**SUPPORTING MATERIALS**

**Supporting Materials I: Diagnosis of major infectious agents**

A detailed report of the data collection and primary analysis can be found in [1]. Cerebrospinal fluid (CSF) was taken from all patients consenting to be included in this study (approximately 2.5 mL for children < 1yo; 3.5 mL for children 1 – 14yo; and 8 mL for patients > 15yo). A venous blood sample was also taken on the same day as the lumbar puncture (approximately 5.5 mL for patients > 15yo; 10 mL for children 1 – 14yo; and 18.5 mL for patients > 15yo). When possible follow-up serum samples were collected between 7 and 10 days post LP. All patient samples were analyzed using a panel of tests, including complete blood count; culture; biochemistry panel; and both serological and molecular assays for a range of fungi, parasites, viruses, and bacteria.

We considered sample size, natural history, and ecology of infections for selecting pathogens for this secondary analysis. Detections of the pathogens included in this analysis are as follows:

* *Japanese encephalitis virus* (JEV) infections were detected using ELISA IgM (Japanese Encephalitis/Dengue IgM Combo ELISA from Panbio) in CSF, and in serum at both admission and follow-up. Patients who were negative at admission but seropositive in a follow-up were classified as confirmed JEV infections. Some JEV infections were also diagnosed by culture or PCR.
* *Cryptococcus* spp. infections were detected using Indian ink stain of CSF; *Cryptococcus* Antigen Latex Agglutination Test with CSF (when HIV infection was suspected); and culture on Sabouraud agar when Indian ink test was positive or HIV infection was suspected.
* *Dengue virus* infections were detected using Hydrolysis probe real time RT-PCR [2] in CSF and serum; NS1 ELISA (Dengue Early ELISA from Panbio) in CSF and serum; and ELISA IgM (Japanese Encephalitis/Dengue IgM Combo ELISA from Panbio) in CSF, and in serum at both admission and follow-up (if negative at admission but seropositive in a follow-up).
* *Flavivirus* infections were detected using nested SYBR Green real-time RT-PCR in CSF and serum [3,4].
* *Rickettsia* spp. infections were detected using Hydrolysis probe RT-PCR in CSF [5,6]; Hydrolysis probe real time PCR and conventional PCR from buffy coat; and genetic sequencing.
* *R. typhi* and *Orientia tsutsugamushi* infections were detected using Hydrolysis probe real time PCR in CSF [5,6]; Hydrolysis probe real time PCR from buffy coat; and IgM and IgG assays from admission and follow-up serum (if there was a > 4-fold rise in antibody at follow-up) [7].
* *Leptospira* spp. infections were detected using hydrolysis probe real-time PCR in CSF [8]; culturing of blood clot on EMJH medium; microscopic agglutination tests at admission and follow-up (if there was a > 4-fold rise in antibody at follow-up)[9]; and hydrolysis probe real time RT-PCR from buffy coat [8].

The final etiology was determined based on the panel of diagnostic tests, including direct detection of pathogens in CSF or blood, IgM in CSF, seroconversion, or a 4-fold increase in antibody titer between the date of admission and follow-up serum samples. When more than 1 pathogen was present, direct tests were prioritized over indirect tests and presence in the CSF was prioritized over presence in the blood.

**Supporting Materials II: Spatial point patterns**

The spatial distribution of villages from which study patients originated were indicated through maps of village locations, standard distance deviations (SDDs), and standard deviational ellipses (SDEs). Both SDDs and SDEs provide a visual representation of the central tendency and spread of points across a landscape [10,11]. SDEs also indicate potential anisotropy.

The SDD gives an indication of how points deviate from the mean center. The formula for the SDD is:

where *xi* and *yi* are geographic references for point *i*;

{*XMC*, *YMC*} is the geometric mean center (MC) for the features.

The SDE differs from the SDD in that the X- and Y-axes are calculated separately and the orientation is not necessarily horizontal/vertical. The Y-axis is rotated clockwise until the sum of the squares of the distances between points (village locations) and axes are minimized. The angle is defined as:

The standard deviation is then calculated along both the shifted X- and Y-axes:

.

