Statistical Analysis Plan

Dominic D. LaRoche

Monday, November 3, 2014

**Project:** Are people at high risk for for poor outcomes from malaria infection being targeted for prophylactic measures in Kenya?

**PI:** Dr. Kacey Ernst

**Statistician** Dominic D. LaRoche

**Version:** 1

# Introduction

Mosquitoes are a problematic transmission vector for a number of infectious diseases in tropical and sub-tropical regions throughout the globe. These diseases, such as malaria, can be particularly deadly for vulnerable populations with restricted access to healthcare. Several promising methods for mitigating the risk of infection from mosquitoes have been deployed recently in high-risk regions. The use of bed nets can reduce the number of encounters with mosquitoes and the diseases they harbor while the use of aerial pesticides can reduce the local mosquito population. Both of these methods have been employed in Kenya in recent years but have been administered under different protocols.

Both bed nets and aerial spraying take time and money to deploy. Therefore, it is critical that treatments are first applied to populations at the highest risk of poor outcome from exposure to mosquitoes and only later applied to populations with low risk of poor outcome from exposure. Some measures have been taken to ensure at-risk populations receive priority when administering treatment such as distributing bed nets to pregnant women. However, it is unclear whether these measures are adequate in prioritizing mitigation for at-risk populations. The purpose of this study is to determine whether individuals with the highest risk are more likely to receive a mitigating treatment.

# Data Description

## Household Data

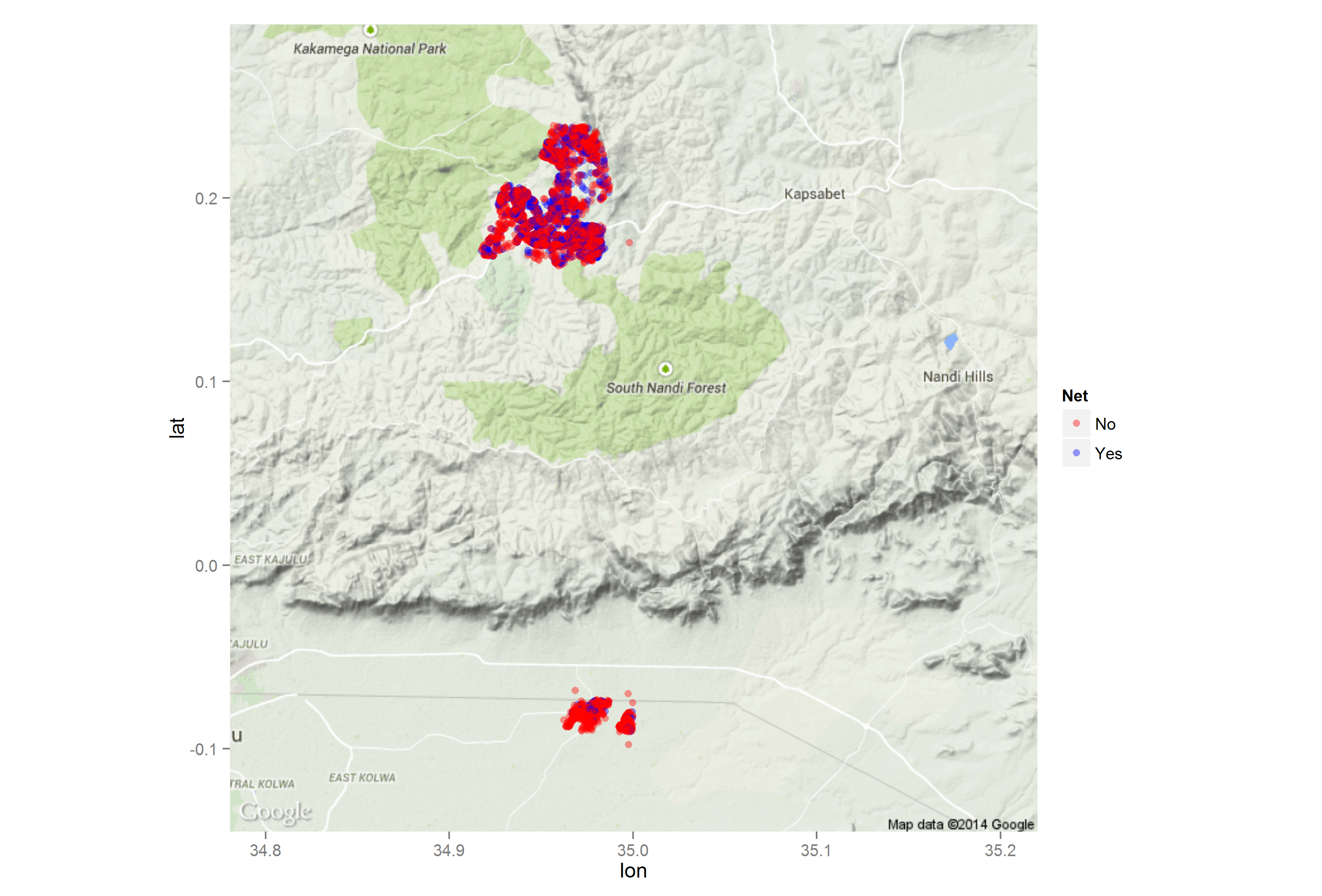
The data are composed of 3,984 households at two sites in Kenya (fig. 1). These two sites represent high elevation and low elevation populations. Both sites have had partial treatment with both bed nets and aerial spraying. The high site has more prevalent bed net usage whereas the low site has more prevalent aerial spraying (figs. 2-5). Each occupant of the 3,984 households was interviewed about bed-net usage and household spraying for a total of 17,823 interviews. Additional information for each participant was also collected such as age, sex, and relation to the head of the household. Table 1 summarizes household information by bed net use and table 2 by spraying.

