

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

*Exploring the role of temperature and other environmental factors in West Nile virus incidence and prediction in California counties from 2017-2022*

Noah Parker<sup>1\*</sup>

<sup>1</sup> Department of Epidemiology, Berkeley School of Public Health, University of California, Berkeley, Berkeley, California, United States of America

\* Corresponding author, email: [noahparker3@berkeley.edu](mailto:noahparker3@berkeley.edu)

21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57

## Abstract

West Nile virus (WNV) is the most common mosquito-borne disease in the United States, resulting in hundreds of reported cases yearly in California alone. The transmission cycle occurs mostly in birds and mosquitoes, making meteorological conditions, such as temperature, especially important to transmission characteristics. Given that future increases in temperature are all but inevitable due to worldwide climate change, determining associations between temperature and WNV incidence in humans, as well as making predictions on future cases, are incredibly important to public health agencies in California. Using surveillance data from the California Department of Public Health (CDPH), meteorological data from the National Oceanic and Atmospheric Administration (NOAA), and vector and host data from VectorSurv, we created GEE autoregressive and zero-inflated regression models to determine the role of temperature and other environmental factors in WNV incidence and predictions. An increase in temperature was found to be associated with an increase in incidence in 11 high-burden Californian counties between 2017-2022 (IRR = 1.06), holding location, time of year, and rainfall constant. A hypothetical increase of two degrees Fahrenheit — predicted for California by 2040 — would have resulted in upwards of 20 excess cases per year during our study period. Using 2017-2021 as a training set, meteorological and host/vector data were able to well-predict 2022 incidence, though the models did overestimate the peak number of cases. The zero-inflated model closely predicted the low number of cases in winter months but performed worse than the GEE model during high-transmission periods. These findings suggests that climate change will, and may be already, altering transmission dynamics and incidence of WNV in California, and provides tools to help predict incidence into the future.

## Author Summary

West Nile Virus is a disease that is spread by mosquitoes. Though it commonly infects birds, transmission to humans is possible and can lead to severe health effects. Temperature is known to affect the transmission cycle of West Nile virus, but it is unclear how global warming might change who, or how many people, may get infected with the virus. In this study, the researchers looked at how climate change may affect West Nile virus in California, and how health officials may better be able to predict future cases. The study found that there could be an increase in West Nile virus cases in humans due to increases in temperature in the next 20 years, but that we already have many tools and sources of data to predict cases. These findings reinforce the possible consequences of climate change on human health, and aid in the understanding in the complex relationship between climate and infectious diseases.

## 58 **B. Background**

59 West Nile virus (WNV) is a vector-borne flavivirus that is responsible for the most  
60 infections of any mosquito-transmitted disease in the United States [1]. West Nile virus was first  
61 detected in Southern California in 2003, before quickly spreading to all 58 Californian counties  
62 within a year [2]. California bears the largest burden of WNV infections, with 4,035 of the  
63 nation's 21,869 infections (18%) between 2009 and 2018 being reported in the state [3]. Only  
64 20% of infected people will show any symptoms, while less than 1% will show neurological  
65 symptoms, manifesting as West Nile neuroinvasive disease (WNND) [3]. Still, despite the low  
66 symptomatic and severity rate, California reported 326 fatalities due to WNND between 2003  
67 and 2018, as the fatality rate of WNND is about 10% [4].

68 Many environmental factors are known to impact WNV transmission, with most  
69 transmission occurring in non-human hosts. Humans become infected through the bite of a  
70 female mosquito from the *Culex* genus, though WNV cannot be spread between or from humans  
71 — they are a dead-end host [3]. The enzootic cycle mainly flows through mosquitoes and certain  
72 bird species, such as corvids, finches, and sparrows, though infections of horses and chickens are  
73 also common [3]. Given that WNV incidence in mosquitoes, bird, horses, and other species are  
74 an indicator of increased transmission in the enzootic cycle, high wildlife prevalence can serve as  
75 a good predictor of the transmission risk to humans [4]. Additionally, certain meteorological  
76 conditions, such as temperature and rainfall, play an important role in mosquito spawning and  
77 behavior and have been shown to be positively associated with human WNV infections [5]. In  
78 particular, temperature has been shown to affect mosquito life cycle traits and interactions with  
79 pathogens, in turn affecting WNV transmission [2,6,7]. Prior research has highlighted the

80 predictive power of these environmental factors on West Nile cases in the US and Canada, with  
81 the possibility that prediction models might be applied in other, higher-burdened areas [1,8].

82 While other studies have modeled the importance of temperature and certain other  
83 environmental factors in WNV transmission in California [9], as well as the ability of the state's  
84 warning system to determine outbreaks [4], there is sparse research that has attempted to  
85 quantify excess cases that may be associated with changing meteorological conditions. Few  
86 studies also expand further to create prediction models using currently collected data sources.  
87 This analysis is especially important given the projected rise in temperature in California due to  
88 worldwide climate change and can help state officials plan for possible increases in future WNV  
89 transmission. To our knowledge, no other papers have used a zero-inflated regression model to  
90 examine WNV in California. In the context of West Nile virus, a zero-inflated regression model  
91 may have advantages over traditional models given the clear seasonality of human infections  
92 and the near certainty of zero cases for much of the year. The effectiveness of such a model over  
93 a normal negative binomial regression model is explored in this paper.

94 In our paper, we used a zero-inflated regression model and a generalized estimating  
95 equation that incorporates autoregressive structure to measure the association between  
96 temperature and WNV incidence, determine the number of excess cases that would result from  
97 an increase in temperature, as well as compare the predictions each make for future WNV  
98 transmission in California. We hypothesize that temperature will be positively associated with  
99 WNV in California after adjusting for rainfall, location, and month of year; meanwhile,  
100 meteorological conditions and environmental factors, such as infected dead birds and mosquito  
101 pool positivity rate, will be able to predict future WNV cases.

102

103 **C. Methods**

104 *I. Data*

105 West Nile virus is a national and state notifiable condition [10]. As such, case data was  
106 collected by the California Department of Public Health as part of their surveillance program.  
107 While weekly case incidence is reported, monthly incidence was used in the study to remove  
108 inconsistent lags in reporting. Only counties with a population over 250,000 residents and a total  
109 of 40 or more cases over the six-year period (2017-2022) were considered high-burden counties  
110 and included in the analysis. This totaled 11 counties: Fresno, Kern, Los Angeles, Merced,  
111 Orange, Riverside, Sacramento, San Bernadino, San Joaquin, Stanislaus, and Tulare. A restriction  
112 was placed on cases due to most counties having a small or zero case count of WNV over the  
113 study period, while a minimum county size was enacted to protect cases' identities in small  
114 counties. We decided to only use the most recent six-year period of case data, as this reflected a  
115 tradeoff between having a large number of observations while maintaining a sufficiently  
116 contained time range where transmission and meteorological conditions expressed similar  
117 characteristics. Average daily temperature (aggregated across the month in Fahrenheit) and  
118 rainfall (summed for the month in inches) were collected from the National Oceanic and  
119 Atmospheric Administration. Mosquito pool positivity rates (percentage of mosquito pools  
120 within a county that tested positive for WNV) and bird infection counts (count of dead birds that  
121 tested positive for WNV) were obtained through California's VectorSurv program under request  
122 number 000067.

