Survival Probability and Contributing Factors of Painted Turtles

Author: Kaiyan Jiang

Project Collaborator: Njal Rollison

Project Advisor: Michael Chong

04/09/2021

ABSTRACT

We researched the living status of painted turtles in Algonquin Provincial Park from 1990 to 2017. This project aims to determine the survival and capture probabilities of painted turtles and how these two probabilities change under the impact of sex, living sites, and plastron length. We used the Cormack-Jolly-Seber (CJS) model with covariates to estimate the survival probability and capture probability by controlling different covariates. We came out with these findings: residing in different sites may differ the survival chances, while the change in plastron length has a little effect on survival rate, especially for turtles with large body size. But for turtles with plastron lengths less than 9.00 cm, the increase in plastron length greatly influences their survival probability. Female turtles have a greater risk of being caught.

Table of Content

SECTION 1: INTRODUCTION

In this long-term project, we have used the data collected by Professor Njal Rollison of The Department of Ecology & Evolutionary Biology at the University of Toronto and his team. From 1978 to 2017, Professor Rollison and his team repeatedly captured painted turtles in Algonquin Provincial Park. They documented the turtles' characteristics, such as sex, site, body size (carapace length and plastron length), and so on. The data contains a total of 10,433 observations of 1,040 painted turtles.

Our primary purpose is to determine these turtles' survival probability and the relationship between survival rate and body size. In addition, we want to see how different sites where the turtle was captured influence the survival rate since we believe that turtle survival rates will vary depending on environmental conditions. Furthermore, sex may be a factor shaping the capture probability that we're looking at. Female turtles are creatures of habit, which may make a difference in how likely they are captured. As a result, our analysis will base on three major contributing variables: sex, sites, and body size.

For these questions, we chose an open-population Cormack-Jolly-Seber (CJS) model[1] to estimate the survival probability and capture probability, along with various covariates. In Section 4: Method, we will present the CJS model and analysis steps. In Section 3: Results, we will interpret our findings and their significance. Finally, in the last section: Conclusions, we will draw a conclusion to our project and address its limitations.

[1] Mu, Jiaqi, "Exploring the Estimability of Mark-Recapture Models with Individual, Time-Varying Covariates using the Scaled Logit Link Function" (2019). Electronic Thesis and Dissertation Repository. 6385. https://ir.lib.uwo.ca/etd/6385

SECTION 2: METHODS

This section will introduce the statistical model and the procedure used to analyze the survival and capture rate. We used the Cormack-Jolly-Seber (CJS) model, commonly used in ecological studies, to estimate survival and capture probabilities for Capture-Recapture (CR) data. Individual turtles can die, be born, or emigrate during our capture process, allowing the population's total number to fluctuate. As a consequence, rather than using a closed-population model, we choose an open-population CJS model.

Two CJS models were constructed in the analysis, the general CJS model and the CJS model with covariates. The general CJS is used to find the general survival probability and capture the probability of painted turtles in this environment. At the same time, the CJS model is mainly used to explore the impact of different body sizes, living sites or sex on survival and capture rates.

First, we need to make some data imputations before constructing the models as there are many missing values in our data.

Section 2.1: Data Imputation

Our turtle research has a total of 1,040 marked turtles, each with a unique notch. According to Professor Rollison, his team has had more helpers and more accurate equipment since 1990. Therefore, in this analysis, we focused on 28 years of data, from 1990 to 2017, which reduced our data from 10,433 to 10,049 observations. We concentrated on three main covariates in our model analysis: sex, site and body size. We dealt with missing values in these three factors one-by-one.

For the missing values in variable sex, I replaced them with the term “Unknown.” Some juvenile turtles are hard to identify gender, so juvenile turtles are normally denoted as “Unknown.” Hence, considering this, replace the missing term as “Unknown” is proper as they are highly likely to be juveniles. Now, we have three categories in sex that are Female, Male and Unknown. The following figure shows the total number of different sexes after the replacement.

