

STA 440 Case Study 3

██████████ Damary ██████████

April 24, 2022

1 Introduction

1.1 Background Information

Lemurs are a family of primates indigenous to Madagascar. As the world’s oldest living primates, lemurs exhibit amazing genetic diversity. Currently there are over 100 known species of lemurs ([8]). Despite their diversity, lemurs are also the most endangered mammal on the planet.

Researchers have long been curious to investigate lemur aging as a possible model for human aging (see [5], [10], [6]). Researchers hypothesize that a variety of factors including birth weight, parental age, caloric content, litter size, hibernating status, sex, and more could be associated with better longevity outcomes in lemurs. However, gaps exist in lemur longevity research literature due to unique challenges, including: 1) the quantity of lemur species, 2) low sample size counts because of their endangered nature, and 3) a lack of aggregated data. Additionally, many studies investigating longevity only study their lemur subjects during the infant stage of life – thus conclusions about the factors analyzed cannot be extrapolated to long-term lemur survivability.

For example, [3] explores neonatal mortality (death within 30 days) and estimates that 21 percent of all Ring-Tailed Lemurs, Red Ruffed Lemurs, and Black and White Ruffed Lemurs die within the neonatal period. Benirschke and Miller conclude that these lemur species have a high incidence of neonatal birth, and find that a low birth weight is associated with higher rates of neonatal death. [9] attempt to search for the "Lansing effect" – a hypothesis among many mammals that pregnancies later in life are associated with lower longevity outcomes for the offspring – within Mouse Lemurs. The study fails to find statistically significant proof of the effect in infant mortality rates. [4] investigates hibernation in lemurs, finding that hibernating fat-tailed dwarf lemurs survive substantially longer than three other similar lemur species without the capacity to hibernate. [11] looked for patterns between litter size and infant mortality, finding instead that having a heavier mother correlates with increased odds of survival among captive Black and White Ruffed Lemurs.

We aim to contribute to the literature of lemur lifespans by analyzing factors contributing longevity across a large data set from [12] of nearly 3,000 observations comprising 27 different lemur species collected at the Duke Lemur Center. Given the high incidence of neonatal death, we choose to split our analysis into 1) investigating factors contributing to whether or not a lemur survives past 30 days (neonatal mortality) and 2) exploring factors contributing to long-term longevity in lemurs surviving longer than 30 days. In both stages, we aim to investigate the factors affecting longevity that have been hypothesized by researchers above.

1.2 Data Description

The datasets used in this analysis were originally collected by the Duke Lemur Center (DLC) and contain the information of 3,627 individual lemurs, lorises, and galagos of 27 different strepsirrhine primate taxa whose information has been verified ([12]). The ages of these individuals range from a birth year of 1960 to 2018 and represent more than 50 years of data collection and breeding. Animals included within this count are animals born at the DLC, animals owned by the DLC but captive born elsewhere, and animals "transferred into the colony as a loan, donation, or trade from another institution" ([12]). Contained within the dataset are variables that refer to each animal’s ID, species, sex, date of birth, the ages of the mother and father at birth (dam and sire), and recorded age at death (if applicable), age at the last data collection (if currently living), and last confirmed age (if it is not clear whether the animal is still living or dead)—in addition to many other variables for a total of 32 variables.

In order to minimize the effects of missing data, we elected to focus only on lemurs born at the DLC, leaving us with 2923 remaining observations. Prior to looking at the differences by species, we first noticed that when looking at a distribution of ages at death (see figure 1), there was a high incidence of newborn death. In fact, of the 2,022

lemurs with known death ages at the Duke Lemur Center, 25.8% died on the day they were born and another 12.1% died within the first month. Only 62.1% of all lemurs survived past their first birthday.

If our main concern is longevity, then the inclusion of these stillborns in our model will likely skew our results. As such, our general model will only take into account lemurs who lived past 30 days (the neonatal period) and a secondary model will explore whether the same factors that contribute to longevity after 30 days also contribute to surviving past 30 days.