123

124 II. *Zero-Inflated and GEE Models for Measure of Association M*

125 **The Models and Variables**

126 Negative binomial regression was conducted in both models instead of Poisson because  
127 the case data had an overdispersed distribution. Due to the time it takes for a person to become  
128 sick enough to seek treatment, as well as the reporting delays of cases to the state health  
129 department, we assumed that meteorological conditions and reported case counts are not  
130 directly related in time. Given that substantial lags between case onset and case reporting — an  
131 average of 5 weeks — have been found in other WNV outbreaks, temperature and rainfall were  
132 both lagged one month, so that the variables *prior month temperature* and *prior month rainfall*  
133 would better reflect the current month's cases [11]. As the reporting of vector data are also  
134 subject to delays in reporting and biological processes, it was assumed that the mosquito pool  
135 positivity rate and dead bird count follow similar delays as cases and therefore were not lagged  
136 [11]. This selective lagging strategy has been used in other work that models WNV [1]. Rainfall,  
137 month of year, and county were adjusted for as confounders in the relationship between  
138 temperature and WNV cases. Controlling for these variables also helps control for the  
139 seasonality and location dependence of West Nile infections. The percentage of positive  
140 mosquito pools and positive dead birds were determined to be on the causal pathway and were  
141 not analyzed in the models testing for a measure of association, the incidence rate ratio.

142

143 **Zero-inflated**

144 We used two different methods to estimate an association between temperature and  
145 West Nile virus incidence. First, we utilized a zero-inflated negative binomial regression model to

146 model the count of cases in the 11 high-burden counties. This method was used to account for  
147 the large number of months with zero cases in the data, as there is little WNV transmission  
148 outside of June through November. 525 of the monthly case counts were zero, representing  
149 66.3% of the observations.

150 We represent the number of cases as  $y_{it}$ , where  $i$  represents the county ( $i = 1 \dots 11$ ) and  $t$   
151 represents the month of observation over the six-year period ( $t = 1 \dots 72$ ). We assume that  
152 structural zeroes appear in the data at time  $t$  and county  $i$  at a probability of  $\pi$ . Therefore, there  
153 are two processes in which data can be generated for each observation: the first, which  
154 generates a case count of 0 with probability of  $\pi$ ; the second, which generates a case count  
155 corresponding to a negative binomial model, represented by  $g(y_{it}|X_{it})$  with a probability of  
156  $1 - \pi$ .

$$157 \quad y_{it} \sim \left\{ \begin{array}{ll} 0 & \text{with probability } \pi_i \\ g(y_{it}|X_{it}) & \text{with probability } 1 - \pi_i \end{array} \right\}$$

158  
159 Therefore, we assume the probability that the expected number of cases,  $Y_{it}$ , is equal to  
160  $y_{it}$  is given by the zero-inflated negative binomial model below.  $X_{it}$  represents the vector of  
161 covariates in the model, while  $Month_{it}$  expresses the calendar month at county  $i$  and time  $t$ ,  
162 which is believed to well predict whether there will be cases in a given observation or not.

$$163 \quad \Pr(Y_{it} = y_{it} | X_{it}, Month_{it}) = \left\{ \begin{array}{ll} \pi(Month_{it}) + \{1 - \pi(Month_{it})\}g(0 | X_{it}) & \text{if } y_{it} = 0 \\ \{1 - \pi(Month_{it})\}g(y_{it} | X_{it}) & \text{if } y_{it} > 0 \end{array} \right\}$$

165

166 The function representing the negative binomial model,  $g(y_{it} | X_{it})$ , is defined below,  
167 where the vector of covariates in the model,  $X_{it}$ , includes the prior month temperature, prior  
168 month rainfall, the county, and the month. Prior month temperature is the exposure of interest  
169 on the outcome, while prior month rainfall, county, and month are all confounders of this  
170 relationship. While county and month are shown here as being expressed by one *beta* term,  
171 they are analyzed as factor variables in the model, meaning that a dummy variable was created  
172 for each category level (10 for county and 11 for month), and a *beta* term was calculated for  
173 each.

$$g(y_{it} | X_{it}) = E(y_{it} | X_{it}) = (\exp^{X_{it}\beta})$$
$$g(y_{it} | X_{it}) = E(y_{it} | X_{it}) = \exp [\beta_0 + \beta_1(\text{PriorMonthTemp}) + \beta_2(\text{PriorMonthRainfall}) + \beta_3(\text{County}) + \beta_4(\text{Month})]$$

176

### 177 **Negative Binomial**

178 Second, we used a negative binomial model with autocorrelation features. We utilized an  
179 “ar1” auto-correlation structure, meaning cases in one month were strongly correlated with  
180 cases in the month before and the month after, but not other months. This structure is thought  
181 to accurately describe infection dynamics in our system, as the cases in county  $i$  at time  $t$  are  
182 dictated by how many mosquito infections, and subsequent human infections, were present at  
183 time  $t - 1$ , as well as influences how many cases there will be at time  $t + 1$  given the same  
184 dynamics. The model follows the same form as the negative binomial part of the zero-inflated  
185 model, but has an added autoregressive term,  $\sigma$ , that follows a distribution of “ar1”, as shown in  
186 the model below.



187

188

$$\log[E(Y_{it} | X_{it})] = X_{it}\beta$$

189

$$\log[E(Y_{it} | X_{it})] = \beta_0 + \beta_1(\text{PriorMonthTemp}) + \beta_2(\text{PriorMonthRainfall}) + \beta_3(\text{County}) + \beta_4(\text{Month}) + \sigma$$

190

$$\sigma \sim AR(1)$$

191

192

### Standardization

193

Model standardization was conducted to retrieve a risk difference (in the number of

194

excess cases) between a hypothetical dataset where temperature was increased by two degrees

195

(with all other variables kept constant) and the observed dataset. We created a population

196

where *prior month temperature* was increased by two degrees Fahrenheit across all

197

observations, later using it to fit a zero-inflated model and find the total number of cases

198

expected in the six-year study. The risk difference is the total number of cases expected in the

199

hypothetical population minus the number of cases expected given the observed data. While this

200

risk difference cannot be interpreted as an association between temperature and West Nile

201

incidence — due to it being a hypothetical population — it can provide insight into possible

202

repercussions of a changing climate in the future. Bootstrapped sampling, using the bias-

203

corrected and accelerated technique (BCA), was subsequently applied to measure the

204

uncertainty in our risk difference calculations.

