Vaccine Effects on Ticks

Jack Cunningham

Goals and Experimental Design

The goal of this data analysis is to determine the effect vaccine status and time have on tick mortality and reproduction when attached to deer.

The experimental design is as follows:

We have 12 experimental units, the deer. Some deer are pre infested, and that data is collected as a baseline before vaccines are applied.

There are three treatments, the vaccination a deer receives. It has three levels: Control, Low, and High.

The measurement unit is a tick, there were plans for 50 ticks for each deer but in the experiment this was not adhered to.

The experiment is repeated three times with a new set of ticks applied to each deer.

The response variables are the mortality, and measurements of egg quality if the tick laid them.

There was one covariate, the weight of the tick before being attached to the deer.

Data Preprocessing

Full details of data preprocessing can be found attached to this submission but I will summarize the main findings and choices I made.

There were a large amount of missing data, particularly the columns tracking egg quantity and quality. The majority of this missing data was found to be associated with a mortality variable being set to one. In that case the data is assumed to be zero, these observations are when a tick dies before giving birth.

I created a variable called result which tracked three cases, a death of a tick, a death of a tick after laying some eggs and when a tick lives. This was not included in the modeling section, but it helps explain why some egg quantity variables have long left tails.

There were 15 rows that remained with NA values. Some had comments indicating the dram spilled which did not allow the researchers to collect egg data. In these cases I used median imputation, given the tick's mortality was reported as alive.

Additionally there were cases where the "% Hatched" variable was missing, before making changes I examined that the "% Hatched" variable was more likely than not a function of two other variables. The number of larvae divided by the number of the number of eggs hatched.

This fact helped fill out cases where either the number of eggs or number of larvae were missing, as I was able to compute the missing value.

Then there were cases where the number of larvae was greater than the number of eggs, after analyzing the rows where this occurred, using the help of the % hatch variable, I determined that most cases were a data collection issue where the two were reported in the opposite column.

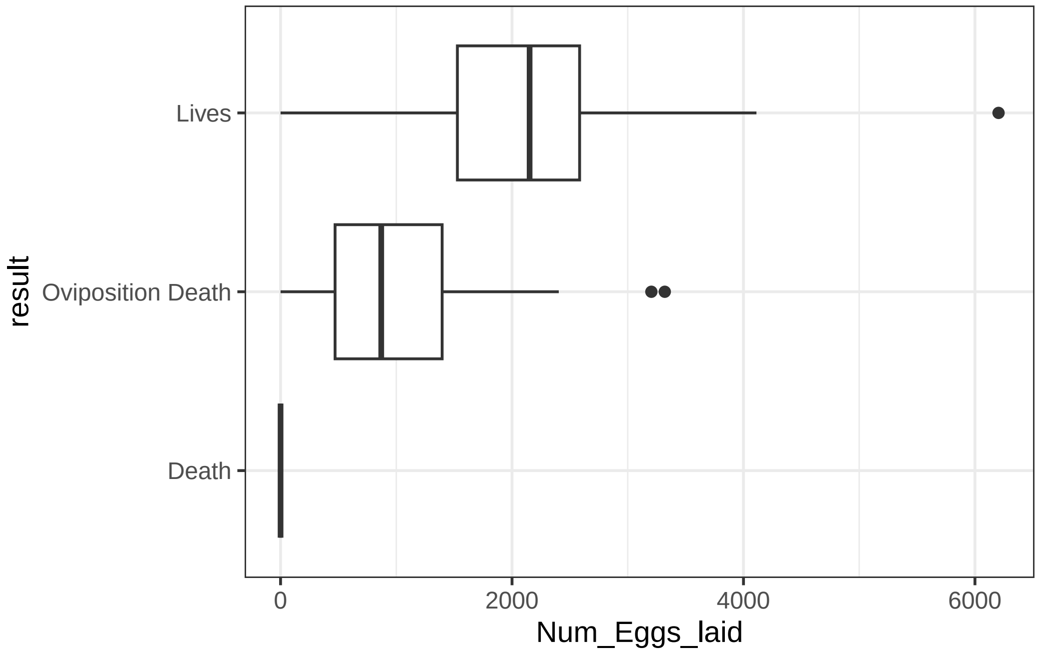
Remaining missing values followed a similar median imputation as previously discussed.

Data was converted to numeric, factor and character data types.

Exploratory Analysis

Response Variables

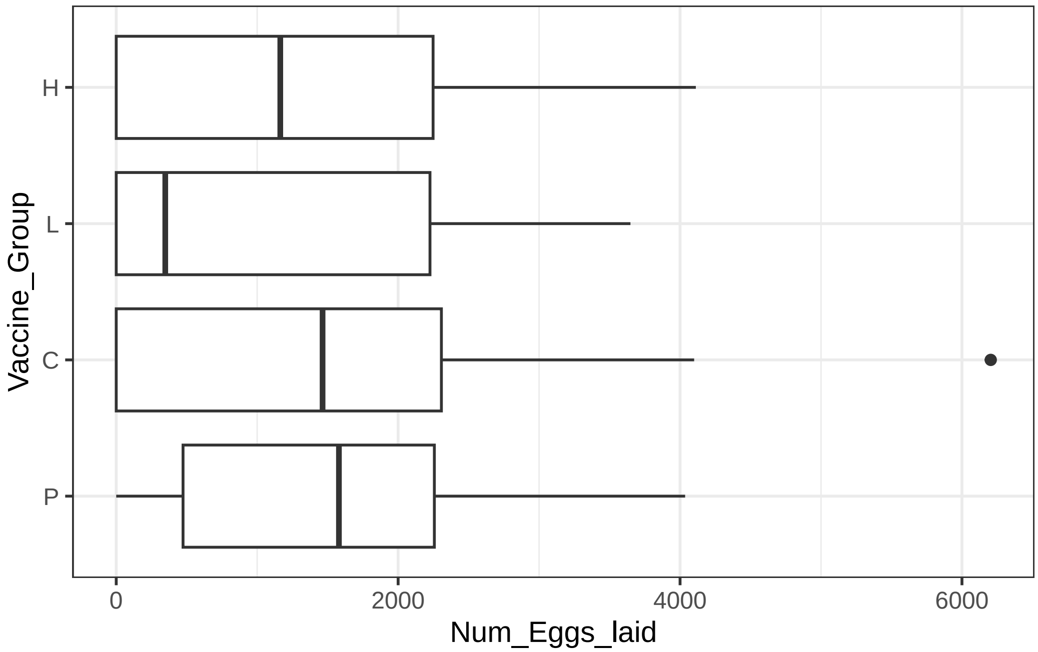
Starting with numeric variables, lets first look at eggs laid.



We can see that the number of eggs laid is a steady zero for when the tick dies. If the tick dies during oviposition the number of eggs laid is generally lower than when the tick lives. The variation in the number of eggs is also greater when the tick lives, compared to oviposition death.