Of the variables contained within this dataset, we identified the following as variables to include in our model according to the literature: litter size, sex, the age of the female and male parents, and the previous number of offspring each dam has had. For some preliminary exploratory data analysis, see the Appendix (specifically figures 3–9).

Since the purpose of this analysis is to identify whether there are systematic differences in longevity among and between lemurs, our exploratory data analysis focuses on differences in lifespans stratified by species. We focused on stratification by species since it seemed like the most natural way to group the lemurs; differences in lifespan and litter size are likely to be best explained by species and not necessarily the environment since all lemurs in our analysis are captive born at the DLC.

Looking only at the animals that survived past 30 days and have a confirmed age at death, we see general trends that point to differences based on species.

In figure 10, we see that for the nine species with the most observations, the litter size and the maximum age reached for each size of litter varies greatly depending on the species. Some lemur species only have litter sizes of one for example, while other lemur species have litter sizes of up to 4 and there is not much variation in the age reached between the different litter sizes. In contrast, a dissimilar trend can be seen in the differences in sex and longevity. Generally, the distribution of ages at death for each species does not seem to vary in any significant way by sex. Furthermore, the differences in the distribution of death age based on the age of the dam and species are very apparent. Some ages are more clustered such as in the species mohol bushbaby, gray mouse lemur, and northern greater galago, while the distribution of ages for other species are more variable and spread out (eulemur hybrid, red ruffed lemur, and black-and-white ruffed lemur). Similar statements can be made regarding sire ages at conception. The variability seen in some variables but not others lends credence to a mixed-effects model with species as our group variable.

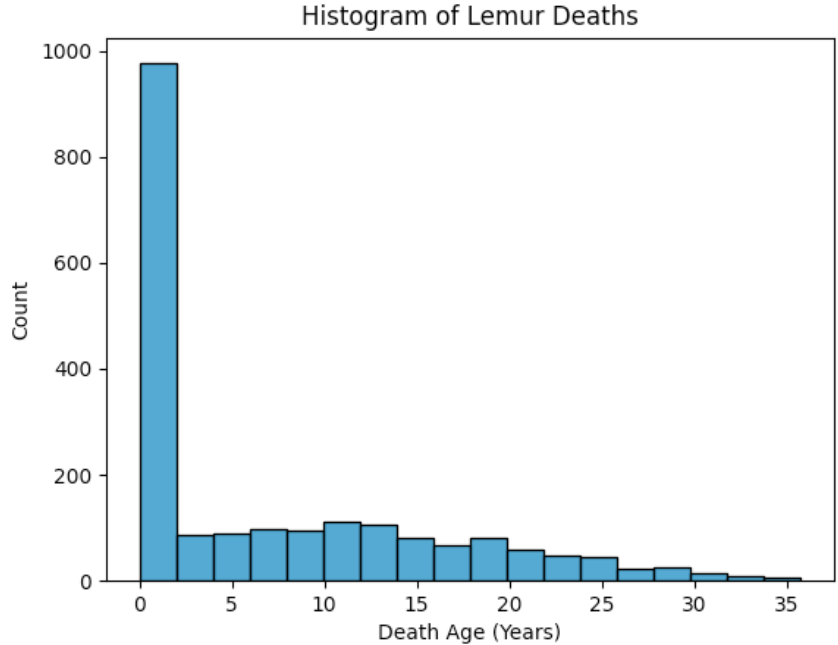
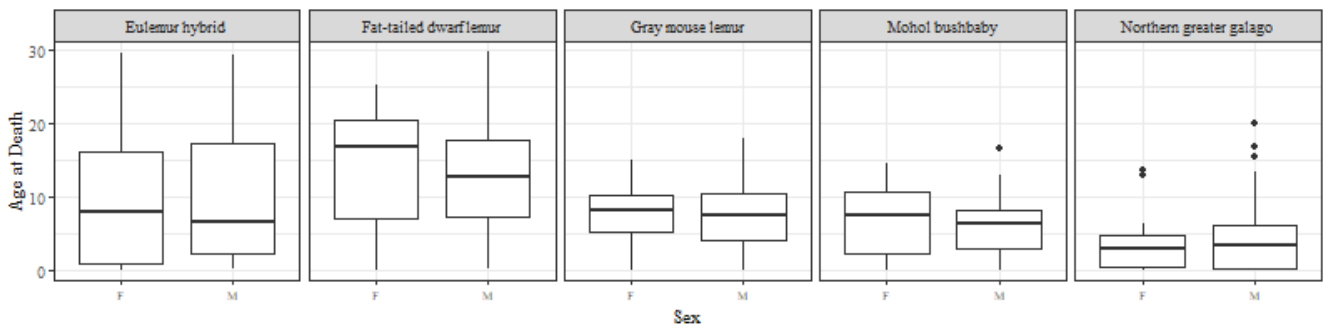
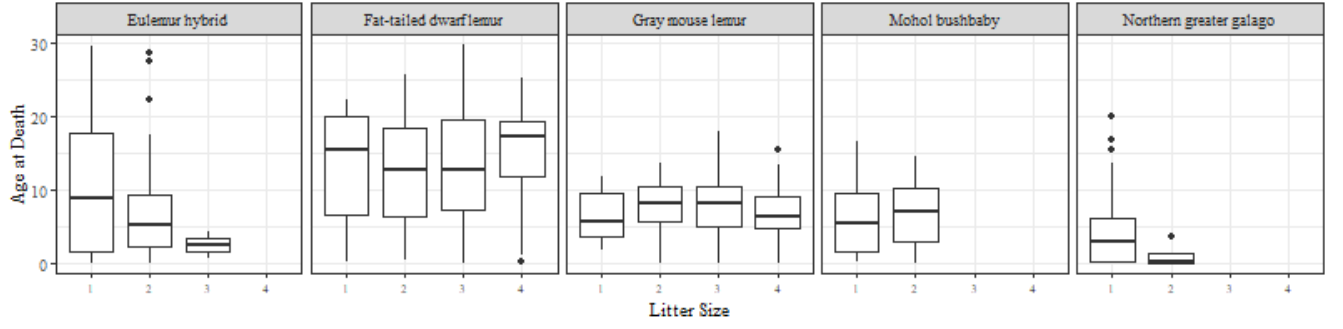