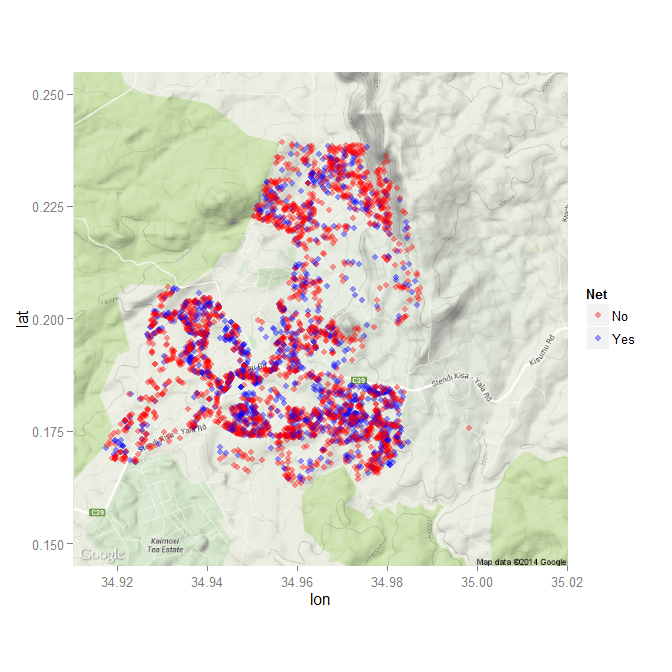
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | No Bed Net | | Bed Net | | Total | |
| Variable | Categories | N | % | N | % | N | % |
| Under 5 | No | 2002 | 72.3% | 753 | 62% | 2755 | 69.2% |
|  | Yes | 768 | 27.7% | 461 | 38% | 1229 | 30.8% |
|  |  |  |  |  |  |  |  |
| Under 1 | No | 2570 | 92.8% | 1005 | 82.8% | 3575 | 89.7% |
|  | Yes | 200 | 7.2% | 209 | 17.2% | 409 | 10.3% |
|  |  |  |  |  |  |  |  |
| Under 18 | No | 2292 | 82.7% | 1019 | 83.9% | 3311 | 83.1% |
|  | Yes | 478 | 17.3% | 195 | 16.1% | 673 | 16.9% |
|  |  |  |  |  |  |  |  |
| Peripheral Family | No | 2626 | 94.8% | 1093 | 90% | 3719 | 93.3% |
|  | Yes | 143 | 5.2% | 121 | 10% | 264 | 6.6% |
|  | Missing | 1 | 0% | 0 | 0% | 1 | 0% |
|  |  |  |  |  |  |  |  |
| House Sprayed | No | 2107 | 76.1% | 1034 | 85.2% | 3141 | 78.8% |
|  | Yes | 663 | 23.9% | 180 | 14.8% | 843 | 21.2% |
|  |  |  |  |  |  |  |  |
| Site | High | 2211 | 79.8% | 1168 | 96.2% | 3379 | 84.8% |
|  | Low | 559 | 20.2% | 46 | 3.8% | 605 | 15.2% |
|  |  |  |  |  |  |  |  |
| House Head Sex | F | 873 | 31.5% | 379 | 31.2% | 1252 | 31.4% |
|  | M | 1780 | 64.3% | 794 | 65.4% | 2574 | 64.6% |
|  | Missing | 117 | 4.2% | 41 | 3.4% | 158 | 4% |
|  |  |  |  |  |  |  |  |
| Number Occ. |  | 4.2 (2.3) |  | 4.8 (2.3) |  | 4.4 (2.3) |  |
| Num. Fem. |  | 2.1 (1.5) |  | 2.5 (1.4) |  | 2.3 (1.5) |  |
| Num. Male |  | 2.1 (1.4) |  | 2.3 (1.5) |  | 2.1 (1.5) |  |
| Age of Oldest Occ. |  | 47.6 (17.8) |  | 44.3 (15.4) |  | 46.6 (17.2) |  |
| Age of Youngest Occ. |  | 14.6 (19.4) |  | 9.1 (14.1) |  | 12.9 (18.1) |  |
| Age of Household Head |  | 47.6 (17.5) |  | 43.9 (14.5) |  | 46.4 (16.7) |  |
| Elevation (m) |  | 1631.9 (382.6) |  | 1729.2 (447.3) |  | 1661.6 (405.8) |  |
| Total |  | 2770 | 69.5% | 1214 | 30.5% | 3984 |  |

