Ecological and socio-economic predictors of transmission assessment survey failure

Ellen M Goldberg

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2018

Committee:

Elizabeth Cromwell

David Pigott

Program Authorized to Offer Degree

Global Health

University of Washington

Abstract

Ecological and socio-economic predictors of transmission assessment survey failure

Ellen M Goldberg

Chair of the Supervisory Committee:

Elizabeth Cromwell

Department of Global Health

The Transmission Assessment Survey (TAS) is recommended to determine whether cessation of mass drug administration (MDA) for lymphatic filariasis (LF) is warranted. Ministries of health typically implement TAS in evaluation units (EUs) that have had more than five rounds of annual MDA. Under TAS guidelines, sample size calculations determine a decision value: if the number of individuals testing positive exceeds this threshold, MDA continues in the EU. The objective of this study was to determine if fine scale geospatial covariates could be used to identify predictors of TAS failure. We geo-referenced 746 TAS EUs of which 65 failed and extracted geospatial covariates using R to estimate odds of failure. We implemented stepwise backwards elimination to select covariates for inclusion in a logistic regression to estimate the odds of TAS failure. Covariates included environmental predictors (aridity, distance to fresh water, elevation, and enhanced vegetation index), cumulative rounds of MDA, measures of urbanicity and access, LF species, and baseline prevalence. Presence of *Brugia* was significantly associated with TAS failure (OR 4.79, 95% CI 2.52 – 9.07), as was population density (OR 2.91, 95% CI 1.06 – 7.98). The presence of nighttime lights was highly protective against failure (OR 0.22,

95% CI 0.10 – 0.50), as was an increase in elevation (OR 0.36, 95% CI 0.18 – 0.732). This work identifies predictors associated with TAS failure at the EU areal level given the data currently available, and also identifies the need for more granular data in order to conduct a more robust assessment of these predictors.

Introduction

Lymphatic filariasis (LF) is a mosquito-borne parasitic infection caused by *Wuchereria bancrofti*, *Brugia malayi*, and *B. timori*. Transmission among humans is largely sustained by mosquitos from the genera *Anopheles*, *Aedes*, *Culex*, and *Mansonia*, and typically observed in rural and underserved communities. *W. bancrofti* are responsible for 90% of human cases, and disease transmission occurs when a mosquito bites a human infected with lymphatic filariasis, acquiring the microfilariae circulating in the blood. The microfilariae develop into larvae, which can enter a new human host when the infected mosquito is feeding. The mosquito-borne larvae then develop into threadlike adult-stage worms and can live 4-6 years in the human lymph system, producing microfilariae and continuing the transmission cycle (1).

Prolonged and repeated infection can result in chronic physical manifestations that are a result of the general impact on the lymphatic system, characterized by swelling of the limbs (lymphedema) or scrotum (hydrocele) (2). Episodes of acute disability occur generally in the form of adenolymphangitis (ADL), often a precursor to lymphedema involving fever, swelling, and malaise. ADL episodes can last for 3 to 15 days at a time, and recur multiple times in a year (3). While the majority of LF-endemic areas are found in Africa and Southeast Asia, the global distribution of LF is widespread, with either evidence of current or previously endemic countries in the Caribbean, South America, the Middle East, and the Pacific Islands. Historical estimates suggest that approximately 119 million people may have been infected with LF globally and another 1.3 billion reside in areas suitable for transmission in 2000 (4, 5). The burden of LF-related disability has also been quantified, with over 1.9 million Disability Adjusted Life Years (DALYs) attributed to hydrocele, lymphedema, and ADL in 2000 (6).

In 1997, the World Health Assembly adopted resolution WHA50.29, the "elimination of lymphatic filariasis as a public health problem" followed by the launch of the Global Programme to Eliminate Lymphatic Filariasis (GPELF) in 2000 (7, 8). Under GPELF, the WHO recommends annual mass drug administration (MDA) with antihelminthic medicine (diethylcarbamazine (DEC), DEC and albendazole, or ivermection and albendazole) for at least five consecutive years reaching a minimum of 65% of the population residing in endemic areas defined as implementation units (IU), typically following second-order administrative boundaries such as districts or counties. The objectives of the GPELF were to interrupt transmission of infection and reduce the incidence of LF-related morbidity through the treatment of at least 845 million individuals globally for at least five years (9).

From 2000 to 2009, more than 2.8 billion treatments were administered to the population at risk of infection (9), and in 2011 the WHO introduced the methodology for the Transmission Assessment Survey (TAS) to determine whether implementation units could cease MDA implementation and begin post-MDA surveillance. Under the TAS design, a sample of approximately 1,500 children ages 6-7 years from either communities or schools are tested for LF infection across an area of space defined as the evaluation unit (EU). EU boundaries can be equivalent to IU boundaries, but can also be formed by combining multiple IUs into a single EU for survey or by subdividing an IU in to multiple EUs. Species and dominant vector type, net primary-school enrolment ratio, total number of 6 and 7 year olds in an EU, and number of primary schools or census enumeration areas (EAs) impact the survey design for a TAS. According to various thresholds for each of these factors as defined in TAS methodology, sampling is either a primary school or EA-based survey, using either a systematic sample or a cluster survey of 6 and 7 year olds (10). Regardless of the type of sampling implemented, the EU is the unit of reporting for TAS surveys and the underlying microdata rarely become publicly available. Sample size calculations dictate a decision rule (called a "critical cut-off value") for the number of children who test positive that would

approximate the minimum prevalence at which LF transmission theoretically could be sustained. These thresholds are species and vector-specific (10), based on models used to determine the impact of MDA on LF transmission (11, 12). If the number of children who test positive exceeds this "critical cut-off value," the EU fails the TAS and MDA should continue. If an EU passes the TAS, annual MDA is no longer recommended and TAS surveys are repeated an additional two more times over a five-year period as part of post-MDA surveillance (10).

To date, 92% of TAS have passed, but 18 countries have experienced at least one failed TAS (13). The objective of this study was to determine if environmental covariates summarized over the geographic boundaries of the EU were associated with an increased odds of failing a TAS. Since the results of TAS surveys implemented at the EU level are areal data, it is possible that areal-level information could be used to characterize these units of space. If predictors of TAS failure could be identified at the EU-level, national program managers could use such information to enhance implementation and supervision of MDA interventions among corresponding IUs. Further, such predictors could be used to intensify the program monitoring of areas at a higher risk of failure.