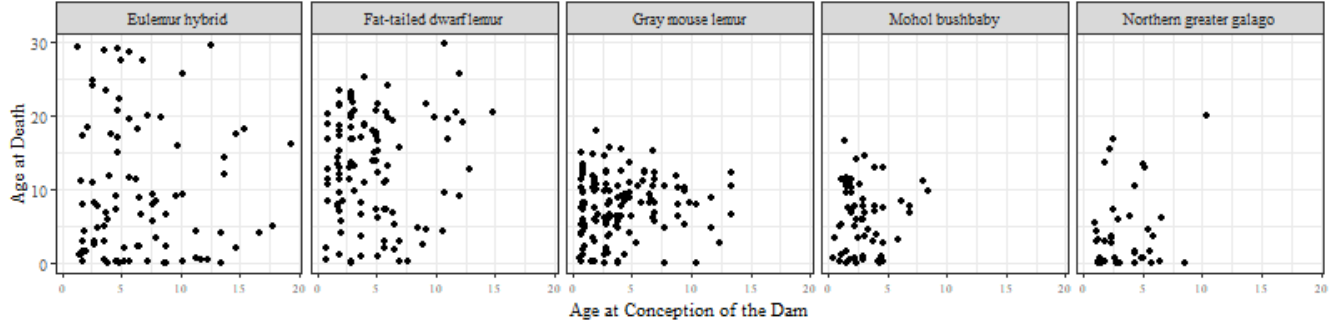
Figure 1: Histogram of death ages among 2,022 lemurs with known death ages at the Duke Lemur Center.