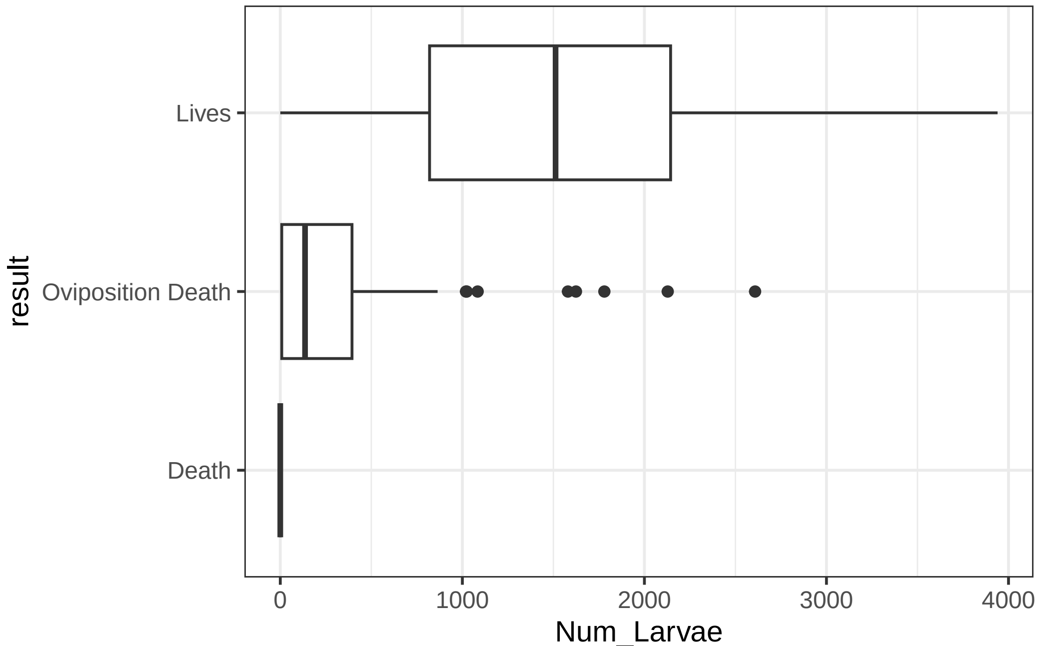
There exists one outlier value from the number of eggs being laid exceeds 6000 but there does not appear to be anything suspicious about this observation, and it may indicate some right skew in the number of eggs laid variable. Generally, this is a symmetric distribution however.

Let's look at the eggs laid across the vaccination groups:

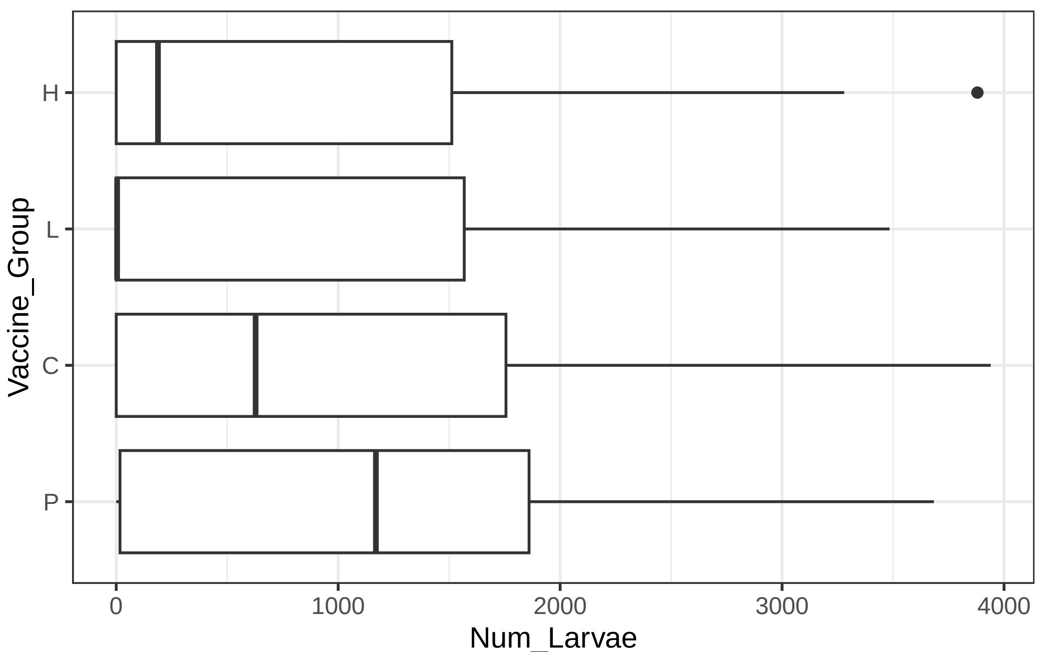


Interestingly enough it appears that when the vaccine dose is low the lower number of eggs are laid. The control and the higher vaccine dose seem quite similar when it comes to this metric.

This motivates me to take a look at another response variable we have, number of larvae. Lets see how this varies across vaccine groups:

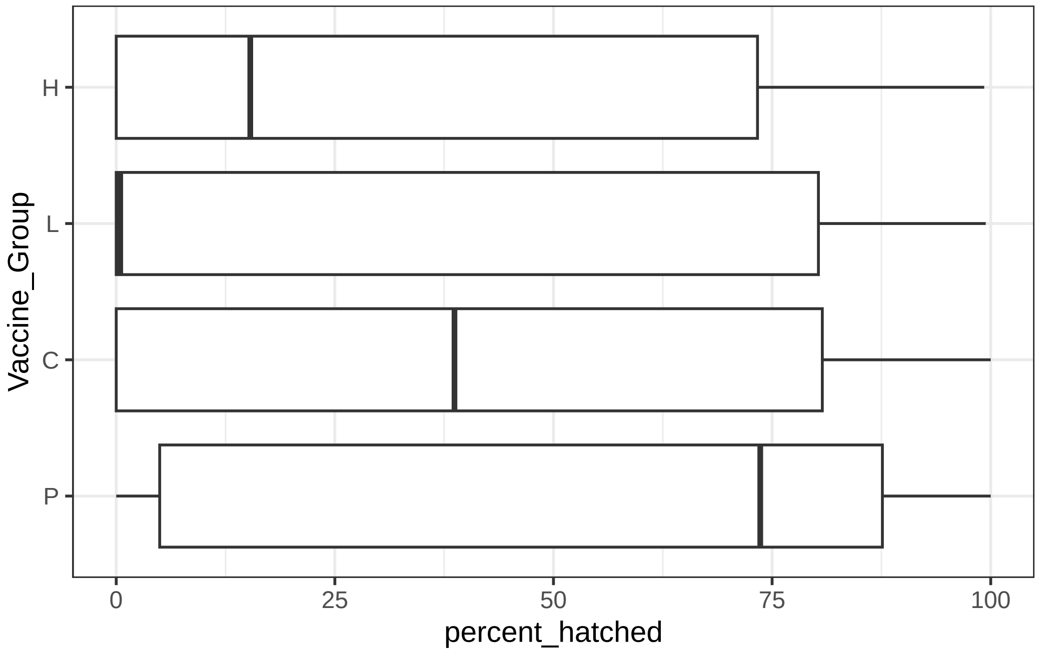


We can see that when the result is Oviposition death there is a lot less larvae than when the result is lives. The distribution of the number of Larvae when oviposition is death is very right skewed however, occasionally there are many Larvae despite the Oviposition Death.



