this setting (the Thailand-Myanmar border). Most individuals who agreed to participate did so across multiple seasons. Further work using GPS loggers in this setting is likely feasible. We found that younger age groups spent more days around their

home compared to older age groups. Older age groups spent almost equal amounts of time both around their home and at places one to three kilometers away from their home. Males spent more nights in the farms and forests, especially those in the 20–40 age group. The resulting human movement characteristics can be incorporated in infectious disease modeling studies in similar regions and the operational and analytic lessons learned from this project are broadly applicable to other studies of human movement and travel.

available at: <https://github.com/SaiTheinThanTun/HumMovPatt>. The time-stamped locational data are restricted for confidentiality reasons (i.e. it would be possible to identify the location of participant homes with these data). Access to sanitized data, through which participant home locations cannot be identified, will be considered on a case-by-case basis. Please contact Daniel M. Parker ([dparker1@hs.uci.edu](mailto:dparker1@hs.uci.edu)) with queries about data access

#### Extended data

All R code for this analysis is available at: <https://github.com/SaiTheinThanTun/HumMovPatt>

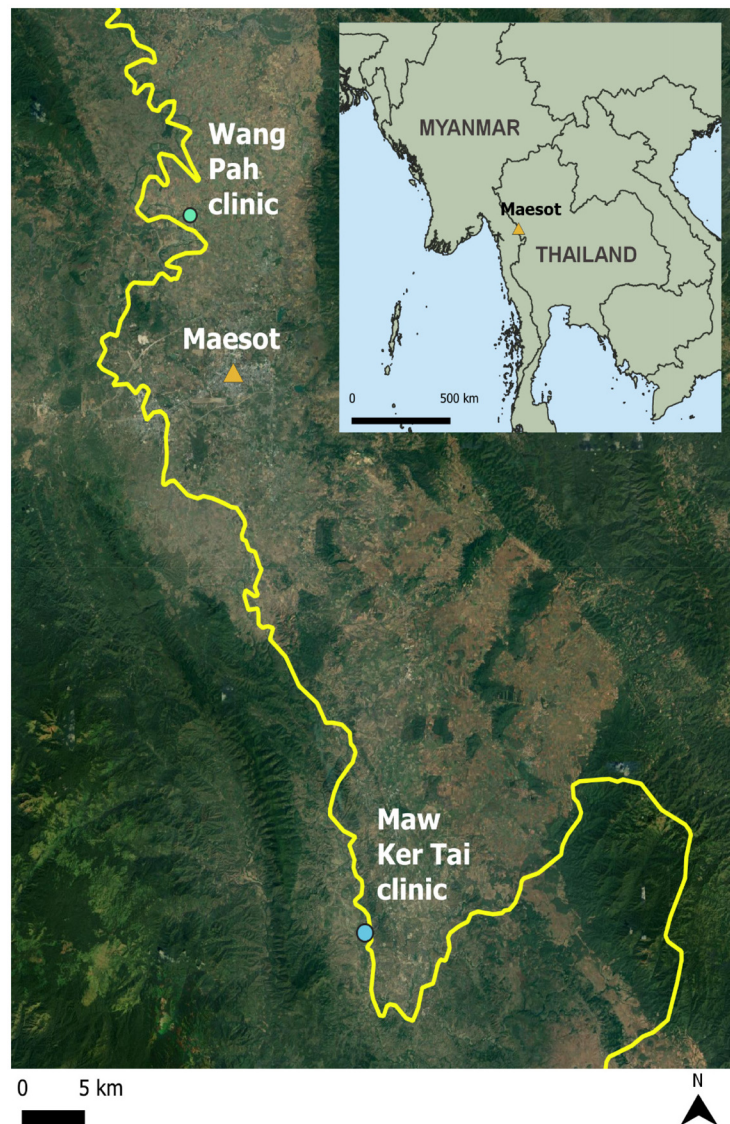
This project contains the following extended data:

- Figure S1 (Study site and location of clinics that were used for recruitment)

#### Data availability

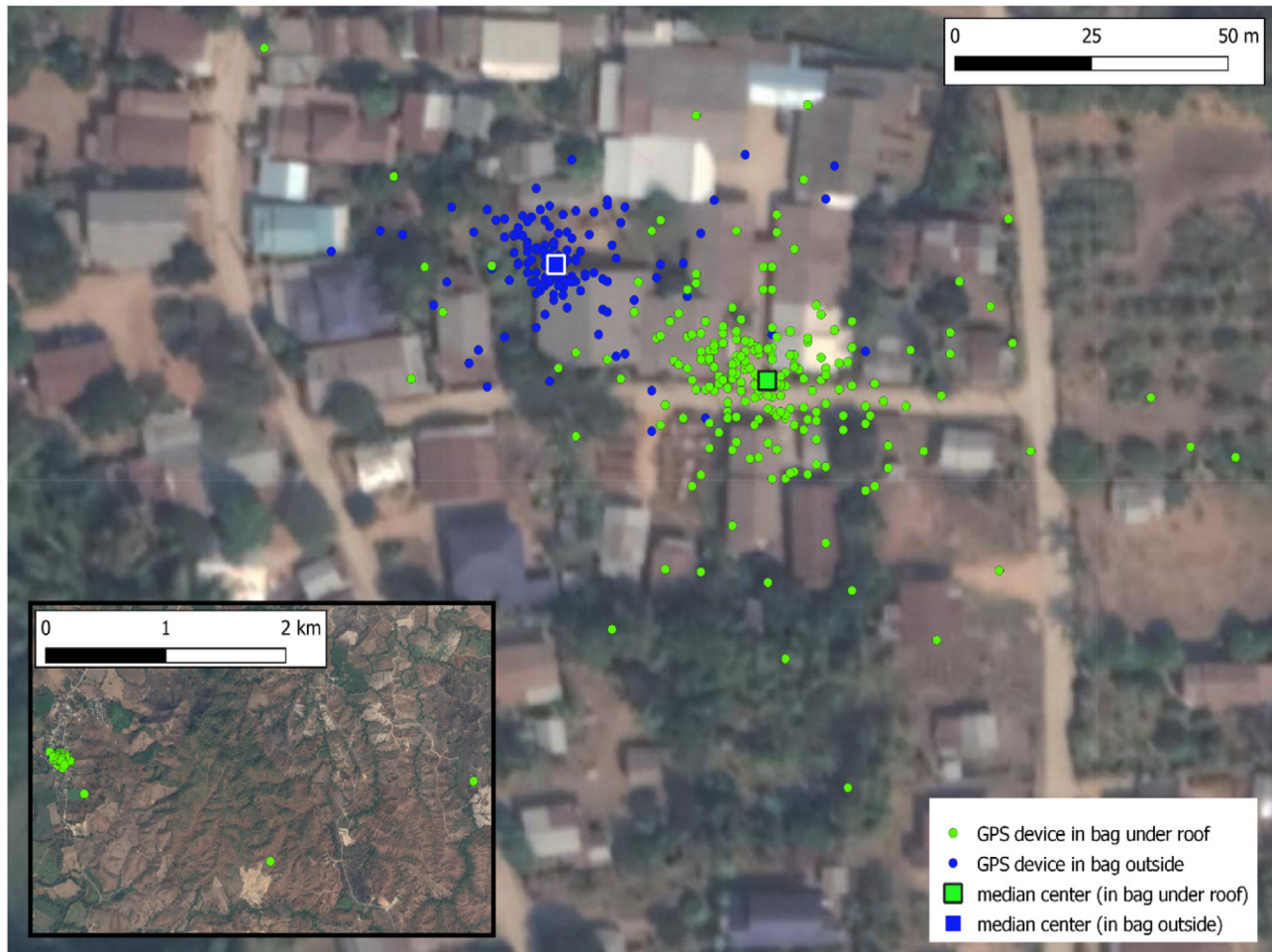
##### Underlying data

Demographic data on participants and participant drop out are provided in the manuscript. All R code used in this analysis is





- Figure S2 (GPS reading errors in stationary devices)