The output of these statistics is traditionally mapped as an ellipse; with 1, 2, or 3 standard deviations (roughly corresponding to 63, 98, or 99 % of all geographic points, respectively). Spatial point patterns that are isotropic will result in an SDE that is equal to the standard distance deviation (SDD), resulting in a circular map layer rather than an ellipse.

Both the SDD and SDE can be weighted (for example, if multiple cases come from a single location).

**Supporting Materials III: Environmental Indices (EIs)**

The photosynthetic components of vegetation (i.e. chlorophyll) absorb visible light, especially in the Red and Blue wavelengths. Conversely, most infrared radiation is reflected by healthy vegetation. The contrast between Red and near-infrared (NIR) responses therefore provides an estimate of healthy vegetation.

One common measure of landscape vegetation is the normalized difference vegetation index (NDVI [12]) which is frequently defined as:

.

This simple measurement is sensitive to atmospheric effects and dense canopy structure [13]. While NIR can pass through multiple layers of canopy structure, Red typically cannot. In areas with high vegetation density NDVI quickly becomes saturated. An improved metric has been developed to account for these problems, referred to as the enhanced vegetation index (EVI [14]). This metric uses the difference between Red and Blue reflectances as an estimator of atmospheric influence level on the vegetation index. EVI is commonly specified as:

; where *L* is the canopy background adjustment;

*C*1 and *C*2 are coefficients of an aerosol resistance term;

and *G* is a scaling factor.

A variety of similar indices have been proposed to measure water content, either within vegetation (i.e. measuring drought conditions or identifying areas that have been burned) or as surface water. In general, indices that use a combination of NIR and shortwave infrared responses (SWIR) have been proposed to measure within-vegetation water content whereas those that use a combination of visible spectral regions (VIS) and SWIR are usually proposed for identifying water bodies.

Almost all include a SWIR component because infrared in these wavelengths are well- absorbed by water (see [15], for example). Following Boschetti et al [16] we use the following normalized flooding index (NFI):

; where SWIR2 is shortwave infrared radiation 2 (~ 1640nm).

**Supporting Materials IV: Statistical model selection**

Formal multivariable analysis was conducted on both the village- and individual-level datasets. Small case numbers for mono-infections limited our multivariable analyses to an analysis of study patient home villages and comparison villages as well as the most commonly diagnosed infection: JEV. The village-level data were coded as a “1” or “0” based on whether or not the village was home to an study patient; and whether or not the village was home to an study patient diagnosed with JEV. The individual-level dataset was likewise coded with a “1” or “0” based on whether or not the individual was diagnosed with a JEV infection (all patients in the individual-level data had an LP).

We began with an exploratory multivariable analysis using generalized additive models (GAMs) with a binomial distribution (logistic GAMs). The use of GAMs allowed us to explore the potentially non-linear shape of the association between continuous environmental predictors (NDVI, EVI, and NFI) and the outcome variables and informed our final model selection and variable specification.

Our first GAMs included both NDVI and EVI, which are considered complimentary to each other [14]. The models showed a high degree of concurvity between these two covariates, almost no added benefit (from model fit statistics), and no detectable effect of the NDVI covariate. In subsequent models we therefore retained only EVI as a measure of vegetation.

Village-level GAMs began with village geographic (elevation, distance to nearest major road) and demographic (village population size) covariates. A second model was then specified including the environmental covariates at the 2km buffer size around each village. Subsequent models tested larger buffer sizes (5km and 10km), investigating overall model fit using the Akaike Information Criterion (AIC) and the explained deviance. The smoothed functions were chosen using restricted maximum likelihood (REML).

The village-level GAM for all study patients (that is, all patients who had an LP regardless of diagnosis) showed statistically significant contributions from NFI, EVI, village population, distance to the nearest major road, and elevation. The effects of NFI, EVI, and village population were positive while the effects for distance to the nearest major road and elevation were negative. The effects for NFI, EVI, village population, and elevation all appeared curvilinear. The village level model for JEV villages (villages from which LP patients who were diagnosed with JEV came) indicated that only village population was a significant predictor, with a curvilinear effect.