We can see that the number of larvae is much lower in the vaccination groups when compared to the two non vaccination groups.

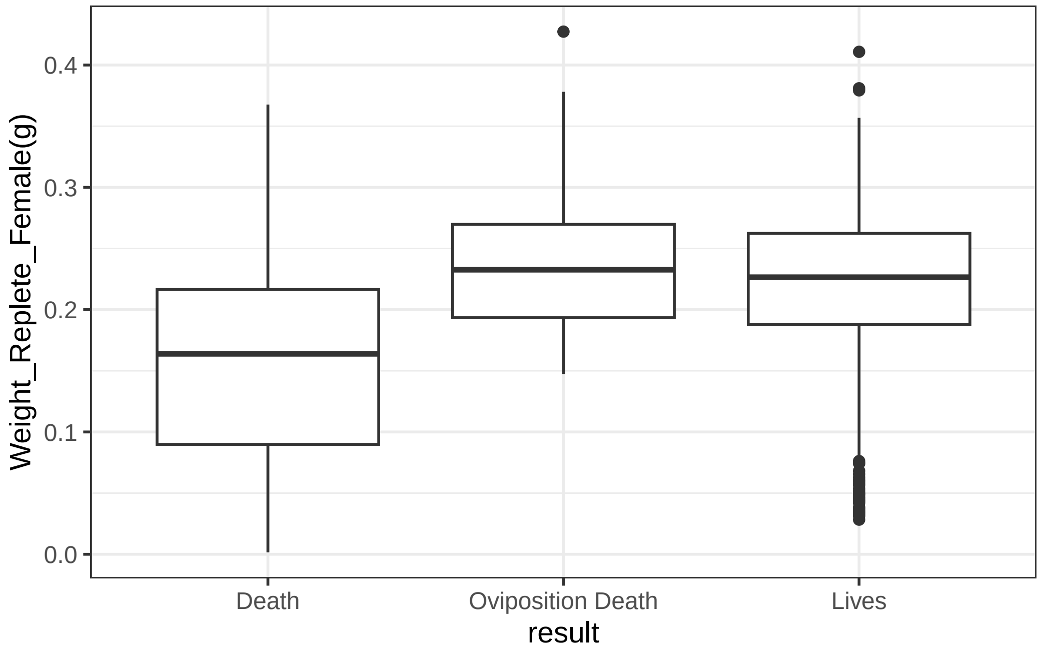
Let's look at % Hatched:



We can see that the vaccination groups L and H both have percent hatched smaller in general, but each distribution is disperse across the full range of reasonable percentages.

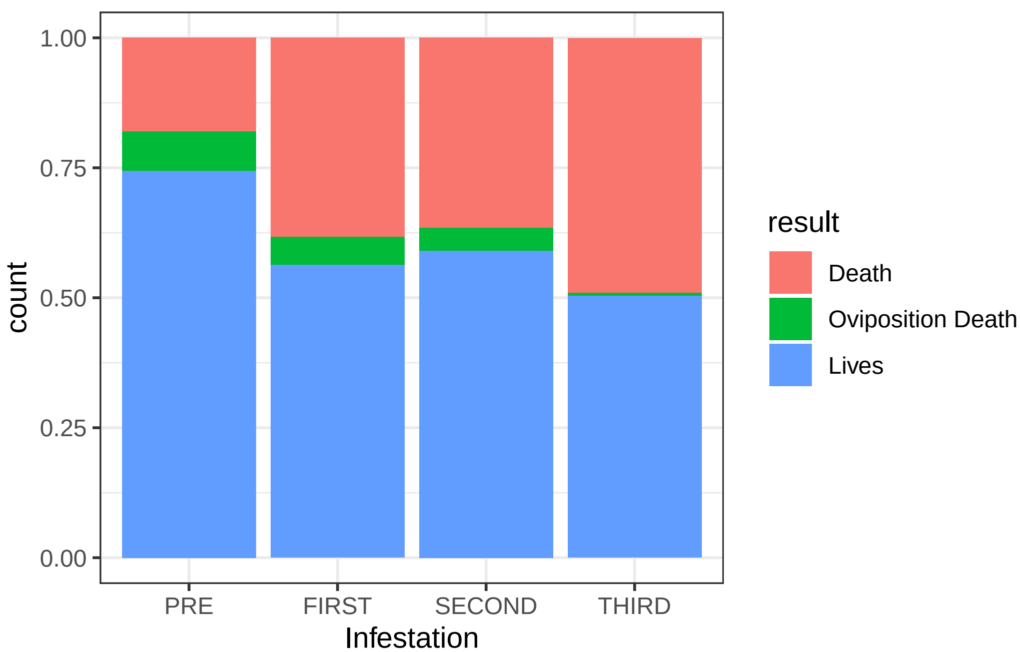
Covariates

Lets see if Weight_Replete_Female has a strong relationship with the result:



The difference is marginal but significant, the range of feasible weights for ticks is not too large and this seems like it will be an important covariate to keep in our model.

Lets see if Infestation is strongly related to the result:



Considering that the pre infestation is confounded with the fact that there were no vaccines administered in that stage, the time of infestation does not initally seem to be too important of a predictor.

Modeling and Hypothesis Testing

Model Considerations

Through the data processing and exploratory data analysis steps I made a few determinations on how to approach modeling.

1. How to treat pre-infestation data

In theory it would be interesting if we had pre-infestation data for each deer, then we would be able to track how tick mortality and egg quality and quantity change after vaccinations are applied.

In reality only six of the deer have this data, so using this data would complicate the analysis and create difficult to interpret effects. For that reason I am focusing the analysis to the experimental data, after vaccinations are applied to each deer.