205

206

### III. Prediction Models

207

Both the zero-inflated model and autoregressive negative binomial model were used as

208

prediction models for WNV cases in 2022. Both models were trained on the data from 2017-

209 2021 and then tested on the data from 2022. The models followed the same assumptions and  
210 equations as the models specified above, respectively, but were made to include different sets of  
211 the covariates in the data. Since they are prediction models, all variables, including percentage of  
212 positive mosquito pools and positive dead birds— which were left out of the measure of  
213 association models due to existing on the causal pathway between temperature and WNV  
214 incidence — were included in the analysis. The four different prediction sets tested were: “All,”  
215 comprising all the covariables listed; “Host/Vector,” comprising dead infected birds and  
216 mosquito positivity rate; “Just Temperature,” comprising just prior month temperature; and  
217 “Weather,” comprising prior month temperature and prior month rainfall. We opted to use  
218 mean squared error (MSE) to compare the models over information criteria tests, such as the  
219 AIC or BIC, as we didn’t have concrete likelihood estimates from the GEE model [12]. Though  
220 mean squared error doesn’t account for and penalize the number of predictor variables in each  
221 model, the lack of a large difference in variables used (4 vs 2 vs 2 vs 1 in “All,” “Host/Vector,”  
222 “Weather,” and “Just Temperature,” respectively) and a clear difference in each model’s mean  
223 squared error highlighted a clear best fit for the data.

224

#### 225 *D. Ethical Dimensions*

226         Given the ecological design of the study and the passively collected surveillance case  
227 data, no data was actively collected on human subjects. No identifiers were included in the data,  
228 and only aggregated data from high-burden, highly populated counties were used in the analysis  
229 to protect individual’s identities. Therefore, there was no risk to any individuals who may have  
230 been included in the data. While this analysis was conducted well after all of the infections and

231 any findings will not be of benefit to the individual, the results may still benefit these community  
232 in the future.

233

## 234 ***E. Results***

### 235 *Descriptive Analysis*

236 Fig 1 plots the mean temperature and West Nile virus cases across all 11 high-burden  
237 counties included in the study. This figure shows the clear seasonality of WNV activity, with the  
238 peak of each year occurring in the summer, when the temperature is highest. This also highlights  
239 the almost complete lack of cases in the colder winter months. There is high variability in the  
240 number of cases in these counties in a single year: there were 515 cases reported in 2017, while  
241 just 77 in 2021.

242

243

244

245

246

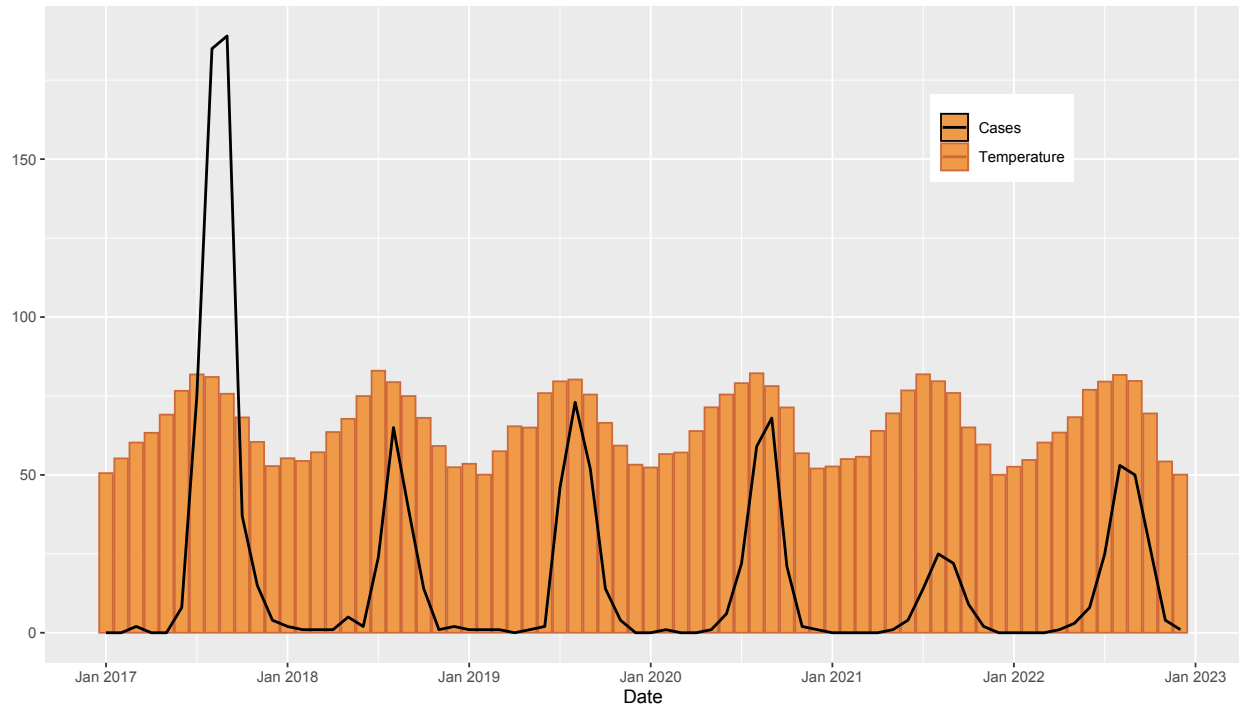
247

248

249

250 ***Fig 1***

Average Monthly Temperature and Total Monthly West Nile Virus Cases in 11 High-Burden California Counties



251

252

These trends are further explored in Table 1, which shows the average cases,

253

temperature, rainfall, dead birds, and mosquito pool positivity per six-month periods in each

254

county over the study period. There are clear seasonal differences in all the variables; the

255

analysis of detangling seasonality and temperature to examine the association between

256

temperature and WNV cases is presented below.

