



RESEARCH ARTICLE

Human movement patterns of farmers and forest workers from the Thailand-Myanmar border [version 1; peer review: 1 approved with reservations]

Sai Thein Than Tun ^{1,2}, Myo Chit Min³, Ricardo Aguas ^{1,2}, Kimberly Fornace ^{4,5}, Gay Nay Htoo³, Lisa J. White ^{2,6}, Daniel M. Parker ^{7,8}

¹Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

²Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

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

⁴Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK

⁵Centre for Climate Change and Planetary Health, London School of Hygiene & Tropical Medicine, London, UK

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

⁷Department of Population Health and Disease Prevention, University of California, Irvine, CA, 92697, USA

⁸Epidemiology and Biostatistics, University of California, Irvine, CA, 92697, USA

V1 First published: 14 Jun 2021, 6:148
<https://doi.org/10.12688/wellcomeopenres.16784.1>

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

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.

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.

Results: Our study showed that it is feasible to use GPS devices to measure travel patterns, though we had difficulty recruiting women

Open Peer Review

Reviewer Status ?

Invited Reviewers

1

version 1

14 Jun 2021

?

report

1. **Steven Stoddard**, San Diego State University, San Diego, USA

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

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

Corresponding author: Daniel M. Parker (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: © 2021 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 1; peer review: 1 approved with reservations]** Wellcome Open Research 2021, 6:148 <https://doi.org/10.12688/wellcomeopenres.16784.1>

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

Introduction

Human movement or travel is important with regard to infectious disease epidemiology and ecology^{1,2}. 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^{3–5}.

A growing number of research projects, some focused on health, are recording human movement patterns^{6–11}. 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 weaknesses¹². 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 populations¹³. 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 well¹⁴ and mobile phone records will not allow for fine-scale mapping of the routes travelled in between locations¹⁵.

Wearable GPS devices offer extremely detailed data, but are labor intensive and dependent on volunteer cohort members. While their use is increasingly common, most studies have been conducted in urban settings in wealthy nations^{16–19}. Diseases such as malaria and other less-well-studied diseases tend to disproportionately affect the rural poor – precisely the type of population that is most often missed in detailed studies of human movement patterns previously. With the cost of the wearable GPS devices getting cheaper, more and more studies such as^{9,20–22} are being conducted with GPS devices in the developing regions over the last decade.

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

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. 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 villages near two clinics on the Thailand-Myanmar border: Wang Pah and Maw Ker Tai Clinics (*Extended data*: Figure S1²³). 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.

The study targeted 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. 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 meters²⁴. 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 ≥ 15 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^{25,26}, using the “sp”, “rgdal”, “raster”, “proj4”, “reshape” and “ggplot2” R packages^{27–32}. 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). 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 S2²³) was created around each home to create a polygon 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, 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^{33,34}. BRB takes the activity time between successive relocations into account rather than the simple local density of individual locations, modelling space utilization as a time-ordered series of points to improve accuracy of space use estimates and adjust 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^{35,36}. Therefore, we split the data on 15th May 2017 and 15th 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

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

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*²³).

Daily movement ranges

The violin plot of the maximum daily Euclidian distances traveled in kilometers in \log_{10} scale (*Figure 1*) shows that there is a bimodal distribution for all three age groups. The first peak was between 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*²³.

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*²³.

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.