(a)



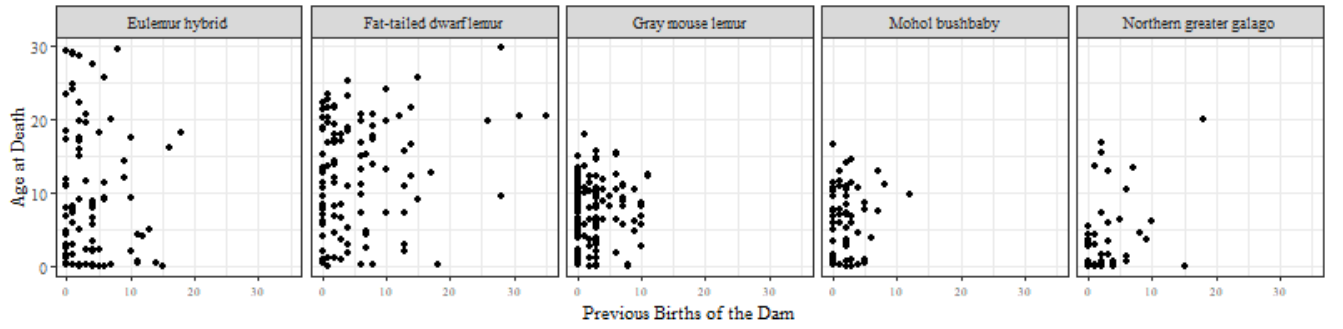
(b)



(c)



(d)



(e)

Figure 2: Distribution of age at death (in years) based on (a) sex, (b) litter size, (c) age of the dam, (d) age of the sire, and (e) previous dam births for lemurs who survived past 30 days and have a confirmed age at death, faceted by the five species with the most observations.

1.3 Variables

1.3.1 Initial variable selection

Of the original 32 variables contained within the dataset, we identified 7 predictor variables that are potentially correlated with longevity of lemurs, through EDA and literature review. They are: sex (Sex), age of mother at conception (Mother age), age of father at conception (Father Age), litter size (Litter size), the type of birth (Birth type), birth institution (Birth institution) and species (Taxon).

From the literature review, we also derived the idea of creating a new variable: number of the mother’s previous births (Prior births). This is derived from counting the number of the births from the same mother (Dam ID) that’s born before the lemur’s birthday (Estimated day of birth).

The response variable is the maximum recorded age of lemur (Age max) in years. We also observed a strong near-zero inflation in the response variable, which is why we divided our approach into 2 models, one investigating neonatal death, and the other investigating long-term longevity. According to previous literature, the neonatal period is typically up to 30 days. Therefore, based on this threshold, we separated the dataset into 2 datasets for the 2 models respectively.

1.3.2 Missing variable treatments

One of the first things we noticed about the data is that it has many missing data values (as seen in table 1). Through our EDA, we found that the observations with the highest missingness tend to be observations collected outside of Duke Lemur Center (DLC). Since the majority of the data points come from DLC (2931 observations), we decided to discard observations outside of DLC. In this way, we avoid having to impute missing data with insufficient information about the lemurs outside of DLC. This may make our model more accurate, since we have limited and incomplete observations for each institution outside of Duke, and there might be geographic differences that we do not have access to across various institutions. However, one drawback is the narrowing scope of our model and hence its interpretation. Overall, it is better to have a more informative model than a more generalizable model given the vast biological and environmental variability of lemurs living in different institutions and in the wild. As a result of focusing our model on lemurs in the DLC, our analysis is limited to captive-born lemurs (“CB” in Birth.Type).

After filtering the birth institution, 3 of the remaining variables - ‘Maternal age’, ‘Father age’, ‘Litter size’ - have missing data. The variable, ‘Sex’, does not have missing data, but has a category representing lemurs for which the sex was “not determined”. Each of them are related to some other variable, which means they are not missing completely at random (MCAR), but rather, either missing at random (MAR) or not missing at random (NMAR). For example, ‘Father age’ is strongly correlated with ‘Father ID’ and the missingness of ‘Mother Age’. It could also be missing depending on the variable itself.