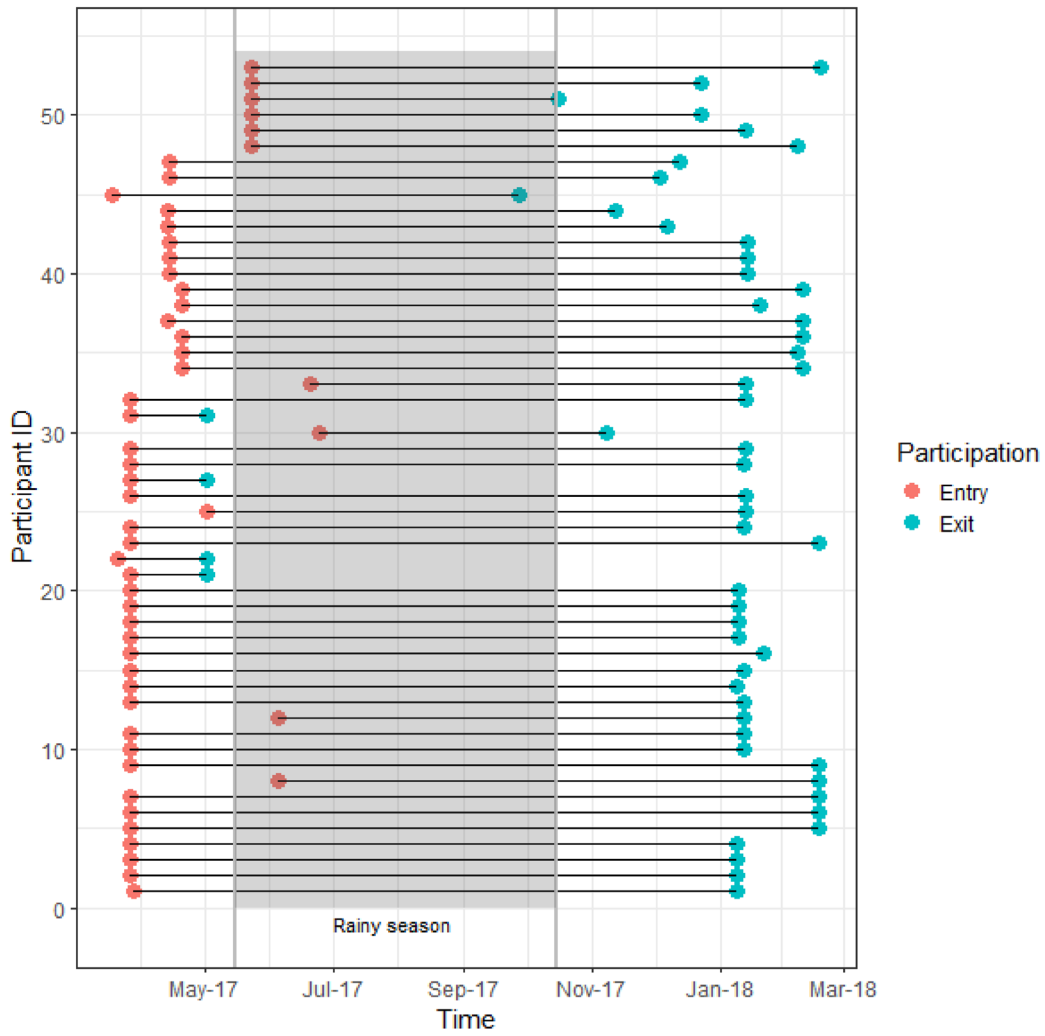


We conducted field tests of GPS device error under stationary conditions (Figure S2). These tests consisted of placing GPS loggers in stationary locations (tied to a bamboo pole, on a shelf in a house), plotting the points from the device over a period of one week, and measuring the geographic distribution of those points from their geographic centre. Devices were placed inside bags, as this would also be likely for carriage/storage by participants.

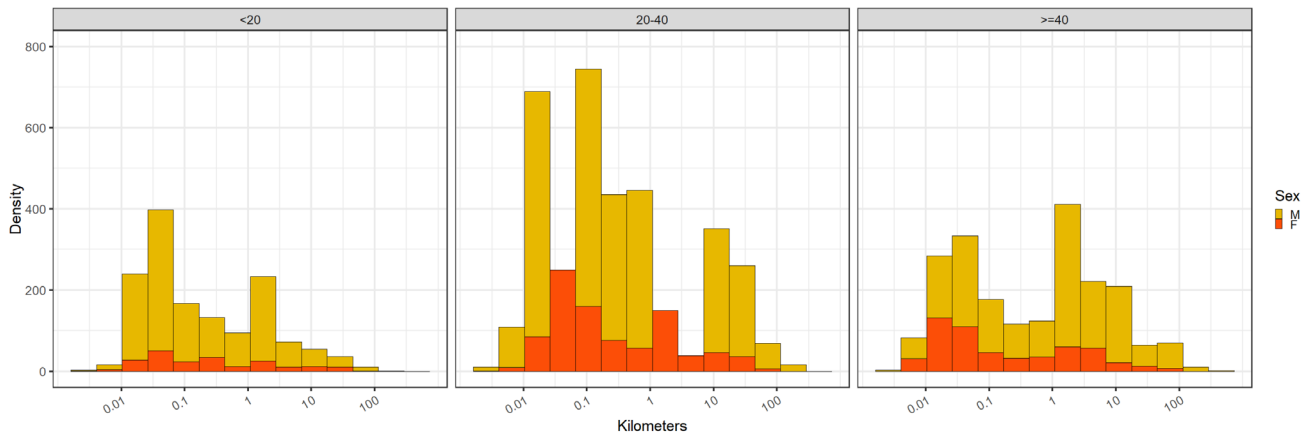
The mean locational error recordings was larger for the indoors device. Most erroneous points were within 50 meters

of the house. However, a few points were far outside of this range (inset map on bottom left). The maximum distance away from the centre for any of the erroneous points was over 3km away. Only one reading was recorded at this distance and the next reading (30 minutes later) was back within the 50m radius around the house. We calculated a standard deviation from the median centre for the worst performing test (in a bag, inside a house) of 266m radius and used this as a basis for judging whether or not a participant's movements were likely real or the result of measurement error. This is a conservative estimate.

- Figure S3 (Duration of participation for each person, over the study period)

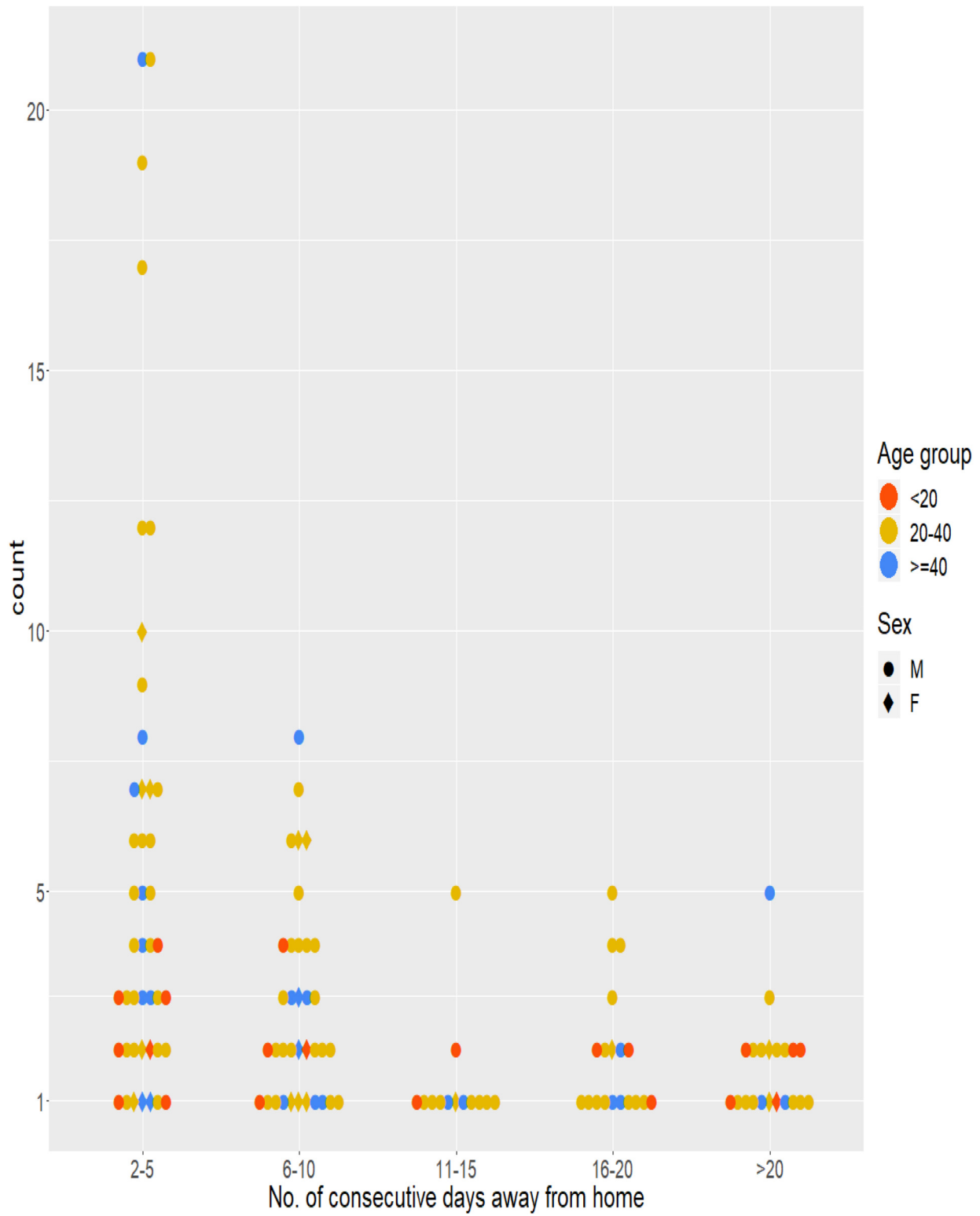


- Figure S4 (Frequency histogram of maximum Euclidian distance travelled by the participants in kilometres)

