

LONDON  
SCHOOL of  
HYGIENE  
& TROPICAL  
MEDICINE



LSHTM Research Online

Cutcher, Z; Williamson, E; Lynch, SE; Rowe, S; Clothier, HJ; Firestone, SM; (2016) Predictive modelling of Ross River virus notifications in southeastern Australia. *Epidemiology and infection*, 145 (3). pp. 440-450. ISSN 0950-2688 DOI: <https://doi.org/10.1017/S0950268816002594>

Downloaded from: <http://researchonline.lshtm.ac.uk/3983501/>

DOI: <https://doi.org/10.1017/S0950268816002594>

**Usage Guidelines:**

Please refer to usage guidelines at <http://researchonline.lshtm.ac.uk/policies.html> or alternatively contact [researchonline@lshtm.ac.uk](mailto:researchonline@lshtm.ac.uk).

Available under license: <http://creativecommons.org/licenses/by-nc-nd/2.5/>

<https://researchonline.lshtm.ac.uk>

1 **Predictive modelling of Ross River virus notifications in south-eastern**  
2 **Australia.**

3  
4 Z. Cutcher <sup>a,b</sup>, E. Williamson <sup>a,c</sup>, S. E. Lynch <sup>d</sup>, S. Rowe <sup>b</sup>,  
5 H. J. Clothier <sup>a</sup>, S. M. Firestone <sup>e\*</sup>  
6

7 <sup>a</sup> Melbourne School of Population and Global Health, The University of  
8 Melbourne, Parkville, Victoria 3010, Australia.

9  
10 <sup>b</sup> Victorian Department of Health and Human Services, Communicable Disease  
11 Epidemiology and Surveillance, Health Protection Branch, Melbourne, Victoria  
12 3000, Australia.

13  
14 <sup>c</sup> London School of Hygiene and Tropical Medicine, London, United Kingdom.

15  
16 <sup>d</sup> Victorian Department of Economic Development, Jobs, Transport and  
17 Resources, Biosciences Research Division, AgriBio Centre, Bundoora, Victoria  
18 3083, Australia.

19  
20 <sup>e</sup> Asia-Pacific Centre for Animal Health, Faculty of Veterinary and Agricultural  
21 Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia.

22  
23  
24  
25 \*Corresponding author: Tel: +61 3 9035 7891; Fax: +61 3 8344 7374;  
26 E-mail: [Simon.Firestone@unimelb.edu.au](mailto:Simon.Firestone@unimelb.edu.au)

27  
28  
29  
30 Running head: Ross River virus modelling in south-east Australia

31 **Summary**

32 Ross River virus (RRV) is a mosquito-borne virus endemic to Australia. The  
33 disease, marked by arthritis, myalgia and rash, has a complex epidemiology  
34 involving several mosquito species and wildlife reservoirs. Outbreak years  
35 coincide with climatic conditions conducive to mosquito population growth.

36

37 We developed regression models for human RRV notifications in the Mildura  
38 Local Government Area, Victoria, Australia with the objective of increasing  
39 understanding of the relationships in this complex system, providing trigger points  
40 for intervention and developing a forecast model. Surveillance, climatic,  
41 environmental and entomological data for the period July 2000–June 2011 were  
42 used for model training then forecasts were validated for July 2011–June 2015.

43

44 Rainfall and vapour pressure were the key factors for forecasting RRV  
45 notifications. Validation of models showed they predicted RRV counts with an  
46 accuracy of 81%. Two major RRV mosquito vectors (*Culex annulirostris* and  
47 *Aedes camptorhynchus*) were important in the final estimation model at proximal  
48 lags.

49

50 The findings of this analysis advance understanding of the drivers of RRV in  
51 temperate climatic zones and the models will inform public health agencies of  
52 periods of increased risk.

53 **1. Introduction**

54 Ross River virus (RRV), Family *Togaviridae* Genus *Alphavirus*, is the most  
55 common mosquito-borne virus in Australia, with the largest burden occurring in  
56 the tropical north [1]. Symptoms in humans include debilitating fatigue, muscle  
57 and joint pain that persist between 3–6 months, and up to a year in some cases [2],  
58 leading to significant morbidity and economic loss [3]. However, 55–75% of  
59 cases are asymptomatic [4].

60

61 In the southeast State of Victoria, RRV is endemic with seasonal incidence. Most  
62 cases occur during the Southern hemisphere summer and early autumn, so  
63 reporting of arbovirus notifiable disease surveillance data typically refers to  
64 Australian financial years (1 July to 30 June the following calendar year) [1]. In  
65 the period July 2005–June 2010, a mean of 214 human cases were notified per  
66 year in Victoria (3.8 per 100,000 people per year), with the majority acquiring  
67 infection in either northern regions of the State (the Murray Valley) or southeast  
68 coastal regions [1]. Outbreaks have occurred in 1992/93, 1996/97, and more  
69 recently in 2010/11 when 1312 cases were notified across the State (23.3 per  
70 100,000 people) [5].