Since the percentage of missing data is less than 16% in each of the above variables, we decided to impute these missing data. Our initial approach was to use kNN imputation, which uses clustering to predict the likely value of the missing data ([2]). This is intuitive because biologically similar lemurs may have similar reproductive characteristics, and vice versa. For example, lemurs of the same species would have similar characteristics. This approach also doesn’t assume the underlying relationship between the variables. However, one drawback of kNN imputation is that it is not suitable for handling categorical variables, which we have a lot of.

Our final imputation approach uses the multivariate imputation by chained equations (MICE) technique. MICE assumes that the missing data is missing at random, which means that whether the data point is missing depends on observed values ([1]). Since we don’t know if the missing variables are missing at random or not missing at random and there is no clear way of imputing not missing at random variables, we decided to proceed with MICE imputation. One advantage of the MICE algorithm is that it can account for categorical variables, which makes it more suitable than kNN imputation in our case. One disadvantage is that by choosing an imputation method, we are assuming the underlying relationship between the missing variable and other predictor variables. However, using the ‘mice’ package in R, we are able to customize the imputation methods for each variable, which gives us flexibility between numerical and categorical variables ([7]). Therefore, we used a mixture of Bayesian linear regression and classification methods in our imputed dataset. Each missing variable is imputed using the rest of

	Missing
DLC ID	0
Taxon	0
Sex	0
Mother age	10
Father age	349
Litter size	10
Previous births	6
Maximum age recorded	0
Censored lemurs	0
Prior births	6

Table 1: The number of missing variables per variable.

the available (not missing) variables in the dataset, excluding the response-related variables, such as "Maximum Age Recorded", "Age at Death", "Age at Living" and "Date of Death".

1.4 Hypothesis

In our literature review, there are a sizable number of studies investigating the factors associated with neonatal deaths. However, there were not many studies on post-neonatal longevity. Combining previous literature with our EDA, we hypothesize that the factors contributing to lemurs' longevity past the neonatal period is similar to those associated with neonatal survival. Namely, we hypothesize that parental age, litter size and sex would have statistically significant associations with both neonatal survival and longevity of lemurs. For some of these factors, such as parental age, the correlation is consistent across various species of lemurs; yet for others, such as sex and litter size, the correlation may vary by a considerable amount across different species of lemurs.

2 Methodology

The goal of our analysis is to identify whether there are systematic differences in longevity among and between lemurs. In order to avoid many of the pitfalls that result out of using data with many missing values, we elected to focus our analysis on animals from the Duke Lemur Center which are all captive-born animals. Due to the abundance of stillborn lemurs and the lack of literature on post-neonatal survival studies, we built 2 models for neonatal and post-neonatal analysis. Both of these models use data imputed through the MICE technique; for a more detailed description of the method (including which variables were used in the imputation), see the missing data section in the Appendix.

2.1 Neonatal Mortality Model

In the neonatal mortality model, we use logistic regression model, because the outcome is binary - we only care about whether the lemur survive past 30 days. We noticed in our EDA that the relationship between litter size and survival varies across different species, but that certain species have few to no observations for all possible litter sizes (which is understandable, as lemur species do not have a uniform litter size). We realized that inferring relationships between litter size and neonatal mortality would be difficult given this lack of data, but we still wanted to account for that variation when estimating other covariates. To do this, we decided to use a random intercept for Taxon and random slope for litter size by Taxon. We do lose some interpretability, but we account for inter-species variation, thereby improving the estimates of our fixed effects. Formally, let $y_{ij} \in \{0, 1\}$ indicate whether a lemur i survived at least 30 days after birth, and let j index the observed Taxa. We specify the following logistic regression:

$$\begin{aligned} \Pr(y_{ij}=1) &= \pi_{ij} \\ \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \theta_j + \beta_0 + \beta_1 \mathbb{1}(S_i = 1) + \beta_2 \mathbb{1}(B_i = 1) + \beta_3 D_i + \beta_{4j} L_i \\ \theta_j &\stackrel{iid}{\sim} N(0, \tau^2), \quad \beta_{4j} \stackrel{iid}{\sim} N(0, \sigma^2) \end{aligned}$$

where for animal i , S_i, B_i, D_i, L_i indicate the sex (0 for female, 1 for male), whether the animal's mother had given birth before (0 for no, 1 for yes), the age of the animal's mother at conception, and the animal's litter size, respectively.