Individual-level models began with village- (village population, elevation, distance to the nearest major road) and individual- (age, gender, admission quarter and year) level variables. Environmental variables were first added at the 2km buffer size and for the year prior to admission. Subsequent models tested larger buffer sizes until the AIC was minimized. The 10km buffer appeared to provide the best model fit. The temporal resolution was then varied at the 10km buffer size, beginning with 1 year mean prior to admission, then 2 months mean prior to admission, and finally within the same month as admission. The best fitting model appeared to be the 10km buffer and measures from the same month as admission.

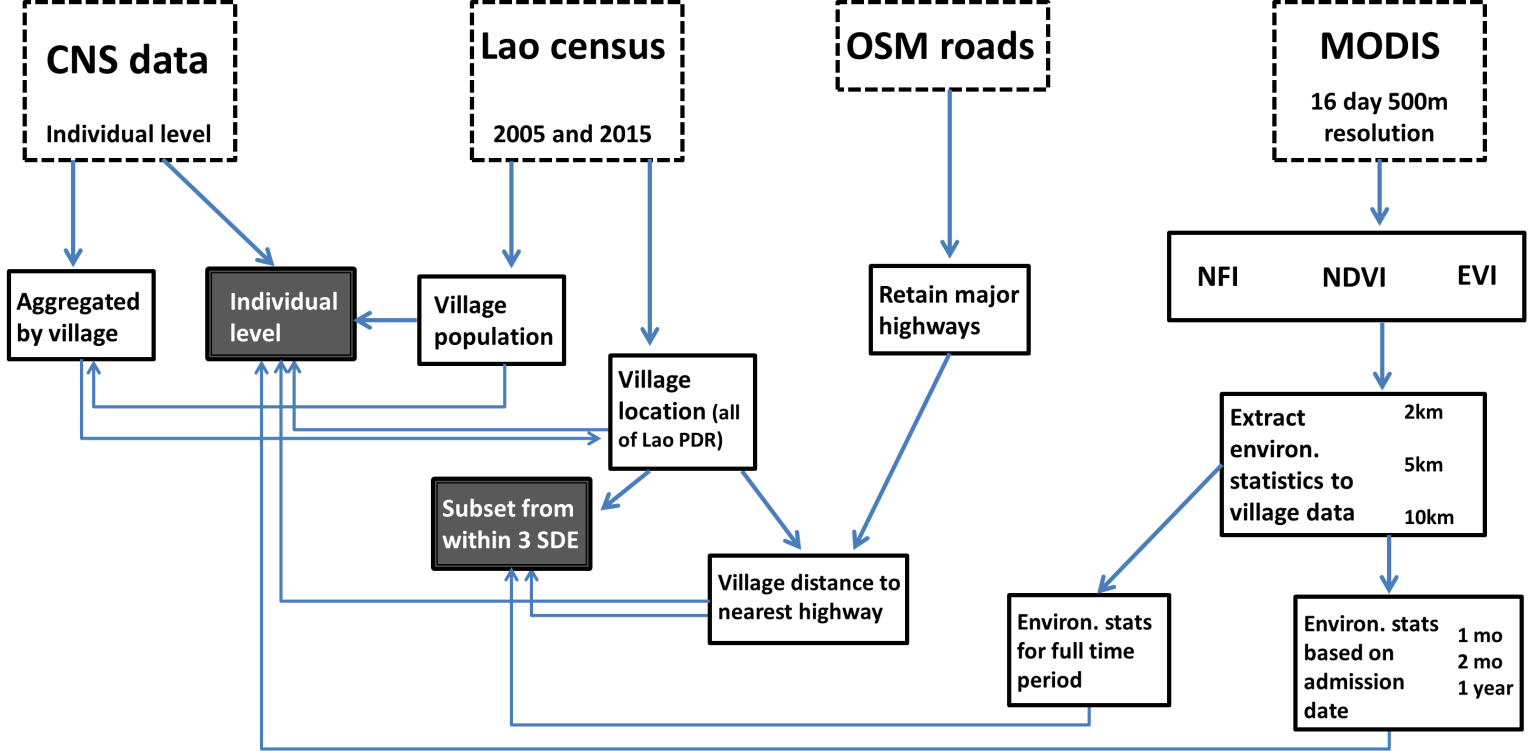
Both NFI and EVI show seasonal variations and calendar month is a strong predictor of NFI. In order to account for concurvity in the GAMs and collinearity in the subsequent logistic generalized linear models, we transformed the environmental variables to quartiles for subsequent models. This transformation allows for easy interpretation of covariate effects, allows for non-linear associations between the covariate and the outcome of interest, and allows the model to simultaneously address both the seasonality in cases (especially JEV) and the apparent associations with surface water (NFI).