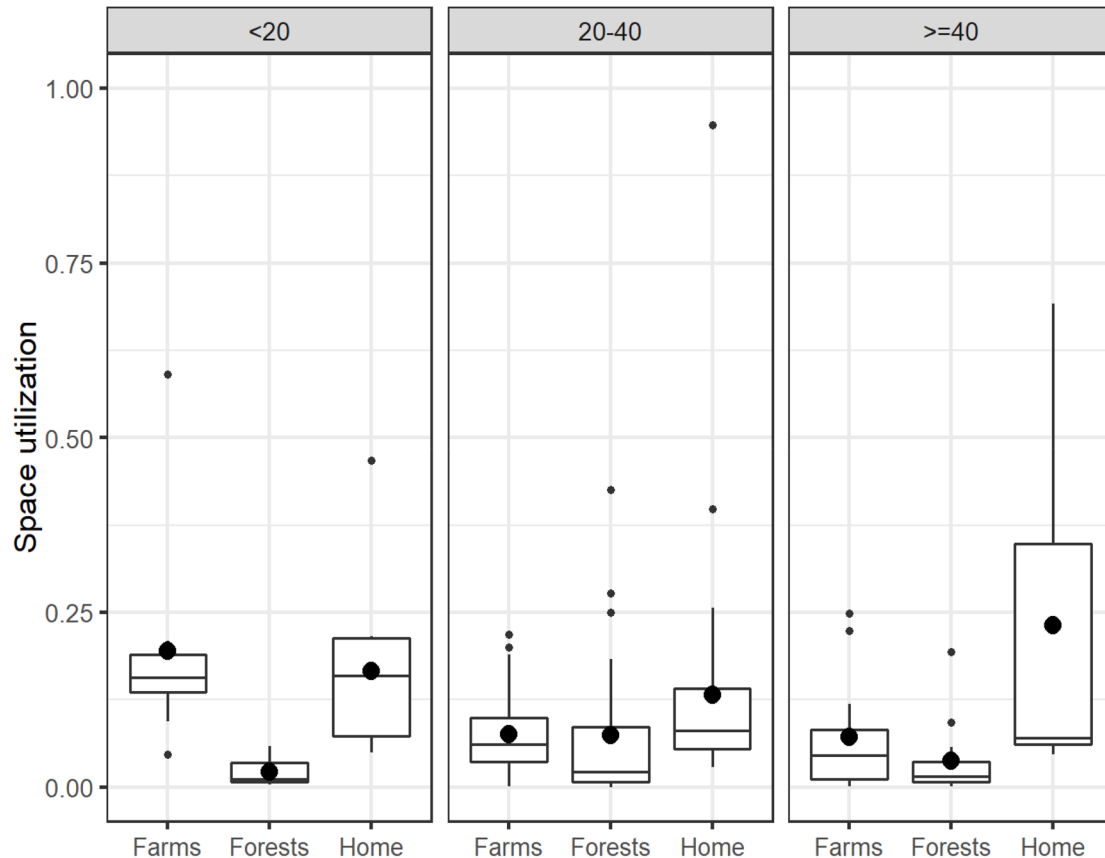




• Figure S5 (Multiday trips made by the participants)



- Figure S6 (Utilization of the farm, forest, and home over the participation period for different age groups)



The bigger dots represent the mean values, while the smaller dots represent the outliers. Usage of Home was underestimated because of the limitation explained in the Methods section.

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

### Code availability

Analysis code available from: <https://github.com/SaiTheinThan-Tun/HumMovPat/tree/v1.0.1>

Archived analysis code at time of publication: <https://doi.org/10.5281/zenodo.4782737><sup>27</sup>.

License: [MIT](#)

### Acknowledgements

We gratefully acknowledge the community members, both research participants and others, who made this work possible. Thanks are also due to Shoklo Malaria Research Unit for allowing us to recruit participants from their border clinics.

### References

1. Reiner RC Jr, Stoddard ST, Scott TW: **Socially structured human movement shapes dengue transmission despite the diffusive effect of mosquito dispersal.** *Epidemics.* 2014; **6**: 30–6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
2. Stoddard ST, Morrison AC, Vazquez-Prokopec GM, *et al.*: **The Role of Human Movement in the Transmission of Vector-Borne Pathogens.** Kittayapong P, editor. *PLoS Negl Trop Dis.* 2009; **3**(7): e481. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
3. Blanford JJ, Kumar S, Luo W, *et al.*: **It's a long, long walk: accessibility to hospitals, maternity and integrated health centers in Niger.** *Int J Health Geogr.* 2012; **11**: 24. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Stock R: **Distance and the utilization of health facilities in rural Nigeria.** *Soc Sci Med.* 1983; **17**(9): 563–70. [PubMed Abstract](#) | [Publisher Full Text](#)
5. Jordan H, Roderick P, Martin D, *et al.*: **Distance, rurality and the need for care: access to health services in South West England.** *Int J Health Geogr.* 2004; **3**(1): 21. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Prothero RM: **Disease and Mobility: A Neglected Factor in Epidemiology.** *Int*

- J Epidemiol.* 1977; **6**(3): 259–67.  
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Tatem AJ, Qiu Y, Smith DL, *et al.*: **The use of mobile phone data for the estimation of the travel patterns and imported *Plasmodium falciparum* rates among Zanzibar residents.** *Malar J.* 2009; **8**: 287.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  8. Bharti N, Tatem AJ, Ferrari MJ, *et al.*: **Explaining Seasonal Fluctuations of Measles in Niger Using Nighttime Lights Imagery.** *Science.* 2011; **334**(6061): 1424–7.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  9. Wesolowski A, Eagle N, Tatem AJ, *et al.*: **Quantifying the Impact of Human Mobility on Malaria.** *Science.* 2012; **338**(6104): 267–70.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  10. Searle KM, Lubinda J, Hamapumbu H, *et al.*: **Characterizing and quantifying human movement patterns using GPS data loggers in an area approaching malaria elimination in rural southern Zambia.** *R Soc Open Sci.* 2017; **4**(5): 170046.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  11. Vazquez-Prokopec GM, Bisanzio D, Stoddard ST, *et al.*: **Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment.** Colizza V, editor. *PLoS One.* 2013; **8**(4): e58802.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  12. Blanford JL, Huang Z, Savelyev A, *et al.*: **Geo-Located Tweets. Enhancing Mobility Maps and Capturing Cross-Border Movement.** *PLoS One.* 2015; **10**(6): e0129202.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  13. Prothero RM: **Malaria, Forests and People in Southeast Asia.** *Singapore J Trop Geo.* 1999; **20**(1): 76–85.  
[Publisher Full Text](#)
  14. Singhanetra-Renard A: **Malaria and mobility in Thailand.** *Soc Sci Med.* 1993; **37**(9): 1147–54.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  15. Saita S, Pan-Ngum W, Phuanukoonnon S, *et al.*: **Human population movement and behavioural patterns in malaria hotspots on the Thai-Myanmar border: implications for malaria elimination.** *Malar J.* 2019; **18**(1): 64.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  16. Pindolia DK, Garcia AJ, Wesolowski A, *et al.*: **Human movement data for malaria control and elimination strategic planning.** *Malar J.* 2012; **11**: 205.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  17. Wesolowski A, Stresman G, Eagle N, *et al.*: **Quantifying travel behavior for infectious disease research: a comparison of data from surveys and mobile phones.** *Sci Rep.* Nature Publishing Group; 2014; **4**: 5678.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  18. Wesolowski A, Eagle N, Noor AM, *et al.*: **The impact of biases in mobile phone ownership on estimates of human mobility.** *J R Soc Interface.* 2013; **10**(81): 20120986.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  19. Williams NE, Thomas TA, Dunbar M, *et al.*: **Measures of Human Mobility Using Mobile Phone Records Enhanced with GIS Data.** *PLoS One.* Public Library of Science; 2015; **10**(7): e0133630.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  20. Elgethun K, Yost MG, Fitzpatrick CTE, *et al.*: **Comparison of global positioning system (GPS) tracking and parent-report diaries to characterize children's time-location patterns.** *J Expo Sci Environ Epidemiol.* Nature Publishing Group; 2007; **17**(2): 196–206.  
[PubMed Abstract](#) | [Publisher Full Text](#)
  21. Elgethun K, Fenske RA, Yost MG, *et al.*: **Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument.** *Environ Health Perspect.* 2003; **111**(1): 115–22.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  22. James P, Weissman J, Wolf J, *et al.*: **Comparing GPS, Log, Survey, and Accelerometry to Measure Physical Activity.** *Am J Health Behav.* 2016; **40**(1): 123–31.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  23. Duncan DT, Kapadia F, Regan SD, *et al.*: **Feasibility and Acceptability of Global Positioning System (GPS) Methods to Study the Spatial Contexts of Substance Use and Sexual Risk Behaviors among Young Men Who Have Sex with Men in New York City: A P18 Cohort Sub-Study.** *PLoS One.* 2016; **11**(2): e0147520.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  24. Stothard JR, Sousa-Figueiredo JC, Betson M, *et al.*: **Investigating the spatial micro-epidemiology of diseases within a point-prevalence sample: a field applicable method for rapid mapping of households using low-cost GPS-dataloggers.** *Trans R Soc Trop Med Hyg.* 2011; **105**(9): 500–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  25. Fornace KM, Alexander N, Abidin TR, *et al.*: **Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysian Borneo.** Cooper B Ferguson NM Cooper B Ruktanonchai N editors. *eLife.* eLife Sciences Publications, Ltd; 2019; **8**: e47602.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  26. Carrasco-Escobar G, Fornace K, Wong D, *et al.*: **Open-Source 3D Printable GPS Tracker to Characterize the Role of Human Population Movement on Malaria Epidemiology in River Networks: A Proof-of-Concept Study in the Peruvian Amazon.** *Front Public Health.* 2020; **8**: 526468.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  27. SaiTheinThanTun: **SaiTheinThanTun/HumMov Patt: second release (Version v1.0.1).** *Zenodo.* 2021.  
<http://www.doi.org/10.5281/zenodo.4782737>
  28. Morris G, Conner LM: **Assessment of accuracy, fix success rate, and use of estimated horizontal position error (EHPE) to filter inaccurate data collected by a common commercially available GPS logger.** *PLoS One.* Public Library of Science; 2017; **12**(11): e0189020.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  29. QGIS Development Team: **QGIS Geographic Information System: Open Source Geospatial Foundation Project.** 2019.  
[Reference Source](#)
  30. R Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria: R Foundation for Statistical Computing; 2019.  
[Reference Source](#)
  31. Pebesma E, Bivand R: **sp: Classes and Methods for Spatial Data.** 2018.  
[Reference Source](#)
  32. Bivand R, Keitt T, Rowlingson B: **rgdal: Bindings for the "Geospatial" Data Abstraction Library.** 2019.  
[Reference Source](#)
  33. Hijmans RJ: **raster: Geographic Data Analysis and Modeling.** 2019.  
[Reference Source](#)
  34. Urbanek S: **proj4: A simple interface to the PROJ.4 cartographic projections library.** 2012.  
[Reference Source](#)
  35. Wickham H: **reshape: Flexibly Reshape Data.** 2018.  
[Reference Source](#)
  36. Wickham H: **ggplot2: Elegant Graphics for Data Analysis.** Springer-Verlag New York; 2009.  
[Reference Source](#)
  37. Benhamou S: **Dynamic Approach to Space and Habitat Use Based on Biased Random Bridges.** *PLoS One.* 2011; **6**(1): e14592.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  38. Calenge C, Fortmann-Roe contributions from S: **adehabitatHR: Home Range Estimation.** 2019.  
[Reference Source](#)
  39. Htway O, Matsumoto J: **Climatological onset dates of summer monsoon over Myanmar.** *Int J Climatol.* 2011; **31**(3): 382–93.  
[Publisher Full Text](#)
  40. Lwin T: **The prevailing synoptic situations in Myanmar.** A report submitted at the Workshop Agricultural Meteorology Department of Meteorology and Hydrology, Yangon, Myanmar, 2000.
  41. Haddawy P, Lawpoolsri S, Sa-ngamuang C, *et al.*: **Effects of COVID-19 government travel restrictions on mobility in a rural border area of Northern Thailand: A mobile phone tracking study.** *PLoS One.* Public Library of Science; 2021; **16**(2): e0245842.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  42. Edwards HM, Sriwichai P, Kirabittir K, *et al.*: **Transmission risk beyond the village: entomological and human factors contributing to residual malaria transmission in an area approaching malaria elimination on the Thailand-Myanmar border.** *Malar J.* 2019; **18**(1): 221.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  43. Parker DM, Carrara VI, Pukrittayakamee S, *et al.*: **Malaria ecology along the Thailand-Myanmar border.** *Malar J.* 2015; **14**: 388.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  44. Nofal SD, Peto TJ, Adhikari B, *et al.*: **How can interventions that target foregoers be tailored to accelerate malaria elimination in the Greater Mekong Subregion? A systematic review of the qualitative literature.** *Malar J.* 2019; **18**(1): 32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
  45. Gao B, Saralamba S, Lubell Y, *et al.*: **Determinants of MDA impact and designing MDAs towards malaria elimination.** *eLife.* 2020; **9**: e51773.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