2.2 Long-term Longevity Model

In the long-term longevity model, we used Cox proportional hazard model, because we care about the probability of survival (hazard of dying) at a certain age. We selected the same variables as the neonatal mortality model under the same mixed-effects structure, because we want to do a comparison between the two periods and the effect of the factors specified below.

Let $\lambda^i(t)$ denote the rate of death of an animal i given that the animal has survived up to time t (i.e., a hazard function) and λ_0 the baseline hazard (hazard function for an animal with the baseline level of all covariates). With

the variables the same as above, for an animal i ,

$$\begin{aligned}\lambda^i(t) &= \lambda_0(t)\exp(\mu_i) \\ \mu_i &= \theta_j + \beta_1 \mathbb{1}(S_i = 1) + \beta_2 \mathbb{1}(B_i = 1) + \beta_3 D_i + \beta_{4j} L_i \\ \theta_j &\stackrel{iid}{\sim} N(0, \tau^2), \quad \beta_{4j} \stackrel{iid}{\sim} N(0, \sigma^2)\end{aligned}$$

3 Results

3.1 Neonatal Mortality

We use a logistic regression model to determine the factors associated with neonatal survival. The response variable is a binary variable indicating 1, for survival past 30 days or 0, for dying before 30 days. The results of our logistic model are displayed in Tables 2 and 3 below.

	Estimate	Std. Error	2.5%	97.5%	p-value
(Intercept)	0.942	0.125	0.698	1.186	< 0.001
Male Sex	-0.006	0.097	-0.196	0.183	0.949
Mother Age	-0.044	0.013	-0.069	-0.020	< 0.001
Mother Prior Births	0.872	0.117	0.642	1.102	< 0.001

Table 2: Main effect estimates and corresponding CIs for our logistic regression on infant mortality.

Groups	Name	Variance	Std. Dev.
Taxon	Intercept	0.063	0.251
	Litter size	0.050	0.022

Table 3: Standard deviation estimates for random effects in our logistic regression on infant mortality. We estimated a Taxon effect and an effect for litter size by Taxon.

From Table 2, we see that the estimates for our intercept, Maternal Age, and Prior Birth are significant at an $\alpha = 0.05$ level.

To interpret our intercept coefficient we say that holding all else constant (sex, mother’s age at conception, Taxon, and whether or not the model had previously given birth), our model estimates the baseline odds of living past 30 days is $e^{0.942} = 2.565$. Similarly, holding all else constant (sex, Taxon, and whether or not the mother had previously given birth), for each additional year older a lemur’s mother is, our model estimates that their odds of living longer than 30 days decreases by a multiplicative factor of $e^{-0.044} = 0.957$. Holding all else constant (sex, Taxon, and mother’s age at conception), when a lemur’s mother has previously given birth to another litter of lemurs, we estimate that their odds of surviving past 30 days increases by a multiplicative factor of $e^{0.872} = 2.392$. That is, there may be a sweet spot between being an older mother whose fertility is in decline, and a younger mother whose reproductive systems are still maturing.

Table 3 shows the variation explained by the random intercept by Taxon and the random slope of litter size by Taxon. We see relatively low variances for both random effects, suggesting that there may not be much variation in neonatal survival by species or by litter size and species.

3.2 Long-term Longevity

We use a Cox proportional hazard model to investigate whether the same factors significantly associated with neonatal survival is also associated with long-term survival. The results of our model are displayed in Tables 4 and 5 below.

In contrast to the neonatal mortality model, our long-term longevity model suggests that the sex, maternal age and material previous births of a lemur do not have statistically significant correlation with long-term longevity.

Notice in Table 5 that the variance explained by the Taxon intercept is $0.882/(0.882 + 0.171)$ is 84%, compared to 16% explained by the litter size by taxon random slope. We cannot directly interpret these standard deviations, but we can note that there appears to be substantial variation in survival likelihood by species.