2. What response variables to model

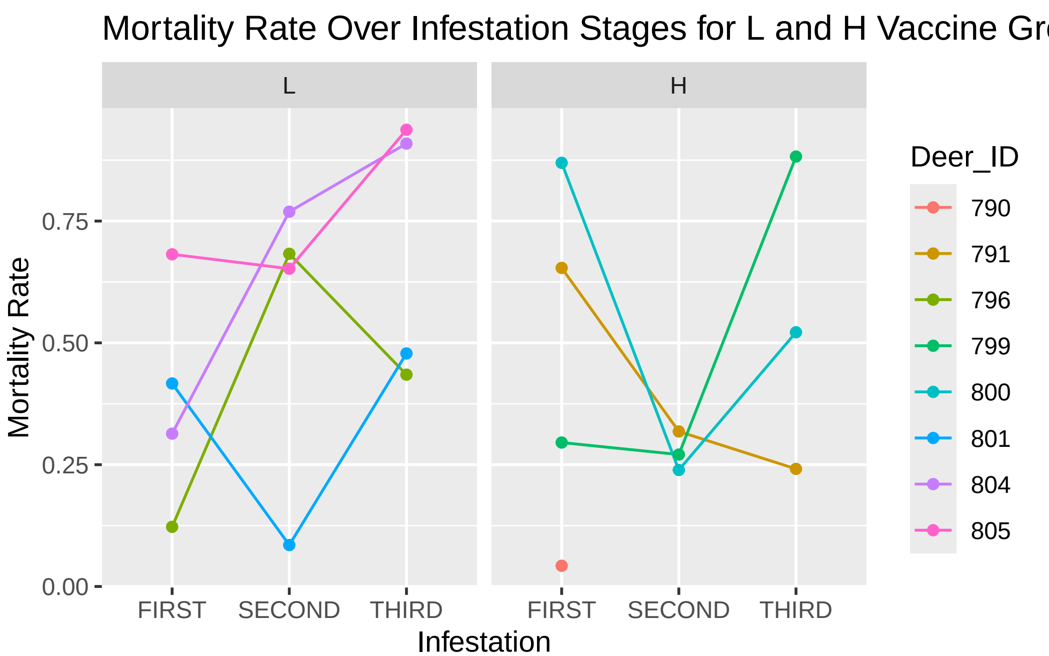
As our client has stated "In order for a vaccine to be effective it must either 1) kill feeding ticks or 2) have significant negative effects on the reproductive parameters in ticks."

So the ideal vaccine would kill the tick before it lays eggs, and if the tick is able to lay eggs we hope the quality will be low.

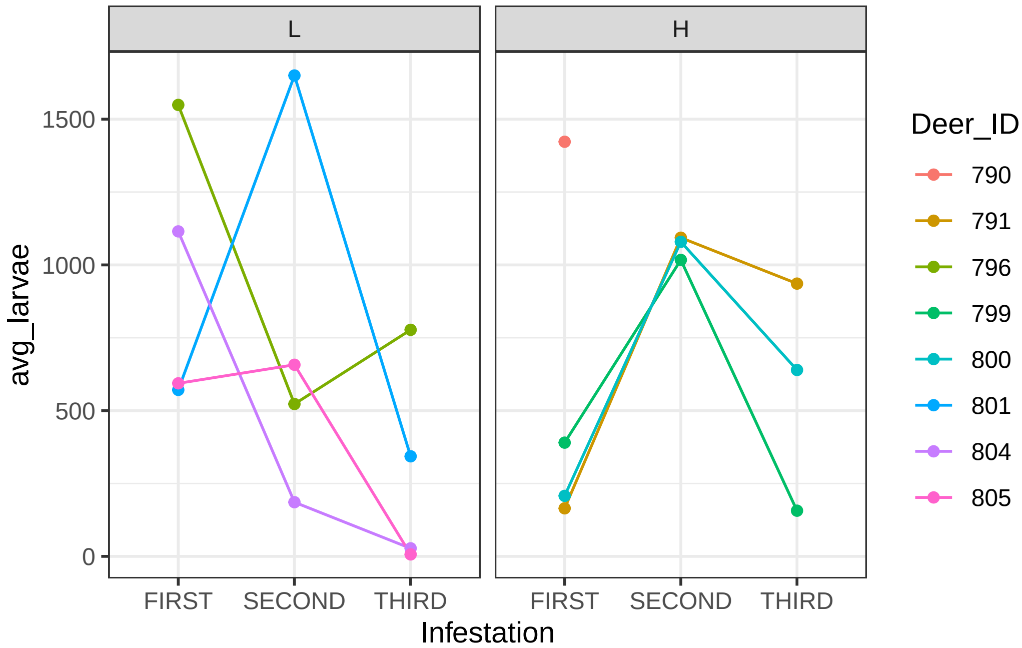
In line with that logic, I first model the mortality of the ticks to assess each vaccine groups effectiveness in killing ticks with all the data. I then model the number of larvae laid by each tick to assess each vaccine groups effectiveness in limiting tick fertility using data on ticks that live long enough to lay eggs.

3. How to examine the time effect

Our client has also stated the effect that time has is of interest. Since we do not know details on the spacing of time and whether the order of examination was randomized, I decided for simplicity to treat the stage of infestation as a fixed effect.



For the low vaccine dose, it appears that ticks die at a higher rate as time passes. While for a high vaccine dose it seems mixed.



In the low vaccination dose group it looks like the vaccine becomes more effective over time at limiting larvae births, in the high vaccination group results are more mixed but it appears the vaccine loses effectiveness.

It does appear there is a time effect in this data. And there also appears to be an interaction between the vaccine group and the infestation time.

Mortality Model

The logistic model I use is:

$$y_i = w_i + v_i + i_i + v_i i_i + e_i$$

Where y_i is the mortality result for tick i, w_i is the weight of tick i, v_i is the vaccine group of tick i, i_i is the infestation stage of tick i, $v_i i_i$ is the interaction between the infestation stage and vaccine group, and e_i is the residual.

We can see the coefficients of the model below:

