

Predicting Malaria in a Highly Endemic Country using Environmental and Clinical Data Sources

Kate Zinszer^{*1}, Ruth Kigozi², Katia Charland¹, Grant Dorsey^{4,2}, Moses Kamya^{3,2} and David Buckeridge¹

¹McGill University, Montreal, QC, Canada; ²UMSP, Kampala, Uganda; ³Makerere University, Kampala, Uganda; ⁴University of California San Francisco, San Francisco, CA, USA

Objective

The objective of the research was to identify the most accurate models for forecasting malaria at six different sentinel sites in Uganda, using environmental and clinical data sources.

Introduction

In 1911, Christophers developed an early-warning system for malaria epidemics in Punjab based on rainfall, fever-related deaths and wheat prices.¹ Since that initial system, researchers and practitioners have continued to search for determinants of spatial and temporal variability of malaria to improve systems for forecasting disease burden. Malaria thrives in poor tropical and subtropical countries where resources are limited.² Accurate disease prediction and early warning of increased disease burden can provide public health and clinical health services with the information needed to implement targeted approaches for malaria control and prevention that make effective use of limited resources. Malaria forecasting models do not typically consider clinical predictors, such as type of antimalarial treatment, in the forecasting models. The objective of the research was to identify the most accurate models for forecasting malaria at six different sentinel sites in Uganda, using environmental and clinical data sources.

Methods

Health clinic data collected by the sentinel site surveillance program, Uganda Malaria Surveillance Project (UMSP), were used to measure the outcome of laboratory-confirmed malaria. Environmental variables considered included satellite-derived rainfall, daytime and nighttime temperature, and vegetation estimates. Clinical predictors considered included treatment variables, screening of suspected cases, and malaria-negative cases. Weekly, site-specific time series were created and examined for all variables. For the long-term models, the AIC was used to guide the model selection through a step-wise selection process whereas the smallest absolute error (AE) was used for the short-term and intermediate-term models combined with the AIC. All potential predictors were incorporated at appropriate lags into Autoregressive Integrated Moving Average (ARIMA) forecast models for laboratory-confirmed malaria. We generated short-term (4 week), intermediate (26 week), and long-term (52 week) forecasts of malaria prevalence at weekly intervals and used a reserved testing portion of the time series to calculate the average forecast error for each model.

Results

The model with the most accurate forecasts varied by site and by forecasting horizon. Treatment predictors were retained in the most accurate models across all clinical sites and forecasting horizons, with appropriate treatment and number of ACTs the most frequently retained treatment predictors. Among environmental predictors, 15 of the 18 most accurate models used total rainfall, while only 10 of the 18 models used vegetation. Generally, the short-term models included more predictors than the intermediate- and long-term models. The weekly error for the short-term models were highest when the observed counts were very low or zero which occurred most often

with one of the sites. The overall short-term error ranged from 20% to 96% over the forecasting period. The long-term models performed best for predicting the cumulative cases during the forecasting period (52 weeks) with error ranging from 2% to 22% across the six sites.

Conclusions

Incorporating clinical predictors such as antimalarial treatment, the proportion of individuals screened for malaria, and the number of individuals testing negative for malaria, improved the forecasting accuracy of several of the models. These results demonstrate the utility of using treatment predictors in conjunction with environmental covariates to predict malaria burden. Despite our gains, considerable variation in the temporal dynamics of malaria remains unexplained. This variation could be due to factors such as measurement error, reporting error, and interventions. Future work includes determining if the influence of environmental and clinical predictors on malaria burden differs by age groups.

Keywords

malaria; forecasting; ARIMA; environment; sentinel site surveillance

References

1. Christophers SR. Epidemic malaria of the Punjab: with a note of a method of predicting epidemic years. *Trans Committee Stud Malaria India* 1911;2:17–26.
2. Sachs J, Malaney P. The economic and social burden of malaria. *Nature*. Nature Publishing Group; 2002;415(6872):680–5.

*Kate Zinszer

E-mail: kate.zinszer@mail.mcgill.ca