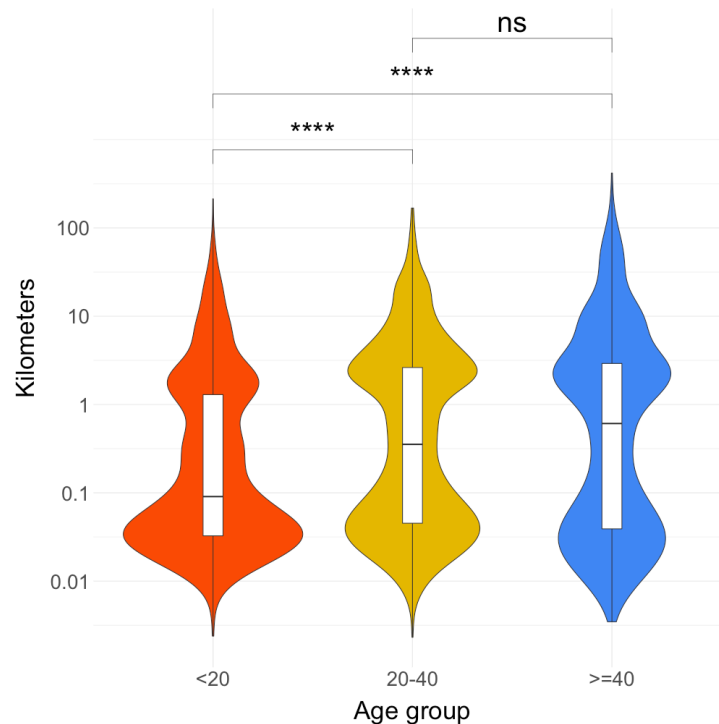


Figure 1. Maximum daily Euclidian distances traveled by participants in kilometers. Distance was calculated from the location a person was at the end of the prior night (most often, this location is their home location). Wilcoxon rank-sum test results are shown on the top of the lines connecting the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

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 S6²³). 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 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).

Participants may spend consecutive nights in the farms or the forests without going back home. Figure 5 quantifies this metric for different age groups and gender. Persons of all age groups and gender spent varying numbers of consecutive nights in the farms. An under-20 male spent the most consecutive