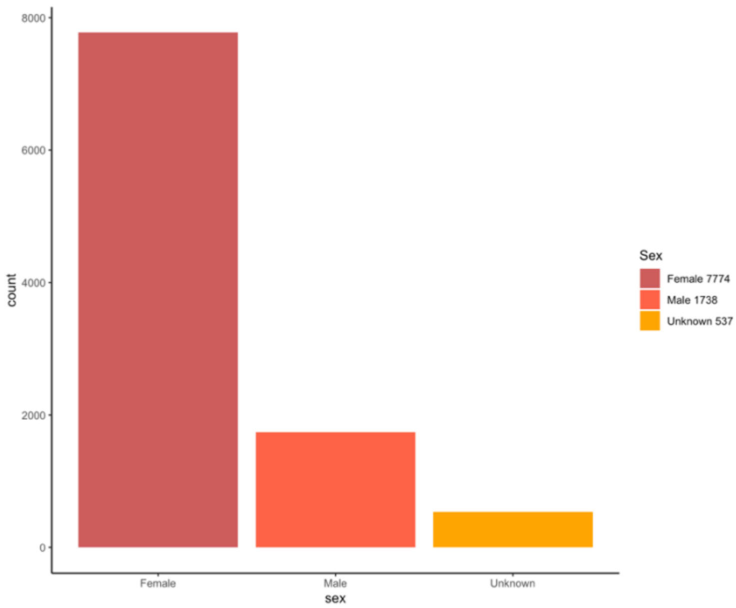


Figure 2.1: Number of Observation and Distribution of Each Sex

From the above figure, in 10,049 observations, there are 7,774 observations belongs to female turtles. Hence, we believe in the turtle population in Algonquin Provincial Park, the portion of female turtles is higher than other sexes. This observation supports that female turtles are creatures of habit. Due to the high portion of females, we think it is necessary to consider sex an important influencing factor for survival probability.

Next, I dealt with the variables for body size. In our dataset, we have the maximum carapace length and maximum plastron length. The maximum plastron length is more accurate in our research since it is difficult for a turtle to turn around when measuring the length beneath the shell. I used the last-observation-carried-forward to deal with the missing values in maximum plastron length; the last-observation-carried forward method uses the last observed value to replace the missing values in previous observations.

Last, I also used the last-observation-carried-forward method to replace the missing values in the variable site. After processing the missing values, there are still missing values, so I replace them as “Unknown” and keep them in the dataset. Keeping “Unknown” in the dataset is because there is the case that some turtles have been captured in none of the known sites. Also, I combined the site "Wolf Howl Pond E." and "Wolf Howl Pond" into "Wolf Howl Pond." Hence, we have nine categories on the site: Arowhon Rd, Sims Creek, Maiden Lake, Road to WHP, Blanding's Bog, Wolf Howl Pond, Road Wrose, West Rose, March Hare Lake and Unknown.

And there is a last step for the data imputation before start to construct the model. We need to reform the data, which can be directly used in the model-building function in the R package “marked.” The sex and site of the latest capture are kept for each turtle. Then the function pivot_wider( ) is used to deal with the time-varying variable plastron length. Each observation corresponds to a marked turtle, and the observation covers the sex, latest capture site and the maximum plastron length at each year from 1990 to 2017. However, at this step, we found a few more problems occurred. A few individuals' plastron lengths have never been recorded, and for other individuals, their plastron lengths are only recorded if they have been captured on that occasion. Hence, we still lost a few records. To overcome these problems, I replaced these new missing values still with the last-observation-carried-forward method.

Next, the two CJS models constructed will be discussed, started with the model assumptions, then to the details of constructions.

<u>Section 2.2: General CJS Model</u>

Every model has a dependent variable and an independent variable. In the CJS models, our dependent variables are always the survival rate ϕ and the capture rate p. Since the survival probability and capture probability are binary responses, a logit link will be used in the model. In the general CJS model, it does not contain any covariates. The input would only be the capture history of individual turtles. Through the general CJS model, the general survival probability and the general capture probability of the turtle population are supposed to be found. In the CJS model, we denote the capture history of individual i as $\omega_i = (\omega_{i1}, \omega_{i2}, \ldots, \omega_{iK})$, where $\omega_{it} = 1$ if the individual was captured on occasion $t = (1 \leq t \leq K)$ and 0 otherwise; n is the number of individuals captured in the study. Below is a table of random sampled 10 individual turtles' capture history from 1990 to 2017.