	Estimate	Std. Error	2.5%	97.5%	p-value
Male Sex	-0.096	0.908	-0.228	0.035	0.153
Mother Age	0.002	0.013	-0.018	0.022	0.828
Mother Prior Births	-0.096	0.090	-0.272	0.080	0.294

Table 4: Main effect estimates and corresponding CIs for our Cox proportional hazards model on lemur longevity.

Groups	Name	Variance	Std. Dev.
Taxon	Intercept	0.882	0.777
	Litter size	0.171	0.029

Table 5: Standard deviation estimates for random effects in our Cox proportional hazards model on lemur longevity. We estimated a Taxon effect and an effect for litter size by Taxon.

4 Discussion

4.1 Conclusions

In order to investigate factors associated with longevity of lemurs in DLC, we first looked at factors associated with neonatal survival in model 1, and compared these same factors with those found significant in the long-term longevity in model 2. Overall, we found a discrepancy in that most of the factors significant in predicting short term neonatal survival were not significant for long-term longevity.

For neonatal survival, we found that older maternal age is associated with decreased odds of survival by a multiplicative factor of 0.957, holder all else constant; while maternal previous births is associated with an increased odds of survival by 2.392. This suggests a sweet spot where the mother is old enough to have some previous births, but not too old. This trend is similar to what we observe in human biology, where there is a arch of fertility in females. Behaviorally, there might be a sweet spot between younger mothers who are inexperienced in taking care baby lemurs and older mother who might be too frail to take care of their children. Lemur caregivers at research facilities could take this into account. In terms of litter size, we found little evidence to suggest that there exists substantial variation in survival probability by litter size per species, which might be a consequence of a lack of data and different supports for different species.

In terms of long-term longevity, we found that factors that are significant in the neonatal survival model - such as the sex, maternal age and material previous births of a lemur - no longer have significant associations with survival. Instead, we found that substantial variation exists in longevity by species. This might suggest that, in contrast to neonatal survival, the long-term longevity of lemurs depends not merely on physiological factors, but also on behavioral and environmental factors specific to lemur species that were not recorded in the data set. More research is needed to understand the factors contributing to post-neonatal longevity in lemurs.

4.2 Limitations and Future Directions

First, one limitation is that we were not able to include all the variables in our model that we hypothesize may be associated with longevity. For example, [3] found birth weight to be highly associated with lowering infant mortality, but our variables did not include birth weight. We tried supplementing the data set with additional data from the Duke Lemur Center measuring birth weights, however there were too many missing values for imputation and inclusion to make sense. Second, we recognize that our data comes specifically from lemurs born in the Duke Lemur Center, which hurts our ability to generalize our conclusions. [12] notes that population weights and characteristics often differ between captive and wild lemur populations. Third, it is impossible for is to know if our censored and missing data is missing at random or not missing at random – a more complete and robust data set would alleviate the need for imputation. Fourth, given some of the small sample sizes associated with few of the lemur taxon, if we had strong prior information we may have been able to use a Bayesian twist on a survival model that decreases variability in the estimates. This research in lemur longevity is not only essential so that conservationists can fight against lemur extinction and maintain populations of lemurs in captivity, but also as a key to understanding longevity in humans. Since lemurs and humans are both primates, future research could be much improved by studying certain species of lemurs instead of laboratory mice.

4.3 Summary

In this paper, we set out to investigate factors affecting lemur longevity using a dataset of lemurs living at the Duke Lemur Center. In our model investigating neonatal mortality, we found statistically significant evidence ($p < 0.05$) that the age of the mother was negatively associated with the odds of surviving past 30 days, while being born to a mother who had previously raised a litter was positively associated with neonatal survival. In contrast, our long-term survival model found that neither mother age or whether the mother had previously given birth was statistically significant.