All other continuous variables were centered on their means and standardized by their standard deviations.

The final village-level model was a logistic regression and the final individual-level model was a mixed effects logistic regression, with a random intercept for village, both using these transformed and standardized variables.

**SUPPORTING FIGURES**

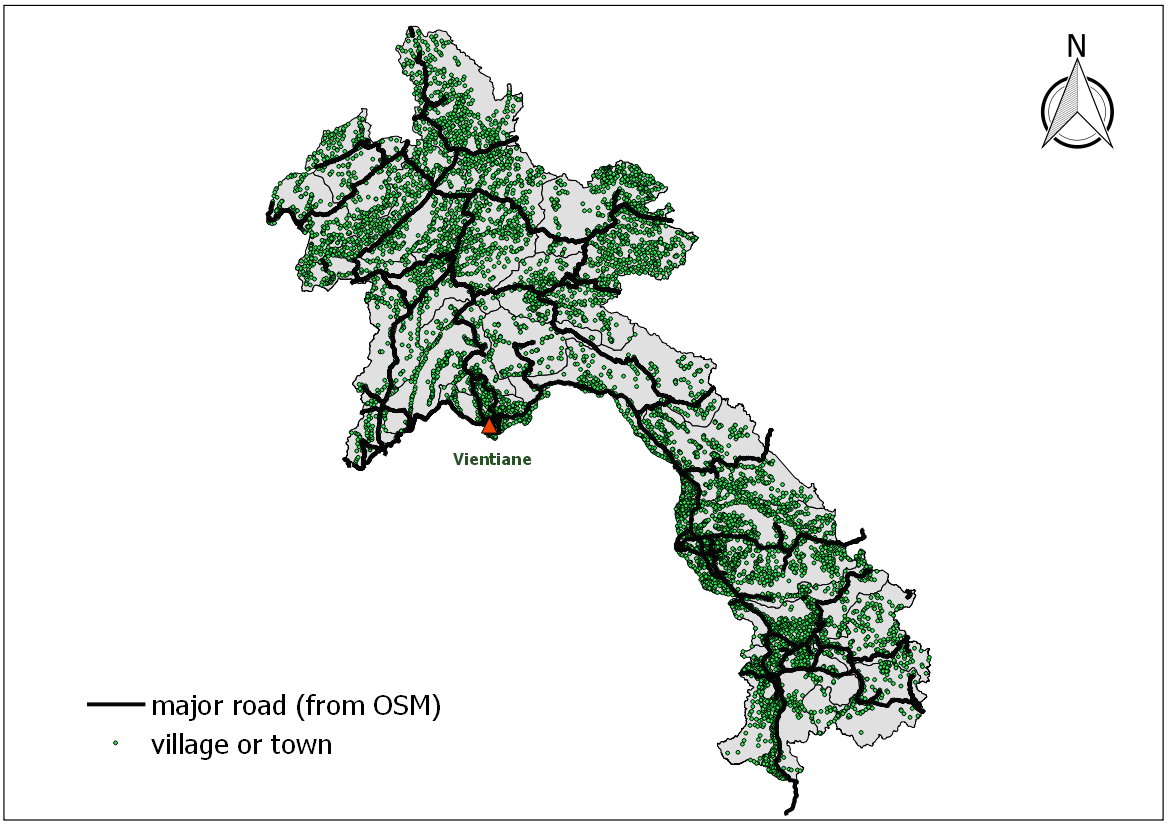
**Supporting Figure 1**: Diagram of data processing and aggregation. Four different data sources are used (indicated by boxes with dashed lines). Two main datasets are created from the combined sources (indicated by boxes shaded in grey): an individual-level dataset (one row per patient) and a village-level dataset (one row per village). NDVI indicates the normalized differential vegetation index, EVI indicates the enhanced vegetation index, and NFI indicates the normalized flooding index.