# Open Peer Review

Current Peer Review Status: ? ✓ ✓ ?

## Version 2

Reviewer Report 20 November 2023

<https://doi.org/10.21956/wellcomeopenres.22276.r68377>

© 2023 Li Y. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Yao Li**

<sup>1</sup> Department of Geographical Sciences, University of Maryland at College Park, College Park, Maryland, USA

<sup>2</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

Q1. Citations are not current,

There are multiple studies studying the human movement in Thailand and Myanmar under the context of malaria very recently, and it is better for authors to cite these papers.

Q2. "These clinics primarily serve migrant and cross-border populations and have connections to village health workers in nearby villages. We reached out to village health workers in the nearby villages to explain the project and to ask if they could help us recruit participants from their respective villages."

Are you recruiting local people, or do you include all the people who are willing to participate, potentially including migrants from Myanmar?

Q3. The conclusion is different from the abstract

In the abstract, the author mentioned the "results from this study are useful for individual-based models of disease transmission and land use." However, in the conclusion paragraph, I didn't see any description of land use.

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

Partly

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

Yes

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

Yes

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

Yes

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

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Infectious diseases, Geographical Sciences, Malaria, human mobility.

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

Reviewer Report 20 November 2023

<https://doi.org/10.21956/wellcomeopenres.22276.r68388>

© 2023 Ruktanonchai N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Nick Ruktanonchai**

Department of Population Health Sciences, Virginia Tech, Blacksburg, VA, USA

I believe the issues brought up by previous reviewers have been sufficiently addressed. I might be inclined to re-introduce the importance of studying the target population--this region is important for malaria elimination as noted by the authors, but it also highlights the importance of the groups that \*weren't\* effectively reached by the study authors, notably women. Certainly I think it is worth noting that there is a bias among participants of research on human mobility that researchers want to understand the travel patterns of only people who travel a lot, and this introduces biases to our understanding of human mobility in many communities across the Global South.

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

Yes

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

Yes

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

Yes

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

Yes

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

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

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

Reviewer Report 23 October 2023

<https://doi.org/10.21956/wellcomeopenres.22276.r67940>

© 2023 Sattenspiel L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Lisa Sattenspiel** 

Department of Anthropology, University of Missouri, Columbia, MI, USA

My previous comments have been addressed in this version. I realize this paper presents the results of a preliminary feasibility study of mobility patterns and I think the research indicates that it is possible to use the methods proposed to collect valuable information on daily mobility patterns in a rural and less developed region. I would have liked to see a little more in-depth discussion of the insights related to the impact of the observed mobility patterns on disease transmission patterns that have been gained from this study and/or would be gained from a more extensive study using the proposed methods.

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

Yes

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

Yes

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

Yes

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



Yes

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

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** human infectious diseases, agent-based modeling, anthropology, mobility and infectious disease

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

---

### Version 1

Reviewer Report 13 January 2023

<https://doi.org/10.21956/wellcomeopenres.18514.r53656>

© 2023 Sattenspiel L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Lisa Sattenspiel** 

Department of Anthropology, University of Missouri, Columbia, MI, USA

This paper presents the results of a pilot study to test the ability of wearable GPS loggers to track the movements of rural farmers in a region along the border of Thailand and Myanmar. The underlying purpose is to assess whether this type of equipment can adequately provide detailed data on human activities that can be used to better inform studies of infectious disease transmission and spread. A second purpose is to assess the feasibility of the technology to collect the detailed data that would be needed to adequately parameterize individual-level models of human activities and disease transmission. The main goal of the paper is primarily to describe the nature of human movement that was measured using the GPS technology, however, and it was not focused on either disease transmission or disease models.

I found the study very interesting and the results encouraging. I agree with the authors that there is a bias in knowledge about human mobility toward data derived from studies of urban populations, and the activities engaged in by the many small-scale societies living under traditional subsistence strategies such as small-scale farming or foraging are much less understood. These types of societies are common in many parts of the world however, and they are especially common in tropical regions with high risks for infectious disease transmission and

very low levels of development and resources. Understanding more about the specific activities that increase potential for disease transmission may help not only the residents of the small-scale societies, but also people in many other regions, because these environments have significant potential for the evolution of new and emerging pathogens.

I have several comments and concerns that I think should be addressed in the manuscript, however. (Page numbers refer to the pdf version of the article)

1. The context of the study needs to be described at the very outset so that the reader can visualize the setting in which the measurements are being taken. What is the geography of the area (e.g., hilly, mountainous, lowland, etc.), what are the houses like, what distinguishes a farm from forest, what activities are performed in these different areas, etc.? Since the ultimate goal is to better understand how the movement patterns may impact infectious disease transmission, it is essential to have some understanding of the setting within which both the movement and potential disease transmission occurs.
2. The authors mention other methods that have been used to collect data on human movement, including surveys and mobile phone records. Important limitations of these methods are mentioned, but there is no discussion of the limitations of the wearable GPS technology. In particular, the authors mention that some types of movement may not be captured by survey methods, especially those of a sensitive nature. This concern holds just as much for wearable GPS devices – people engaging in illegal or sensitive activities may either not wear the devices on certain days or refuse to engage in the study altogether. The article would benefit by a more substantive discussion of both the pros and cons of each of the methods, not just the cons of other methods and the pros of the wearable technology.
3. Have any procedures been implemented to ensure that the GPS loggers are always being worn by the intended persons?
4. The authors mention that they gathered information on illnesses during their monthly meetings. Although the goal of this paper is to discuss the data that were collected on movement patterns, since these data are being collected to eventually help assess disease risks and patterns, it would be useful to have a brief discussion of the kind of illness information that was collected during the pilot study and the underlying reasons for the collection of that information.
5. The description of the two methods used to estimate the utilization of places was a little unclear, especially for someone like myself who is not familiar with the methods. Some clarification would be helpful.
6. The authors mention the strong sex bias in the participants. The potential implications of this bias should be addressed fully in the discussion section. Was it at all related to whether a request for participation was made by a male vs. female researcher? How does the lack of female participation influence the results and the prospects for a full-scale study? If such a bias cannot be sufficiently overcome, what are the implications of this for the data that have been and will be collected?
7. The authors should provide a brief explanation of violin plots and how they should be examined. For example, it would help to explain that the "heights" of the peaks are the

amount they extend out to the left or right, not the vertical distance. In addition, the values mentioned in the results on daily movement ranges (i.e., the first peak is between 10 and 100 km and the second between 1 and 10 km – p. 5) seem inconsistent with what is shown in Figure 1. To me it looks like the first was between 1 and 10 km and the second between 0.01 and 0.1 (although the authors actually refer to the first peak being the larger one for the <20, so that should be the one between 0.01 and 0.1 and the second peak should be the one between 1 and 10 km). These results are mentioned again in the discussion section on p. 9, but in that case they do seem consistent with what is observed in Figure 1.