71

72 The epidemiology of RRV is complex with the disease maintained in wildlife  
73 reservoirs and transmitted to humans by mosquitoes, with human-mosquito-  
74 human transmission potentially occurring during epidemics [4]. The virus has  
75 been isolated from over 40 different mosquito species however only a small

76 number are thought to be competent vectors [6]. The predominant mosquito  
77 vector species vary by location and season. Macropods are thought to be the major  
78 wildlife reservoir, which also vary by ecological niche. Other marsupials,  
79 rodents and flying foxes may also be involved [6], particularly in urban areas [4].  
80 Horses can also be clinically infected [7], however their role in amplifying the  
81 virus is unclear.

82

### 83 **1.1. Arboviral surveillance and intervention in Victoria**

84 Ross River virus is a notifiable human disease under the Public Health and  
85 Wellbeing Regulations (2009). In Victoria, doctors and/or pathology laboratories  
86 must notify all laboratory confirmed cases to the Department of Health and  
87 Human Services (DHHS) within five days of diagnosis. According to the  
88 nationally agreed case definition [1] laboratory definitive evidence confirming a  
89 case requires either:

- 90 • isolation of RRV, or
- 91 • detection of RRV nucleic acid, or
- 92 • immunoglobulin G (IgG) seroconversion or a significant increase in  
93 antibody level or a  $\geq$ fourfold rise in titre to RRV, or
- 94 • detection of RRV-specific IgM, in the absence of Barmah Forest virus  
95 IgM, unless Ross River virus IgG is also detected, or
- 96 • detection of RRV-specific IgM in the presence of Ross River virus IgG.

97

98 Control of arboviruses relies on early detection of increased levels of mosquitoes  
99 and/or virus activity, prompting public health interventions including vector  
100 control and public education for bite prevention [8]. Under the Victorian  
101 Arbovirus Disease Control Program (VADCP) local governments across Victoria  
102 implement surveillance and control strategies on vector mosquito populations  
103 during the peak season between November and April each year when most human  
104 arbovirus notifications are received [9]. This program has been providing  
105 standardized adult mosquito monitoring and sentinel chicken surveillance targeted  
106 at Murray Valley encephalitis (MVE) and other endemic arboviruses since 1991  
107 in a One Health model of collaboration. The Victorian Department of Economic  
108 Development, Jobs, Transport and Resources (DEDJTR) provides virological and  
109 entomological support to the VADCP, funded equally by the DHHS and the local  
110 governments involved, overseen by a multidisciplinary Task Force. Surveillance  
111 involves weekly mosquito trapping using carbon dioxide and light-baited traps in  
112 eight local government areas across Victoria. Mosquitoes are counted and  
113 identified by species and viral isolation is attempted in an effort to detect the  
114 presence of RRV.

115

116 Before and during each peak season for arboviral activity, the VADCP analyses  
117 three broad environmental indicators [9-11] of conditions suitable for increased  
118 MVE virus activity in southeast Australia. Meteorological data (rainfall in the  
119 catchment basins of the four main river systems in Eastern Australia and proxy  
120 measures for the Southern Oscillation Index (SOI) and La Niña events) are

121 considered by DHHS and councils to inform of likely disease occurrence and  
122 when to instigate interventions. No models are currently available to combine  
123 these data for RRV prediction, with public health interventions being informed by  
124 routine notifiable disease surveillance and mosquito monitoring through the  
125 VADCP.

126

## 127 **1.2. Modelling and prediction**

128 Due to the climatic dependence of wildlife and mosquito populations, models  
129 using climate and/or entomological variables to predict RRV incidence may be  
130 helpful for informing disease control activities and forecasting the impact of  
131 climate change. A detailed review [3] describes previous models for RRV. Most  
132 predictive models for RRV have used logistic regression to estimate the odds or  
133 probability of an outbreak within a season, using seasonal variables at fixed points  
134 in time [12-16]. Others have explored prediction of disease using time-series  
135 analysis techniques [12], such as seasonal autoregressive integrated moving  
136 average and polynomial distributed lag (PDL) time-series models [17], and also  
137 negative binomial regression [18], to predict rates of disease, rather than simply  
138 whether or not an outbreak might occur in a season. Models tailored to conditions  
139 at the local level have tended to have better predictive capacity than broader  
140 geographic models [13]. All previous models based on RRV surveillance data for  
141 Southern Australia have estimated associations with annual case counts, with only  
142 two incorporating both entomological and climatic variables (for the southwest  
143 region of Western Australia [13] and southern South Australia [15]). None of the

144 models for RRV in southern Australia have attempted to model monthly counts  
145 and none have explicitly undertaken out-of-sample validation (forecasting),  
146 however their outputs have informed surveillance and control activities.

147

148 Models combining mosquito count and climate data have produced better results  
149 than models considering climatic variables alone [13, 17]. For example, Woodruff  
150 et al. (2006) developed early and late warning models for RRV outbreak years in  
151 14 statistical local areas of Western Australia and found climate data alone had  
152 64% sensitivity for an early warning model, and the addition of mosquito  
153 surveillance data increased the sensitivity to 85%. Previous models for predicting  
154 RRV in Victoria [16] have used only climatic data at one time point per season  
155 (total rainfall in July, maximum temperature in November) to estimate the  
156 probability of an outbreak during peak transmission season for two adjacent areas  
157 in the Murray Valley, achieving in-sample sensitivity (internal ‘rotational’  
158 validation) of between 64–96% for predicting an outbreak season.