Methods

Input Data

Transmission assessment surveys were implemented by ministries of health and reported to the WHO as part of monitoring guidelines for the Global Program for the Elimination of Lymphatic Filariasis. In this analysis, TAS failure was defined by the number of children who test positive for LF across an entire EU exceeding the "critical cut-off value." Sample size and cut-off values were calculated using a lot quality assurance sampling method and could be determined using a *Survey sample builder* (10). These values

were designed to ensure that an EU has at least a 75% chance of passing the TAS if the true prevalence of LF is half the threshold, and no more than a 5% chance of falsely passing the TAS if the true prevalence is greater than or equal to the threshold. The thresholds are species- and vector-specific, and are described in Table 1. If the critical cut-off value was not reported in the original data, the TAS was assumed to have failed if the proportion of children testing positive exceeded the theoretical prevalence thresholds as defined by LF species and vector. In this analysis, if both LF species were co-endemic, the TAS was classified as a failure if the number positive in a TAS exceeded at least one critical cut-off value. Children who tested positive using diagnostics including the circulating antigen tests rapid immunochromatographic test (ICT) and Filariasis Test Strip (FTS) and filarial antibody detection tests for *Brugia spp.*, such as the Brugia Rapid™ test, were considered to be positive for LF infection (10).

Geo-referencing TAS EUs

Geographic information for EUs and IUs were reviewed to first determine if the geographic boundaries were equivalent to IUs for areas that implemented MDA. If so, they were geo-referenced to administrative boundaries by reviewing country maps and fuzzy matching IU names to polygons in either GAUL or GADM shapefiles in ArcMap (10.4.1). If GAUL or GADM administrative boundaries did not reflect programmatic boundaries, then an implementation unit shapefile maintained by the Expanded Special Project for Elimination of Neglected Tropical Diseases (ESPEN) was used for Africa or NTD implementation unit outside of Africa (14, 15). For EUs that covered multiple IUs, country program reports and published scientific literature were reviewed to create custom geography. If IUs followed standard administrative boundaries, GAUL or GADM shapefiles were joined in ArcMap to create custom geographies. 231 (31.0%) EUs were comprised of IUs that did not correspond to standard administrative boundaries, so program reports and publications in peer-reviewed journals were reviewed to identify maps of custom geographies which were then created in ArcMap.

Covariates

Geo-referenced TAS data were linked to a range of environmental and socio-economic covariates, selected after reviewing the literature for their potential to be associated with LF transmission or impact effectiveness of implementation (16). A summary is presented in Table 2 and more information on the source and definitions of these covariates is presented in the SI (Table S1). We compiled these covariates as 5 x 5-km raster layers, and extracted mean or maximum values over each evaluation unit associated with the geo-referenced TAS data. To extract geospatial covariates, we extracted every pixel value encompassed within each EU geography and calculated the respective summary statistic.

In order to account for the spatial distribution of different covariates over an entire EU, as well as avoid sampling locations without human habitation, we overlaid each EU geography with the population raster and sampled 1,000 pixels with replacement, weighted by population density. We took these 1,000 sampled covariate values for each EU-year combination, as illustrated in Figure 1. Extraction of covariate values was performed using R version 3.4.3.

We initially tested predictors in their raw continuous form (results not presented), but needed to be categorized in order to produce comprehensible odds ratios. We converted these continuous covariates into binary categorical variables chosen to correspond to thresholds that could be translated into programmatic recommendations or for identifying areas at a higher risk of failure. Reference groups were defined either by the distribution of extracted values where we selected thresholds producing a relatively even distribution, or by previously determined limits for urbanicity and access in published literature (17).

Exclusion criteria

Since the TAS guidelines took effect in 2011, this analysis only includes TAS implemented from 2011 and onwards reported to the WHO monitoring program or published in the peer-reviewed literature through September 2017. In addition, we excluded TAS observations implemented in historically non-endemic EUs for the purposes of confirmatory mapping, had incomplete data on the number of rounds of MDA, or EUs that were unable to be georeferenced, and TAS in which a community was defined as an EU.

Model

EU Mean Regression and Pixel-Level Median Regression

We tested for an association between environmental and socio-economic covariates and TAS failure using a logistic regression to estimate odds ratios. We implemented backwards elimination to select covariates for inclusion, with a p-value ≤ 0.15 for retention in the model and a p-value ≤ 0.05 to determine statistical significance of associations (18). We also tested a generalized estimating equations (GEE) model to account for dependency due to repeat measures within the same location as TAS are repeated up to three times for a single EU (e.g., TAS1, TAS2, TAS3). Lastly, we conducted a sensitivity analysis to determine whether the thresholds used to create categorical variables biased our results, testing three separate values for each independent variable, retaining the original covariate definitions as presented in the main analysis (see SI Table 2a). We repeated these three regressions and sensitivity analysis for the pixel-level medians (see SI Table 2b).

EU Pixel-Level Draws Regression

In order to simulate the uncertainty associated with aggregating covariate values over space, we ran 1,000 logistic regressions, one for each set of draw values, and generated another 1,000 draws from the betas and standard errors. This resulted in a total of 1,000,000 draws representing the distribution of

the association between environmental and socio-economic covariates and TAS failure. We extracted the exponentiated mean, 2.5th, and 97.5th percentile values for each covariate, generating the mean and confidence intervals associated with the pixel-level draw simulation. Statistical analysis was conducted using STATA 13.

Results

Input Data

A total of 936 TAS survey records were reported to WHO from 2011 to 2017 across 39 countries in South America, Africa, Asia, and Latin America. We excluded data that were unable to be georeferenced (N = 108), and during the covariate extraction process, we were unable to extract all covariates across 10 additional TAS EUs due to small island geography and missingness in the covariate rasters, and so excluded these observations from the final dataset. TAS surveys implemented in EUs to confirm endemicity status or in areas otherwise considered non-endemic and never received MDA were also excluded (N = 32). We also excluded observations that had incomplete data on the number of rounds of MDA (less than a maximum of 4 rounds reported) (N = 40). A total of 75 TAS EUs were matched to IU boundaries using GAUL, 440 were matched using GADM, and for 231 EUs the ESPEN or other custom geography shapefile was used to match IUs.