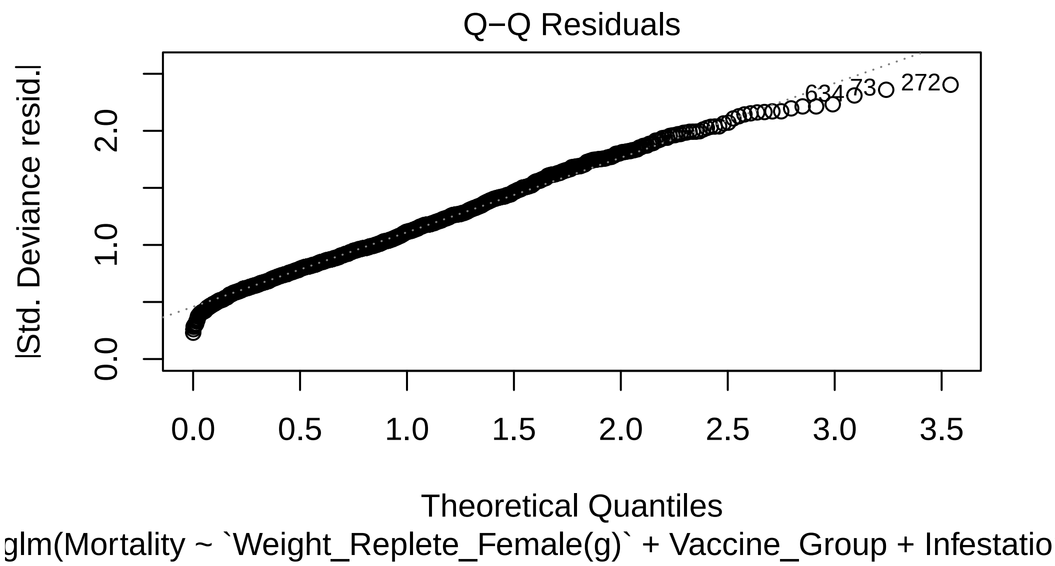
```
Call:
glm(formula = Mortality ~ `Weight_Replete_Female(g)` + Vaccine_Group +
```

```
    Infestation + Vaccine_Group * Infestation, family = "binomial",
    data = data_no_pre)

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)    
(Intercept)                        1.9025     0.2638   7.211 5.54e-13 ***
`Weight_Replete_Female(g)`       -13.9354     1.0227 -13.627  < 2e-16 ***
Vaccine_GroupL                     0.3597     0.2691   1.336   0.1814    
Vaccine_GroupH                     0.5460     0.2662   2.051   0.0403 *  
InfestationSECOND                  0.2080     0.2814   0.739   0.4599    
InfestationTHIRD                  -0.5856     0.3009  -1.946   0.0517 .  
Vaccine_GroupL:InfestationSECOND   0.7557     0.3708   2.038   0.0415 *  
Vaccine_GroupH:InfestationSECOND  -0.8342     0.3904  -2.137   0.0326 *  
Vaccine_GroupL:InfestationTHIRD    1.0184     0.4142   2.459   0.0139 *  
Vaccine_GroupH:InfestationTHIRD    0.4042     0.4099   0.986   0.3241    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1695.5  on 1257  degrees of freedom
Residual deviance: 1394.8  on 1248  degrees of freedom
AIC: 1414.8

Number of Fisher Scoring iterations: 4
```

We see the interaction between vaccine group and infestation stage is significant. So we must discuss contrasts at each different stage.

As a note, the normality of residuals look appropriate indicating a good fit.

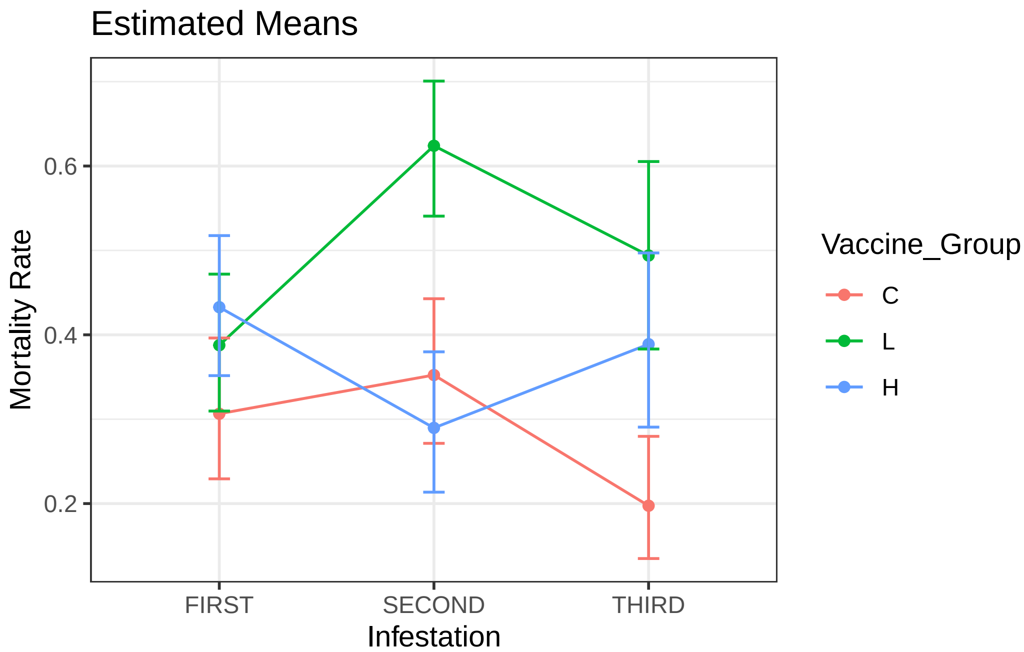
```r
plot(model_mortality_1,which = 2)
```



Contrasts and Model Results

Time Effect

The model produces the following estimated means, which plotted in this way can help us investigate the time effect on the mortality of ticks.



Each point represents a treatment combination and the error bar is the 95% confidence interval for each mean.

We can observe that it appears that low vaccination dose is the most effective dose at killing ticks, especially at the later two infestation levels where it improves performance.

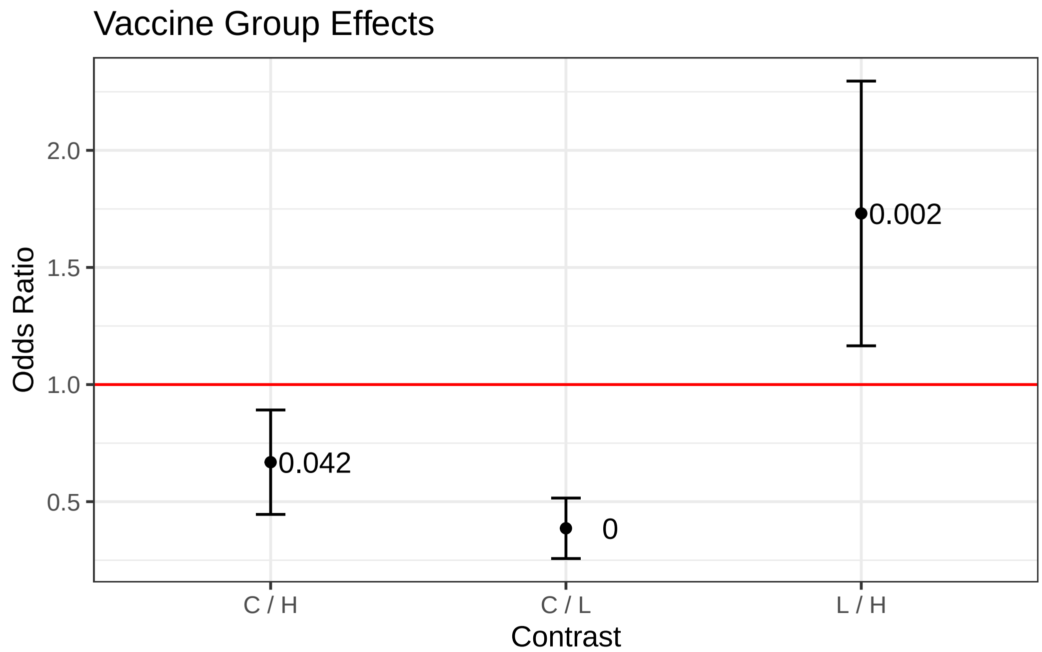
The control group seems to struggle to fight off ticks, which may indicate the immune system of the deer deteriorating as it deals with more ticks.