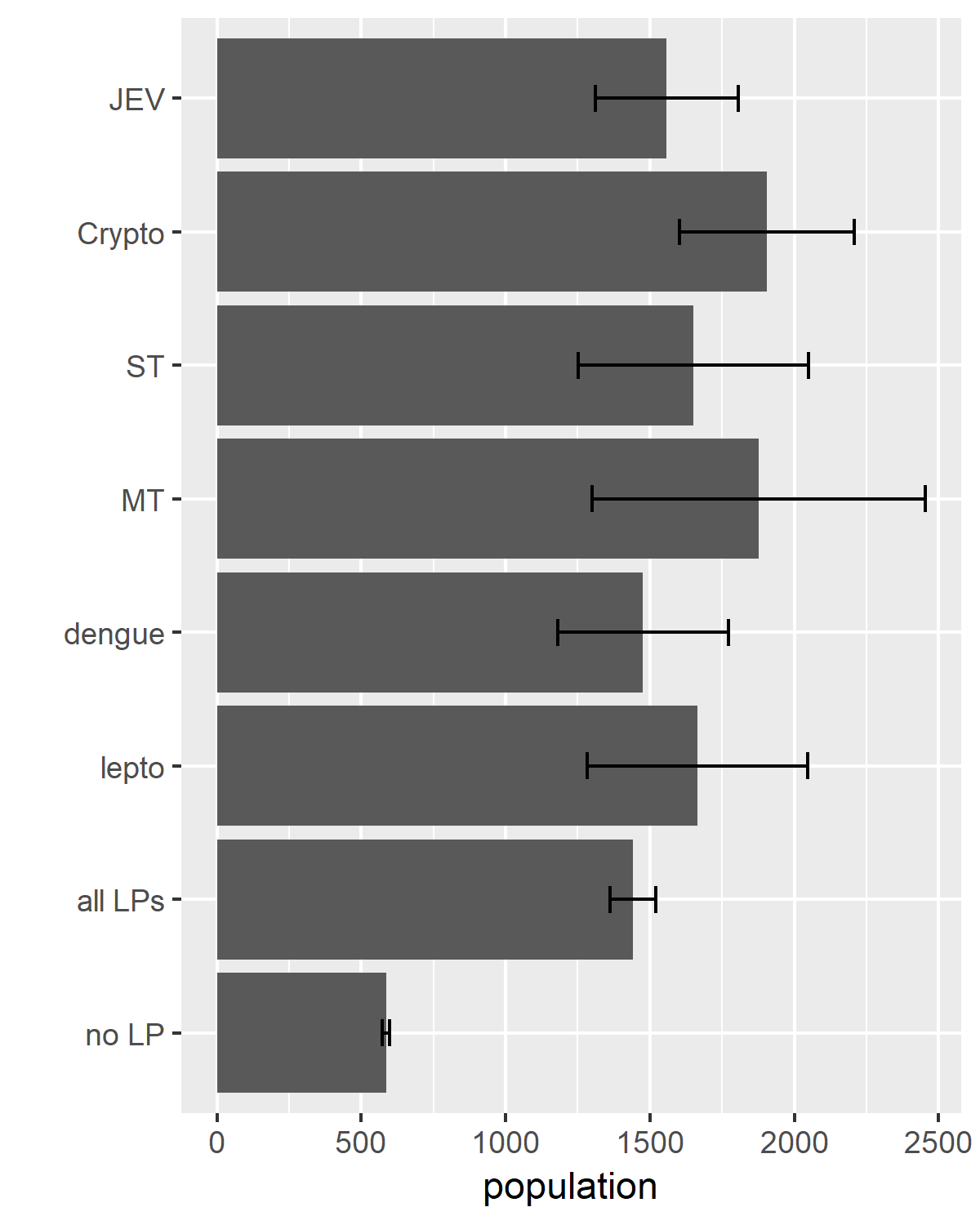
**Supporting Figure 2**: Subset of villages selected for village level analysis. A standard deviational ellipse ((SDE) with 3 standard deviations) was drawn around the home villages of all LP patients. All villages within that SDE were selected for the village level analysis.