Of the 746 total observations included in the analysis, 65 (8.7%) failed TAS, 531 (71.2%) were completed using ICT as the diagnostic tool, 125 (16.8%) used Filariasis Test Strip (FTS), 63 (8.4%) used the Brugia Rapid test, and 6 (0.8%) used the identification of microfilariae in a blood smear. 59.1% of the total observations had a mean value 60 minutes or less travel time to nearest settlement of >50,000 inhabitants, while 75.3% had a pixel-level median draw value of 60 minutes or less. 43.0% of the total

observations had a mean value of an elevation more than 200 meters, and only 34.3% had a pixel-level median draw value of more than 200 meters. Population density also differed significantly between the mean EU value and the pixel-level draw median, as 53.9% of observations had a mean value of more than 5,000 people per pixel and 61.3% had a pixel-level median draw value of more than 5,000 people per pixel. Only 17.4% recorded fewer than 6 rounds of MDA, and 46.1% were located in areas endemic to *Brugia spp* in addition to *W. bancrofti*. A summary of the covariate distribution for the analytical dataset is presented in Table 3.

Model

Logistic regression: EU Mean

Nighttime lights, LF species type, elevation, and population density were significantly associated with TAS failure ($p < 0.05$) in the logistic regression with backwards elimination to select covariates for inclusion, using the mean values across each EU. Table 4 provides a summary of associations between the covariates and TAS failure for all models tested. The four significant predictors were statistically significant ($p < 0.05$) across the full and reduced logistic regressions, as well as in the GEE model, and the direction of their associations remained the same. Of the covariates that were retained in the reduced logistic regression, presence of *Brugia* was significantly associated with TAS failure (OR 4.79, 95% CI 2.52 – 9.07), as was population density (OR 2.91, 95% CI 1.06 – 7.98). The presence of nighttime lights (greater than a mean index of 1.5) was highly protective against failure (OR 0.22, 95% CI 0.10 – 0.50), as was a mean elevation of greater than 200 meters (OR 0.36, 95% CI 0.18 – 0.732. Distance to rivers and access were both retained in the reduced logistic regression, but were not statistically significant in the logistic or GEE regressions.

Logistic regression: EU Pixel-Level Draws

When using the median pixel-level draw value for each TAS, nighttime lights, LF species, and elevation were again significantly associated with TAS failure ($p < 0.05$) in the logistic regression with backwards elimination and in the GEE logistic regression. Aridity and population density were also retained, but only aridity was statistically significant in the reduced logistic regression, and was no longer statistically significant in the GEE logistic regression. Presence of *Brugia* was significantly associated with TAS failure in the reduced logistic regression (OR 6.88, 95% CI 3.44 – 13.73), while the presence of nighttime lights and mean elevation of greater than 200 meters were both protective against failure (OR 0.08 95% CI 0.04 – 0.17 and OR 0.37 95% CI 0.17 – 0.83, respectively). The odds ratio associated with presence of nighttime lights decreased when compared to the estimates using the covariate means. Aridity was protective against failure (OR 0.53 95% CI 0.27 – 0.99). Presence of *Brugia*, nighttime lights, and elevation greater than 200 m were all significantly associated with TAS failure in the GEE logistic regression.

In the aggregation of the 1,000 logistic regressions run on the 1,000 pixel-level draws, presence of *Brugia* was significantly associated with TAS failure (OR 5.19 95% CI 2.49 – 11.06), nighttime lights was protective against failure (OR 0.23 95% CI 0.08 – 0.67), and elevation of greater than 200 m was slightly protective (OR 0.38 95% CI 0.14 – 0.99). Table 3 shows a summary of associations and odds ratios for all model inputs and methods tested.

Sensitivity analysis

To determine how sensitive our findings were to the thresholds used to create categorical variables from continuous values extracted from geospatial covariates, we tested three separate categorical thresholds for each independent variable within the reduced logistic regression, using covariates as

defined in the main analysis (Table S2 in the SI). Access and population density, covariates with thresholds defined through previously published literature, did not vary in this sensitivity analysis.

When using covariate means as the model input, species type and elevation remained predictive in every variation of the model and the directionality of their effects did not change. Nighttime lights was retained and statistically significant in every iteration except for one, and population density was included in 9 of the 12 models tested and was statistically significant in 5. Access and distance to rivers were included in 7 iterations, and aridity, EVI, and maximum baseline prevalence were each included twice across the 12 models tested. All of these covariates were only statistically significant once. When using the median pixel-level draw value, nighttime lights and species type remained significantly protective and predictive of failure, respectively, in every variation. Elevation was retained in all but one iteration and was significantly protective against failure in all but two, as was aridity in 9 and 7 out of the 12 models, respectively. Population density was retained in 11 models, but was statistically significant in only one, and maximum baseline prevalence was included in 3 models, and was significantly predictive of failure in one. Table S2 in the SI provides a summary of associations between covariates and TAS failure for each iteration of this sensitivity analysis.

To address the potential bias introduced with the population-weighted sampling of pixels, we tested the pixel-level analysis without the population weighting but excluded all pixels that had a population density of fewer than ten people per 1 x 1 km and were classified as “barren or sparsely vegetated” (19). Presence of nighttime lights and increased elevation were both statistically significant and protective while population density and species type were predictive of TAS failure across the full, reduced, GEE, and pixel-level logistic regression analyses. Aridity was also protective against failure in the full, reduced,

and GEE logistic regressions, but was no longer statistically significant at the pixel-level analysis. Table S3 in the SI provides a full summary of associations across the models tested.

We also tested the sensitivity of our exclusion of observations with incomplete MDA data (excluding surveys done in EUs with fewer than 5 rounds of recorded MDA) in the reduced logistic regression. Presence of *Brugia*, nighttime lights, and elevation were all retained and statistically significant, maintaining the same directionality as in the primary analysis.

Discussion

At this stage in the LF elimination program, the implementation of TAS across a range of geographic regions offers an opportunity to examine whether the areal classification of spatial covariates is associated with TAS failure. This study provides the first analysis on environmental and socio-economic predictors of TAS failure. Our work identifies predictors that are associated with TAS failure at the EU areal level given the data currently available, and also identifies the need for more granular data in order to conduct a more robust assessment of these predictors.