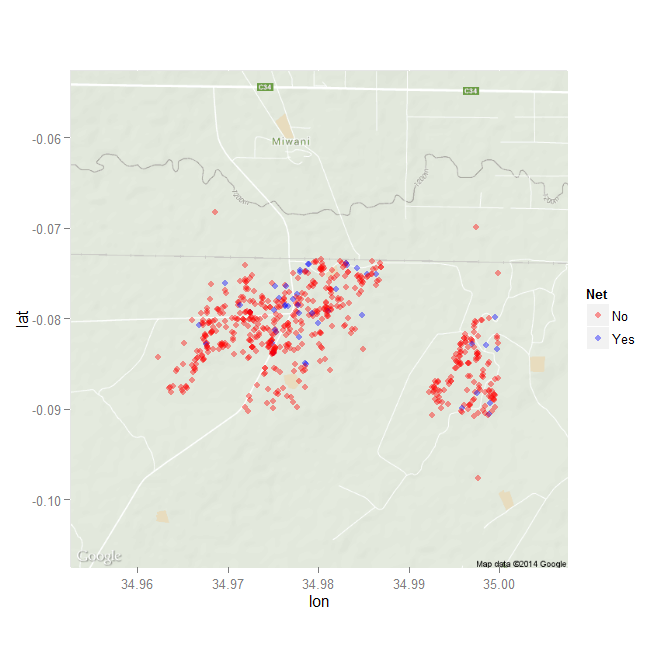
Table 1. Summary of household attributes by houses with and without nets. Frequencies are given for categorical variables and means and standard deviations are given for continuous variables.