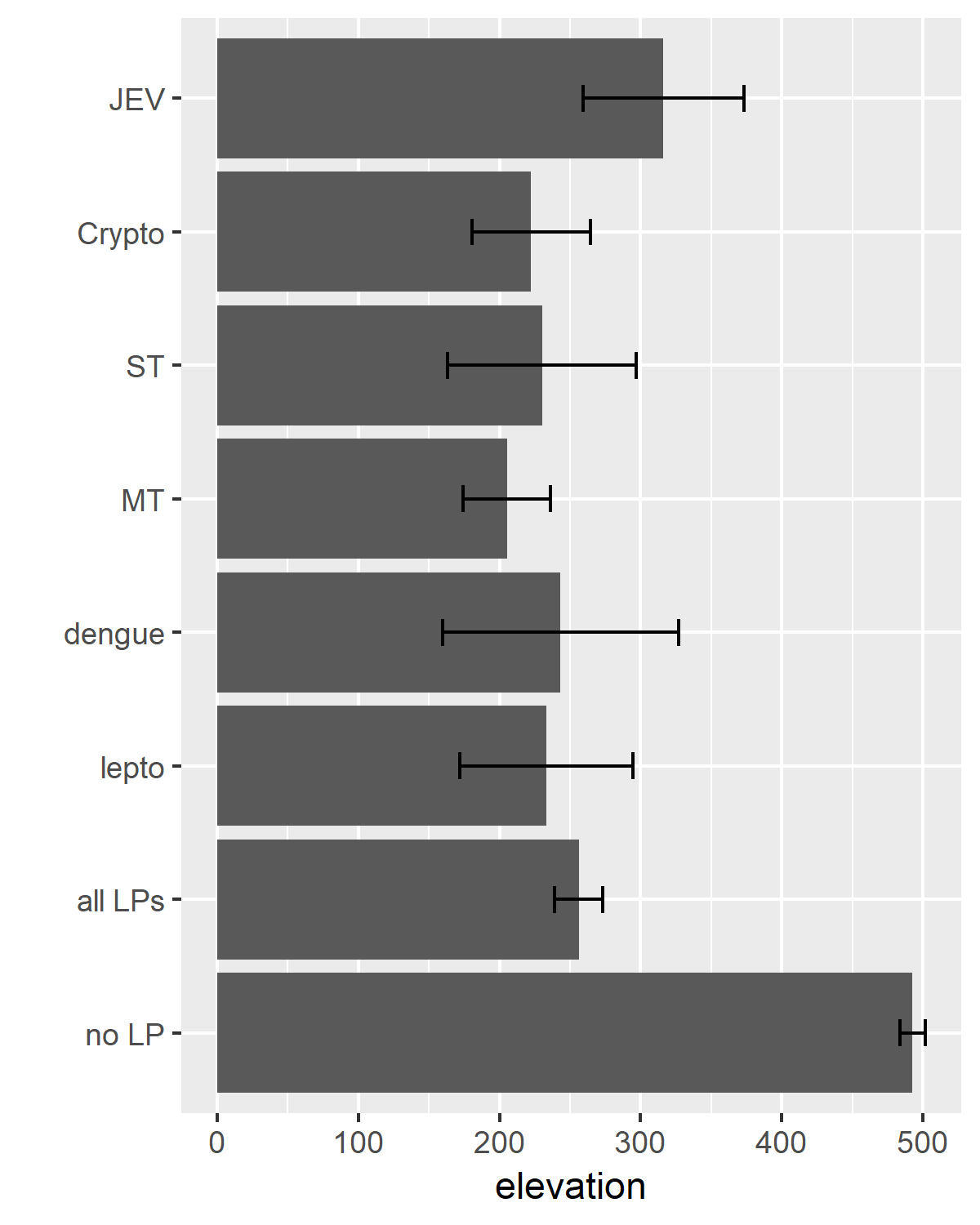
**Supporting Figure 3**: Major roads (dark black lines) downloaded from OpenStreetMaps (2017) for use in calculating the Euclidian distance from each village to the nearest major road. Roads included “primary”, “secondary”, and all major connecting roads.