8. Was any information gathered about the reasons for multiday trips? Besides being longer (and presumably further) were the activities engaged in similar or different to the single day trips? What implications do multiday trips have for understanding the potential for infectious disease transmission?
9. Some additional discussion of the reasons for the movement patterns that were observed should be added to the discussion section of the paper. For example, in the results section it is noted that spending more time at the farms usually meant less time in the forests (p. 5) and it is also noted that 20-40 year olds spend the most time in the forest (p. 6). Why is this? How does it relate to the potential for disease transmission?
10. Explain why the pictures in Figure 2 and Figure S6 differ from each other, since both are measuring the utilization of different places.
11. The discussions of the implications of the observed movement patterns for transmission of diseases prevalent in the study environment needs to be fleshed out. Although this paper is designed to present the results derived from the GPS loggers, there should be a full discussion of what the preliminary data from this study might suggest about how the mobility in the region directly affects disease transmission. Assuming the authors can use their results to derive hypotheses to test in future studies, the discussion is a good place to present those ideas.
12. The paper mentions the importance of the data for “calibrat[ing] human movement patterns in models of infectious diseases”, but there is no discussion at all of what such models are like, what kinds of data are needed in such models, and how the data from the GPS loggers can help calibrate those models. There should either be a substantive discussion of this issue or the use of the data for modeling disease spread should be omitted from the paper at this time.
13. Finally, on p. 9 the authors mention some limitations related to occupation. However nowhere previous to this in the paper has there been any discussion of occupation. A discussion of the normal activities/occupations of the participants should be included in the introductory material suggested in #1 above. Then, the discussion of occupation as a limitation of the pilot study should be elaborated so it is clear why this is an important issue.

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

Yes

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

Yes

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

Partly

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

Partly

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

Partly

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Human infectious diseases, agent-based modeling, anthropology, the impact of daily mobility on infectious disease transmission

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

Author Response 23 Oct 2023

**Daniel Parker**

Dear Editors and Reviewers; Thank you for your helpful comments and suggestions. We would also like to take this opportunity to apologize for our slow movement on addressing your comments and critiques. In the last few years there has been a military coup in Myanmar, on top of the extreme challenges that the pandemic of SARS-CoV-2 has presented, and this has delayed us in our focus on this manuscript and in many other ways. We have also been experiencing some debilitating health issues that have slowed our work. We very much appreciate the time and effort you put into reviewing our manuscript and we hope that we've made it clear that our delays in addressing these comments are because of external factors.

We have attempted to address each of your comments and suggestions and we believe that the manuscript is now much improved. Please see our line-by-line responses below: 1. The context of the study needs to be described at the very outset so that the reader can visualize the setting in which the measurements are being taken. What is the geography of the area (e.g., hilly, mountainous, lowland, etc.), what are the houses like, what distinguishes a farm from forest, what activities are performed in these different areas, etc.? Since the ultimate goal is to better understand how the movement patterns may impact infectious disease transmission, it is essential to have some understanding of the setting within which both the movement and potential disease transmission occurs.

We have added a new section named "Context of the study area":

"The study area is on the Thailand-Myanmar border. Participants were recruited from clinics that serve rural, mostly underdeveloped, and low population density communities. Most participants were of the Karen ethnic group. Villages were made up of a few dozen of mostly multigeneration families living in stilt houses made of wood and thatched roof. Villages didn't normally have schools, clinics, or sanitary toilets. The houses are normally located along the main dirt road of the village. The dirt roads then continued to connect to other villages and small towns through a hilly and rugged terrain with occasional watersheds and river basins which made traversing difficult, especially during the rainy season. 2. Villagers made their living mostly through agriculture, but they have to undertake various types of jobs throughout the year for their subsistence. They developed land in and around their villages into farms to cultivate rice and vegetables. They farm poultry and pigs under their stilt houses. Some villagers go into the forests for hunting, or for foraging wood, and to collect wild edibles. They would go to the farms and forests overnight occasionally, and sleep without much protection from mosquitos and insects. We focused on farms and forests as places of interest in this study since apart from their homes, farms and forests might be the places the villagers spend significant amount of their time while being vulnerable to infectious diseases such as malaria."The authors mention other methods that have been used to collect data on human movement, including surveys and mobile phone records. Important limitations of these methods are mentioned, but there is no discussion of the limitations of the wearable GPS technology. In particular, the authors mention that some types of movement may not be captured by survey methods, especially those of a sensitive nature. This concern holds just as much for wearable GPS devices – people engaging in illegal or sensitive activities may either not wear the devices on certain days or refuse to engage in the study altogether. The article would benefit by a more substantive discussion of both the pros and cons of each of the methods, not just the cons of other methods and the pros of the wearable technology.

We agree with the limitations of the wearable GPS technology you have pointed out. They have been added to the Discussion section in appropriate places. 3. Have any procedures been implemented to ensure that the GPS loggers are always being worn by the intended persons? When the field managers met with the participants on a monthly basis during the study, their general movement patterns were asked and checked against the information from the GPS loggers. Apart from this monthly validation, we cannot enforce the loggers to be worn only by the intended persons.

4. The authors mention that they gathered information on illnesses during their monthly meetings. Although the goal of this paper is to discuss the data that were collected on movement patterns, since these data are being collected to eventually help assess disease risks and patterns, it would be useful to have a brief discussion of the kind of illness information that was collected during the pilot study and the underlying reasons for the collection of that information.

During the study period, a 39-year old male was diagnosed with *Plasmodium vivax* malaria. A 45-year old male sustained a non-fatal gunshot injury while going to his farm one early morning. A 30-year old male passed away from the "cancer inside stomach", the illness that was made known to the field supervisor seven months after being in the study when his

mobility became restricted. 5. The description of the two methods used to estimate the utilization of places was a little unclear, especially for someone like myself who is not familiar with the methods. Some clarification would be helpful. We have revised the description of the two methods for better clarity as follows:

"Utilization of places (home, farms or forests) for each person was estimated using two different approaches. The first method was by checking whether more than two temporally-consecutive GPS points of a person fall within a polygon (a GIS object with a series of x and y coordinate pairs that represents an enclosed area on a map) designated for the person's home (for this particular measurement, the polygon is a circle of radius 266 meters around the person's home location), farms, or forests on each day. This is equivalent to checking if a person spent at least an hour within the same polygon. For each participant, the number of days spending in each category of place (home, farm, forest) was divided by the total number of days participated during the study period to obtain the proportion of being at the respective places. The second method estimated the utilization of places by a biased random bridge (BRB) technique. Unlike prior methods for estimation of utilization of places such as location-based kernel density estimations (LKDE), BRB takes the activity time between successive relocations into account and models space utilization as a time-ordered series of points to improve accuracy and biological relevance while adjusting for missing values. BRB estimate the probability of an individual being in a specific location during the study time period and can be used to estimate home range (the area where individuals spend a defined percentage of their time)."

6. The authors mention the strong sex bias in the participants. The potential implications of this bias should be addressed fully in the discussion section. Was it at all related to whether a request for participation was made by a male vs. female researcher? How does the lack of female participation influence the results and the prospects for a full-scale study? If such a bias cannot be sufficiently overcome, what are the implications of this for the data that have been and will be collected?

Request for participation was made by male researchers only and this could have influenced the recruitment. We did attempt to boost female participation, but potential female participants responded that they rarely go beyond village boundaries which is one of the exclusion criteria of the study. As such, we are quite limited in analyzing female movement patterns and in future studies would put more effort into recruiting more women participants (also through recruiting female staff to work on the project).

7. The authors should provide a brief explanation of violin plots and how they should be examined. For example, it would help to explain that the "heights" of the peaks are the amount they extend out to the left or right, not the vertical distance. In addition, the values mentioned in the results on daily movement ranges (i.e., the first peak is between 10 and 100 km and the second between 1 and 10 km – p. 5) seem inconsistent with what is shown in Figure 1. To me it looks like the first was between 1 and 10 km and the second between 0.01 and 0.1 (although the authors actually refer to the first peak being the larger one for the <20, so that should be the one between 0.01 and 0.1 and the second peak should be the one between 1 and 10 km). These results are mentioned again in the discussion section on p. 9, but in that case they do seem consistent with what is observed in Figure 1.

We have now added an explanation of the violin plots and how they should be read. On page 5, the first peak we have is in meters (10-100 meters that correspond to 0.01-0.1



kilometers) which has a different measurement unit from the second peak. We've added the corresponding value in kilometers for the first peak to prevent the confusion.

8. Was any information gathered about the reasons for multiday trips? Besides being longer (and presumably further) were the activities engaged in similar or different to the single day trips? What implications do multiday trips have for understanding the potential for infectious disease transmission?