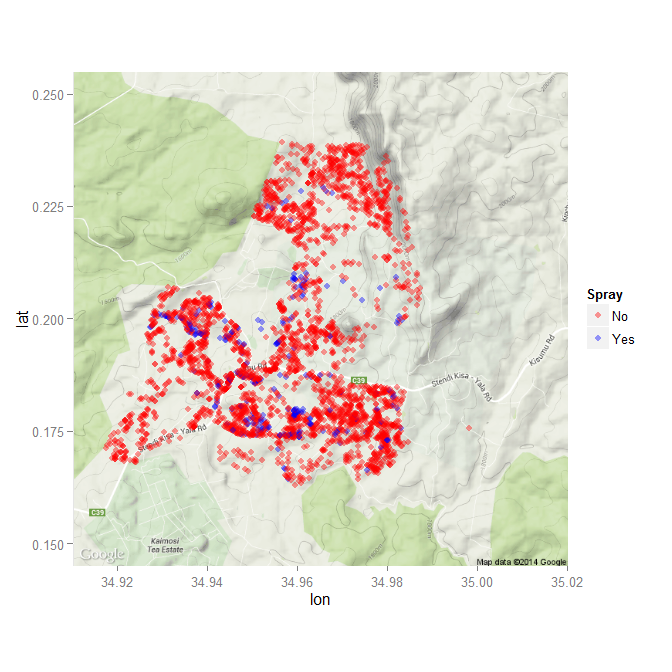
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | No | No % | Yes | Yes % | Total | Total % |
| Variable | Categories | N | % | N | % | N | % |
| Under 5 | No | 2157 | 68.7% | 598 | 70.9% | 2755 | 69.2% |
|  | Yes | 984 | 31.3% | 245 | 29.1% | 1229 | 30.8% |
|  |  |  |  |  |  |  |  |
| Under 1 | No | 2819 | 89.7% | 756 | 89.7% | 3575 | 89.7% |
|  | Yes | 322 | 10.3% | 87 | 10.3% | 409 | 10.3% |
|  |  |  |  |  |  |  |  |
| Under 18 | No | 2613 | 83.2% | 698 | 82.8% | 3311 | 83.1% |
|  | Yes | 528 | 16.8% | 145 | 17.2% | 673 | 16.9% |
|  |  |  |  |  |  |  |  |
| Peripheral Family | No | 2913 | 92.7% | 806 | 95.6% | 3719 | 93.3% |
|  | Yes | 227 | 7.2% | 37 | 4.4% | 264 | 6.6% |
|  | Missing | 1 | 0% | 0 | 0% | 1 | 0% |
|  |  |  |  |  |  |  |  |
| Bed Net | No | 2107 | 67.1% | 663 | 78.6% | 2770 | 69.5% |
|  | Yes | 1034 | 32.9% | 180 | 21.4% | 1214 | 30.5% |
|  |  |  |  |  |  |  |  |
| Site | High | 3124 | 99.5% | 255 | 30.2% | 3379 | 84.8% |
|  | Low | 17 | 0.5% | 588 | 69.8% | 605 | 15.2% |
|  |  |  |  |  |  |  |  |
| House Head Sex | F | 994 | 31.6% | 258 | 30.6% | 1252 | 31.4% |
|  | M | 1996 | 63.5% | 578 | 68.6% | 2574 | 64.6% |
|  | Missing | 151 | 4.8% | 7 | 0.8% | 158 | 4% |
|  |  |  |  |  |  |  |  |
| numocc |  | 4.3 (2.3) |  | 4.6 (2.3) |  | 4.4 (2.3) |  |
| numFem |  | 2.2 (1.5) |  | 2.3 (1.5) |  | 2.3 (1.5) |  |
| numMale |  | 2.1 (1.5) |  | 2.2 (1.4) |  | 2.1 (1.5) |  |
| oldest |  | 46.1 (16.7) |  | 48.6 (18.7) |  | 46.6 (17.2) |  |
| youngest |  | 12.8 (17.9) |  | 13.2 (19) |  | 12.9 (18.1) |  |
| headage |  | 45.9 (16.1) |  | 48.2 (18.7) |  | 46.4 (16.7) |  |
| elev |  | 1745.2 (399.3) |  | 1350.1 (247.7) |  | 1661.6 (405.8) |  |
| Total |  | 3141 | 78.8% | 843 | 21.2% | 3984 |  |