In this analysis, we sampled from pixel-level values to account for the potential heterogeneity of geospatial covariates that exists within evaluation units. In comparison to the logistic regression results without pixel-level sampling, the point estimates across all data inputs and methodological approaches do not vary significantly, which suggests that extreme variation in the covariate values did not contribute substantial bias. Additionally, we used the GEE logistic regression on the mean and pixel-level median draw values so we could account for dependency between rounds of TAS that are done in the same EU. This also was not significantly different, suggesting that dependency due to repeated

measures did not contribute much bias to the results, largely due to the few number of EUs with more than one observation.

We identify environmental and socio-economic predictors of transmission assessment survey failure. Presence of nighttime lights is consistently and significantly protective of failure across a variety of models and data input types. We selected nighttime lights as a covariate as it can be indicative of socio-economic factors, acting as a proxy for spatial distribution of quality living conditions, which has been shown to be protective against LF transmission (20). The odds ratio for nighttime lights decreases when using the pixel-level median value and the covariate means across all models in the main and sensitivity analyses, likely due to the skewed distribution of nighttime lights and the difference between extracted mean and median values.

We also found that presence of *Brugia spp*. in addition to *W. bancrofti* at the country level is associated with failure. *Brugia spp*. differs from *W. bancrofti* with non-periodic biting patterns and presence of animal reservoirs, which could impact the effectiveness of intervention programs beyond MDA, such as insecticide-treated net (ITN) coverage (21). Elevation was shown to be protective, which is in agreement with previously published literature on disease transmission which consistently finds that LF transmission is negatively associated with increasing elevation (16, 22), likely because increased elevation is less suitable for vector survival (23).

Population density was significantly associated with failure when using EU means, but was no longer significant when using the pixel-level draws or medians. It is often a predictor of LF transmission in some settings (16), and its fluctuating significance could be due to the different vector-specific transmission

patterns, with *Culex* (primarily in East Africa and the Nile Delta) known for its urban transmission, and *Anopheles, Aedes,* and *Mansonia* for rural transmission (16).

While we have sought to leverage any available covariates in testing for predictors of TAS failure, we recognize a number of limitations in this analysis. We included LF species type, but were unable to include dominant vector classification in the regression due to incomplete vector data across all geographies. LF is different from other mosquito-borne diseases, such as malaria or dengue, in that its vectors belong to multiple genera, and experience a variety of breeding habitats and biting patterns which can be influenced by both environmental factors and human-related activity (24, 25). Additionally, an ITN coverage map was not available outside of Africa (26), which could have been indicative of the effectiveness of control and elimination efforts beyond MDA.

Lastly, the TAS data are aggregated over EUs, and characterizing geographic variables over a large unit of space can be problematic and biased with varying amounts of heterogeneity and sensitive to any errors in geo-referencing. We try to account for this with the pixel-level population-weighted draws analysis, but leveraging the specific survey clusters from individual schools or communities is necessary in identifying other drivers of TAS failure. For EUs that failed their TAS, it is important to understand the distribution of LF within these geographic regions. Future analysis should examine associations between the number of children testing positive at the community or school level and geospatial covariates from those locations. If the number of children who test positive for LF infection are clustered within a few locations across an EU, then such results could inform more targeted surveillance and monitoring at the sub-IU level.

Future research should examine predictors at the local level, categorizing areas in need of enhanced supervision. Through the Global Program to Eliminate Lymphatic Filariasis, 18 countries have completed LF interventions and are undergoing surveillance to validate elimination, and an additional 22 countries have delivered MDA to all endemic areas (27). Geospatial information related to TAS failure could result in the more efficient use of program resources during this critical post-MDA surveillance phase of the program.

Acknowledgements

This analysis was commissioned by the World Health Organization Department of Control of Neglected Tropical Diseases to support the Global Programme to Eliminate Lymphatic Filariasis. We acknowledge WHO collaboration and leadership in the project. We also acknowledge the individuals who participated in Transmission Assessment Surveys as part of routine programme evaluation and as well as the national LF elimination programs that reported these data to WHO.

References

1. Ottesen, Eric A. “Lymphatic Filariasis: Treatment, Control and Elimination.” Advances in Parasitology Control of Human Parasitic Diseases, 2006, pp. 395–441., doi:10.1016/s0065-308x(05)61010-x.
2. Taylor, Mark J, et al. “Lymphatic Filariasis and Onchocerciasis.” *The Lancet*, vol. 376, no. 9747, 2010, pp. 1175–1185., doi:10.1016/s0140-6736(10)60586-7.
3. Mues, Katherine E., et al. “Changes in Antibody Levels during and Following an Episode of Acute Adenolymphangitis (ADL) among Lymphedema Patients in Léogâne, Haiti.” Plos One, vol. 10, no. 10, 2015, doi:10.1371/journal.pone.0141047.
4. World Health Organization. Global Programme to Eliminate Lymphatic Filariasis: Progress report on mass drug administrations in 2005. Weekly Epidemiological Record. 2006;22:221–232
5. Michael E, Bundy DA, Grenfell BT. “Re-assessing the global prevalence and distribution of lymphatic filariasis.” Parasitology vol. 112, 1996, pp. 409–428.
6. Institute for Health Metrics and Evaluation (IHME). GBDCompareDataVisualization. Seattle, WA: IHME, University of Washington, 2016. Available from http://vizhub.healthdata.org/gbd-compare. (Accessed April 2018)
7. Ottesen EA, Duke BO, Karam M, Behbehani K. “Strategies and tools for the control/elimination of lymphatic filariasis.” *Bulletin of the World Health Organization*. 75.6 (1997): 491–503.
8. Global programme to eliminate lymphatic filariasis: progress report, 2014. Weekly Epidemiological Record, 2015, 90:489-504.
9. World Health Organization. Progress report 2000-2009 and strategic plan 2010-2020 of the global programme to eliminate lymphatic filariasis: halfway towards eliminating lymphatic filariasis. Geneva : World Health Organization, 2010. http://www.who.int/iris/handle/10665/44473