Notch	**Capture History**
0916	0001001001111111111000000000
1245	0000000001111111110000000000
0278B	0000000000000000111011111111
2254	0000000000000000001100000111
0803	0111111110111111111011000000
0709	1110111111111111111111111111
1275	0000000000010000000000010000
2102	0000000000000000000000000010
1531	0000000000000011010000110111
1209	0000000000100000000000000000

Table 2.1: Random Sample of Capture History

Also, we denote the survival history of individual i as $d_i = d_{i1}, d_{i2}, \cdots, d_{iK}$, where $d_{it} = 1$ if an individual i is alive and $\Sigma_{i=1}^{K} \omega_{it} > 0$, and $d_{it} = 0$ otherwise.

The general CJS model has a key assumption that all individual in the population has equal chance to being captured and is equal likely to survive in the next capture occasion. Moreover, the model also assumes that[2]:

[2] Williams, B. K., Nichols, J. D., and Conroy, M. J. (2003). Analysis and management of animal populations. 67(3):654–655.

1. Capture probability is the same for all individuals on each occasion.
2. Survival probability is the same for all individuals on each occasion.
3. Marks are neither lost nor overlooked and are recorded correctly throughout the study.
4. Sampling periods are instantaneous and recaptured individuals are released immediately.
5. Emigration from the study area is permanent.
6. Individuals are independent from each other.
7. Captures of the same individual on different occasions are independent.

However, from the capture procedure, some of the assumptions are violated. For example, the marks will lose during the process. As mentioned before, the different environments may affect turtles' capture rate and survival rate. Hence, the capture probability and survival probability of each individual turtle are not constant on each occasion. The assumption 1 & 2 may not hold in our case.

Section 2.3: CJS Model with Covariates

Compared to the basic CJS model, the CJS model with covariates has a different assumption that the survival probability is no longer assumed to be equal for each individual in each capture occasion. The change in covariates may differ the individual's survival rate.

It is natural to use a logit link in the CJS model with covariates since survival probability is binary response. The general equation for survival probability ϕ looks like below, it is similar as the generalized linear regression[3]:

$$\phi(z_{it}|\beta) = \frac{exp(\beta_0 \ + \beta_1 \, z_{it})}{1 + \ exp(\beta_0 + \beta_1 \, z_{it})}$$

Where z_{it} is the value of covariate for individual i at occasion t. β_0 and β_1 are the coefficients for baseline and covariate z_{it}, respectively. Since capture probability is also binary response, we apply the logit link on capture probability p.

$$p(z_{it}|\beta) = \frac{exp(\beta_0 \ + \beta_1 \, z_{it})}{1 + \ exp(\beta_0 + \beta_1 \, z_{it})}$$

The characters represent the same meaning as in the survival probability equation. z_{it} is the value of covariate for individual i at occasion t. β_0 and β_1 are the coefficients for baseline and covariate z_{it}, respectively.

And in the CJS model with covariates, we have two types of covariates: static and time varying. In our data, site and sex are turtles' characteristics and geographic factors, both static. And body size, maximum plastron length, is a time varying covariate, which will change as the years differ.

Based on the research questions, the survival probability may be associated with the plastron length and site. Because the environment, such as the ponds' depth, will threaten turtles' life, and the natural predators of painted turtles tend to prey on smaller turtles. Thus, the body size of the turtle varies the survival chances. The

[3] Mu, Jiaqi, "Exploring the Estimability of Mark-Recapture Models with Individual, Time-Varying Covariates using the Scaled Logit Link Function" (2019). Electronic Thesis and Dissertation Repository. 6385. https://ir.lib.uwo.ca/etd/6385

capture probability may be correlated with the static variable sex. Different sex may differ the chance being captured as some sex like females are creatures of habitat, and some are not. Therefore, in this model, we put site as the static covariate and plastron length as the time-varying covariate for survival probability ϕ. And we put sex as the static covariate for capture probability p.

In the next section, we will discuss the results of our CJS models and answer the research questions.

SECTION 3: RESULTS

Section 3.1: General CJS Model

I constructed a basic CJS model first. This general model does not contain any covariates but only contains the capture history. Here are the results of the basic CJS model:

	Survival Prob.	Standard Error	Confidence Interval
Survival Probability	0.9533	0.0020	[0.9493,0.9571]
Capture Probability	0.7978	0.0039	[0.7900,0.8054]

Table 3.1: Estimates of General Survival Probability and General Capture Probability

As we used a logit link in the CJS model, I used the predict function to inverse the logit link to obtain probabilities. From the above table, we can see that the survival probability of these turtle populations is up to 95.33%, which is very high. The survival probability has a 95% chance to be predicted in this interval [94.93%,95.71%]. While the capture probability of these populations is 79.78%, and the capture probability has a 95% chance to be predicted in the interval [78.80%, 80.54%].