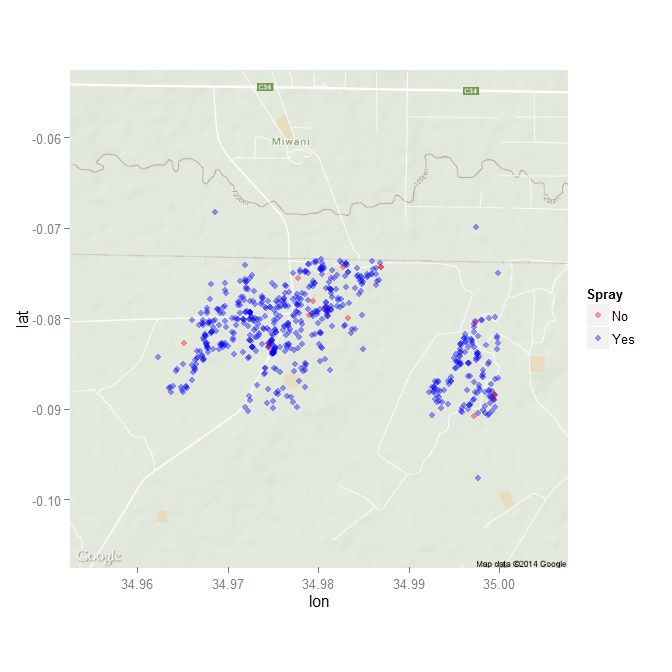
Table 2. Summary of household attributes by houses with and without aerial spraying. Frequencies are given for categorical variables and means and standard deviations are given for continuous variables.

 Figure.1 Locations of households in study. Blue dots represent households with nets.

 Figure 2. Households with and without nets at the high site.

 Figure 3. Households with and without nets at the low site.

 Figure 4. Households with and without spraying at the high site.

 Figure 5. Households with and without spraying at the low site.

## Geo-Spatial Data

The Biodiversity Monitoring Transect Analysis in Africa network [(BIOTA)](http://www.biota-africa.de/East_GISWeb2_ba.php) has created a high resolution (10m) digital terrain model of the high elevation site in this study. Lower resolution data (30m) is available for both the high and low site from the NASA Shuttle Radar Topography Mission [(SRTM)](http://www2.jpl.nasa.gov/srtm/cbanddataproducts.html). The high resolution model was developed from topographic maps as part of BIOTA sub-project E02 but access to this data is currently in question.

# Study Objectives

## Primary Objective

The primary objective of this research is to determine whether populations which are both at high combined risk for 1) a poor outcome from malaria, and 2) encountering a mosquito hosting malaria, are receiving mitigating treatments at a higher rate than those with a low combined risk. Specifically we will test the following alternative hypotheses:

High-risk households are not more likely to receive treatment than low-risk households.  
 High-risk households are more likely to receive treatment than low-risk households.

## Secondary Objectives

As a secondary objective, we will determine if populations which are only at risk either for a poor health outcome from a malaria infection, *or* for a mosquito encounter, are preferentially receiving mitigating treatments. The secondary objective will allow us to determine if either risk is currently being addressed by existing protocols. By combining measures from the primary and secondary outcomes we will evaluate the potential benefit from modifying treatment administration protocols to incorporate more information on household risk in addition to the current strategy.

# Study Variables and Covariates

The survey data are composed of 17,823 interviews at 3,984 households. The unit of study for this analysis is the household. Therefore, I will summarize the information from the surveys into household attributes. Each household will vary with respect to both the risk of exposure to mosquitoes and the number of at risk individuals in the household. We will assign a health risk to each household by adding the number of children in the household under age 5 and the number of adults in the household over age 65.