10. Monitoring and epidemiological assessment of mass drug administration for eliminating lymphatic filariasis: a manual for national elimination programmes. Geneva, World Health Organization, 2011. Available at http://apps.who.int/iris/ handle/10665/44580, accessed April 2018.
11. Swaminathan S et al. “Mathematical models for lymphatic filariasis transmission and control: challenges and prospects.” Parasites & Vectors, 2008, 1:2 (doi: 10.1186/1756-3305-1-2).
12. Michael E et al. “Mathematical modeling and the control of lymphatic filariasis.” Lancet, vol. 4, 2004, pp. 223–234.
13. World Health Organization. “Global programme to eliminate lymphatic filariasis: progress report, 2016.” Weekly Epidemiological Record, vol. 92, no. 40, 2017, pp. 589-608.
14. World Health Organization. Expanded Special Project for Elimination of Neglected Tropical Disease: Lymphatic Filariasis
15. Flueckiger RM, Nikolay B, Gelderblom HC, Smith JL, Haddad D, et al. “Integrating Data and Resources on Neglected Tropical Diseases for Better Planning: The NTD Mapping Tool (NTDmap.org).” PLoS Negl Trop Dis vol. 9, no. 2, 2015, e0003400
16. Cano, Jorge, et al. “The Global Distribution and Transmission Limits of Lymphatic Filariasis: Past and Present.” Parasites & Vectors, vol. 7, no. 1, 2014, pp. 466., doi:10.1186/preaccept-4166218931344497.
17. Burstein, R., et al. “High Spatial Resolution Mapping of Changing Inequalities in Child Mortality Across Africa between 2000 and 2015.” Annals of Global Health, vol. 83, no. 1, 2017, pp. 62–63., doi:10.1016/j.aogh.2017.03.135.
18. Bursac, Z., et al. “Purposeful selection of variables in logistic regression.” *BioMed*, vol. 3, no. 17, 2008.

19. Lloyd, C. T., Sorichetta, A. & Tatem, A. J. High resolution global gridded data for use in population studies. Sci. Data 4, sdata20171 (2017). World Pop. Get data.
20. Bonfim, C., et al. “A socioenvironmental composite index as a tool for identifying urban areas at risk of lymphatic filariasis.” Tropical Medicine and International Health, vol. 14, no. 8, 2009, pp. 877–884., doi:10.1111/j.1365-3156.2009.02317.
21. John, D.T., William A.P. "Brugia malayi." Medical Parasitology. 9th ed. St. Louis, Missouri: Saunders Elsevier, 2006. 285-289.
22. Ngwira B.M.M, et al. “The geographical distribution of lymphatic filariasis infection in Malawi.” Filaria Journal, vol. 6, no. 12, 2007.
23. Manguin S., et al. “Review on global co-transmission of human *Plasmodium* species and *Wuchereria bancrofti* by *Anopheles* mosquitos.” Infection, Genetics, and Evolution, vol. 10, no. 2, 2010, pp. 159-177.
24. Sinka ME, et al. “A global map of dominant malaria vectors.” Parasit Vectors, vol. 5, 2012.
25. Brady OJ, et al. “Refining the global spatial limits of dengue virus transmission by evidence-based consensus.” PLoS Negl Trop Dis., vol. 6, 2012.
26. Bhatt, S., et al. “Coverage and system efficiencies of insecticide-treated nets in Africa from 2000 to 2017.” eLife, vol. 4, 2015, e09672.
27. World Health Organization. “Global Progress Towards Elimination. *NTD Roadmap Target: Global Elimination as a Public Health Problem by 2020.*”
28. Shamsuzzaman, A.K.M., et al. “The significant scale up and success of Transmission Assessment Surveys ‘TAS’ for endgame surveillance of lymphatic filariasis in Bangladesh: One step closer to the elimination goal of 2020.” PLoS Negl Trop Dis., vol. 11, no. 1, 2017, e0005340.

Tables and Figures

Table 1: Species- and vector-specific thresholds for transmission assessment surveys

Species type	**Dominant vector**	**Threshold**
W. bancrofti	*Anopheles* and/or *Culex*	< 2%
W. bancrofti	*Aedes*	< 1%
Brugia spp.	All vectors	< 2%

In areas where *W. bancrofti* is endemic and *Anopheles* and/or *Culex* are the principal vectors the threshold is less than 2%; less than 1% in areas where *W. bancrofti* is endemic and *Aedes* is the principal vector, and less than 2% in areas where *Brugia spp.* is endemic. These thresholds approximate the minimum prevalence at which LF transmission theoretically could be sustained, and a transmission assessment survey fails if it exceeds the threshold.

Table 2: Summary of covariates used in logistic regression

Covariate	**Description**	**Reference Group**	**Summary Statistic**
Access	Travel time to nearest settlement of >50,000 inhabitants	≤60 minutes	Mean
Aridity	Index from the Climatic Research Unit Time-Series	≤ 1	Mean
Distance to rivers	Distance to rivers	≤ 25 km	Mean
Nighttime lights	Nighttime light index from 0-63	≤ 1.5	Mean
Elevation	Elevation measured in meters	≤ 200 m	Mean
EVI	Enhanced Vegetation Index	≤ 0.3	Mean
Irrigation	Mean percentage per pixel equipped for irrigation	≤ 5	Mean
Species	Presence of *W. bancrofti* only versus presence of *Brugia spp.* or presence of *Brugia spp.* and *W. bancrofti*	*Bancrofti* only	Binary value
MDA	Maximum number of recorded rounds of Mass Drug Administration	≤ 5 rounds of MDA	Maximum
Max baseline prevalence	Maximum baseline prevalence observed within the EU	≤ 5% prevalence	Maximum

Table 3: Characteristics of the TAS data and extracted geospatial covariates

	EU Mean	Draw Median
	N (%)	N (%)
TAS observations	746	
Pass*	681 (91.3)	
Fail	65 (8.7)	
Covariates		
Access		
≤ 60 minutes	441 (59.1)	562 (75.3)
> 60 minutes	305 (40.9)	184 (24.7)
Aridity		
≤ 1*	375 (50.3)	382 (51.2)
> 1	371 (49.7)	364 (48.8)
Distance to Rivers		
≤ 25 km*	274 (36.7)	293 (39.3)
> 25 km	472 (63.3)	453 (60.7)
Nighttime lights		
≤ 1.5*	373 (50.0)	411 (55.1)
> 1.5	373 (50.0)	335 (44.9)
Elevation		
≤ 200 m*	425 (57.0)	490 (65.7)
> 200 m	321 (43.0)	256 (34.3)
Enhanced Vegetation Index		
≤ 0.3*	249 (33.4)	259 (34.7)
> 0.3	497 (66.6)	487 (65.3)
Population density		
≤ 5,000 people per pixel	344 (46.1)	289 (38.7)
> 5,000 people per pixel	402 (53.9)	457 (61.3)
Max baseline prevalence		
≤ 5%*	556 (74.5)	
> 5%	190 (25.5)	
MDA		
≤ 5 rounds*	130 (17.4)	
> 5 rounds	616 (82.6)	
Species		
W. bancrofti only*	402 (53.9)	
Brugia and *Brugia* + *W. bancrofti*	344 (46.1)	