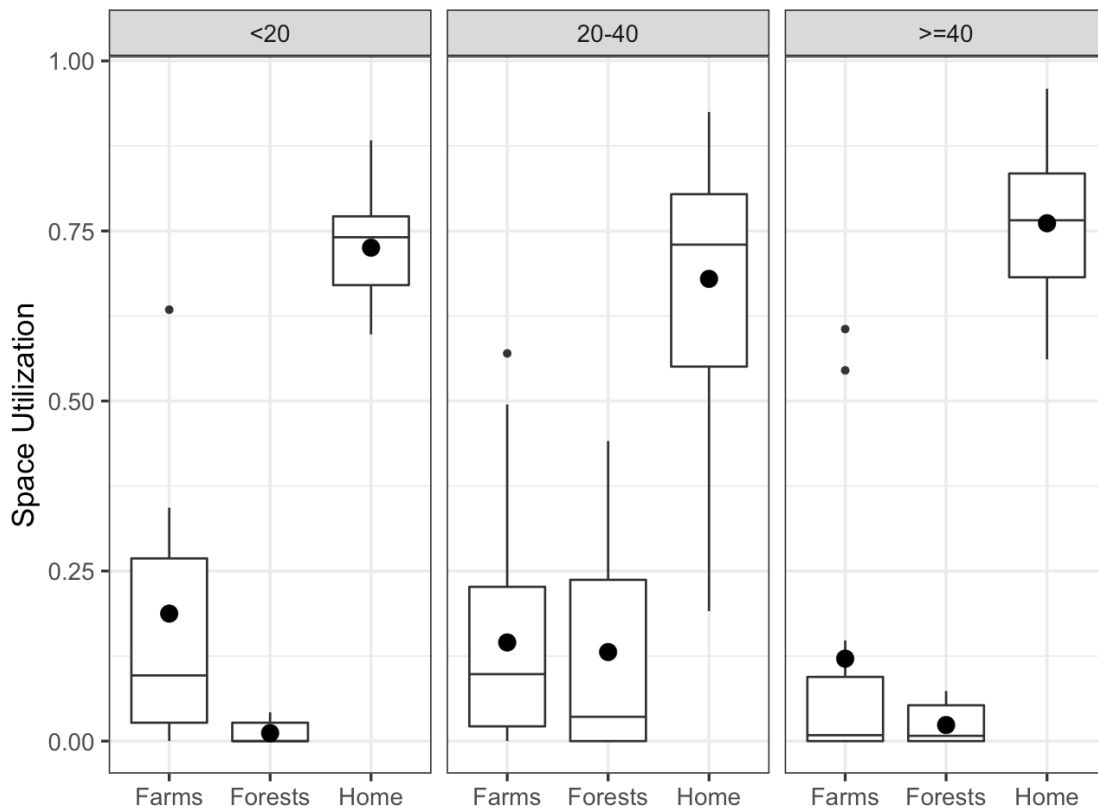


Figure 2. Utilization of the farm, forest, and home (calculated as a proportion of number of days being at the respective places) over the participation period for different age groups. The bigger dots represent the mean values, while the smaller dots represent the outliers.

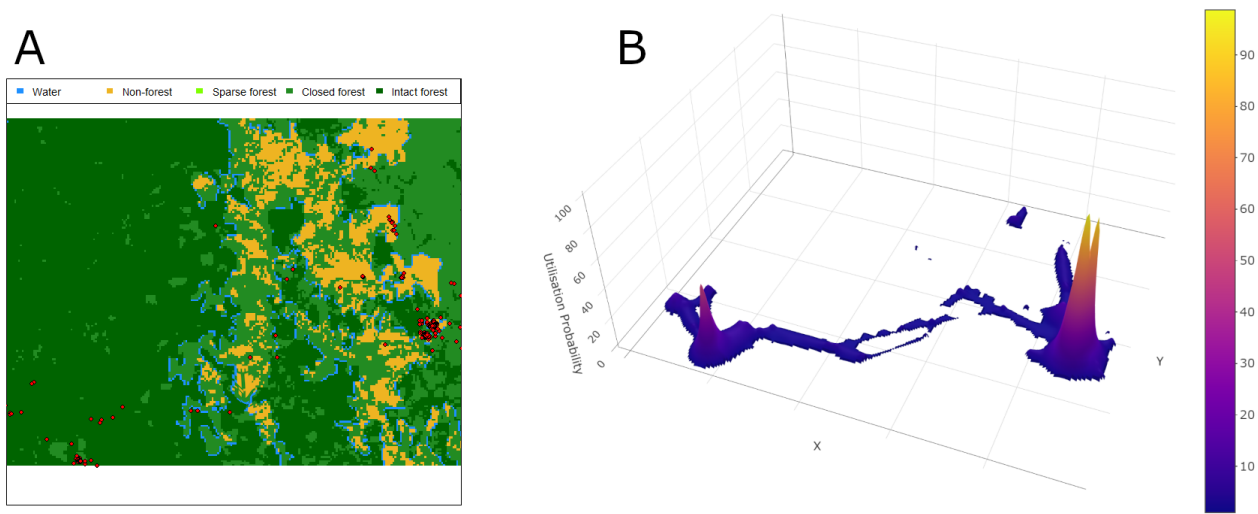


Figure 3. **A.** Example of GPS points (red points) recorded for a person. **B.** Corresponding utilization probability calculated from the GPS points. Its 3D version can be found [here](#).