159

160 The aim of this analysis was to develop predictive models for monthly counts of  
161 human RRV notifications in a highly affected inland location. Specific objectives  
162 included estimating the association between notified case counts and explanatory  
163 climatic, environmental and entomological variables, evaluating the usefulness of  
164 mosquito count data for informing public health interventions by estimating  
165 trigger points for action and, lastly, developing a forecasting tool.

166



167 **2. Methods**

168 **2.1. Data**

169 Mildura Local Government Area (LGA), located inland in northwest Victoria  
170 (Figure 1) was selected for this analysis as it has the highest RRV disease burden  
171 in the State. RRV notifiable disease surveillance data for the period July 2000–  
172 June 2015 were provided by the DHHS including the following variables:  
173 estimated date of onset, 5 year age-group, sex and residential address (or exposure  
174 address where ascertained at interview by health officials). These data were  
175 geocoded utilising the Google Maps® application programming interface,  
176 aggregated by month of onset and divided by annual Australian Bureau of  
177 Statistics estimates of the resident LGA population.

178

179 Weekly mosquito trapping count data were provided by the Victorian Department  
180 of Economic Development, Jobs, Transport and Resources (DEDJTR) for the  
181 same time period, for four traps in the Mildura LGA. Six species of interest were  
182 investigated for predictive value, including two thought to play a major role in  
183 Victoria in RRV transmission [4] (*Aedes camptorhynchus* and *Culex*  
184 *annulirostris*), two mosquito species with possible roles in transmission (*Ae.*  
185 *notoscriptus*, *Coquillettidia linealis*) and two further species with unknown  
186 importance for RRV transmission (*Cx. australicus*, the principal vector for MVE,  
187 and *Cx. globicoxitus*). Mosquitoes are only counted for the months November to  
188 April of each year. The median, mean and maximum counts across the four traps  
189 located in the Mildura LGA were calculated each month and categorized as

190 follows for each species: “no mosquitoes trapped” (the reference category), “1–9  
191 mosquitoes”, “10–99 mosquitoes”, “100–999 mosquitoes”, “≥1000 mosquitoes”.

192

193 Climatic and environmental variables were selected following a review of  
194 previous models, and are summarized by source in Table 1. Weather station data  
195 were obtained from the Australian Bureau of Meteorology weather station with  
196 the most complete data in Mildura LGA (Mildura airport; Bureau of Meteorology  
197 Station Number: 076031; geo-coordinates 142.0867°E, -34.2358°S, see Figure 1).

198

## 199 **2.2. Descriptive and univariable statistical analyses**

200 The distribution of each variable was examined and described, using contingency  
201 tables for categorical variables, collapsing categories where appropriate. Summary  
202 statistics and histograms were inspected for continuous variables and these  
203 transformed as required.

204

205 Data for the period July 2000–June 2011 were used to train the model. Owing to  
206 over-dispersion, negative binomial regression models were constructed to predict  
207 the monthly count of notified RRV cases each month for Mildura LGA ( $y$ ), of the  
208 form:

$$Y \sim \text{Poisson}(\mu^*)$$

$$\ln(\mu^*) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + v$$

$$\exp(v) \sim \text{Gamma}\left(\frac{1}{\alpha}, \alpha\right)$$

209

210 where the  $p$  predictor variables  $x_1, x_2, \dots, x_p$  are given, and the population  
211 regression coefficients  $\beta_0, \beta_1, \dots, \beta_p$  are estimated, applying a dispersion  
212 parameter ( $\alpha$ ) to represent the ratio of the variance of the expected counts to their  
213 mean. The dispersion parameter affects the variance of the expected counts, not  
214 the expected counts themselves. Exponentiation allows expression of the  
215 coefficients as incidence rate ratios (IRR).

216

217 Climatic and entomologic variables were lagged by 1–12 months and screened for  
218 entry into multivariable modelling. For each putative predictor variable, the lag  
219 with the strongest statistical association was selected using Akaike's Information  
220 Criterion (AIC) [19] – as this criterion may be applied to non-nested models – and  
221 entered into multivariable models if they were crudely statistically associated with  
222 RRV case count based on a liberal  $P$ -value threshold ( $P < 0.25$ ). The linearity of  
223 the univariable relationship with the outcome variable was assessed graphically  
224 for each continuous variable and by comparing the AIC of univariable models  
225 including a linear term versus those with the variable categorized into quintiles.  
226 Where appropriate categorized variables were retained for further analyses and  
227 category levels collapsed.

228

229 All continuous covariates were tested for collinearity in pairs by calculating  
230 Spearman's correlation coefficient ( $\rho_s$ ). Among pairs of highly correlated  
231 predictors ( $\rho_s \geq |0.70|$ ), only the variable with the strongest statistical association  
232 with the outcome was retained for further analysis [20].

233

### 234 **2.3. Multivariable analyses**

235 Multivariable models were constructed including all retained variables and  
236 trimmed for parsimony using manual backwards-stepwise regression to  $P < 0.20$ .  
237 Each removed variable was re-entered individually into the preliminary main  
238 effects model and retained if  $P < 0.15$ . At this point, pairwise interactions were  
239 tested among all retained terms, categorising continuous variables as required, and  
240 the model was reconstructed as a generalized linear model to implement  
241 regression diagnostics (deviance-based goodness-of-fit to the training data,  
242 assessment of residuals, influence and leverage). Maximum likelihood  $R^2$  was  
243 used as a robust measure of fit (no universally accepted adjusted- $R^2$  measure is  
244 available for negative binomial models [21]). The final ‘estimation’ model was  
245 checked for serial auto-correlation (AC) by including case counts in immediately  
246 preceding months [22] after testing for non-stationarity and trend in the time  
247 series following the Dickey-Fuller (DF) approach [23].