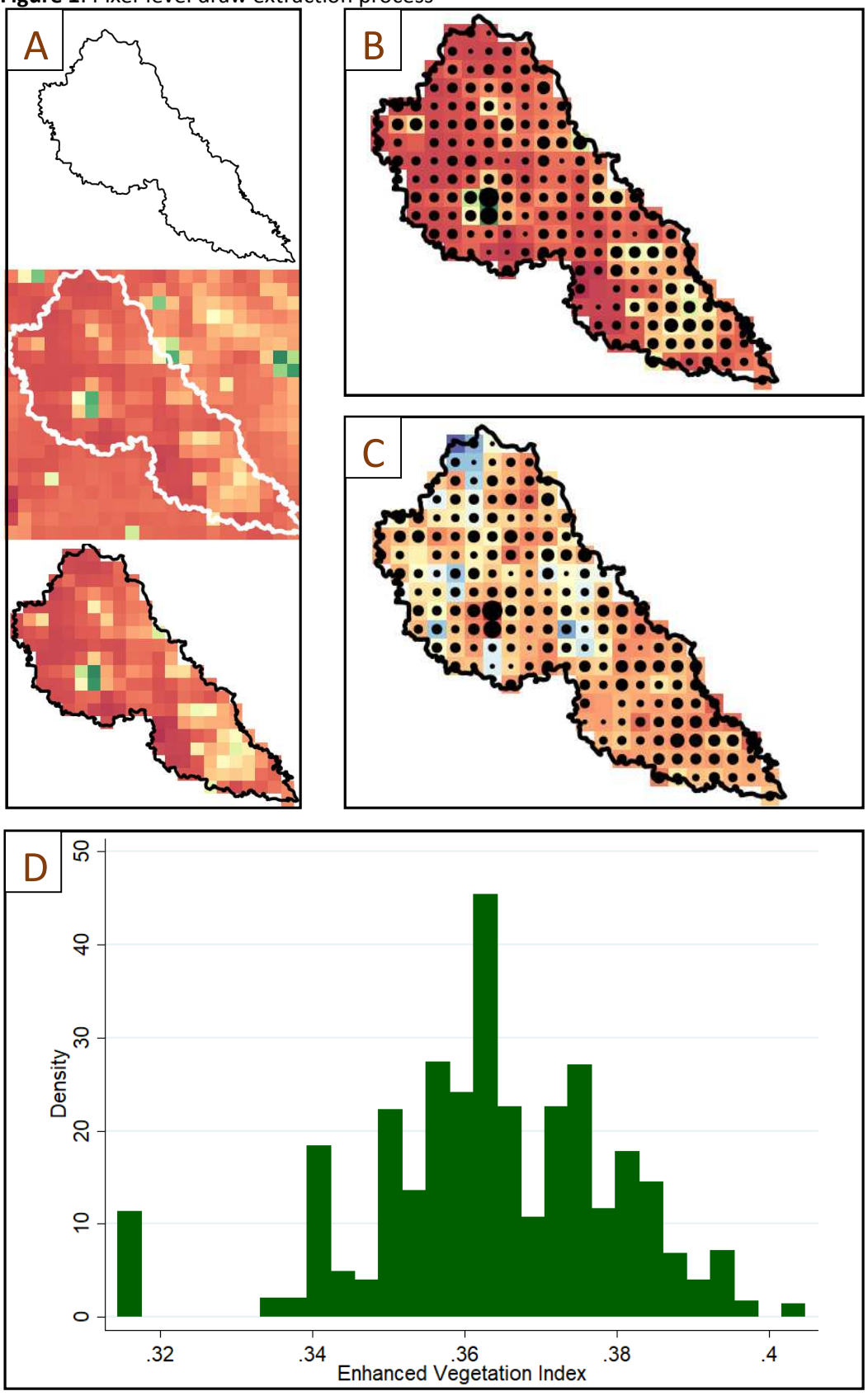
*Denotes reference value

Table 4: Association between geospatial covariates and transmission assessment survey failure in a logistic regression with backwards elimination covariate selection

	Logistic Regression						
	EU Mean			EU Draw Median			Pixel-Level Draws
	Full	Reduced*	GEE	Full	Reduced*	GEE	
Covariate	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)	OR (95% CI)
Access	2.29 (0.84 - 6.30)	2.15 (0.78 - 5.93)	2.15 (0.74 - 6.26)	0.97 (0.44 - 2.14)	--	--	1.00 (0.35 - 2.86)
Aridity	0.59 (0.31 - 1.12)	--	--	0.49 (0.25 - 0.98)	0.53 (0.28 - 0.99)	0.57 (0.29 - 1.09)	0.61 (0.28 - 1.30)
Distance to rivers	1.68 (0.90 - 3.14)	1.66 (0.90 - 3.07)	1.56 (0.81 - 3.00)	1.27 (0.66 - 2.43)	--	--	1.48 (0.68 - 3.37)
Nighttime lights	0.23 (0.10 - 0.53)	0.22 (0.10 - 0.50)	0.25 (0.11 - 0.58)	0.08 (0.04 - 0.17)	0.08 (0.04 - 0.17)	0.09 (0.04 - 0.19)	0.23 (0.08 - 0.67)
Elevation	0.33 (0.16 - 0.67)	0.36 (0.18 - 0.72)	0.38 (0.18 - 0.78)	0.36 (0.16 - 0.81)	0.37 (0.17 - 0.83)	0.39 (0.17 - 0.88)	0.38 (0.14 - 0.99)
Enhanced Vegetation Index	1.58 (0.77 - 3.23)	--	--	1.02 (0.50 - 2.11)	--	--	1.35 (0.56 - 3.40)
Population density	3.03 (1.11 - 8.26)	2.91 (1.06 - 7.98)	3.05 (1.07 - 8.73)	1.89 (0.88 - 4.05)	1.73 (0.91 - 3.29)	1.81 (0.93 - 3.53)	1.22 (0.51 - 2.96)
MDA	0.80 (0.40 - 1.59)	--	--	0.95 (0.46 - 1.96)	--	--	0.84 (0.40 - 1.74)
Max baseline prevalence	1.37 (0.73 - 2.58)	--	--	1.50 (0.78 - 2.89)	--	--	1.42 (0.73 - 2.77)
Species	5.61 (2.78 - 11.33)	4.79 (2.52 - 9.07)	4.87 (2.49 - 9.51)	7.35 (3.59 - 15.03)	6.88 (3.44 - 13.73)	6.76 (3.32 - 13.78)	5.19 (2.49 - 11.06)