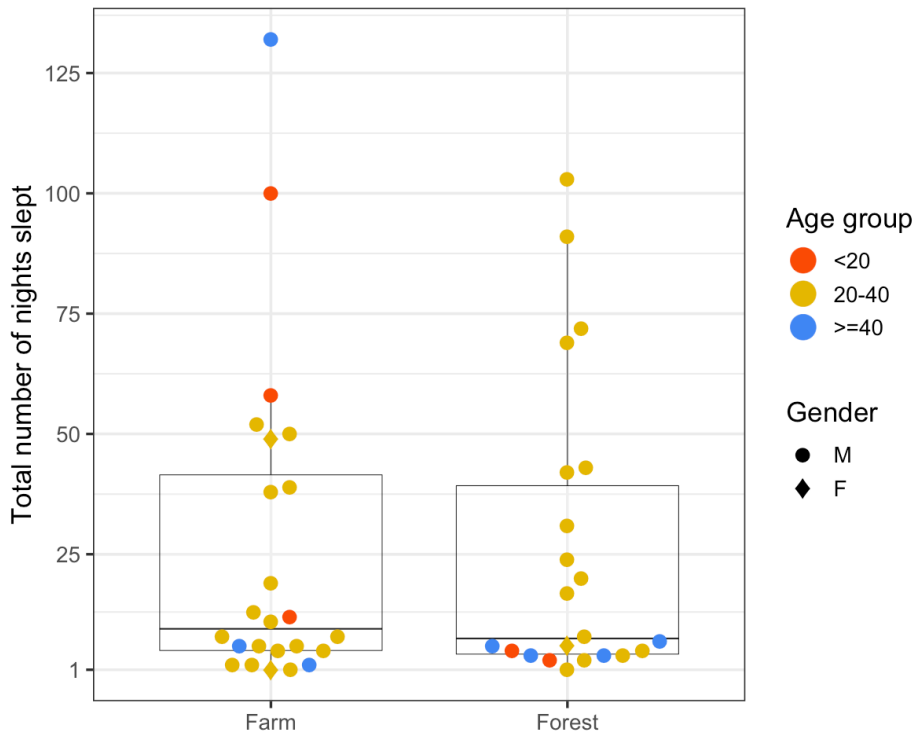


Figure 4. Total number of nights spent in the farms and the forests by each person over the participation period.

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

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]	0.005
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]	0.043
Gender [F]: comparator				
Gender [M]	14.80 [1.07 – 205.70]	0.045	46.34 [2.47 – 869.07]	0.010
Season [Dry]: comparator				
Season [Rainy]	1.20 [0.66 – 2.19]	0.554	0.74 [0.35 – 1.58]	0.435
Random effects				
σ^2	1.35		1.66	
τ_{00}	6.62 _{pid}		5.66 _{pid}	
ICC	0.83		0.77	
N	47 _{pid}		47 _{pid}	
Observations	89		89	
Marginal R ² / Conditional R ²	0.210 / 0.866		0.341 / 0.850	