The high vaccine dose is relatively steady over time, performing between the control group and the low vaccination dose group.

Overall, the takeaway from this plot is that as time passes the low vaccination dose performs better, peaking at the second stage with a 60% mortality rate, while the control group gets worse at killing ticks dropping to a low of a 20% mortality rate in the final stage. The high vaccine dose does not appear to have a strong time interaction, and kills ticks around 40% of the time at all stages.

Which vaccine performs best?

```
NOTE: Results may be misleading due to involvement in interactions
```

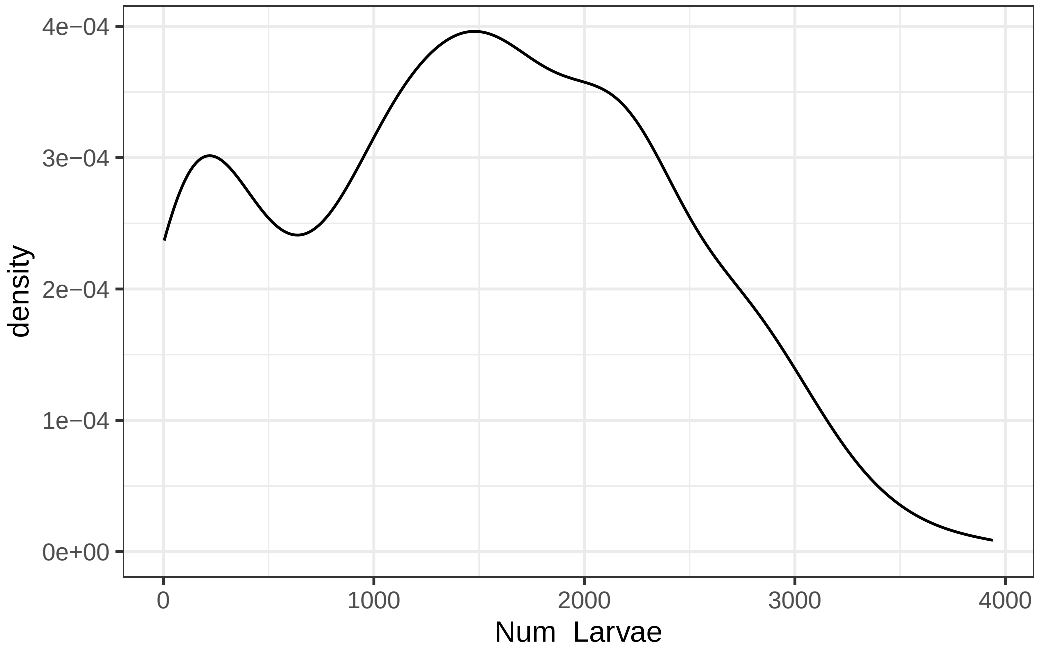


We must answer the question of which vaccination group performs the best. This plot shows pairwise comparisons in the odds ratio between the vaccination groups. The standout is that the low vaccination dose has an odds ratio of about 1.75 when compared to the high vaccination group. In other words, we are 95% confident that the low vaccination dose group kills ticks 75% more than the high vaccination dose group. This makes it the clear favorite when it comes to killing ticks.

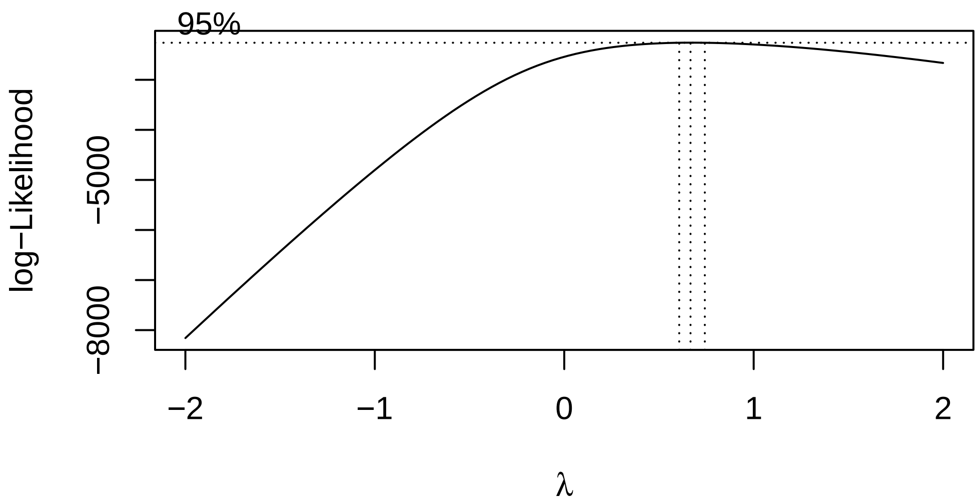
Larvae Model

The first goal of vaccination is for to ticks to die. If ticks do not die, the hope is that vaccination will limit the number of larvae that are produced by the ticks.

The first concern from a modeling perspective is the distribution of the number of larvae.

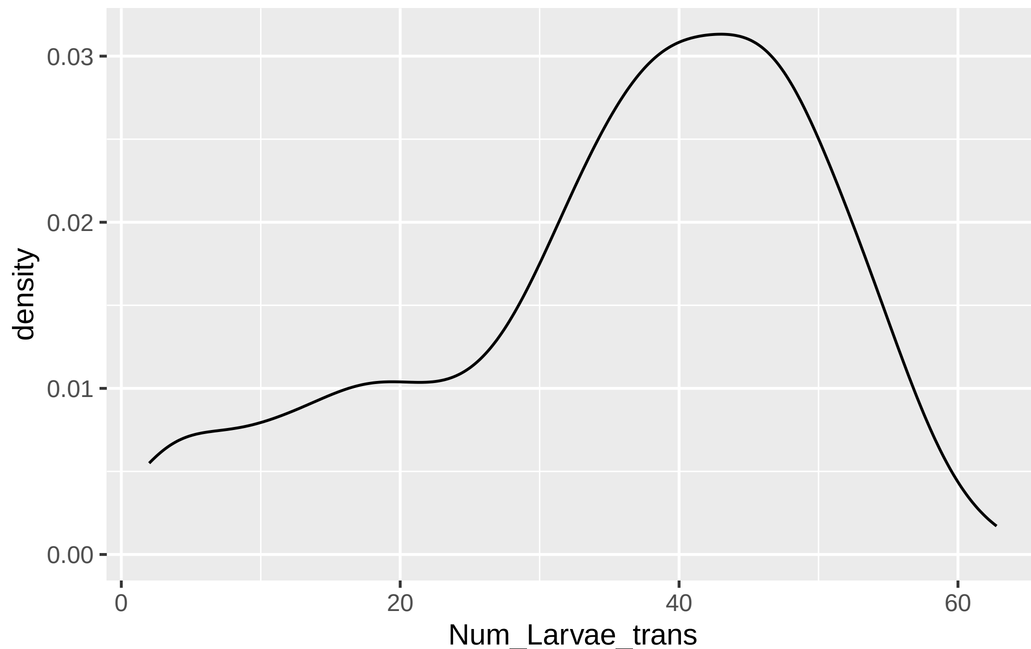


We can see a very long left tail, which motivates me to use a transformation. I opted for a box-cox:




```
[1] 0.6666667
```

With a $\lambda = .667$ I used a square root transformation.