* Backwards selection used a p-value ≤ 0.15 for retention in the model

Figure 1: Pixel-level draw extraction process



(A) An example EU shapefile (from Bangladesh) overlaid with a population density raster and cropped to the size of the shapefile (28). (B) Sampling of 1,000 pixels with replacement, weighted by population density. The size of the dot on each pixel represents the number of times it was sampled, with larger dots covering higher population density (green) areas. (C) Pixel-level draws overlaid with the enhanced vegetation index raster, and extracted to produce (D) a distribution of values over the EU, representing geographic heterogeneity.

Supplementary Information

Table S1: Covariate information and source

Covariate	Temporal resolution	Source	Reference
Access	Static	Malaria Atlas Project, University of Oxford	Weiss, D. J. et al. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature 533, 333-336 (2018)
Aridity	Annual	Climatic Research Unit Time-Series (CRUTS)	Harris, I., Jones, P. d., Osborn, T. j. & Lister, D. h. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 dataset. Int. J. Climatol. 34, 623–642 (2014). University of East Anglia. Climatic Research Unit TS v. 3.24 dataset. Available at: https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.24.01/.
Distance to rivers	Static	Natural Earth Data (derived)	Natural Earth. Rivers and lake centerlines dataset. Available at: http://www.naturalearthdata.com/downloads/10mphysical-vectors/10m-rivers-lake-centerlines/.
Nighttime lights	Annual	Moderate Resolution Imaging Spectroradiometer (MODIS)	USGS & NASA. Land surface temperature and emissivity 8-day L3 global 1km MOD11A2 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2.
Elevation	Static	National Oceanic & Atmospheric Administration	National Centers for Environmental Information Available at: https://www.ngdc.noaa.gov/mgg/topo/gltiles.html
Enhanced Vegetation Index (EVI)	Annual	MODIS	Huete, A., Justice, C. & van Leeuwen, W. MODIS vegetation index (MOD 13) algorithm theoretical basis document. (1999). USGS & NASA. Vegetation indices 16-Day L3 global 500m MOD13A1 dataset. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13a1. Weiss, D. J. et al. An effective approach for gapfilling continental scale remotely sensed timeseries. Isprs J. Photogramm. Remote Sens. 98, 106–118 (2014).
Population density	Annual	WorldPop	Lloyd, C. T., Sorichetta, A. & Tatem, A. J. High resolution global gridded data for use in population studies. Sci. Data 4, sdata20171 (2017). World Pop. Get data.

			Available at: http://www.worldpop.org.uk/data/get_data/.
Species	Static	World Health Organization (WHO)	Global Program to Eliminate Lymphatic Filariasis
Maximum rounds of Mass Drug Administration (MDA)	Annual	WHO	Global Program to Eliminate Lymphatic Filariasis
Maximum baseline prevalence	Static	WHO	Global Program to Eliminate Lymphatic Filariasis

Table 2a: Reduced logistic regression using EU means with varying categorical thresholds

	Original	Aridity		Distance to rivers		Nighttime lights		Elevation		EVI		Max baseline prevalence	
	stepwise	≤ 0.8	≤ 1.3	≤ 15 km	≤ 40 km	≤ 1	≤ 5	≤ 150	≤ 500	≤ 0.25	≤ 0.5	≤ 2%	≤ 10%
Access	2.15 (0.78 - 5.93)	2.15 (0.78 - 5.93)	2.15 (0.78 - 5.93)	--	--	2.21 (1.19 - 4.09)	2.21 (0.89 - 5.47)	--	--	2.15 (0.78 - 5.93)	--	2.41 (0.88 - 6.59)	2.15 (0.78 - 5.93)
Aridity	--	--	--	--	--	--	0.43 (0.22 - 0.82)	--	--	--	0.6 (0.32 - 1.15)	--	--
Distance to rivers	1.66 (0.91 - 3.07)	1.66 (0.91 - 3.07)	1.66 (0.91 - 3.07)	--	--	--	2.84 (1.40 - 5.78)	1.63 (0.88 - 3.03)	--	1.66 (0.91 - 3.07)	--	1.65 (0.89 - 3.06)	1.66 (0.91 - 3.07)
Nighttime lights	0.22 (0.10 - 0.50)	0.22 (0.10 - 0.50)	0.22 (0.10 - 0.50)	0.27 (0.15 - 0.49)	0.27 (0.15 - 0.49)	--	0.04 (0.02 - 0.10)	0.20 (0.09 - 0.44)	0.22 (0.10 - 0.48)	0.22 (0.10 - 0.50)	0.17 (0.08 - 0.39)	0.23 (0.10 - 0.50)	0.22 (0.10 - 0.50)
Elevation	0.36 (0.18 - 0.72)	0.36 (0.18 - 0.72)	0.36 (0.18 - 0.72)	0.39 (0.20 - 0.73)	0.39 (0.20 - 0.73)	0.43 (0.22 - 0.83)	0.25 (0.12 - 0.51)	0.53 (0.30 - 0.95)	0.51 (0.22 - 1.14)	0.36 (0.18 - 0.72)	0.41 (0.21 - 0.80)	0.33 (0.16 - 0.68)	0.36 (0.18 - 0.72)
Enhanced Vegetation Index	--	--	--	--	--	1.74 (1.19 - 4.09)	--	--	--	--	2.16 (0.98 - 4.77)	--	--
Population density	2.91 (1.06 - 7.98)	2.91 (1.06 - 7.98)	2.91 (1.06 - 7.98)	--	--	--	3.84 (1.52 - 9.71)	2.03 (0.95 - 4.35)	1.97 (0.94 - 4.09)	2.91 (1.06 - 7.98)	2.18 (0.94 - 5.08)	2.97 (1.09 - 8.13)	2.91 (1.06 - 7.98)
MDA	--	--	--	--	--	--	--	--	--	--	--	--	--
Max baseline prevalence	--	--	--	--	--	--	1.71 (0.87 - 3.34)	--	--	--	--	2.26 (1.29 - 3.94)	--
Species	4.79 (2.52 - 9.07)	4.79 (2.52 - 9.07)	4.79 (2.52 - 9.07)	4.83 (2.57 - 9.07)	4.83 (2.57 - 9.07)	3.45 (1.88 - 6.34)	8.85 (4.41 - 17.75)	5.17 (2.73 - 9.77)	5.02 (2.63 - 9.58)	4.79 (2.52 - 9.07)	4.95 (2.49 - 9.82)	5.69 (2.90 - 11.18)	4.79 (2.52 - 9.07)