Hence, from the general CJS model results, we would conclude that every painted turtle living in Algonquin Provincial Park has a survival probability of 95.33% and has a 79.78% probability of being captured, not accounting for any variation based on the turtle characteristics or external factors.

However, from the research group's experience, we know that the model assumptions of equal capture probability and equal survival probability do not held. Therefore, we doubt the truthiness of our model results. Thus, we may not rely on this result and process a further step analysis: constructing a CJS model with covariates.

Section 3.2: CJS Model with Covariates

From the previous discussion, the survival probability may be associated with the plastron length and site. The capture probability may be correlated with the static variable site. Therefore, in this model, we put site as the static covariate and plastron length as the time-varying covariate for survival probability ϕ . And we put sex as the static covariate for capture probability p. A predict function is used to inverse the logit link to compare the survival probability and capture probability with covariates.

First, I want to show the survival probability affected by sites. Thus, I filtered a few rows of different sites with a common maximum plastron length to compare the results. However, due to lack of data, we can only select the results from similar plastron length, not exactly from the same plastron lengths. Here is the figure of results and their 95% prediction interval:

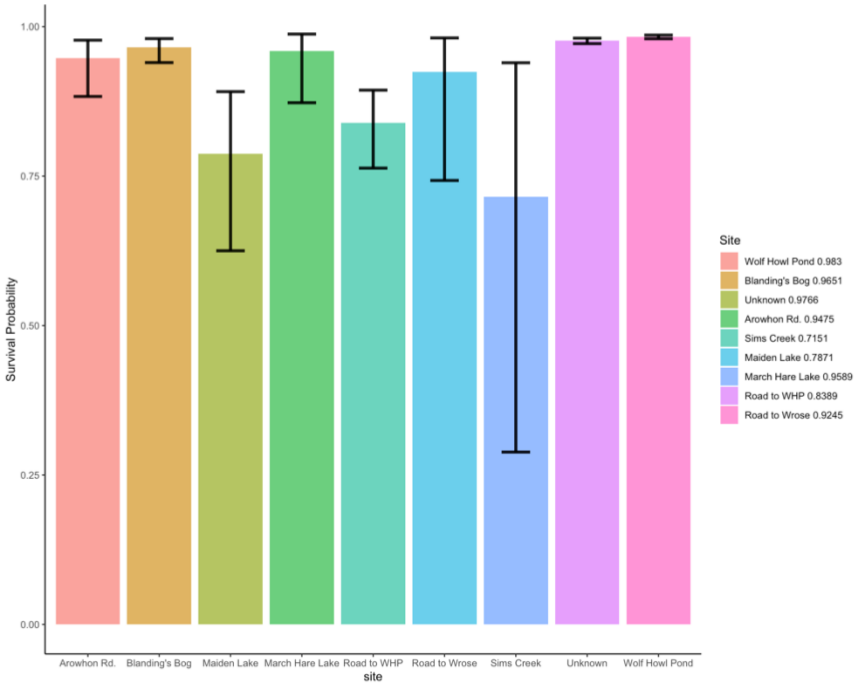


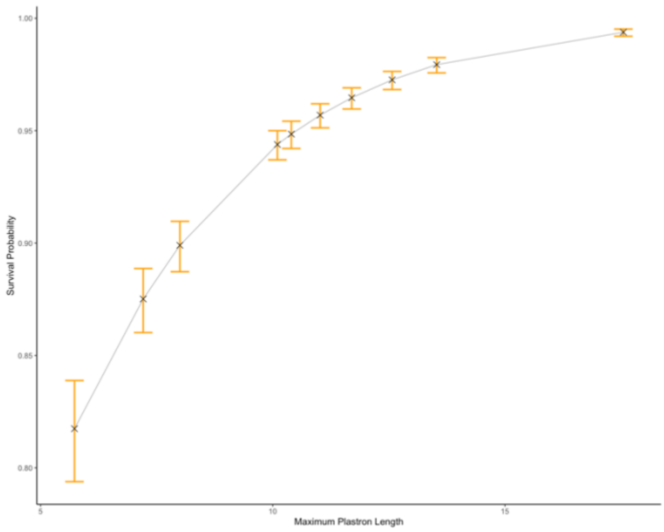
Figure 3.1: Estimates and Confidence Interval of Sites

From Figure 3.1, we can see that the survival probabilities in different sites are all surpassed 50%. It seems that turtles lived in Wolf Howl Pond have the highest survival rate of 98.30% among all sites (Data See Appendix). And we can see that prediction intervals for different sites vary a lot. Maiden Lake and Sims Creek have the largest prediction interval, which means we are uncertain about the estimated survival probability. Because we have a wider range, which our prediction might be in, the estimated survival probability we made now would be far away from the actual probability.