257

258

259

260

261

262

263

264

265

266

267

268

269

270 Table 1

*Table 1. Characteristics of WNV Cases and Environmental Factors per Month By County from 2017-2022*

\* All variables are average per month

	Total	Fresno	Kern	Los Angeles	Merced	Orange	Riverside	Sacramento	San Bernardino	San Joaquin	Stanislaus	Tulare
Cases	0.12	0.06	0.08	0.75	0.03	0.06	0.11	0.03	0.14	0.03	0.03	0.03
Nov - Apr	3.15	3.61	2.92	13.33	1.39	2.03	1.72	1.14	2.11	1.28	3.14	2.03
Temperature (F)	56.5	55.2	56.1	61.3	52.5	61.1	59.3	54.7	59.5	53.7	53.5	54.9
Nov - Apr	75.2	78.4	79.3	72.3	74.2	71.8	75.6	74.5	76.1	73.6	74.3	76.6
May - Oct	Rainfall (In.)	1.47	0.87	1.98	1.60	1.57	1.29	2.82	1.74	1.96	1.65	1.81
Nov - Apr	0.15	0.12	0.10	0.10	0.11	0.09	0.09	0.34	0.10	0.24	0.14	0.18
May - Oct	Dead Birds	0.00	0.00	0.50	0.00	0.47	0.06	0.00	0.00	0.00	0.00	0.00
Nov - Apr	3.80	0.64	0.17	11.42	0.25	4.67	1.58	16.83	1.75	2.81	0.50	1.14
May - Oct	Mosquito Positivity Rate (%)	0.00	0.00	1.12	0.00	0.21	0.11	0.00	0.17	0.00	0.00	0.00
Nov - Apr	7.83	13.55	14.97	10.46	4.37	3.70	4.10	2.32	3.26	9.76	8.63	11.03
May - Oct												

271 *Statistical Models and Standardization*

272 Both the zero-inflated and GEE with autoregressive correlation models show a positive  
273 correlation between prior month temperature and WNV cases in a month (Table 2). Each model  
274 was adjusted for the variables prior month rainfall, county, and month, which were determined  
275 to be confounders in the relationship of interest.

276

277 *Table 2*

Association Between Incidence of WNV and Prior Month Temperature in High-Burden California Counties

Characteristic	Zero-Inflated Regression		GEE with Autoregressive Correlation	
	IRR <sup>1,2</sup>	95% CI <sup>2</sup>	IRR <sup>1,2</sup>	95% CI <sup>2</sup>
Prior Month Temperature	1.057	1.013, 1.103	1.048	1.012, 1.086

<sup>1</sup> Adjusted for Prior Month Rainfall, County, and Month  
<sup>2</sup> IRR = Incidence Rate Ratio, CI = Confidence Interval

278

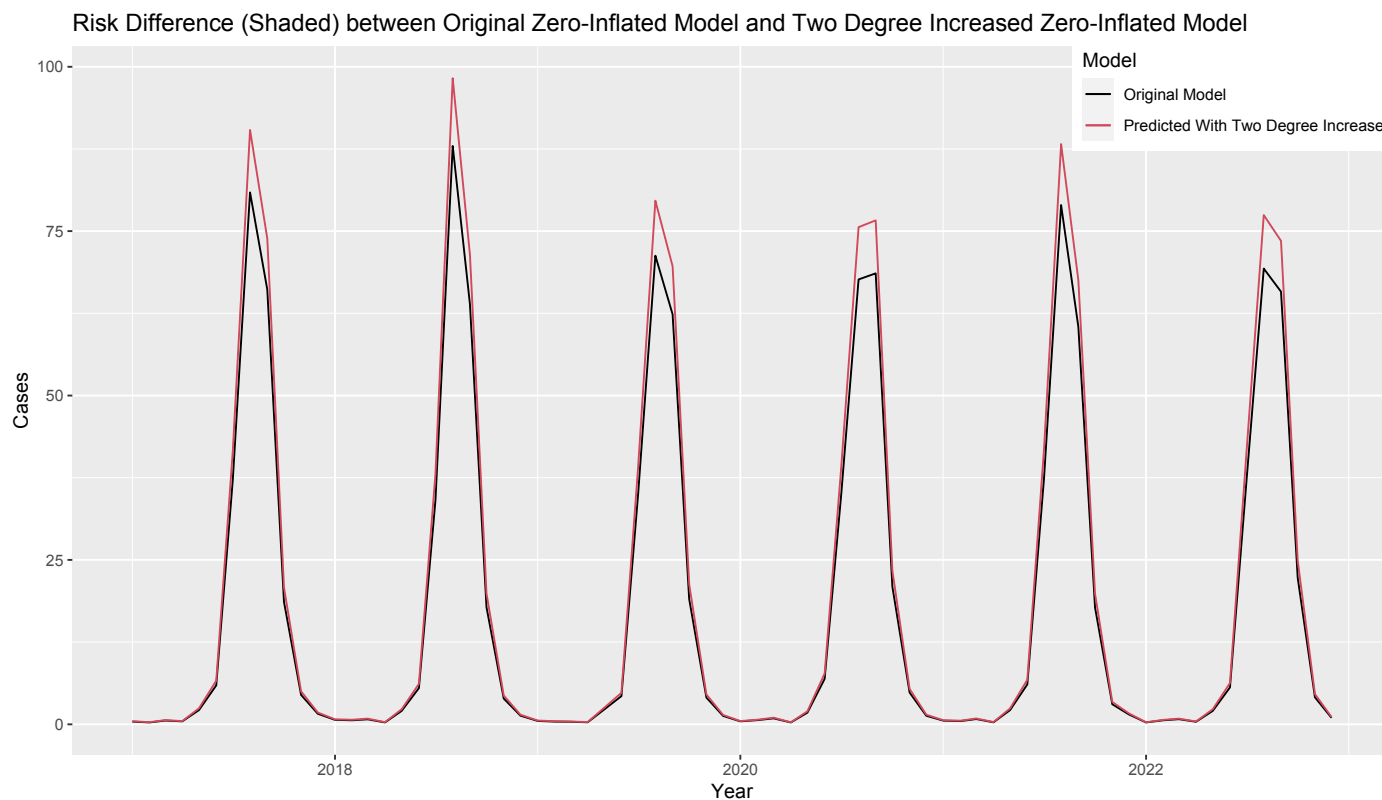
279 Both models had nearly the same estimation of 1.05 for the incidence rate ratio — this  
280 represents that a one degree increase in prior month temperature is associated with a 1.05  
281 times greater incidence of West Nile virus in a month, controlling for prior month rainfall,  
282 location, and the time of year. The IRR estimates from both models were statistically significant,  
283 though the autoregressive GEE model had a narrower 95% confidence interval (Table 2).

284 The models can also be extrapolated to larger changes in temperatures, as is predicted  
285 by some climatologists [13–15]. Some estimates suggest that California’s average temperatures  
286 may rise by two degrees Fahrenheit by 2040 [13,14]. Given a two degree increase in  
287 temperature, and holding all other variables constant, the zero-inflated models suggest that this

288 would be associated with a 1.12 times greater incidence rate, while the GEE model reports this  
289 number to be 1.10.