Trip information was recorded during the monthly meeting with the field managers regardless of the duration of the trip. Multiday trip may have implications on infectious disease transmission depending on what activities were performed and the location where the participant spent the night. There's a section later on about consecutive nights spent at farms/forests which is a subset of this multiday trip metric and have more relevance for disease transmission. 9. Some additional discussion of the reasons for the movement patterns that were observed should be added to the discussion section of the paper. For example, in the results section it is noted that spending more time at the farms usually meant less time in the forests (p. 5) and it is also noted that 20-40 year olds spend the most time in the forest (p. 6). Why is this? How does it relate to the potential for disease transmission?

We have added/expanded the discussion points in the relevant sections such as the following: "We surmise that the males of 20-40, most likely being the breadwinners of the family, are subject to any possible subsistence activities and are regarded as the most suitable persons to venture into the forests overnight despite the danger from the wildlife and the harsh living conditions." 10. Explain why the pictures in Figure 2 and Figure S6 differ from each other, since both are measuring the utilization of different places. We have explained it in the last paragraph of the discussion section: "Finally, the estimation of land utilization regardless of the method used is imperfect. Having two consecutive GPS points to constitute usage of the land area provide too crude a result (Figure 2). While BRB method provide more accurate and precise estimates (Extended data: Figure S6), it is not without its caveats. BRB assumed that consecutive points that were more than three hours apart were uncorrelated. Since the GPS logger went into sleep mode while stationary, the current land utilization estimation under-estimates the time spent motionless (e.g., sleeping) and hence resulting in lower usage of home in Extended data: Figure S6 compared to that in Figure 2."

11. The discussions of the implications of the observed movement patterns for transmission of diseases prevalent in the study environment needs to be fleshed out. Although this paper is designed to present the results derived from the GPS loggers, there should be a full discussion of what the preliminary data from this study might suggest about how the mobility in the region directly affects disease transmission. Assuming the authors can use their results to derive hypotheses to test in future studies, the discussion is a good place to present those ideas.

We have added a few discussion points on this. "There were differences in these movement patterns by demography, with under-20s staying close to home on the majority of the days and both 20-40 and over-40 age groups tending to move farther away each day. We hypothesize that the reason for this difference is that over-20 age groups are more

heavily involved in subsistence activities (e.g., farming and foraging which are conducted further away from home) than the under-20 age group." "Farming is one of the major forms of subsistence for rural families and it must be regarded as relatively safe compared to subsistence activities in the forests that all age groups partake in it." "We surmise that the males of 20-40, most likely being the breadwinners of the family, are subject to any possible subsistence activities and are regarded as the most suitable persons to venture into the forests overnight despite the danger from the wildlife and the harsh living conditions." "Spending several consecutive nights in the farms and forests may increase the chances of vector-borne diseases such as malaria. Since major malaria vectors in the area such as *Anopheles dirus*, and *Anopheles minimus* are found in the deep forests, forest edges, plantations and even in the rice fields. Studies have found that the increased risk of malaria in forest-goers is contributed by inconsistent bed net usage, misconception that alcohol consumption or blankets provides protection against mosquito bites, non-participation in the malaria prevention activities held at the villages."

12. The paper mentions the importance of the data for "calibrat[ing] human movement patterns in models of infectious diseases", but there is no discussion at all of what such models are like, what kinds of data are needed in such models, and how the data from the GPS loggers can help calibrate those models. There should either be a substantive discussion of this issue or the use of the data for modeling disease spread should be omitted from the paper at this time.

We have added a few discussion points on this. "Results from this study, particularly the space utilization data, would be useful in spatially explicit individual-based infectious disease model such as [Gao et. al. 2020] which models the malaria elimination in the rural South East Asian region. Human mobility is a crucial part of many disease transmission dynamics, yet it has been ignored in many infectious disease models because of constraints on data and computational capacity. Compartmental models assume homogeneous mixing of individuals in their respective compartments. While they are quick to set up, they are not suitable for the disease elimination settings. Their homogeneous nature limits the modelers from exploring the impact of multiple interventions tailored towards different risk groups such as forest-goers in malaria intervention. Individual-based models could have individual specific properties and their related movement patterns thus achieving a heterogeneous population. Calibrating on the space utilization data of this study, such models could become more realistic in terms of transmission dynamics. They could provide more accurate and precise estimations to tackle infectious diseases cost-effectively."

13. Finally, on p. 9 the authors mention some limitations related to occupation. However nowhere previous to this in the paper has there been any discussion of occupation. A discussion of the normal activities/occupations of the participants should be included in the introductory material suggested in #1 above. Then, the discussion of occupation as a limitation of the pilot study should be elaborated so it is clear why this is an important issue.

Thank you for this important point/question. We've added some text to address this: "The most commonly reported occupation was farming and most people in this study area, indeed, farm for at least part of the year. However, people in the study area usually perform different types of work according to the season and assigning a single occupation to a person may not be appropriate. Employment in this region is almost entirely informal, and

most working-age men will work in agriculture for part of the year and in other types of labor during other parts of the year. Responses to surveys about employment will therefore vary by the time of year, even within a single research participant.”

**Competing Interests:** We declare that we have no competing interests

Reviewer Report 16 July 2021

<https://doi.org/10.21956/wellcomeopenres.18514.r44413>

© 2021 Stoddard S. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Steven Stoddard

Graduate School of Public Health, San Diego State University, San Diego, CA, USA

The manuscript presents data from a pilot study of human mobility, measured using GPS, for a rural population in Myanmar. The data are interesting, but are too limited to support much analysis that would be of general value (a mistake made in the manuscript). True, mobility patterns are relevant for infectious disease transmission—and many things—but the context provided in the manuscript is too broad to understand the design of the study and justify how movements were sampled. I do think there is great value in exploring these data and thinking about what they suggest about certain movements, but more specificity about the objectives and description of the cultural and environmental context would help understand the data and the importance of collecting them.

There are too many aspects of ‘human movement’ and not all of them can be perfectly observed, there is always a tradeoff. Even focusing on infectious diseases, there are too many ways different types of movement play a role in transmission, depending on the system. You implicitly made assumptions about which movements were of interest and didn’t justify these. There is theory that provides direction as to which types of movement might matter when, and it would be helpful to this analysis to refer to that theory.

Altogether, I see value here but recommend revising the introduction to clarify and justify your assumptions, reconsidering some analyses and figures, better describing uncertainty in your data, and editing your discussion and conclusions accordingly.

- Consider discussing Prothero’s work in your lit review, which did focus on movements in rural communities like yours. I’m not sure the gap is that movements in rural areas are understudied, as you claim, but really it is that we still don’t have a full understanding of the nature and diversity of movement behavior across landscapes and populations, nor how these play into disease transmission. Certainly, the cell-phone data papers have fallen into many of the high-impact journals of late (and your review seems biased toward them), but you don’t do enough of a review to make this claim. But you don’t need to, as the point of the paper is to describe seasonal movements of a rural population using GPS.
- What is the scale of relevant landscape features? How far apart are homes, for instance? For

a person who is just stepping out of the house and undergoing regular daily activities, such as going to a school or a shop, how far would they go and how long would it take? Same for other features that people are moving to/from. Put another way, the layout of the landscape will shape the movement kernel, both distance traveled and time spent in residence and in transit.

- Consider trying to describe your data in terms of a movement kernel. This is particularly useful for infectious disease modeling. Basically, you would construct a probability distribution of stays by distance and time spent on a patch. Though your sample is limited, you then have an idea of the probability an individual will move  $X$  distance and stay there for  $dT$ , which are the metrics one can use to estimate exposure across a population. Your analyses do touch on this idea, but as presented it would be challenging to extract the probability from your paper.
- Please describe the composition of the villages. What types of structures are there? Where do people stay and what are they like? What are the roof materials?
- Please describe the community. Who are the people and what are their lifestyles and livelihoods? What do they farm? Why are they going to the forest? What illnesses are relevant in this area? Malaria? Arboviruses? Why do you think movement will change seasonally?
- Why did you focus on farms, forests, and “homes” (presumably homes were always in a village)?
- How many villages were included?
- How far apart were the clinics? Is there anything different about the catchments they serve? Did you look to see if there was movement variation across clinics? i.e., did the population using clinic 1 move differently than the population using clinic 2?
- When people stay at farms or in the forest, where do they stay? Where do they sleep? Would they be at higher risk to vector borne illnesses? A little more information on the local ecology would be good. The forest isn't necessarily the high risk area for eg. malaria, the farms might be. But if the villages are small and you have more anthropophilic mosquitoes like *P falciparum*, the villages may be the risk area. And conversely, respiratory illnesses will be more of a risk in the village.
- It would help the paper if you defined what epidemiological risk you are interested in. What motivated the study? Surely not just any infectious disease, because you made specific sampling choices that may or may not be as informative for directly transmitted or other pathogens.
- Was monitoring continuous over the time of participation? Does figure S3 show the time that participants carried the devices or the time that the devices were actually recording data? You say the charge was roughly 1-1.5 months and you rotated monthly. I hardly expect they all worked perfectly. Please say something about monitoring fidelity, i.e., proportion of the observation period over which you actually have tracks.