Table 2b: Reduced logistic regression using EU pixel-level median draws with varying categorical thresholds

	Original	Aridity		Distance to rivers		Nighttime lights		Elevation		EVI		Max baseline prevalence	
	stepwise	≤ 0.8	≤ 1.3	≤ 15 km	≤ 40 km	≤ 1	≤ 5	≤ 100	≤ 400	≤ 0.25	≤ 0.5	≤ 2%	≤ 10%
Access	--	--	--	--	--	--	--	--	--	--	--	--	--
Aridity	0.53 (0.28 - 0.99)	--	0.53 (0.27 - 0.99)	--	0.53 (0.28 - 0.99)	0.53 (0.28 - 0.99)	0.47 (0.25 - 0.90)	0.53 (0.28 - 1.02)	--	0.53 (0.28 - 0.99)	0.53 (0.28 - 0.99)	0.54 (0.28 - 1.03)	0.53 (0.28 - 0.99)
Distance to rivers	--	--	--	--	--	--	--	--	--	--	--	--	--
Nighttime lights	0.08 (0.04 - 0.17)	0.10 (0.05 - 0.20)	0.08 (0.04 - 0.17)	0.16 (0.08 - 0.33)	0.08 (0.04 - 0.17)	0.08 (0.04 - 0.17)	0.05 (0.02 - 0.12)	0.07 (0.03 - 0.15)	0.09 (0.05 - 0.19)	0.08 (0.04 - 0.17)	0.08 (0.04 - 0.17)	0.07 (0.03 - 0.16)	0.08 (0.04 - 0.17)
Elevation	0.37 (0.91 - 3.29)	0.47 (0.22 - 1.01)	0.38 (0.17 - 0.84)	0.41 (0.19 - 0.89)	0.37 (0.17 - 0.83)	0.37 (0.17 - 0.83)	0.39 (0.17 - 0.90)	0.50 (0.26 - 0.95)	--	0.37 (0.17 - 0.83)	0.37 (0.17 - 0.83)	0.36 (0.16 - 0.81)	0.37 (0.17 - 0.83)
Enhanced Vegetation Index	--	--	--	--	--	--	--	--	--	--	--	--	--
Population density	--	1.86 (0.98 - 3.52)	1.69 (0.88 - 3.22)	--	1.73 (0.91 - 3.29)	1.73 (0.91 - 3.29)	1.61 (0.86 - 3.03)	1.86 (0.98 - 3.53)	2.21 (1.19 - 4.09)	1.73 (0.91 - 3.29)	1.73 (0.91 - 3.29)	1.76 (0.91 - 3.41)	1.73 (0.91 - 3.29)
MDA	--	--	--	--	--	--	--	--	--	--	--	--	--
Max baseline prevalence	--	--	--	1.77 (0.90 - 3.48)	--	--	1.77 (0.91 - 3.42)	--	--	--	--	2.34 (1.21 - 4.19)	--
Species	6.88 (3.44 - 13.73)	5.81 (3.02 - 11.18)	7.04 (3.52 - 14.07)	4.70 (2.34 - 9.43)	6.88 (3.44 - 13.73)	6.88 (3.44 - 13.73)	8.22 (4.10 - 16.49)	8.10 (4.10 - 16.01)	7.41 (3.99 - 13.77)	6.88 (3.44 - 13.73)	6.88 (3.44 - 13.73)	5.61 (2.90 - 10.84)	6.88 (3.44 - 13.73)

Table S3: Association between geospatial covariates and transmission assessment survey failure in a logistic regression with backwards elimination covariate selection using non-population-weighted pixel-level draws

Covariate	Full OR (95% CI)	Reduced* OR (95% CI)	GEE OR (95% CI)	Pixel-level Draws OR (95% CI)
Access	1.09 (0.49 - 2.45)	--	--	1.05 (0.38 - 2.92)
Aridity	0.40 (0.20 - 0.82)	0.42 (0.22 - 0.83)	0.44 (0.22 - 0.87)	0.62 (0.28 - 1.40)
Distance to rivers	1.80 (0.91 - 3.56)	1.84 (0.95 - 3.57)	1.85 (0.94 - 3.67)	1.66 (0.73 - 3.98)
Nighttime lights	0.04 (0.02 - 0.10)	0.04 (0.02 - 0.11)	0.04 (0.02 - 0.11)	0.18 (0.05 - 0.57)
Elevation	0.38 (0.17 - 0.85)	0.38 (0.17 - 0.82)	0.38 (0.17 - 0.84)	0.37 (0.14 - 0.94)
Enhanced vegetation index	1.41 (0.64 - 3.10)	--	--	1.21 (0.49 - 3.13)
Population density	2.85 (1.16 - 6.99)	2.46 (1.21 - 5.03)	2.56 (1.23 - 3.67)	1.52 (0.62 - 3.91)
MDA	1.10 (0.54 - 2.26)	--	--	0.90 (0.43 - 1.88)
Max baseline prevalence	1.62 (0.83 - 3.17)	--	--	1.46 (0.75 - 2.86)
Species	9.05 (4.49 - 18.24)	8.25 (4.19 - 16.24)	8.13 (4.08 - 16.22)	5.30 (2.56 - 11.18)

* Backwards selection used a p-value ≤ 0.15 for retention in the model