This looks a lot better, but the left tail is still long.

However, due to the fact n is large we can lean on the CLT to help us with multiple comparisons.

I use a linear model of the form:

$$\sqrt{y_i} = w_i + v_i + i_i + v_i i_i + e_i$$

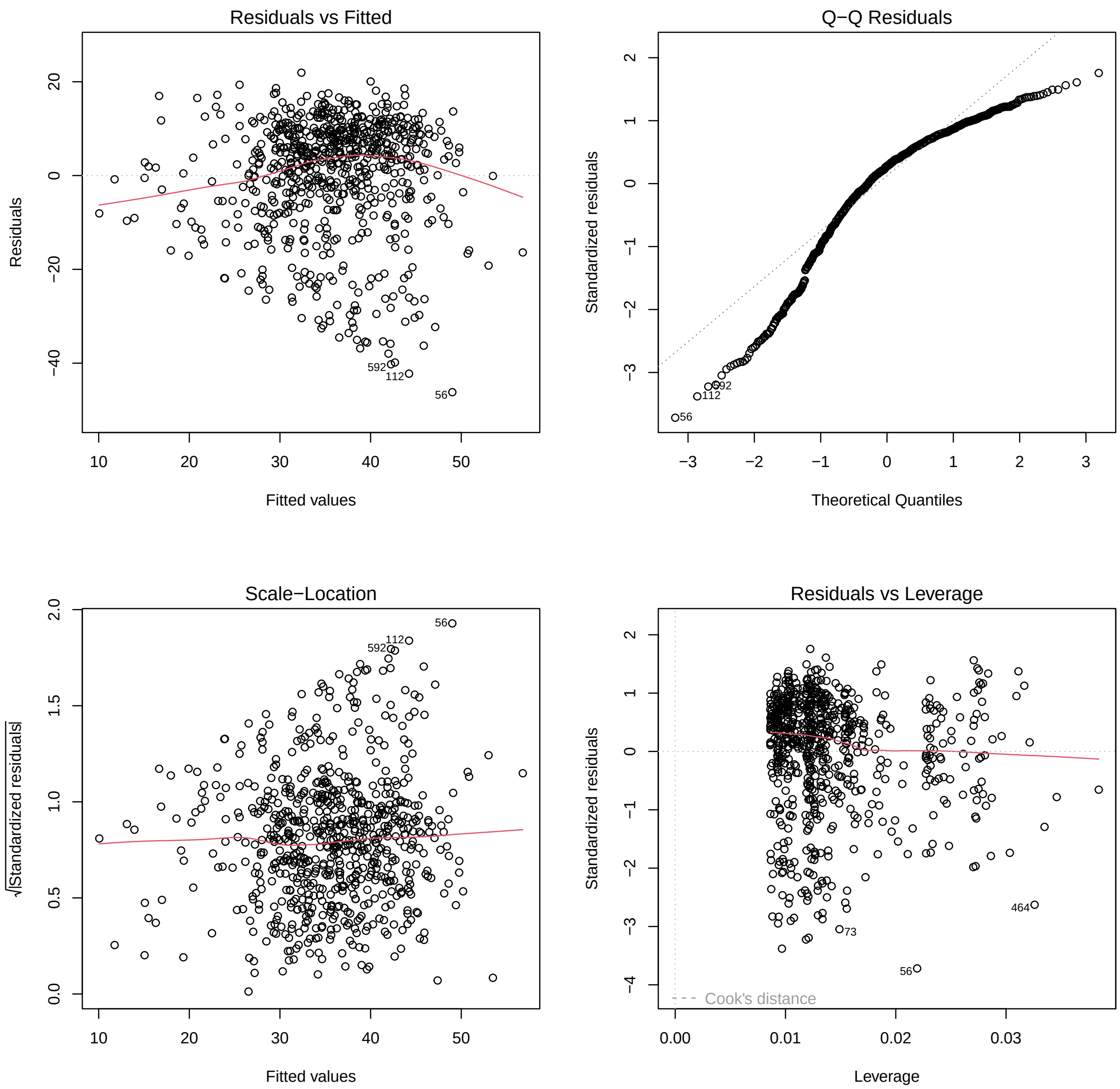
Where $\sqrt{y_i}$ is the number of larvae laid for tick i, w_i is the weight of tick i, v_i is the vaccine group of tick i, i_i is the infestation stage of tick i, $v_i i_i$ is the interaction between the infestation stage and vaccine group, and e_i is the residual.

It is also important to note, as previous discussed this model is only considering larvae that live long enough to give birth.

```
Call:
lm(formula = sqrt(Num_Larvae) ~ `Weight_Replete_Female(g)` +
    Vaccine_Group + Infestation + Vaccine_Group * Infestation,
    data = data_no_death)
```