****

**Supporting Figure 4:** Mean village population (and 95% CI) by diagnosis



**Supporting Figure 5:** Mean village elevation (and 95% CI) by LP diagnosis

****

**SUPPLEMENTAL REFERENCES**

1. Dubot-Pérès A, Mayxay M, Phetsouvanh R, Lee SJ, Rattanavong S, Vongsouvath M, et al. Management of Central Nervous System Infections, Vientiane, Laos, 2003–2011. Emerg Infect Dis. 2019;25:898–910.

2. Leparc-Goffart I, Baragatti M, Temmam S, Tuiskunen A, Moureau G, Charrel R, et al. Development and validation of real-time one-step reverse transcription-PCR for the detection and typing of dengue viruses. J Clin Virol. 2009;45:61–6.

3. Moureau G, Temmam S, Gonzalez JP, Charrel RN, Grard G, de Lamballerie X. A real-time RT-PCR method for the universal detection and identification of flaviviruses. Vector Borne Zoonotic Dis. 2007;7:467–77.

4. Moureau G, Ninove L, Izri A, Cook S, De Lamballerie X, Charrel RN. Flavivirus RNA in phlebotomine sandflies. Vector Borne Zoonotic Dis. 2010;10:195–7.

5. Jiang J, Chan T-C, Temenak JJ, Dasch GA, Ching W-M, Richards AL. Development of a quantitative real-time polymerase chain reaction assay specific for Orientia tsutsugamushi. Am J Trop Med Hyg. 2004;70:351–6.

6. Jiang J, Stromdahl EY, Richards AL. Detection of Rickettsia parkeri and Candidatus Rickettsia andeanae in Amblyomma maculatum Gulf Coast Ticks Collected from Humans in the United States. Vector-Borne and Zoonotic Diseases. 2011;12:175–82.

7. Phetsouvanh R, Blacksell SD, Jenjaroen K, Day NPJ, Newton PN. Comparison of Indirect Immunofluorescence Assays for Diagnosis of Scrub Typhus and Murine Typhus Using Venous Blood and Finger Prick Filter Paper Blood Spots. The American Journal of Tropical Medicine and Hygiene. 2009;80:837–40.

8. Thaipadunpanit J, Chierakul W, Wuthiekanun V, Limmathurotsakul D, Amornchai P, Boonslip S, et al. Diagnostic Accuracy of Real-Time PCR Assays Targeting 16S rRNA and lipl32 Genes for Human Leptospirosis in Thailand: A Case-Control Study. PLOS ONE. 2011;6:e16236.

9. Cole JR, Sulzer CR, Pursell AR. Improved Microtechnique for the Leptospiral Microscopic Agglutination Test1. Appl Microbiol. 1973;25:976–80.

10. Mitchell A. The ESRI Guide to GIS Analysis. ESRI Press; 2005.

11. Levine N. Spatial Distribution. CrimeStat III: a spatial statistics program for the analysis of crime incident locations (version 30) [Internet]. Houston, TX; Washington DC: Ned Levine & Associates; National Institute of Justice; 2004. Available from: https://www.nij.gov/topics/technology/maps/documents/crimestat-files/CrimeStat%20IV%20Chapter%204.pdf

12. Rouse J, Hass R, Deering D, Sehell J. Monitoring the vernal advancement and retrogradation (Green wave effect) of natural vegetation [Internet]. 1974 p. 8. Report No.: E74-10676, NASA-CR-139243, PR-7. Available from: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19740022555.pdf

13. Huete AR. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. 1988;25:295–309.

14. Huete A, Didan K, Miura T, Rodriguez E, Gao X, Ferreira L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment. 2002;83:195–213.

15. Hale GM, Querry MR. Optical Constants of Water in the 200-nm to 200-μm Wavelength Region. Appl Opt, AO. 1973;12:555–63.

16. Boschetti M, Nutini F, Manfron G, Brivio PA, Nelson A. Comparative Analysis of Normalised Difference Spectral Indices Derived from MODIS for Detecting Surface Water in Flooded Rice Cropping Systems. Schumann GJ-P, editor. PLoS ONE. 2014;9:e88741.