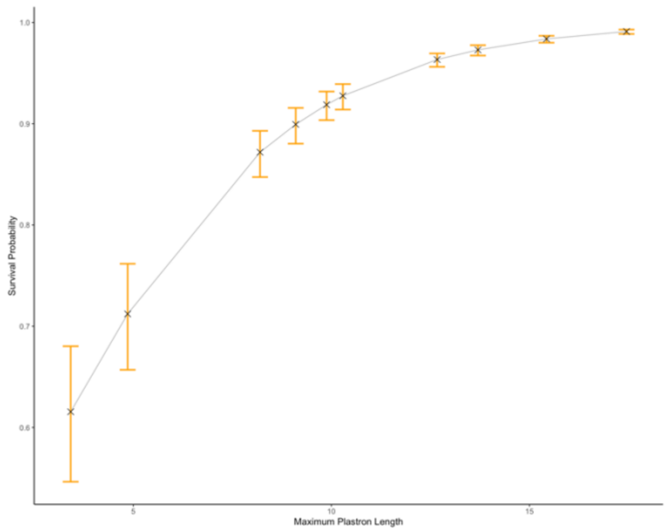
Compare this figure to the figure of the number of observations and distribution of sites in the appendix; we can find that the sites with the largest prediction interval are the sites with the fewest captures. Thus, I think that may be due to the captured sample size; for the sites with many captures, like Wolf Howl Pond, the estimated survival probabilities have a narrow variation range, so the estimates are reliable. And for the sites with few residents, the estimates are very likely to change and not reliable.

In general, we can say that there is an association between different sites and survival probability. In sites with more residents they usually have a smaller prediction interval. Those estimates are more reliable and closer to the actual survival probability, such as Wolf Howl Pond. For sites with few residents, the estimates are hard to justify as the estimate have a large interval to vary.

Next, I would like to compare how the survival probability would change with the change in plastron length. I randomly selected two samples of 10 combinations of different plastron lengths in Wolf Howl Pond and in Unknown site (Model results see appendix). The scatterplots are shown below. From both plots, there is an apparent upward trend. We may conclude that generally, the survival probability increases as plastron length increases.



(a)



(b)

Figure 3.2: Estimated Survival Probability for different Plastron Length in Wolf Howl Pond

However, it seems that when plastron length surpassed a certain value, 10.00 cm and 9.00 cm in these two samples, respectively, the 95% prediction interval would be narrower. A narrow prediction interval means we are more certain with the survival probability we predicted. And the estimated survival probability is highly reliable and may close to the true probability.

As we know, any turtle whose plastron length smaller than 9.00 cm is a juvenile turtle. Thus, from the above figure, we may conclude that the survival probability for turtles whose plastron lengths less than 9.00 cm increases dramatically with the increase of plastron length. While, for turtles with plastron lengths greater than 9.00 cm, the increase of plastron length would lead to a little increase in survival probability. For the last cross point in the graphs, the prediction interval's upper limit is very close to 100%, which means the survival probability would be high enough once the turtles' body size is large enough. It is not unusual because huge turtles will not die for many reasons, such as they are less threatened by natural predators.

Lastly, I would like to see how the capture probability change in different sex groups. From the below table, female groups have the highest capture probability, which is 81.72%. At the same time, the male groups and unknown sex groups have the capture probability of 73.76% and 65.23%, respectively.

	Capture Prob.	Standard Error	Prediction Interval
Female	0.8172	0.0042	[0.8088,0.8253]
Male	0.7376	0.0100	[0.7175,0.7568]
Unknown	0.6523	0.0379	[0.5749,0.7224]

Table 3.2: Estimated Capture Probability for Different Gender Group

Therefore, the female turtles are creatures of habit, supported by our model results since the capture probability is much higher than the rest two gender groups.