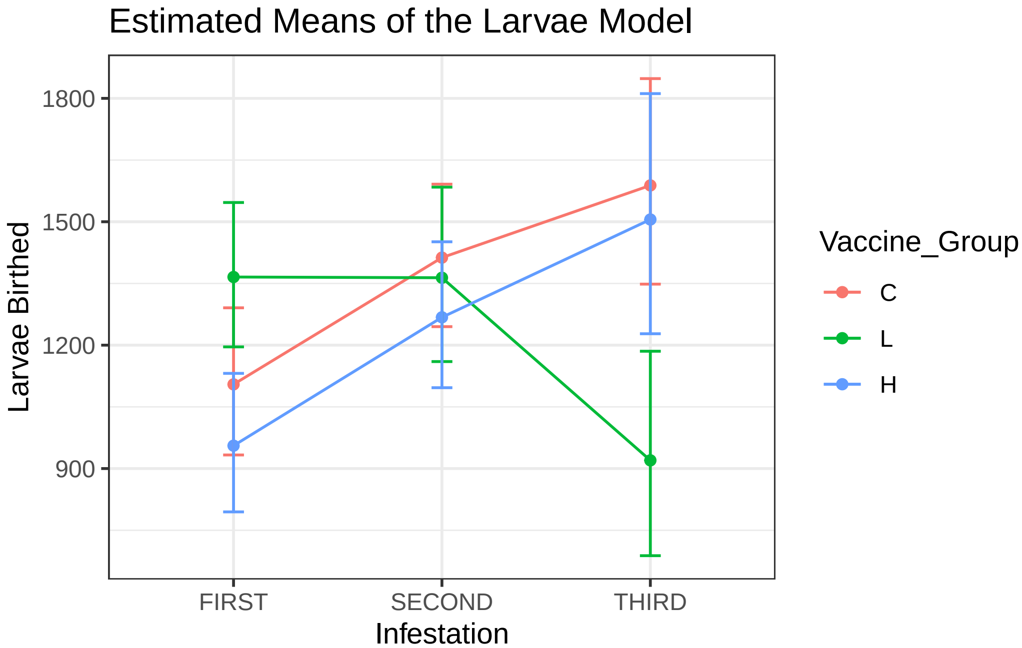
```
Residuals:
    Min      1Q  Median      3Q     Max 
-46.187  -5.840   3.419   8.936  21.931 

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)                        9.4446     2.3861   3.958 8.32e-05 ***
`Weight_Replete_Female(g)`       104.6575     8.4933  12.322  < 2e-16 ***
Vaccine_GroupL                     3.7184     1.8271   2.035  0.04221 *  
Vaccine_GroupH                    -2.3236     1.9494  -1.192  0.23370    
InfestationSECOND                  4.3512     1.8024   2.414  0.01603 *  
InfestationTHIRD                   6.6175     2.1076   3.140  0.00176 ** 
Vaccine_GroupL:InfestationSECOND  -4.3751     2.6038  -1.680  0.09334 .  
Vaccine_GroupH:InfestationSECOND   0.3407     2.6040   0.131  0.89594    
Vaccine_GroupL:InfestationTHIRD  -13.2420     3.1663  -4.182 3.25e-05 ***
Vaccine_GroupH:InfestationTHIRD    1.2700     3.1325   0.405  0.68528    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.56 on 700 degrees of freedom
Multiple R-squared:  0.2264,	Adjusted R-squared:  0.2164 
F-statistic: 22.76 on 9 and 700 DF,  p-value: < 2.2e-16
```

We can see that the interaction between vaccination group and infestation remains significant.



The residuals stray for the normal distribution a bit, but as previously discussed the number of observations is quite high so I have some protection from CLT to make pairwise comparisons. The other assumptions look well met.

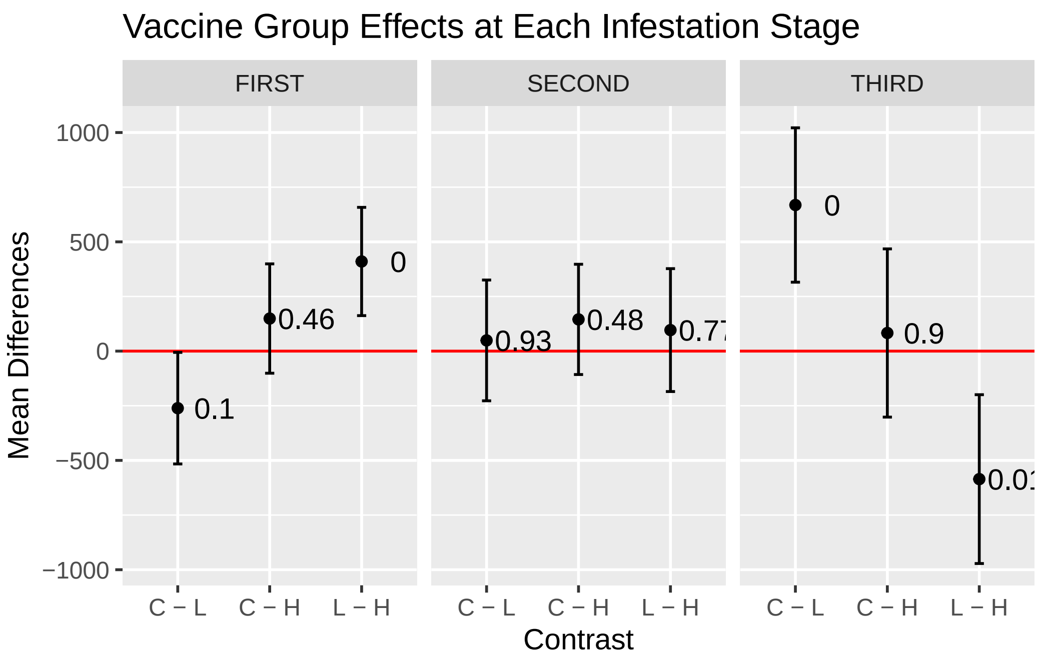
Time Effect



We can observe that the control and high dose vaccine groups both track each other and the average larvae birthed increases in the later infestation stages above 1500 larvae birthed on average.

On the other hand, the low dose vaccine group seems to perform better at limiting larvae births as the infestation stage gets later dropping to a low of 900 larvae birthed on average.

Which vaccine performs best?



This is a pairwise comparison of each vaccination group at each Infestation Group, next to each interval is the adjusted p-value of the comparison rounded to two digits.

In stage 1, the only significant comparison is between the low vaccination group and the high vaccination group. The low vaccination group has about 400 more larvae births on average.

In stage 2 no comparison is significant.

In stage three two comparisons are significant, both our previous discussion that the low dose vaccine is more effective at preventing larvae births over time. There are about 600 less births when compared to the high vaccination group and control group.

The overall conclusion for larvae births is mixed. However, since in real life you could feasibly vaccinate a deer and wait a few days before you allow them back into the wild (and that there would be a natural gap before they encounter a tick as opposed to this experiment) I would say the low vaccination dose is better at preventing larvae births.

Conclusion

The low vaccination dose performs the best out of the group. It kills the most ticks while arguably also preventing the most larvae deaths.

The time effect depends on the vaccination dose group, the control group quickly fails to kill ticks and surviving ticks lay more surviving larvae on average. The high dose vaccination group kills ticks at a steady rate over time but surviving ticks begin to lay more surviving larvae. The low dose vaccination group gets significantly better at each metric, killing ticks more effectively and preventing larvae births, as time passes.

Suggestions

A follow up study could consider tracking the deer's vitals over time, it would be interesting to see if the deer is suffering from side effects from the high vaccination dose or if its immune system is compromised later on in the experiment in the control group. These facts could help explain why the low vaccination dose is the most effective option.

If this were to be restudied, it would help if there were more consistent data collection methods. Particularly the fifty new observations on each deer was not consistently done, leading to some treatments with low observation counts.

From a data analysis perspective, it would be interesting to consider how mixed effect models would deal with within deer variance and whether it would negatively affect our ability to easily interpret results. A different way of considering the time effect could also be chosen. There could also be non-parametric methods implemented for modeling non-normal distributions that struggle to be transformed to normality.

Appendix

Please see attached files, there is a data preprocessing file and a file including the code which produced this report.