290 Fig 2 shows the risk difference of WNV cases conducted through model standardization,  
291 comparing the predicted cases under the observed data versus a hypothetical dataset where all  
292 temperature observations were increased by two degrees Fahrenheit. As shown in the figure, a  
293 two degree increase in temperature, holding all other variables constant, results in more cases (a  
294 positive risk difference) in the zero-inflated model. This increase is heavily concentrated in  
295 summer months, with at least eight excess cases predicted every year during peak months. An  
296 increase in temperature barely affects winter months where there is little to no transmission.  
297 Across the six-year study period, the risk difference between the hypothetical data, with a two  
298 degree temperature increase, and the observed data is 148.7 (95% confidence interval of 28.8 -  
299 292.8) cases for the zero-inflated model. This averages out to 24.8 (4.8 - 48.8) excess cases per  
300 year.

301 *Fig 2*



302

303

304 *Prediction Models*

305 Table 3 shows the output of the prediction model with all variables for the zero-inflated  
306 model and GEE with autoregressive correlation model. As in the measure of association model,  
307 both types of models showed fairly similar results, as prior month temperature, mosquito pool  
308 positivity rate, and dead birds positively predicted WNV cases in each model, while prior month  
309 rainfall negatively predicted WNV cases. The models' predictions slightly differed in their  
310 estimates of the beta for each variable, and while small in absolute size, the differences were  
311 significant, as seen in later analysis.

312



313 **Table 3**

Prediction Model Output Between Environmental Factors and WNV Incidence				
Characteristic	Zero-Inflated Regression		GEE with Autoregressive Correlation	
	Beta	95% CI <sup>1</sup>	Beta	95% CI <sup>1</sup>
Prior Month Temperature	0.08	0.04, 0.11	0.07	0.02, 0.12
Prior Month Rainfall	-0.21	-0.48, 0.07	-0.24	-0.38, -0.11
Mosquito Positivity Rate	0.05	0.03, 0.06	0.04	0.02, 0.05
Dead Birds	0.06	0.04, 0.08	0.02	0.01, 0.04

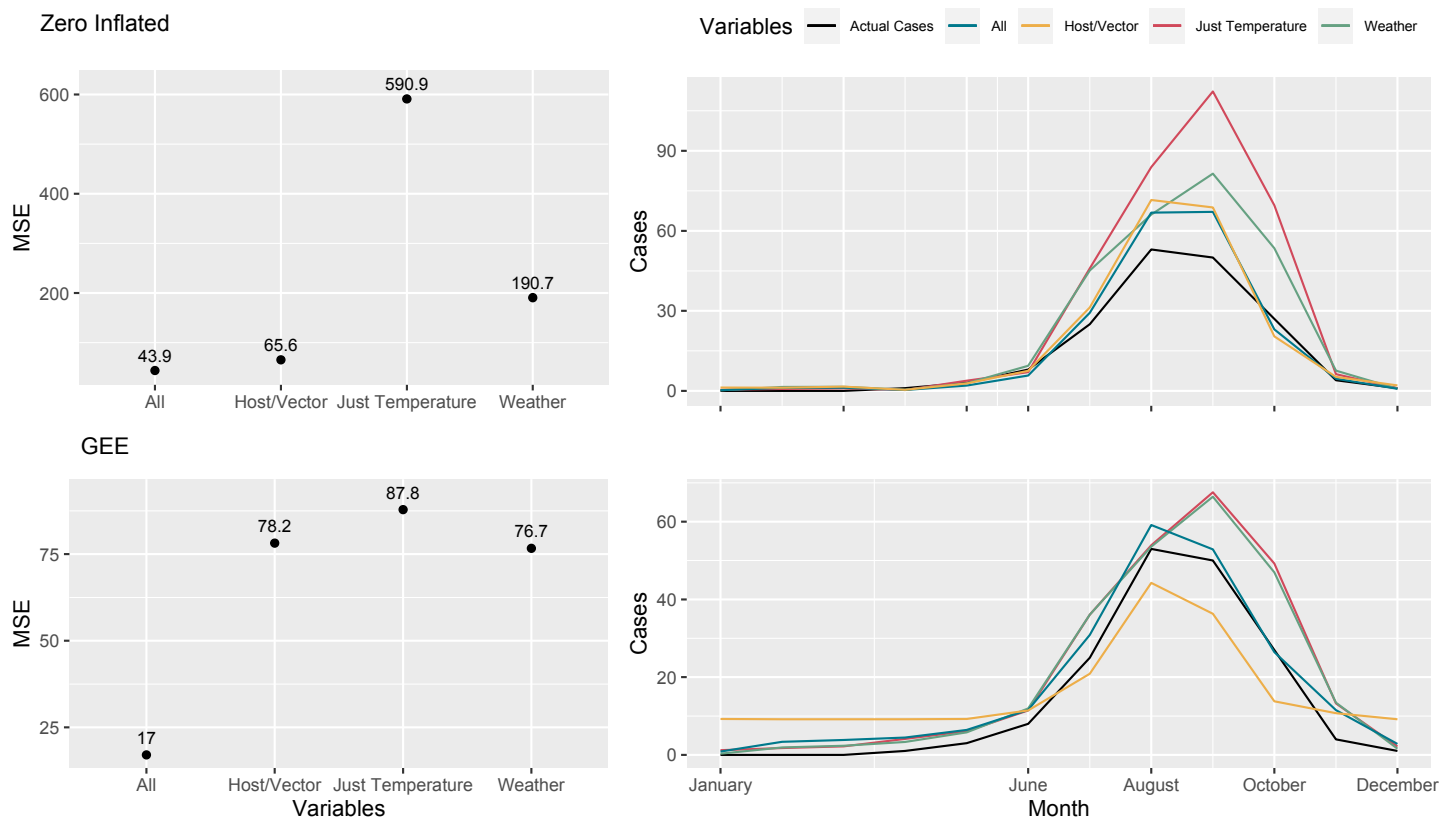
<sup>1</sup> CI = Confidence Interval

314

315 Fig 3 plots the predicted cases in the 11-high burden counties in 2022 for each variable  
316 set — across both model types — against the actual number of reported cases, as well as each  
317 model’s MSE. All variable sets of the zero-inflated model fit the actual data well during zero case  
318 count months (November-June) but over-estimated the number of cases in the peak summer  
319 months. “Just Temperature” and “Weather” both greatly over-estimated the peak summer  
320 months and incorrectly predicted the shape of the incidence curve. It’s possible that since the  
321 zero-inflated model already uses the month of year, which is very strongly tied to weather  
322 patterns, to determine non-zero case counts, adding weather data to the prediction model did  
323 not increase the strength of the fit. “All” variables provided the best fit — the lowest MSE — to  
324 the data out of the zero-inflated subset.