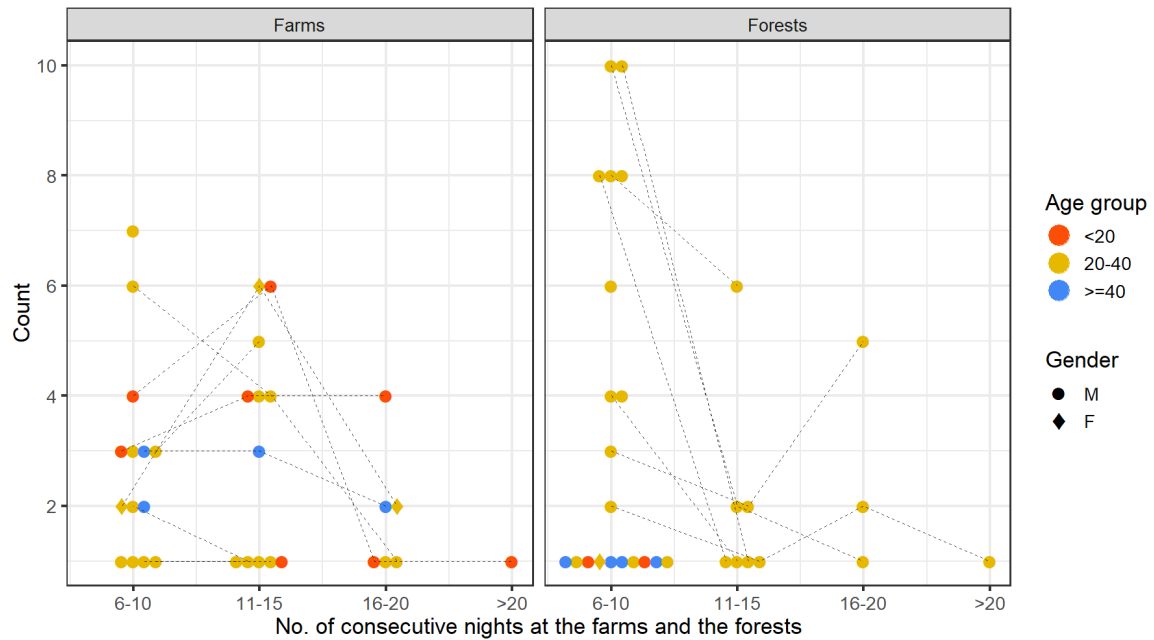


Figure 5. Number of consecutive nights spent in the farms and forests. In each panel, each of the points in each column represents a person of a specific age group and gender, defined in the legend. A single person may contribute a point in each of the columns (e.g., In the panel named Farms, a single 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.

area that has been under-studied with regard to human ecology (though do see 37,38). 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 than the under-20 age group.

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. There was no seasonal variation in the number of nights spent at the farms in these data.

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. No seasonal variation was found in the number of nights of sleeping in the forest.

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. By understanding the utilization of space and place in this population, we can calibrate human movement patterns in models of infectious diseases.

The study has several limitations. It was a pilot study and had a limited sample size. Most participants were adult males. 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.

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. During the study period, participants may have failed to carry the GPS device (intentionally 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 2²³). 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.

Finally, the estimation of land utilization 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).

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.

Data availability

Underlying data

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. An example of a case where a data request might be accepted is when the researcher presents a study protocol through which it is obvious that confidentiality of participant will be maintained, where there is proof of ethical approval of the study, and after signing a data access agreement. Please contact Daniel M. Parker (dparker1@hs.uci.edu) with queries about data access.

Extended data

Zenodo: SaiTheinThanTun/HumMovPatt: second release. <https://doi.org/10.5281/zenodo.4782737>²³.

This project contains the following extended data in the file 'Supplementary materials.docx':

- Figure S1 (Study site and location of clinics that were used for recruitment)
- Figure S2 (GPS reading errors in stationary devices)
- 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)

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>²³.

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

- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- Blanford JJ, 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- 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](#)
- Fornace KM, Alexander N, Abidin TR, *et al.*: **Local human movement patterns and land use impact exposure to zoonotic malaria in Malaysia 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](#)
- 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](#)
- SaiTheinThanTun: **SaiTheinThanTun/HumMovPat: second release (Version v1.0.1).** *Zenodo.* 2021. <http://www.doi.org/10.5281/zenodo.4782737>
- 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](#)
- QGIS Development Team: **QGIS Geographic Information System: Open Source Geospatial Foundation Project.** 2019. [Reference Source](#)
- R Core Team: **R: A Language and Environment for Statistical Computing.** Vienna, Austria: R Foundation for Statistical Computing; 2019. [Reference Source](#)
- Pebesma E, Bivand R: **sp: Classes and Methods for Spatial Data.** 2018. [Reference Source](#)
- Bivand R, Keitt T, Rowlingson B: **rgdal: Bindings for the "Geospatial" Data Abstraction Library.** 2019. [Reference Source](#)
- Hijmans RJ: **raster: Geographic Data Analysis and Modeling.** 2019. [Reference Source](#)

30. Urbanek S: **proj4: A simple interface to the PROJ.4 cartographic projections library**. 2012.
[Reference Source](#)
31. Wickham H: **reshape: Flexibly Reshape Data**. 2018.
[Reference Source](#)
32. Wickham H: **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York; 2009.
[Reference Source](#)
33. 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](#)
34. Calenge C, Fortmann-Roe contributions from S: **adehabitatHR: Home Range Estimation**. 2019.
[Reference Source](#)
35. Htway O, Matsumoto J: **Climatological onset dates of summer monsoon over Myanmar**. *Int J Climatol*. 2011; **31**(3): 382–93.
[Publisher Full Text](#)
36. Lwin T: **The prevailing synoptic situations in Myanmar**. A report submitted at the Workshop Agricultural Meteorology Department of Meteorology and Hydrology, Yangon, Myanmar. 2000.
37. 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](#)
38. 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](#)

Open Peer Review

Current Peer Review Status: ?

Version 1

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.