248

### 249 **2.4. Prediction, validation and adjustment for over-fitting**

250 The final estimation model was used to predict monthly notified human RRV case  
251 counts notified in each month in the 4 year validation dataset (July 2011–June  
252 2015) for Mildura LGA, and 95% prediction intervals (PI) were estimated  
253 adapting the method of Farrington et al [24] to the negative binomial distribution.  
254 External (‘out-of-sample’) forecasts and their 95% PIs were then compared to  
255 observed data (not used in model development) using Pearson’s correlation

256 coefficient ( $\rho_p$ ) [25], and models were tested for their proportional agreement with  
257 subjectively defined ‘outbreak alerts’ ( months with >2 notified cases and where  
258 the count of cases exceeded the 5-year mean plus 1 SD for that month estimated  
259 excluding known outbreak years, i.e. 2010/11, assuming a negative binomial  
260 distribution) [26]. The final estimation model was pruned to account for over-  
261 fitting by removing variables sequentially, and the comparisons repeated, to arrive  
262 at the final ‘prediction’ model, selected based on its forecasting ability.

263

264 Analyses were undertaken using Stata (StataCorp Texas, version 14.0) and the R  
265 statistical package version 3.1.1 [27] using the libraries ‘MASS’ [28] and ‘epiR’  
266 [29].

267

### 268 **3. Results**

269 There were 479 notified cases of RRV in Mildura LGA during the study period.  
270 The outbreak during the 2010/11 financial year accounted for 251 notifications  
271 (52.4%) (Figure 2). The mean notification rate (excluding 2010/11) was 63.9 per  
272 100,000 person years (32.6 per 100,000). Cases were notified year-round however  
273 87% had estimated dates of onset between November and April. There were 31  
274 outbreak alerts in the study period, six of these in 2010/11 and sixteen in the  
275 model validation period.

276

277 Amongst those species investigated, the predominant mosquito species trapped in  
278 Mildura LGA during the study period were *Culex annulirostris* (n=142,638),

279 *Aedes camptorhynchus* (n=24,349), *Cx. australicus* (n=6,768) and *Coquillettidia*  
280 *linealis* (n=5,249). Univariable associations between RRV incidence in Mildura  
281 LGA and lagged counts of the mosquito species and climatic and environmental  
282 variables are provided in supplementary material (Tables S1-2).

283

284 The final estimation model for Mildura LGA is presented in Table 2. A doubling  
285 of maximum vapour pressure was associated with a 3.5-fold rise in the rate of  
286 notifications in the following month (IRR=3.47; 95% CI: 1.57, 7.66). Mean trap  
287 counts of *Cx. annulirostris*  $\geq 1000$  were associated with a seven-fold increase in  
288 the rate of RRV notifications in the following month. When the mean *Ae.*  
289 *camptorhynchus* was  $\geq 10$ , RRV notifications 2 months later were increased 55%.

290 A doubling of precipitation and more rain days, were associated with 25% and 8%  
291 rises in RRV notifications, 4 and 6 months later, respectively. Two interaction  
292 terms were retained in the final model. The main effect of Murray River flows in  
293 the highest quintile (maximum daily flow in a month  $\geq 16,268$  ML) was an 85%  
294 reduction in RRV notifications 3 months later (IRR=0.15; 95% CI: 0.03, 0.81),  
295 whereas when the Southern Oscillation Index (measured 6 months prior) was  
296 greater than its median across the study period ( $\geq 1.7$  units) Murray River flows in  
297 the highest quintile were associated with a 5.7-fold increase in the rate of RRV  
298 notifications 3 months later. The main effect of Pacific Ocean sea surface  
299 temperatures  $\geq 26.8$  °C was a 68% reduction in notifications 2 months later,  
300 whereas when minimum monthly sea levels (measured 7 months prior) were

301  $\geq 13.2$  cm and sea surface temperatures  $\geq 26.8$  °C were associated with a 4-fold rise  
302 in RRV notifications 2 months later.

303

304 There was no long term trend in the time-series ( $P=0.14$ ) and the null hypothesis  
305 of non-stationary was rejected (DF test statistic=-5.856, degrees of freedom=132,  
306  $P<0.001$ ). Moderate serial auto-correlation was detected (Lag 1, AC=0.61) with  
307 each case one month prior being associated with a 12% increase in RRV  
308 incidence the following month (IRR=1.12, 95% CI: 1.05, 1.19). An  
309 autocorrelation term was included then eliminated (owing to  $P>0.20$ ) from the  
310 final estimation model.

311

312 Forecast ability of the model was improved by pruning to the final forecasting  
313 model (presented in Table 3 with a comparison of observed data and forecasts).