We will assign each household a risk for exposure to mosquitoes by deriving a continuous risk surface over the study area. This risk will result from a Topographical Wetness Index (TWI) derived from the DEM data. The TWI combines the total basin area (the area from which water will flow to a particular point) with the slope at that point to determine the amount of water likely to accumulate and provide breeding habitat for mosquitoes. We will create a continuous risk surface over the landscape by interpolating between points of high TWI.

# Statistical Methods

Since the two sites have dramatically different rates of spraying and bed net usage we will analyze the high and low sites separately with regards to these outcomes. We will also analyze the spraying and bed net usage separately within each site since these are known to be distributed to households under different protocols and are therefore likely to have different patterns.

## Primary Analysis

To be at risk for a poor outcome a person must 1) come in contact with a malaria hosting mosquito, and 2) be inherently vulnerable to malaria infection (i.e. very young or very old). We will create two risk scores representing each of these household risks. The household health risk will be equal to the number of individuals at high risk (i.e. under age 5 or over age 65). The household exposure risk will be equal to the score of the risk surface at the house location. Since these risks will be calculated on different scales we will center at 0 and standardize risk scores so that they are scale-independent.

We will add the standardized household health risk with the standardized household exposure risk to create a combined risk. We will then determine if high risk households are more likely to have received either a bed-net or aerial spraying with a logistic model;

where p = Probability of a house having a treatment. If is > 1 and statistically significant () then high-risk households are more likely to receive treatment. If is < 1 and statistically significant then high-risk households are less likely to receive treatment. If is not statistically different from 1 then either there is no difference in spraying or bed net usage between high and low risk households or the test did not have enough power to detect a difference (see sensitivity analysis).

## Secondary Analyses

Existing protocols may be adequate at addressing household risk due to either inherent health risk or mosquito exposure risk, but not both. We will conduct the same primary analysis but separate out these two risk scores as separate predictors:

where p = Probability of a house having a treatment. The interpretation of and is the same as from the primary analysis, but specific to a risk type.

## Sensitivity Analyses

It is important to note that the result of a type II error from the test of using logistic regression could result in an unnecessary change in protocol because we fail to find a preferential administration of treatments to the high risk households when there really is one. This failure to find a difference could be due solely to poor power in the test. In the logistic regression the null hypothesis, , is no difference in the administration of treatments. Under this null the probability of finding a difference when there really is none is controlled at . However, since protocols are in place to provide bed nets and spraying to at risk households, we feel it is also important to directly test whether high risk households are equally likely to receive a treatment as low-risk households. We will test this hypothesis by modifying the null and alternative hypotheses:

High-risk households are more likely to receive treatment than low-risk households.  
 High-risk households are not more likely to receive treatment than low-risk households.

Therefore, it is appropriate to set the null hypothesis to test whether lies outside an "equivalence margin" (a difference which policy makers would consider trivial). Unfortunately, it is unclear what size of equivalence margin would be considered appropriate by policy makers. Therefore, we will test equivalence with 3 different margins: OR=0.95-1.05, 0.90-1.10, and 0.85-1.15. We will use a two-one-sided test (TOST) procedure to test for equivalence by treating as the measured difference between high-risk and low-risk households.

# List of Tables and Figures

Table 1: Parameter estimates and 95% confidence intervals from tests of primary objectives.  
Table 2: Parameter estimates and 95% confidence intervals from tests of secondary objectives.  
Table 3: Parameter estimates and 95% confidence intervals from sensitivity analyses.  
Figure 1: Map of high site showing risk level and treatment type for each household.  
Figure 2: Map of low site showing risk level and treatment type for each household.

## Shells for Tables

Tables 1-3 will all follow the same format.

|  |  |  |  |
| --- | --- | --- | --- |
| Effect | N | Estimate | 95% CI |
| High Net | 3379 |  |  |
| High Spray | 3379 |  |  |
| Low Net | 605 |  |  |
| Low Spray | 605 |  |  |