- By sampling every 30 min, you are choosing to sample movements that are longer in time (and often distance). Why? Understandably this extends the sampling period because the devices' battery charge lasts longer, but you miss many small-scale movements that will be highly relevant for most pathogen transmission events. On the upside, you do a better job of sampling day-to-day variability and capture more longer distance movements. Please provide justification for your sampling focus. Note that stopping the tracking at 6 pm also cuts out many potential relevant movements at a time when many vectors are active.
- Note also that letting the devices 'go to sleep' after being still for an hour inevitably leads to missed points, especially when these stays are in indoor locations. You didn't do any validation of your GPS tracking, so it is worth discussing.
- I'd like to see more discussion of your definition of 'home' and whether your approach was valid. Given the uncertainty in any single GPS reading, I don't see how you can be sure that the last point of the day - at 6 pm no less (especially without some explanation for why people would not move at all after this time) - is the point indicating where they stayed the night. The accuracy of single GPS readings is variable and tends to be worse in and around structures but much better when outside and in movement. So identifying 'stays' would require a 'cloud' of points indicating that a specific site is repeatedly visited and used. Also, a radius of 266 m for a home buffer seems quite large. What fits within that radius? Aren't these small villages? What is the scale of the villages? I would think you would know the location of the home a priori based on recruitment and then could better define a person's immediate 'home range' based on point density (with the caveat that you weren't really sampling frequently enough to capture fine scale movement).
- Why Euclidean distance? Is 'as the crow flies' really the most parsimonious approach or just the easiest? Did you consider other options? A friction surface based on landcover and/or DEM, for instance? Or based on roads/tracks?
- Why use a violin plot for Figure 2? Did you intend to show distributions for males and females? Otherwise, why not just overlay them so the discrepancies could be more clear?
- What were the actual ages of the <20? Your methods state you 'targeted adults'. What ages? 18 and up? 16 and up?
- Why do you break down your age data into the bins you selected? Your data are really quite limited, it might be better to not bother at all.
- How did movement vary across individuals originating in different villages?
- Figure 3 A is illegible when printed.
- I'm not sure what you are trying to do with figure 5 and recommend you either remove it or reconceive. It would be useful to see the distribution of residence time, but boxplots or similar might be more illustrative.
- pg 6 you state "the negative binomial regression provided strong evidence..." Not sure this can be the case, especially since you have a convenience sample that by your admission was

short for females. You can't make any inferences with these data for this reason, so should stick to descriptive statistics or otherwise be careful about how you use any models.

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

Yes

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

Yes

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

Yes

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

Partly

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

Partly

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** infectious disease epidemiology and ecology; human mobility and disease transmission

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

Author Response 23 Oct 2023

**Daniel Parker**

Dear Editors and Reviewers; Thank you for your helpful comments and suggestions. We would also like to take this opportunity to apologize for our slow movement on addressing your comments and critiques. In the last few years there has been a military coup in Myanmar, on top of the extreme challenges that the pandemic of SARS-CoV-2 has presented, and this has delayed us in our focus on this manuscript and in many other ways. We have also been experiencing some debilitating health issues that have slowed our work. We very much appreciate the time and effort you put into reviewing our manuscript and we hope that we've made it clear that our delays in addressing these comments are because of external factors.

We have attempted to address each of your comments and suggestions and we believe that the manuscript is now much improved. Please see our line-by-line responses below:



Reviewer 1: Steven Stoddard, Graduate School of Public Health, San Diego State University, San Diego, CA, USA

The manuscript presents data from a pilot study of human mobility, measured using GPS, for a rural population in Myanmar. The data are interesting, but are too limited to support much analysis that would be of general value (a mistake made in the manuscript). True, mobility patterns are relevant for infectious disease transmission—and many things—but the context provided in the manuscript is too broad to understand the design of the study and justify how movements were sampled. I do think there is great value in exploring these data and thinking about what they suggest about certain movements, but more specificity about the objectives and description of the cultural and environmental context would help understand the data and the importance of collecting them.

There are too many aspects of 'human movement' and not all of them can be perfectly observed, there is always a tradeoff. Even focusing on infectious diseases, there are too many ways different types of movement play a role in transmission, depending on the system. You implicitly made assumptions about which movements were of interest and didn't justify these. There is theory that provides direction as to which types of movement might matter when, and it would be helpful to this analysis to refer to that theory.

Altogether, I see value here but recommend revising the introduction to clarify and justify your assumptions, reconsidering some analyses and figures, better describing uncertainty in your data, and editing your discussion and conclusions accordingly.

- Consider discussing Prothero's work in your lit review, which did focus on movements in rural communities like yours. I'm not sure the gap is that movements in rural areas are understudied, as you claim, but really it is that we still don't have a full understanding of the nature and diversity of movement behavior across landscapes and populations, nor how these play into disease transmission. Certainly, the cell-phone data papers have fallen into many of the high-impact journals of late (and your review seems biased toward them), but you don't do enough of a review to make this claim. But you don't need to, as the point of the paper is to describe seasonal movements of a rural population using GPS.

Thank you for your suggestions. We have included some brief discussion about Prothero's work in our revised manuscript and have dropped the statement about empirical movement studies being relatively rare in resource poor settings.

- What is the scale of relevant landscape features? How far apart are homes, for instance? For a person who is just stepping out of the house and undergoing regular daily activities, such as going to a school or a shop, how far would they go and how long would it take? Same for other features that people are moving to/from. Put another way, the layout of the landscape will shape the movement kernel, both distance traveled and time spent in residence and in transit.

Thanks for these important comments and points. Landscape features, including house locations, are certainly important in assessing movement processes. In this study we did not recruit from villages or houses, but rather from clinics located on the Thai side of the

international border. We don't have GPS points per house. From our experience working in this setting, we know that these types of features vary quite a lot from village to village – and so it really isn't possible to say how far someone would normally have to travel to go to a school or shop. For example, relatively few villages will have a school, even fewer would have a secondary school, etc. Indeed, one of the major goals of doing this type of work in this setting would be to develop an understanding the size and shape of the movement kernel.

- Consider trying to describe your data in terms of a movement kernel . This is particularly useful for infectious disease modeling. Basically, you would construct a probability distribution of stays by distance and time spent on a patch. Though your sample is limited, you then have an idea of the probability an individual will move X distance and stay there for dT, which are the metrics one can use to estimate exposure across a population. Your analyses do touch on this idea, but as presented it would be challenging to extract the probability from your paper.

Thanks for this suggestion. This seems very similar to the approach we took with the biased random bridges, with kernels shown for one individual in Figure 3B. Each individual has a different utilization probability, so we aren't sure that providing plots for all would be optimal for the paper. If researchers would like to work with these data, we are open to that possibility (see data access statement) but we aren't sure that there is an appropriate way to accurately provide such information in the manuscript. We would essentially need 50+ probability distributions and readers would still need to rely on interpreting the plots to estimate the different probability distributions (with multiple kernels per research participant).

- Please describe the composition of the villages. What types of structures are there? Where do people stay and what are they like? What are the roof materials?

We have provided some text about normal household structures now, but this was not a village or house-based study so we do not have information about participant households. We have attempted to make our recruitment process more clear in the text now. Please see the section on the "Context of the study area" in the manuscript and copied underneath. "Context of the study area The study area is on the Thailand-Myanmar border. Participants were recruited from clinics that serve rural, mostly underdeveloped, and low population density communities. Most participants were of the Karen ethnic group. Villages were made up of a few dozen of mostly multigeneration families living in stilt houses made of wood and thatched roof. Villages didn't normally have schools, clinics, or sanitary toilets. The houses are normally located along the main dirt road of the village. The dirt roads then continued to connect to other villages and small towns through a hilly and rugged terrain with occasional watersheds and river basins which made traversing difficult, especially during the rainy season. Villagers made their living mostly through agriculture, but they have to undertake various types of jobs throughout the year for their subsistence. They developed land in and around their villages into farms to cultivate rice and vegetables. They farm poultry and pigs under their stilt houses. Some villagers go into the forests for hunting, or for foraging wood, and to collect wild edibles. They would go to the farms and forests overnight occasionally, and sleep without much protection from mosquitos and

insects. We focused on farms and forests as places of interest in this study since apart from their homes, farms and forests might be the places the villagers spend significant amount of their time while being vulnerable to infectious diseases such as malaria.” • Please describe the community. Who are the people and what are their lifestyles and livelihoods? What do they farm? Why are they going to the forest? What illnesses are relevant in this area? Malaria? Arboviruses? Why do you think movement will change seasonally?

Please see the section on the “Context of the study area” in the manuscript.

- Why did you focus on farms, forests, and “homes” (presumably homes were always in a village)?

One motivating factor for this work is that the ecology of malaria in this setting remains poorly understood. There is quite a bit of literature about ‘forest malaria’ – even going back to Prothero – with relatively scant data about the amount of time spent in forests. Most of the people living in this setting are farmers for at least part of the year. We therefore focus on villages, farms (which are normally near villages), and forests (which tend to be on the outskirts of farms) as these tend to be the places where folks spend most of their time and since forests have long been suspected as important places for *Plasmodium falciparum* transmission.

Home location in the study is derived from the median center of all the GPS points where an individual spent the night, which in turn is derived from the last GPS point of the day between 6pm to 12 midnight.

- How many villages were included? We have added the fact that 10 villages were included in the study:

“Participants were recruited from 10 villages (the lowest level of administrative division in Myanmar) near two clinics on the Thailand-Myanmar border: Wang Pah and Maw Ker Tai Clinics.”