314 Total observed annual counts were within forecast prediction intervals in all four  
315 validation years (Figure 2), and at a monthly resolution observed counts were  
316 within the forecast prediction intervals in 39 of 48 months in the validation period  
317 (81%), in comparison to 129 of 132 months in the model training period (98%). In  
318 two of the validation years (2011/2012 and 2013/2014) there was excellent  
319 agreement between forecast and observed case counts and outbreak alerts,  
320 proportional agreement of 0.92 and 0.83, respectively. The model under-predicted  
321 case counts in 2012/2013 and 2014/2015, all 9 months with observed counts  
322 above the forecast prediction interval occurred in these two years, resulting in

323 poorer proportional agreement (0.50 in both cases) with observed outbreak alerts  
324 in these two years.

325

#### 326 **4. Discussion**

327 Climate, environmental and entomologic variables were used to develop  
328 prediction models for monthly RRV incidence rates for the Victorian inland Local  
329 Government Area with the highest notification rates. To our knowledge, this study  
330 was the first to integrate mosquito count data into Victorian RRV predictive  
331 modelling and the first to attempt out-of-sample forecasting of monthly counts of  
332 RRV for a location in Southern Australia.

333

334 The most robust way to assess predictive model accuracy is to review a graphical  
335 representation of observed versus predicted events using external data [30], as  
336 adopted for assessing the current models. The final forecasting model performed  
337 extremely well at tracking the observed counts in the validation period, and  
338 clearly fit the data well (differentiating between the outbreak year 2010/11 and  
339 other years with relatively low counts). Forecast prediction intervals encompassed  
340 the observed monthly counts in 39 of 48 months in the validation period. Of the  
341 nine months with observed counts falling above the predicted interval, five in  
342 2012/13 and two in 2014/15 had very low notified case counts ( $\leq 4$ ) and raised  
343 outbreak alerts merely on the basis that these low counts were well outside the  
344 typical RRV activity season (when typically  $\leq 1$  case was observed in most other  
345 years). The subjectively defined outbreak alert threshold is likely to be



346 oversensitive, so direct comparisons can only be interpreted cautiously. Raising  
347 the alert threshold to 2 SD greater than the long-term mean did not resolve the  
348 issue, as such a threshold was largely insensitive at detecting months that  
349 appeared to be clearly in excess of normal.

350

351 Statistical epidemiological modelling is often applied to address questions of  
352 causality (estimation and hypothesis testing) with fewer examples where the  
353 explicitly-stated aim is modelling for prediction of future observations [22]. When  
354 forecasting (predicting into the ‘out-of-sample’ future), a modified approach may  
355 be required, as was the case in this study, reducing the focus on the relationships  
356 between individual variables. While model fit remains important there is a trade-  
357 off, external validity is paramount (models constructed based on historical data  
358 must hold into the near future) and over-fitting to training data may well come at  
359 the expense of robust future prediction [22]. For this reason the final ‘estimating’  
360 model, used for assessing the relationships between variables, was pruned to  
361 produce a more parsimonious ‘forecasting’ model.

362

363 Other models of RRV in Southern Australia have been restricted to providing  
364 early warning of outbreak years, rather than attempting to forecast monthly  
365 counts. As presented, the forecasting model will be utilized each year to provide  
366 forecasts to the DHHS. Further modelling will be required to refine the variable  
367 selection and improve the robustness of forecasts. Other more complex

368 approaches may be required [25], perhaps following the PDL modelling approach  
369 that Hu et al. (2006) implemented for Brisbane, Queensland.

370

371 Rainfall and vapour pressure were key factors for forecasting RRV notifications  
372 in Mildura LGA. Rainfall has been included as an important predictor in all  
373 previous Ross River virus models for Southern Australia [12, 13, 15, 16], and  
374 underlies one of the broad early warning indicators [10] considered by DHHS for  
375 years of increased MVE activity. Vapour pressure is a measure of air humidity  
376 that depends on temperature and air pressure, similar variables have been included  
377 in all previous prediction models [12, 15, 16] developed for regions along the  
378 Murray River (that forms a natural border between the States of Victoria and New  
379 South Wales). It is biologically plausible that these variables are related to  
380 arbovirus transmission, as mosquitoes require a minimum temperature and  
381 moisture for breeding. The lags of these variables likely reflect effects of water,  
382 temperature and climatic conditions on local ecology, for example through their  
383 effects on vegetation and wildlife reservoir host populations along with their  
384 direct effect on mosquito populations. Whilst it is difficult to identify causal links  
385 between distally-lagged precipitation variables and the timescales of vector  
386 development and transmission of RRV, the main purpose of the models developed  
387 here was as predictive tools rather than to draw explicit conclusions regarding  
388 causation. Including rainfall parameters with lags between 4 and 6 months  
389 provided the model with the best predictive performance at a monthly resolution.  
390 When we evaluated rainfall variables over lags of 1 to 3 months (in univariable

391 analysis), very similar estimates were obtained as those included in the final  
392 model (for total monthly precipitation lagged 4 months, and number of days with  
393 greater than 1 mm rainfall lagged 6 months). There were only low levels of  
394 temporal auto-correlation observed between these variables, so these were  
395 included in multivariable estimation and prediction models at shorter lags (as  
396 secondary effects of rainfall over different time-scales). However, these variables  
397 representing shorter lags of rainfall were subsequently eliminated. Owing to weak  
398 correlations between climatic variables (rainfall, vapour pressure, humidity and  
399 temperature) in our data, it is also likely that some of the proximal effect of  
400 rainfall is represented by other variables in the final models.