In the final section, I will conclude all results we find and discuss our analysis's limitations.

Section 4: Conclusion & Discussion

In the introduction, we have these few research questions:

- The survival probability of painted turtles in Algonquin Provincial Park
- The correlation between survival probability and turtles' body size
- Do different living sites have impact on survival probability?
- And does gender influence the painted turtles' capture rate?

Through our previous analysis, we can answer these questions one by one. First of all, we carried a general CJS model to find the general survival probability and capture probability. But the results are not reliable as many of the assumptions are violated. Therefore, we can use the survival probability predicted by the CJS model with covariates instead. When presenting these results, the covariates need to be clarified.

Secondly, after analyzing the survival probability of turtles living in the same site with different plastron lengths, we found a general upward trend between plastron length and survival probability. To be more specific, the increase of plastron length would slightly increase the survival probability for an adult turtle. But for juvenile turtles, whose plastron length is less than 9.00 cm, their survival probability increases as the plastron length increases. Hence, we can conclude that adults are more likely to survive than juveniles as they have longer plastron length.

Thirdly, from the CJS model results with covariates, we know that different sites influence turtles' survival probability. And the site with more living turtles has a more accurate estimation; its confidence interval is narrower. Estimates of sites with very few residents usually have a wider confidence interval, which means a high uncertainty with the true survival probability.

Finally, we found that female turtles have 81.72% to be captured. Compared to the

73.76% and 65.23% of male turtles and unknown sex, we can say that female turtles are not easy to change their habitats, so they have a higher chance of being captured.

In the rest of this report, I would like to discuss some of the limitations we are concerned about during the analysis. First, we face the problem of lacking data. When trying to control the maximum plastron lengths constant same for each site, we found it is difficult to find such a length. Because for some sites like Sims Creek, they only a very few records. So in the comparison, we chose similar plastron lengths. Hence, our comparison results would be inaccurate to some extent. Also, the lack of data will cause a very wide prediction interval. For Sims Creek, which only has two capture records, the 95% prediction interval of survival probability at maximum plastron length equals 14.19 cm is [0.2884,0.9396], which is an extremely large range. Therefore, in this case, we have enough evidence to doubt the validity of the estimation.

Second, in our capture process, it is hard to say we captured real juvenile turtles because the turtles we captured have already passed the most dangerous time of living. On the other hand, that means these turtles we captured have been survived for a time so that they may have a comparative high survival probability. The turtles just born are very rare to be found and captured, and due to their weakness, they face a great threat from predators. The death rate of just-born turtles is very high. Hence, our data only includes a part of the turtle population, and we have overestimated the survival probability in our analysis.

Lastly, the CJS model is widely used in survival questions. And it is a convenient model to output the survival probability and capture probability. However, it is hard to test the model's accuracy and validity. For example, we know that for some covariates, it is hard to find the actual estimates from the model results because of the wide prediction interval. For CJS models, we cannot carry out an ANOVA test or likelihood test, which are both efficient methods to testify the accuracy of the model in regression cases. As it is hard to carry out the tests, we do not know the model's validation or compare whether the predicted data fits the actual data.

Hence, in further research, a way to verify the assumptions and test the model's validation should be found. My suggestion would be to use case-control that divides the data into two groups, treatment and control. Fit a model with the treatment group and test with the control group. Moreover, further analysis of these relationships should be carried out; we should explore more possibilities, such as the interaction between sex and plastron length and how this interaction impacts survival rate. Other contributing factors of survival and capturing probabilities of painted turtles should also be researched.

Reference

1. Mu, Jiaqi, "Exploring the Estimability of Mark-Recapture Models with Individual, Time-Varying Covariates using the Scaled Logit Link Function" (2019). Electronic Thesis and Dissertation Repository. 6385. https://ir.lib.uwo.ca/etd/6385
2. Williams, B. K., Nichols, J. D., and Conroy, M. J. (2003). Analysis and management of animal populations. 67(3):654–655.