325 **Fig 3**

Predicted Cases for 2022 for Zero-Inflated Model (Top Row) and GEE Model (Bottom Row) Using Different Sets Of Variables



\*All = Prior Month Temp, Prior Month Rainfall, Dead Birds, Mosquito Positivity Rate; Host/Vector = Dead Birds and Mosquito Positivity Rate; Weather = Prior Month Temp and Prior Month Rainfall

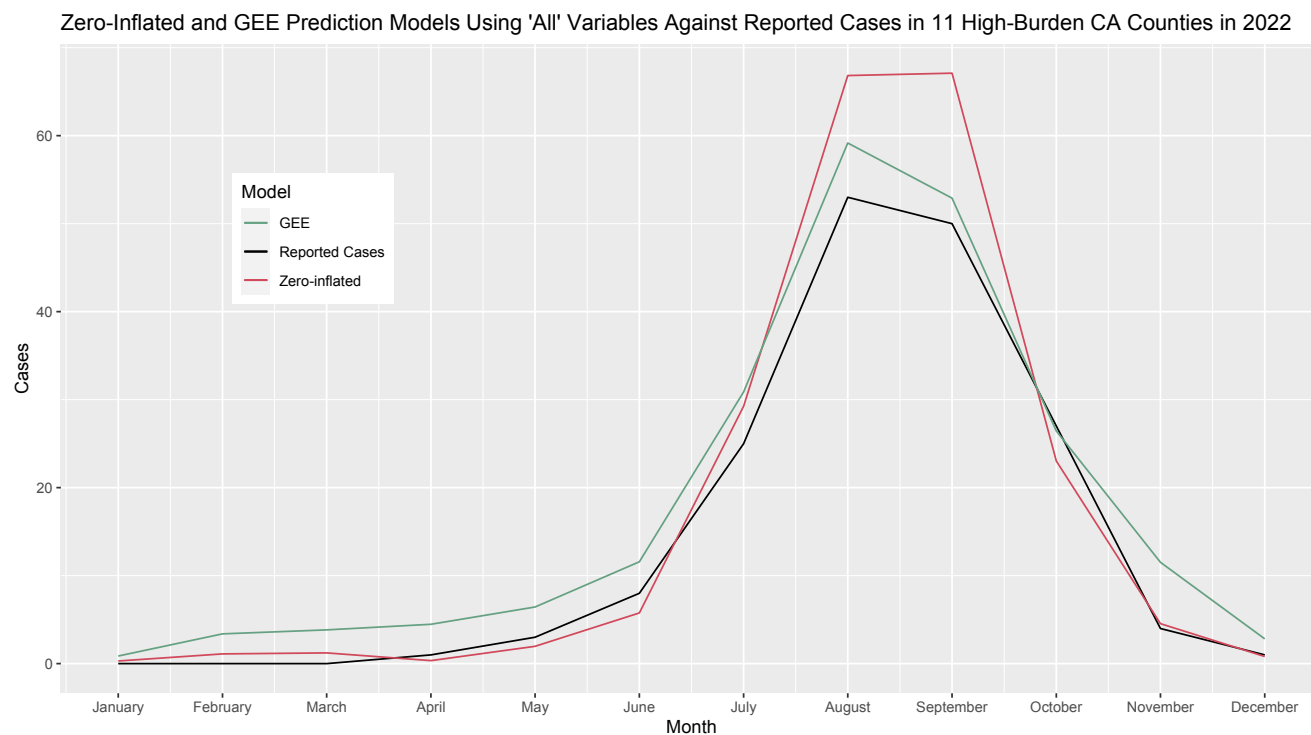
326

327 The GEE subset of models showed more similar fits, again most likely due to the model  
 328 not having the month predictor for zero case counts. In contrast to the zero-inflated models, the  
 329 GEE prediction models followed the peak of cases in August much more closely, though it also  
 330 overestimated the actual reported number of cases slightly. The GEE model greatly  
 331 overestimated the number of cases in months with low case counts, however, often at scales of  
 332 three to four times larger. The “Host/Vector” model especially produced a skewed fit, as even  
 333 with a mid-range mean squared error, it predicted 10 cases a month in the winter months and  
 334 overall had a non-normal shape. The “All” GEE model resulted in the best fit across all of the  
 335 prediction models with a mean squared error of 17 and a closely following curve.

336            Though it had the lowest MSE, the “All” GEE model still predicted more total cases in  
337 2022 than were actually reported, by a count of 214 to 172, respectively. The best performing  
338 model in this regard was in fact the skewed “Host/Vector” GEE model (192 cases), followed by  
339 the “All” zero-inflated model (202 cases), and the “Host/Vector” zero-inflated model (214 cases).

340            In a simpler view, Fig 4 shows both the zero-inflated and GEE “All” models — the two-  
341 best performing models according to MSE — plotted against the actual reported cases in the 11-  
342 high burden counties in 2022. This more clearly shows the differences between the prediction  
343 models discussed above.

344 **Fig 4**



345

#### 346 **F. Discussion**

347            Our study highlights an association between temperature and West Nile virus cases in 11  
348 high-burden California counties from 2017-2022; holding month of year, location, and prior

349 month rainfall constant, a one unit increase in prior month temperature was associated with  
350 1.05 times the incidence of WNV. Given a hypothetical two degree Fahrenheit increase in  
351 average monthly temperature, which has been predicted for California in the next few decades  
352 due to climate change [13,14], the zero-inflated model predicted 24.8 excess cases per year over  
353 our study period. In the prediction models, using all variables — weather and host/vector  
354 combined — resulted in the best fit, and lowest mean squared error, for 2022 case counts. The  
355 GEE approach had overall better fitting prediction models, but the zero-inflated model was able  
356 to predict the total case count more closely for the year.

357 Our estimates on the significantly positive association between temperature and WNV  
358 incidence at the group level is well-supported in other studies in California and the US [9,16]. The  
359 magnitude of our association is in line with findings from Hahn et. al, who observed an odds ratio  
360 of 1.1 nationally [16]; alternatively, Hernandez et. al calculated an odds ratio of 10.5 in  
361 California's San Joaquin Valley, though this finding did have a wide confidence interval [9]. The  
362 connection between temperature and mosquito life traits, [6] such as development time and  
363 survival, as well as mosquito and viral interaction, such as the extrinsic incubation period — the  
364 time it takes for an infected mosquito to become infective to hosts — have been documented in  
365 numerous other studies [7,17]. These associations are complex and can often vary at different  
366 temperatures: e.g. *Culex* development rates rise with temperature until a certain point, when  
367 increasing temperature then causes rising mortality [18]. While this paper doesn't go into the  
368 biological mechanisms that mediate this relationship, these previous findings provide a causal  
369 basis for the observed association.