401

402 *Culex annulirostris* and *Ae. camptorhynchus* are the two major mosquito vectors  
403 for Ross River virus in Victoria [4]. Their inclusion in the final estimation model  
404 at proximal lags is consistent with their role in transmitting virus to humans from  
405 wildlife reservoirs and the time taken for mosquitoes to develop, the ~2 week  
406 extrinsic and 1-2 week intrinsic incubation periods of RRV [17]. The univariable  
407 associations presented in supplementary Table S1 represent useful trigger points  
408 for action by the local council (such as mosquito larvicidal treatments and public  
409 announcements about the risk and appropriate preventative actions). Risk of RRV  
410 is likely to be greatly increased in months subsequent to those when mean weekly  
411 trap counts of *Cx. annulirostris* and *Ae. camptorhynchus* exceed 100 and 10  
412 mosquitoes, respectively. Contrary to the findings of previous modelling studies  
413 of RRV notifications in other Australian States [13, 17], we found that inclusion

414 of variables representing mosquito numbers provided no improvement in model  
415 forecasting ability (although strongly statistically significant associations were  
416 observed between lagged mosquito count variables and RRV notifications in the  
417 final estimation model). Hu et al. (2006) noted the limitations of including  
418 mosquito count data in early warning forecasting models (cost of collection and  
419 proximal lags limiting the extent of early warning).

420

421 Two interesting interactions were present in the final estimation model, both of  
422 which appear indicative of periods of extreme climatic conditions. Elevated SOI  
423 (i.e. a La Niña event) 6 months earlier and maximum Murray River flow 3 months  
424 prior were associated with increased rates of notification for RRV. A severe  
425 flooding event affecting the Murray River valley occurred in the 2010/11 outbreak  
426 year. Interestingly, on its own, high maximum Murray River flows (indicative of  
427 low amounts of irrigation) were associated with substantially decreased rates of  
428 RRV notification.

429

430 Weather patterns in the study region are heavily influenced by the development  
431 and intensity of El Niño/La Niña events in the Pacific Ocean [31]. Across eastern  
432 Australia, El Niño events are often associated with drier than normal conditions  
433 while La Niña events are associated with wetter than normal conditions. Lower  
434 sea surface temperatures in the Niño 3.4 region (SST) are an indicator of La Niña  
435 events and in this analyses were associated with increased rates of RRV  
436 notification, which is biologically plausible as wetter conditions favour mosquito

437 larval development. Sea surface temperature was considered as a potential model  
438 covariate, even for this inland study area, as it was identified by Woodruff et al  
439 [16] as a predictor in their model of RRV for the Murray region in Victoria, and  
440 for its role in the El Niño Southern Oscillation phenomenon that influence  
441 weather patterns across Australia.

442

443 Of interest, another biologically plausible and statistically significant interaction  
444 was detected, between SST and sea levels (when both were increased, rates of  
445 notification of RRV cases were also likely to be increased). Sea level changes are  
446 driven by complex processes including thermal expansion of water, input of water  
447 into the ocean from glaciers and ice sheets, and changed water storage on land  
448 [32]. Variables representing sea level were considered for inclusion in these  
449 models because sea levels are correlated with SST and the SOI [33]. Again, this  
450 interaction term may indicate periods of extreme climatic conditions, with  
451 extremes in sea levels and sea surface temperature being a feature of cyclones (as  
452 experienced in the 2010/11 outbreak year when cyclones in Queensland caused  
453 major flooding in the Murray-Darling river basin immediately preceding  
454 extremely high arbovirus activity). The DHHS utilizes another sea surface  
455 temperature measure, the Indian Ocean Dipole (IOP), which is based on the  
456 difference between sea surface temperature in the Western and Eastern tropical  
457 Indian Ocean, as a predictor for MVEV activity in south-eastern Australia [9].  
458 Negative IOP events generally coincide with La Niña events.

459

460 The study was subject to a number of limitations: notification data may be  
461 undoubtedly understated and biased toward cases with typical clinical symptoms -  
462 those with less severe illness may not seek medical help or may be misdiagnosed.  
463 For this reason model outputs are interpreted as notification rates (rather than  
464 incidence rates). Residential location was accepted as a proxy for place of  
465 infection as this information was not available for a majority of cases.  
466 Misclassification of place of infection for some cases may have altered the  
467 measured associations between model covariates and disease, thus reducing  
468 predictive accuracy. The model did not account for mosquito control activities, as  
469 a reliable, consistent measure of these activities was unavailable. It is likely this  
470 omission has reduced the predictive accuracy of the models and ideally these  
471 should be accounted for in future research. Despite these limitations, the model  
472 presented appears a useful forecasting tool for RRV in region investigated with  
473 81% of observed monthly counts in the validation period falling within forecast  
474 prediction intervals.

475

476 Changing climatic conditions over the coming decades are likely to alter the  
477 current patterns of arboviral disease in Australia [3, 34], although the nature of  
478 this change is controversial [35]. The effect on arbovirus transmission is likely to  
479 vary regionally. For example, the impact will differ in arid compared to  
480 temperate, and coastal versus inland regions, reflecting variation in the effect of  
481 climate change on local ecological conditions [34]. Advanced tools, such as the  
482 models presented here, will be required to monitoring the changing relationship

483 between notified cases and local conditions, and to provide early warning of  
484 periods of high arbovirus activity.

