26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Predictive modelling of Ross River virus in NSW

Suhasini Sumithra [a, b], **Angus McLure** [b], **Aparna Lal** [b], **Kishen Lachireddy** [a]

[a] *Environmental Health Branch, NSW Health, St Leonards, New South Wales, Australia,* [b] *Australian National University, Canberra, Australian Capital Territory, Australia*
Email: Suhasini.Sumithra@health.nsw.gov.au

1. INTRODUCTION

Ross River virus (RRv) is Australia's most common, endemic mosquito-borne disease. RRv transmission is influenced by mosquito abundance, reservoir host populations, human behaviour and climatic, environmental, and socio-economic factors. The aim of this study was to develop models to predict RRv human notifications in NSW at a broader region and local government area (LGA) level using climate data.

2. METHODS

De-identified RRv notifications data from 1991 to 2021 were obtained from the NSW Notifiable Conditions Information Management System. Climate data for the same period were obtained from the Scientific Information for Land Owners database. LGAs were clustered into three broad regions using K-means clustering based on distance from the coast, annual rainfall, temperature, and humidity. For each cluster, negative binomial regression was used to regress weekly RRv human notifications for each LGA against time-lagged weather variables that are considered important variables for influencing mosquito and arbovirus activity. These included rainfall, maximum and minimum temperature, relative humidity, actual and potential evapotranspiration, mean sea level pressure and vapour pressure. Mean, minimum and maximum tidal height were also included for coastal LGAs. Dependence of cases on prior weather data (minimum lag: 4, 8, or 12 weeks; maximum lag: 52 weeks) was captured using polynomial distributed lags to manage high collinearity. Combinations of variables that achieved the lowest error on cross-validation were retained as final models.

3. RESULTS

Three broad regional clusters were identified: Coastal, South-Central and Western. At the cluster level, twelve-week forecasts were able to predict years with higher than average RRv notifications with good sensitivity and specificity (Coastal: 0.75/0.89, South-Central: 1.0/1.0, Western: 1.0/1.0). Twelve-week forecasts for predicting years with higher than average RRv notifications were less precise when analysis was done at the LGA level, then aggregated to the cluster level (sensitivity/specificity – Coastal: 0.67/0.89, South-Central: 0.50/0.67, Western: 0.67/1.0). Predictions were better for the Coastal and Western regions compared to the South-Central region. Specificity tended to be better than sensitivity for each region and across all time lags. At the LGA level, the best performing models were more variable and did not always align with the final model retained at the cluster level.

4. CONCLUSIONS

Our study found that years with higher than average RRv notifications in NSW can be predicted well with climate data alone and that cluster level predictions were more precise than at the LGA level. The best performing models at the cluster level included rainfall and at least one temperature variable which are two of the most influential climate variables for mosquito and arbovirus dynamics. Each regional model included a unique combination of predictors, aligning with findings from other studies that variables customised to local conditions can predict RRv disease more effectively than using the same set of variables across areas. Only one other study in NSW has attempted to predict RRv disease at the regional level using multiple climate variables. Our study has expanded on that study by using a larger dataset, climate variables that had not previously been explored in NSW studies (e.g. evapotranspiration) and weekly versus monthly unit of analysis. Our findings may be incorporated into NSW Health's existing arbovirus surveillance and early warning system to support and enhance prediction of RRv disease at a local level in NSW.

Keywords: *Vector-borne disease, predictive modelling, environmental data analysis*