370           Given the results of both types of prediction models, it appears that having weather data  
371 alone provides little predictive power for WNV cases in California. Host and vector data provides  
372 a pretty good fit in the absence of weather data for the zero-inflated model, showing that simple  
373 enzootic data from local vector departments can be enough to give useful information on  
374 current and future WNV cases in humans. However, given how readily available weather data is,  
375 it seems unnecessary and uninformative to use a prediction model with just host and vector  
376 data. This mostly aligns with the results of previous work. Wimberly et. al shows that while  
377 meteorological and mosquito infection rate data can predict WNV cases better than just  
378 historical data, it takes the combination of the two to best predict future cases [8]. Davis et. al's  
379 prediction model also showed that meteorological and mosquito data together can well predict  
380 future WNV cases, but again, disagreed with our model that mosquito infection data alone does  
381 nearly as well as all the data [19].

382           Our study revealed the power of the zero-inflated model to correctly predict the number  
383 of cases in low transmission months compared to more traditional models. However, the utility  
384 of this feature is limited, as low transmission months are of much smaller worry to health  
385 departments than summer months. The more traditional GEE model followed the shape and  
386 case counts of the peak summer months much closer and may provide health departments with  
387 a much more accurate picture of what may come month to month. Though it greatly  
388 overpredicts the number of cases during the summer, the zero-inflated model more accurately  
389 predicted the total number of cases in 2022, highlighting its potential utility in helping the state  
390 plan on a macro level. These models will need to be tested on a larger geographic scale with

391 more counties and with newer years of data to make any final statements on their ability to be  
392 used by health departments.

393 A great strength of the study is its potential to be easily interpretable and implementable  
394 to public health agencies and the public: the data was easily accessible, and the model has just a  
395 few parameters that can be quickly integrated. These models allow for analyses that don't  
396 require complex knowledge of the many relationships that underlie the observed associations,  
397 including the mechanisms by which temperature is associated with incidence. It also lends itself  
398 to future use by public health departments, which don't need mountains of data to help predict  
399 WNV cases. We also utilized standardization to estimate a risk difference in cases between the  
400 observed data and a hypothetical California with an increased temperature of two degrees  
401 Fahrenheit, an underutilized strategy that can provide crucial insights into the future of WNV  
402 transmission in the state. Though a hypothetical scenario, it provides more easily digestible and  
403 understandable estimates (in number of excess cases per month) than an odds ratio, or even our  
404 calculated rate ratio.

405 The data in this study has inherent limitations, however. Cases were aggregated on a  
406 monthly level, and other variables were in turn averaged or summed to correspond to this time  
407 frame. The lack of granularity in case data may bias the association between temperature and  
408 incidence, especially considering that weather can be highly volatile, a characteristic that is lost  
409 when averaged over a monthly time frame. Surveillance data, which comprises the case counts,  
410 dead bird counts, and mosquito pool positivity rate, also carries its own shortcomings, as it  
411 misses most, if not all, of asymptomatic infections and may not be representative of the  
412 infection dynamics in the system as a whole. We also lost granularity and possible variation in

413 weather patterns and transmission dynamics across counties by analyzing the counties together;  
414 future work would benefit by looking at the effect of temperature on WNV incidence within  
415 counties, possibly at the zip code or census tract level.

416         There are also limitations based on the way the models were constructed, especially the  
417 zero-inflated model. Based on data from the last decade, we assumed that the month of year  
418 was a quality predictor of whether there would be non-zero case counts. However, studies have  
419 shown that a rise in temperature can also shift the seasonality of WNV in California [20]. This  
420 change would challenge our assumption and possibly bias our estimate of the association  
421 between temperature and WNV. Sensitivity analyses can be run to determine if other predictors,  
422 or sets of predictors, would be better suited as the zero count predictor in the zero-inflated  
423 model.

424         While this aggregated ecological study has limitations, it provides an easy and convenient  
425 study design using pre-existing surveillance data. As previously mentioned, it is mandatory to  
426 report WNV disease, creating a passive surveillance system; as a result, no new data on the  
427 outcome needed to be assessed. The type of exposure — environmental — also pointed to using  
428 an ecological study, as temperature is a group-level exposure. It not only would be hard to  
429 measure the temperature experienced by an individual in a cohort, but generally uninformative  
430 to attempt to discern how a change in temperature affects a singular person's risk of WNV  
431 disease, which is influenced by many different possible factors, including type of job and  
432 socioeconomic status.

433         Therefore, we don't believe to be committing any ecological fallacy — inferring that the  
434 ecological association equals the individual association — as we don't attempt to determine or

435 make assumptions about the effect of temperature on the individual risk of WNV disease. Our  
436 study simply examines the risk aggregated at the county level. This is probably more informative  
437 to Californian public health agencies, as the risk of WNV disease to the average Californian is  
438 very low. In all, an ecological study was sufficient and better applicable to calculate the  
439 association of interest at the county level, while remaining easily replicable to future researchers  
440 and public health departments.

441 More types of climatic analysis can be conducted using this same data, such as the effect  
442 of prior seasonal rainfall and temperature on future WNV incidence, the importance of which  
443 has been shown for the length of the WNV season in the US as a whole [21,22]. Exploring the  
444 mechanisms that mediate the association between temperature and West Nile virus would also  
445 be beneficial to further analyze different intervention strategies. Finally, more work is required  
446 to understand the interface between humans and the enzootic carriers of WNV in California, and  
447 how climate change may alter the ways people interact and are exposed to these hosts and  
448 vectors.

449

#### 450 ***Conclusion***

451 Increasing temperatures is associated with a higher incidence of West Nile virus infection  
452 in 11 high-burden California counties. Along with mosquito and bird infection data,  
453 meteorological data can well-predict future WNV incidence in these counties. This will be of  
454 extreme importance to public health agencies as WNV continues to become a larger health  
455 burden due to global warming and climate change.

456



457 ***Acknowledgements***

458           I would like to thank Dr. Jack Colford for his continued help with the structure, revisions,  
459 and publishing process of this work, Dr. Mary Beth Danforth for her guidance with the CDPH  
460 data, and Lincoln Wells and the team at VectorSurv for providing the host and vector data. This  
461 work was supported by the Department of Epidemiology at the Berkeley School of Public Health  
462 and was only possible with the assistance of numerous researchers and mentors in Berkeley.  
463