Appendix

Figure of The Observation and Distribution of Different Site

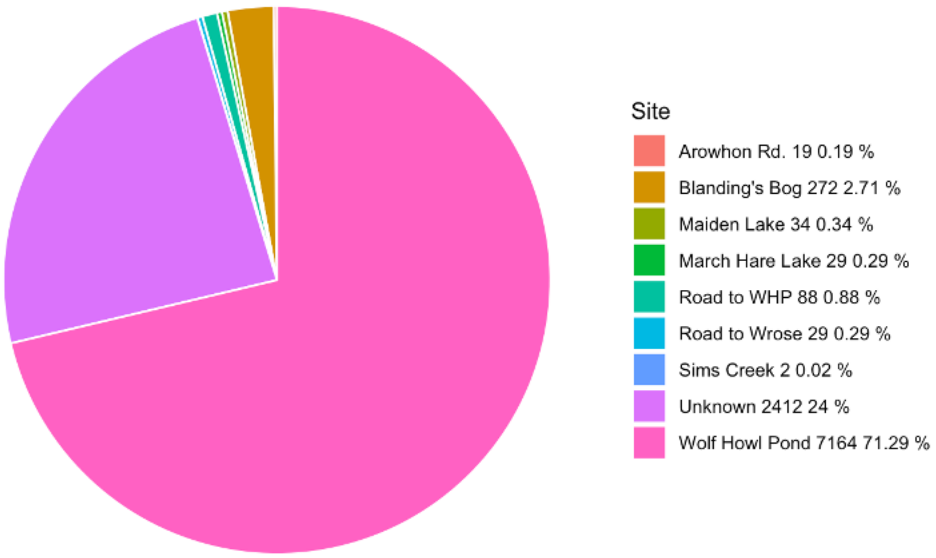


Table of Estimated Survival Probability for Different Sites

Site	Maximum Plastron Length	Estimated Survival Probability	Standard Error	95% Confidence Interval
Wolf Howl Pond	14.190	0.9830	0.0015	[0.9798,0.9858]
Road to WHP	14.170	0.8389	0.0330	[0.7633,0.8937]
Blanding’s Bog	14.190	0.9651	0.0098	[0.9399,0.9800]
Unknown	14.190	0.9766	0.0023	[0.9716,0.9807]
March Hare Lake	14.170	0.9589	0.0246	[0.8727,0.9876]
Road to Wrose	14.260	0.9245	0.0515	[0.7427,0.9811]
Maiden Lake	14.160	0.7871	0.0681	[0.6251,0.8913]
Arowhon Rd.	14.190	0.9475	0.0221	[0.8832,0.9773]
Sims Creek	14.190	0.7151	0.1896	[0.2884,0.9396]

Table of Estimated Survival Probability for Different Plastron Length (a)

Site	Maximum Plastron Length	Estimated Survival Probability	Standard Error	95% Confidence Interval
Wolf Howl Pond	5.730	0.8174	0.0115	[0.7938,0.8388]
Wolf Howl Pond	7.210	0.8751	0.0073	[0.8602,0.8887]
Wolf Howl Pond	8.000	0.8990	0.0057	[0.8872,0.9097]
Wolf Howl Pond	10.100	0.9439	0.0033	[0.9370,0.9500]
Wolf Howl Pond	10.400	0.9485	0.0031	[0.9421,0.9542]
Wolf Howl Pond	11.020	0.9569	0.0027	[0.9512,0.9620]
Wolf Howl Pond	11.700	0.9647	0.0024	[0.9696,0.9691]
Wolf Howl Pond	12.570	0.9726	0.0021	[0.9683,0.9764]
Wolf Howl Pond	13.530	0.9793	0.0017	[0.9766,0.9825]
Wolf Howl Pond	17.550	0.9938	0.0008	[0.9920,0.9952]

Table of Estimated Survival Probability for Different Plastron Length (b)

Site	Maximum Plastron Length	Estimated Survival Probability	Standard Error	95% Confidence Interval
Unknown	3.420	0.6155	0.0343	[0.5464,0.6803]
Unknown	4.858	0.7122	0.0268	[0.6570,0.7617]
Unknown	8.200	0.8719	0.0116	[0.8473,0.8930]
Unknown	9.100	0.8994	0.0090	[0.8803,0.9157]
Unknown	9.880	0.9188	0.0072	[0.9036,0.9318]
Unknown	10.290	0.9276	0.0064	[0.9140,0.9392]
Unknown	12.670	0.9634	0.0034	[0.9562,0.9695]
Unknown	13.700	0.9729	0.0026	[0.9674,0.9776]
Unknown	15.430	0.9838	0.0017	[0.9976,0.9869]
Unknown	17.450	0.9911	0.0011	[0.9887,0.9931]