485

## 486 **5. Acknowledgements**

487 The authors gratefully acknowledge the following for preparing and providing  
488 data necessary for these analyses: the Victorian Government Department of  
489 Health and Human Services (particularly Nicola Stephens, Rebecca Feldman and  
490 Katherine Gibney), and Department of Economic Development, Jobs, Transport  
491 and Resources (particularly Karen Brown, Joe Azuolas, Kim Andrews and Elwyn  
492 Wishart), The Mildura Rural City Council (Dale Hutchinson and Stuart Maher),  
493 the Victorian Arbovirus Taskforce and the Australian Bureau of Meteorology's  
494 National Climate Data Service. Human notifiable surveillance data was provided  
495 by the Department of Health and Human Services in accordance with its privacy  
496 act after ethical approval for all procedures used in this study was obtained from  
497 the Human Research Ethics Committee of the University of Melbourne  
498 (1339794).

499

## 500 **6. Conflict of interest statement**

501 The authors have no competing interests to declare.

502

## 503 **7. Financial support**

504 This research received no specific grant from any funding agency, commercial or  
505 not-for-profit sectors. Dr Simon Firestone is supported by an Australian Research

506 Council Discovery Early Career Researcher Award (project number

507 DE160100477).

508



509 **8. References**

- 510 1. Knope, K., et al., *Arboviral diseases and malaria in Australia, 2010-11:*  
511 *annual report of the National Arbovirus and Malaria Advisory Committee.*  
512 Communicable diseases intelligence quarterly report, 2013. **37**(1): p. E1.
- 513 2. Westley-Wise, V.J., et al., *Ross River virus infection on the north coast of New*  
514 *South Wales.* Australian and New Zealand Journal of Public Health, 1996.  
515 **20**(1): p. 87-92.
- 516 3. Jacups, S.P., P.I. Whelan, and B.J. Currie, *Ross River virus and Barmah Forest*  
517 *virus infections: a review of history, ecology, and predictive models, with*  
518 *implications for tropical northern Australia.* Vector-Borne and Zoonotic  
519 Diseases, 2008. **8**(2): p. 283-298.
- 520 4. Harley, D., A. Sleight, and S. Ritchie, *Ross River virus transmission, infection,*  
521 *and disease: a cross-disciplinary review.* Clinical Microbiology Reviews,  
522 2001. **14**(4): p. 909-932.
- 523 5. Australian Government Department of Health, *National notifiable diseases*  
524 *surveillance system (NNDSS).* [http://www9.health.gov.au/cda/source/cda-](http://www9.health.gov.au/cda/source/cda-index.cfm)  
525 [index.cfm](http://www9.health.gov.au/cda/source/cda-index.cfm) (accessed 11/2/2015). 2015.
- 526 6. Russell, R.C., *Ross River virus: ecology and distribution.* Annual Review of  
527 Entomology, 2002. **47**(1): p. 1-31.
- 528 7. El-Hage, C.M., M.J. McCluskey, and J.K. Azuolas, *Disease suspected to be*  
529 *caused by Ross River virus infection of horses.* Australian Veterinary Journal,  
530 2008. **86**(9): p. 367-370.

- 531 8. van den Hurk, A.F., et al., *Evolution of Mosquito-Based Arbovirus Surveillance*  
532 *Systems in Australia*. Journal of Biomedicine and Biotechnology, 2012. **2012**.
- 533 9. Lynch, S., et al., *Victorian Arbovirus Disease Control Program Annual Report*  
534 *2014-2015*. 2015, The State of Victoria Department of Economic  
535 Development, Jobs, Transport and Resources: Melbourne.
- 536 10. Forbes, J.A., *Murray Valley Encephalitis 1974, Also, The Epidemic*  
537 *Variance Since 1914 and Predisposing Rainfall Patterns*. 1978: Australasian  
538 Medical Publishing Company.
- 539 11. Nicholls, N., *A Method For Predicting Murray Valley Encephalitis In*  
540 *Southeast Australia Using The Southern Oscillation*. Australian Journal of  
541 Experimental Biology and Medical Science, 1986. **64**(6).
- 542 12. Bi, P., et al., *Defining the local ecology of Ross River virus infections as a*  
543 *basis for risk assessment and better prevention strategies in South Australia :*  
544 *final report prepared for the S.A. Department of Health*. 2006, The University  
545 of Adelaide: Adelaide.
- 546 13. Woodruff, R.E., et al., *Early warning of Ross River virus epidemics:*  
547 *combining surveillance data on climate and mosquitoes*. Epidemiology, 2006.  
548 **17**(5): p. 569-575.
- 549 14. Gattton, M.L., B.H. Kay, and P.A. Ryan, *Environmental predictors of Ross*  
550 *River virus disease outbreaks in Queensland, Australia*. The American journal  
551 of tropical medicine and hygiene, 2005. **72**(6): p. 792-799.
- 552 15. Williams, C.R., S.R. Fricker, and M.J. Kokkinn, *Environmental and*  
553 *entomological factors determining Ross River virus activity in the River*