464 References

- 465
- 466 1. Albrecht L, Kaufeld KA. Investigating the impact of environmental factors on West Nile virus  
467 human case prediction in Ontario, Canada. *Front Public Health* [Internet]. 2023 [cited 2023 Oct  
468 31];11. Available from: <https://www.frontiersin.org/articles/10.3389/fpubh.2023.1100543>
- 469 2. Reisen W, Lothrop H, Chiles R, Madon M, Cossen C, Woods L, et al. West Nile virus in  
470 California. *Emerg Infect Dis*. 2004;10:1369–78.
- 471 3. Snyder RE, Feiszli T, Foss L, Messenger S, Fang Y, Barker CM, et al. West Nile virus in California,  
472 2003–2018: A persistent threat. *PLoS Negl Trop Dis*. 2020;14:e0008841.
- 473 4. Danforth ME, Snyder RE, Lonstrup ETN, Barker CM, Kramer VL. Evaluation of the effectiveness  
474 of the California mosquito-borne virus surveillance & response plan, 2009–2018. *PLoS Negl Trop*  
475 *Dis*. 2022;16:e0010375.
- 476 5. Soverow JE, Wellenius GA, Fisman DN, Mittleman MA. Infectious Disease in a Warming World:  
477 How Weather Influenced West Nile Virus in the United States (2001–2005). *Environ Health*  
478 *Perspect*. 2009;117:1049–52.
- 479 6. CIOTA AT, MATAACCHIERO AC, KILPATRICK AM, KRAMER LD. The Effect of Temperature on Life  
480 History Traits of *Culex* Mosquitoes. *J Med Entomol*. 2014;51:55–62.
- 481 7. Richards SL, Mores CN, Lord CC, Tabachnick WJ. Impact of Extrinsic Incubation Temperature  
482 and Virus Exposure on Vector Competence of *Culex pipiens quinquefasciatus* Say (Diptera:  
483 Culicidae) for West Nile Virus. *Vector Borne Zoonotic Dis* Larchmt N. 2007;7:629–36.
- 484 8. Wimberly MC, Davis JK, Hildreth MB, Clayton JL. Integrated Forecasts Based on Public Health  
485 Surveillance and Meteorological Data Predict West Nile Virus in a High-Risk Region of North  
486 America. *Environ Health Perspect*. 2022;130:087006.
- 487 9. Hernandez E, Torres R, Joyce AL. Environmental and Sociological Factors Associated with the  
488 Incidence of West Nile Virus Cases in the Northern San Joaquin Valley of California, 2011–2015.  
489 *Vector-Borne Zoonotic Dis*. 2019;19:851–8.
- 490 10. Data and Maps | West Nile Virus [Internet]. CDC. 2023 [cited 2023 Dec 26]. Available from:  
491 <https://www.cdc.gov/westnile/statsmaps/data-and-maps.html>
- 492 11. DeFelice NB, Birger R, DeFelice N, Gagner A, Campbell SR, Romano C, et al. Modeling and  
493 Surveillance of Reporting Delays of Mosquitoes and Humans Infected With West Nile Virus and  
494 Associations With Accuracy of West Nile Virus Forecasts. *JAMA Netw Open*. 2019;2:e193175.
- 495 12. 12.1 - Introduction to Generalized Estimating Equations [Internet]. Penn State Eberly Coll. Sci.  
496 [cited 2023 Dec 26]. Available from: <https://online.stat.psu.edu/stat504/book/export/html/797>

- 497 13. Cayan D, Kalansky J, Pierce D, Gershunov A. FAQ: Climate Change in California | Scripps  
498 Institution of Oceanography [Internet]. UC San Diego. [cited 2023 Dec 26]. Available from:  
499 <https://scripps.ucsd.edu/research/climate-change-resources/faq-climate-change-california>
- 500 14. Pierce D, Kalansky J, Cayan D. Climate, Drought, and Sea Level Rise Scenarios for California’s  
501 Fourth Climate Change Assessment. California Energy Commission; 2018 Aug p. 17–20. Report  
502 No.: 006.
- 503 15. Gershunov A, Guirguis K. California heat waves in the present and future. *Geophys Res Lett*  
504 [Internet]. 2012 [cited 2023 Dec 26];39. Available from:  
505 <https://onlinelibrary.wiley.com/doi/abs/10.1029/2012GL052979>
- 506 16. Hahn MB, Monaghan AJ, Hayden MH, Eisen RJ, Delorey MJ, Lindsey NP, et al. Meteorological  
507 Conditions Associated with Increased Incidence of West Nile Virus Disease in the United States,  
508 2004–2012. *Am J Trop Med Hyg*. 2015;92:1013–22.
- 509 17. Reisen WK, Fang Y, Martinez VM. Effects of temperature on the transmission of west nile  
510 virus by *Culex tarsalis* (Diptera: Culicidae). *J Med Entomol*. 2006;43:309–17.
- 511 18. DeFelice NB, Schneider ZD, Little E, Barker C, Caillouet KA, Campbell SR, et al. Use of  
512 temperature to improve West Nile virus forecasts. *PLoS Comput Biol*. 2018;14:e1006047.
- 513 19. Davis JK, Vincent G, Hildreth MB, Kightlinger L, Carlson C, Wimberly MC. Integrating  
514 Environmental Monitoring and Mosquito Surveillance to Predict Vector-borne Disease:  
515 Prospective Forecasts of a West Nile Virus Outbreak. *PLOS Curr Outbreaks* [Internet]. 2017 [cited  
516 2023 Dec 21]; Available from: [http://currents.plos.org/outbreaks/article/integrating-  
517 environmental-monitoring-and-mosquito-surveillance-to-predict-vector-borne-disease-  
518 prospective-forecasts-of-a-west-nile-virus-outbreak/](http://currents.plos.org/outbreaks/article/integrating-environmental-monitoring-and-mosquito-surveillance-to-predict-vector-borne-disease-prospective-forecasts-of-a-west-nile-virus-outbreak/)
- 519 20. Hartley DM, Barker CM, Le Menach A, Niu T, Gaff HD, Reisen WK. Effects of Temperature on  
520 Emergence and Seasonality of West Nile Virus in California. *Am J Trop Med Hyg*. 2012;86:884–  
521 94.
- 522 21. Gorris ME, Randerson JT, Coffield SR, Treseder KK, Zender CS, Xu C, et al. Assessing the  
523 Influence of Climate on the Spatial Pattern of West Nile Virus Incidence in the United States.  
524 *Environ Health Perspect*. 2023;131:047016.
- 525 22. Ward MJ, Sorek-Hamer M, Henke JA, Little E, Patel A, Shaman J, et al. A Spatially Resolved  
526 and Environmentally Informed Forecast Model of West Nile Virus in Coachella Valley, California.  
527 *GeoHealth*. 2023;7:e2023GH000855.
- 528 23. Erdman D, Jackson L, Sinko A. Zero-Inflated Poisson and Zero-Inflated Negative Binomial  
529 Models Using the COUNTREG Procedure. SAS Institute. 2008;  
530
- 531 24. WestNile.ca.gov [Internet]. Vector-Borne Disease Section | CA Dept of Public Health. [cited  
532 2023 Dec 26]. Available from: <https://westnile.ca.gov/>

- 533
- 534 25. Negative Binomial Regression. NCSS Documentation. NCSS Statistical Software; 2023.
- 535
- 536 26. Zero-Inflated Negative Binomial Regression. NCSS Documentation. NCSS Statistical Software;
- 537 2023.
- 538
- 539