- How far apart were the clinics? Is there anything different about the catchments they serve? Did you look to see if there was movement variation across clinics? i.e., did the population using clinic 1 move differently than the population using clinic 2?

The clinics are approximately 57km apart from each other using a straight line distance. There are some environmental differences between the catchment areas. The southern clinic is nearer to the mountain range and forests/jungles whereas the northern clinic is surrounded by hilly agricultural fields. We did include the recruitment place in the generalized linear mixed effect model. We dropped the variable because it didn’t improve the model.

- When people stay at farms or in the forest, where do they stay? Where do they sleep? Would they be at higher risk to vector borne illnesses? A little more information on the local ecology would be good. The forest isn’t necessarily the high risk area for eg. malaria, the farms might be. But if the villages are small and you have more anthropophilic mosquitoes

like *P falciparum*, the villages may be the risk area. And conversely, respiratory illnesses will be more of a risk in the village.

Thank you for this important comment/question. More information on this have been added as a context of the study area: "Villagers made their living mostly through agriculture, but they have to undertake various types of jobs throughout the year for their subsistence. They developed land in and around their villages into farms to cultivate rice and vegetables. They farm poultry and pigs under their stilt houses. Some villagers go into the forests for hunting, or for foraging wood, and to collect wild edibles. They would go to the farms and forests overnight occasionally, and sleep without much protection from mosquitos and insects. We focused on farms and forests as places of interest in this study since apart from their homes, farms and forests might be the places the villagers spend significant amount of their time while being vulnerable to infectious diseases such as malaria."

- It would help the paper if you defined what epidemiological risk you are interested in. What motivated the study? Surely not just any infectious disease, because you made specific sampling choices that may or may not be as informative for directly transmitted or other pathogens.

The study was not designed to estimate epidemiological risk for any specific disease. It was a pilot study to see if it would be possible to use GPS loggers to study human movement patterns in this setting – which is notable because this is a difficult setting to work in because of the political and civil conflict situation, both historically and in the present. A potential next step could be to assess risk of malaria infection, and in some ways this motivated the present study – but the present study was not designed to specifically study risk of malaria infection. Results from studies like this should be relevant for other environmentally transmitted diseases as well (especially vector-borne diseases).

- Was monitoring continuous over the time of participation? Does figure S3 show the time that participants carried the devices or the time that the devices were actually recording data? You say the charge was roughly 1-1.5 months and you rotated monthly. I hardly expect they all worked perfectly. Please say something about monitoring fidelity, i.e., proportion of the observation period over which you actually have tracks.

Yes, monitoring was continuous over the time period. Figure S3 only showed the first and the last time stamp of the GPS device that the participant carried. Unless the battery of the GPS device ran out, or the participant left the device forgotten somewhere, the monitoring was continuous. We had one technical difficulties with only 1 device in the study, which was replaced by a working device in the next 1-1.5 month interval. The mean percentage of days GPS points were actually observed for each participant over the study is 86.17% (median: 91.21, IQR: 79.54 to 96.93).

- By sampling every 30 min, you are choosing to sample movements that are longer in time (and often distance). Why? Understandably this extends the sampling period because the devices' battery charge lasts longer, but you miss many small-scale movements that will be highly relevant for most pathogen transmission events. On the upside, you do a better job of sampling day-to-day variability and capture more longer distance movements. Please

provide justification for your sampling focus. Note that stopping the tracking at 6 pm also cuts out many potential relevant movements at a time when many vectors are active.

The primary goal of this study was to assess the feasibility of doing GPS logger work in this setting. We also wanted to see if it would be possible to assess seasonal variations in movement behavior (i.e. could a longitudinal GPS logger study on human movement patterns work in this setting?) We chose 30 minute intervals as a balance between capturing short-term movements and having data over a long period of time, with a need to keep interview/visits to a minimum because of funding and labor constraints. Shorter time intervals would have meant either more frequent visits with the research participants to replace devices with fully charged ones, or we would have had to significantly shorten the study period (losing the longitudinal nature of the data collection). Also, the GPS tracking doesn't stop at 6pm. GPS points recorded between 6pm to 12 midnight were used to estimate the home location of the participant or the location where the participant spent the night.

- Note also that letting the devices 'go to sleep' after being still for an hour inevitably leads to missed points, especially when these stays are in indoor locations. You didn't do any validation of your GPS tracking, so it is worth discussing.

The devices 'awaken' when the devices are moved so that dormant GPS devices do restart recoding upon detection of movement. "Figure S2: GPS reading errors in stationary devices." discusses the error compared to the actual GPS location under different scenarios.

- I'd like to see more discussion of your definition of 'home' and whether your approach was valid. Given the uncertainty in any single GPS reading, I don't see how you can be sure that the last point of the day - at 6 pm no less (especially without some explanation for why people would not move at all after this time) - is the point indicating where they stayed the night. The accuracy of single GPS readings is variable and tends to be worse in and around structures but much better when outside and in movement. So identifying 'stays' would require a 'cloud' of points indicating that a specific site is repeatedly visited and used. Also, a radius of 266 m for a home buffer seems quite large. What fits within that radius? Aren't these small villages? What is the scale of the villages? I would think you would know the location of the home a priori based on recruitment and then could better define a person's immediate 'home range' based on point density (with the caveat that your weren't really sampling frequently enough to capture fine scale movement).

The GPS tracking didn't stop at 6pm. The last GPS point of the day is actually selected from the timeframe between 6pm to 12 midnight. We agree that there's a 'cloud' of points around a site that is visited frequently (Figure S2). Participant's home location (a single point) is derived from the median center of the cloud of all the GPS points where s/he spent the night, each of which in turn is derived from the last GPS point of the day between 6pm to 12 midnight. The 266 meters radius buffer comes from the standard deviation from the median centre for the worst performing test (in a bag, inside a house) as explained in Figure S2: GPS reading errors in stationary devices. It is a conservative estimate to judge whether or not a participant's movements were likely real or the result of measurement error. The buffer usually only covers a section of the village.

- Why Euclidean distance? Is 'as the crow flies' really the most parsimonious approach or just the easiest? Did you consider other options? A friction surface based on landcover and/or DEM, for instance? Or based on roads/tracks? In some ways, Euclidian distance is the most parsimonious for this setting for a variety of reasons, even though it is an oversimplification of how actual movements would work. A friction surface would be more sophisticated, but would inherently rely on assumptions that make even less sense in this setting than in others. This setting is an international border in a conflict setting. Movements that might make sense from a ease of travel perspective often don't align with what is safe. Roads are often lined with check points, for example, and so while they may seem to be the most logical travel route many people will actually avoid them. Using Euclidian distances also allows for calculating distances that are intuitive for comparing across individuals or demographic groups.
- Why use a violin plot for Figure 2? Did you intend to show distributions for males and females? Otherwise, why not just overlay them so the discrepancies could be more clear? Violin plot is used in Figure 1 to show the location of the peaks together with the boxplot, as well as whether or not the distributions are statistically different at the 5% level between the three age groups. In this case, all three groups have two peaks, but the one in <20 age group is different from the other two ( $P \leq 0.0001$ ) since its first peak is much higher/broader than the other two. We believe that the difference could still be discerned from the current violin plot.
- What were the actual ages of the <20? Your methods state you 'targeted adults'. What ages? 18 and up? 16 and up? Yes, we targeted individuals who were 18 years of age and above. We have now added this into the text: "The study targeted individuals who were 18 years of age and above adults from the Karen or Burmese ethnic groups, who stated that they would be able to keep track of the GPS device, who were capable of walking outside of village boundaries at recruitment, and who were willing to provide written consent to the study."
- Why do you break down your age data into the bins you selected? Your data are really quite limited, it might be better to not bother at all. We use the bins largely because of the small sample sizes. Most of the analyses and plots would not be possible with age as a continuous variable, especially with such a small cohort. Using age bins is also common practice in both epidemiology and demography, partially for dealing with problems around age-heaping. It is even more important in smaller sample sizes.
- How did movement vary across individuals originating in different villages? Since we did not recruit by village, we did not look at potential variations between villages and since the cohort size is quite small (roughly 50 participants from 10 villages) comparisons aren't really possible between villages.
- Figure 3 A is illegible when printed. We have now increased the font size. The new figure is [linked here](#).
- I'm not sure what you are trying to do with figure 5 and recommend you either remove it



or reconceive. It would be useful to see the distribution of residence time, but boxplots or similar might be more illustrative. Fig. 5 shows how certain age group (20-40) of males consistently frequent the forest or stay the longest time in the forest. Boxplots alone would not be able to describe the age group and gender. We would prefer to keep the figure in the text.

- pg 6 you state “the negative binomial regression provided strong evidence...” Not sure this can be the case, especially since you have a convenience sample that by your admission was short for females. You can't make any inferences with these data for this reason, so should stick to descriptive statistics or otherwise be careful about how you use any models. We agree that caution should be used in interpreting the results from this analysis, in large part because of the convenience sample. We do believe that this analysis is applicable to the cohort though. We have added some text to soften this statement: “The negative binomial regression provided strong evidence that males in this cohort were more likely to spend nights in farms ( $p=0.045$ ) and in forests ( $p=0.01$ ) compared to females, and that young adults (the 20-40 age group) were more likely to spend nights in the forest compared to the under-20 age group ( $p=0.043$ ), after controlling for the remaining variables (Table 2).”

**Competing Interests:** We declare that we have no competing interests

---