- 554 *Murray Valley of South Australia*. Australian and New Zealand Journal of  
555 Public Health, 2009. **33**(3): p. 284-288.
- 556 16. Woodruff, R.E., et al., *Predicting Ross River virus epidemics from*  
557 *regional weather data*. Epidemiology, 2002. **13**(4): p. 384-393.
- 558 17. Hu, W., et al., *Rainfall, mosquito density and the transmission of Ross*  
559 *River virus: A time-series forecasting model*. Ecological Modelling, 2006.  
560 **196**(3): p. 505-514.
- 561 18. Werner, A., et al., *Environmental drivers of Ross River virus in*  
562 *southeastern Tasmania, Australia: towards strengthening public health*  
563 *interventions*. Epidemiology and Infection, 2012. **140**(02): p. 359-371.
- 564 19. Akaike, H., *Information theory and an extension of the maximum*  
565 *likelihood principle*, in *Selected Papers of Hirotugu Akaike*. 1998, Springer. p.  
566 199-213.
- 567 20. Armitage, P., G. Berry, and J.N.S. Matthews, *Statistical Methods in*  
568 *Medical Research, 4th Edition*. 2001: Wiley-Blackwell. pp 356-360.
- 569 21. Grimshaw, J.M., et al., *Applying psychological theories to evidence-based*  
570 *clinical practice: identifying factors predictive of lumbar spine x-ray for low*  
571 *back pain in UK primary care practice*. Implementation Science, 2011. **6**(1): p.  
572 55.
- 573 22. Dohoo, I., W. Martin, and H. Stryhn, *Veterinary Epidemiologic Research,*  
574 *2nd ed*. 2009: VER Inc., Charlottetown, Prince Edward Island, Canada.

- 575 23. Dickey, D.A. and W.A. Fuller, *Distribution of the estimators for*  
576 *autoregressive time series with a unit root*. Journal of the American statistical  
577 association, 1979. **74**(366a): p. 427-431.
- 578 24. Farrington, C., et al., *A statistical algorithm for the early detection of*  
579 *outbreaks of infectious disease*. Journal of the Royal Statistical Society. Series  
580 A (Statistics in Society), 1996: p. 547-563.
- 581 25. Axelsen, J.B., et al., *Multiannual forecasting of seasonal influenza*  
582 *dynamics reveals climatic and evolutionary drivers*. Proceedings of the  
583 National Academy of Sciences, 2014. **111**(26): p. 9538-9542.
- 584 26. Yohannes, K., et al., *Australia's notifiable diseases status, 2004, annual*  
585 *report of the National Notifiable Diseases Surveillance System*. Commun Dis  
586 Intell, 2006. **30**(1): p. 1-79.
- 587 27. R Core Team, *R: A language and environment for statistical computing*. R  
588 *Foundation for Statistical Computing, Vienna, Austria*. Available at:  
589 <http://www.R-project.org/> (accessed 15/9/2014). 2016.
- 590 28. Venables, W.N. and B.D. Ripley, *Modern Applied Statistics with S, fourth*  
591 *edition*. 2002, New York: Springer.
- 592 29. Stevenson, M., et al., *epiR: An R package for the analysis of*  
593 *epidemiological data*. R package version 0.9-69. 2015.
- 594 30. Moons, K.G., et al., *Risk prediction models: I. Development, internal*  
595 *validation, and assessing the incremental value of a new (bio) marker*. Heart,  
596 2012. **98**(9): p. 683-690.

- 597 31. Australian Government Bureau of Meteorology. *El Niño - Detailed*  
598 *Australian Analysis*. 2015; Available from:  
599 <http://www.bom.gov.au/climate/enso/enlist/> (accessed 20/11/15).
- 600 32. Rahmstorf, S., *A semi-empirical approach to projecting future sea-level*  
601 *rise*. *Science*, 2007. **315**(5810): p. 368-370.
- 602 33. Lombard, A., et al., *Contribution of thermal expansion to present-day sea-*  
603 *level change revisited*. *Global and Planetary Change*, 2005. **47**(1): p. 1-16.
- 604 34. Woodruff, R.E. and H. Bambrisk, *Climate change impacts on the burden*  
605 *of Ross River virus disease*, in *Garnaut Climate Change Review*, R. Garnaut,  
606 Editor. 2008, Cambridge University Press: Port Melbourne.
- 607 35. Russell, R.C., *Mosquito-borne disease and climate change in Australia:*  
608 *time for a reality check*. *Australian Journal of Entomology*, 2009. **48**(1): p. 1-7.  
609

610 **9. Figure captions**

611

612 **Figure 1:** Study extent of predictive modelling of Ross River virus cases in the  
613 Mildura Local Government Area (shaded grey), Victoria, Australia, for the period  
614 1 July 2000 to 30 Jun 2015. Black circle represents the location of the Mildura  
615 airport weather station. The Murray River forms the northern border of Mildura  
616 local government area.

617

618

619 **Figure 2:** Monthly time-series, predictions and forecasts of notified Ross River  
620 virus cases in the Mildura Local Government Area, Victoria, Australia, for the  
621 period 1 July 2000 to 30 Jun 2015. Data for the Australian financial year 2010/11  
622 have been rescaled by a factor of 3. Dotted lines represent upper 95% prediction  
623 intervals.