5 Response to Peer Review

5.1 Introduction

Thank you for your comments on the literature review! While we were worried that we might run out of space if we made the graphs too large, we had space leftover and overlooked fixing the size of the plots prior to our first submission. Fixing the size of our graphs is a good suggestion, and we updated the size of the graphs in order to take your suggestion into account and improve the overall flow of the manuscript. Including raw variable names was also an oversight on our part, so we also went through our report and replaced those with more descriptive names.

5.2 Methodology

Taking your suggestion into account, we also included a sentence referring to the variables that were used to impute values for each of the missing variables. This will aid in future efforts for reproducibility. We also better clarified the models beyond just the types used.

5.3 Results

As you suggested, we replaced all p-values of 0.000 with the more accurate <0.001 and separated each of the variables by their “class”. In terms of only using the interaction effects and not including species as a main effect, this was done in order to account for the fact that litter size varies wildly by species, which did make our interpretations a bit indecipherable. As such, we made the interaction term a random effect for litter size by Taxon. While we can no longer interpret the litter size by Taxon effects anymore since they’re random, we can still interpret the other fixed effects. This also takes into “fixes” the pseudo-separation you mention.

5.4 Discussion

The inclusion of real-world ramifications is a good suggestion as well, and we have included these ramifications in the limitations and future directions section of the discussion which allows us to better explain the impact of our findings.

6 Appendix

6.1 Extra EDA

Prior research indicated that certain parent characteristics may be tied to animal longevity, so we investigated the ten parent variables given in our dataset, taking care to address any problems with missing data. First, we discarded the variables giving an animal’s mother’s (dam) and father’s (sire) name, DOB, and Taxon, as these are contained within the variables giving the dam and sire ID (DLC-assigned), the age of the dam and sire at conception, and the Taxon of the animal (the Taxon variable has special designations for hybrids). We noticed the remaining variables tended to be missing together, as shown in Figure 3.

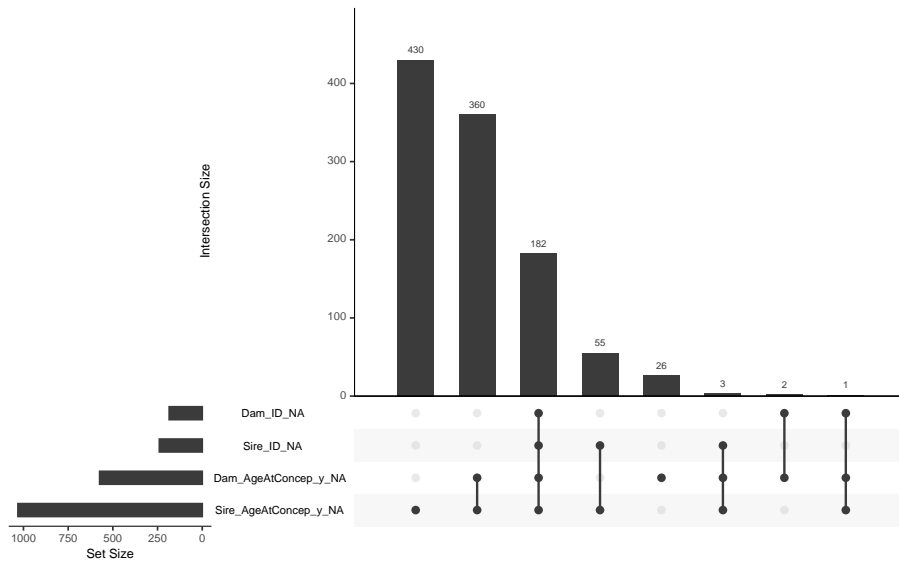


Figure 3: Missingness relationships between relevant parent variables (dam and sire DLC ID and dam/sire age at conception).

In nearly all cases, if an individual’s dam’s ID was missing, then the remaining variables were missing as well. Moreover, the missing dam variables corresponded exactly with whether or not the animal was born at the DLC (4), with less than five exceptions. For animals born at the DLC, we had (nearly) complete dam information, but sire ID and age at conception were still missing in about 450 cases (≈ 15 percent of all observations from the DLC).

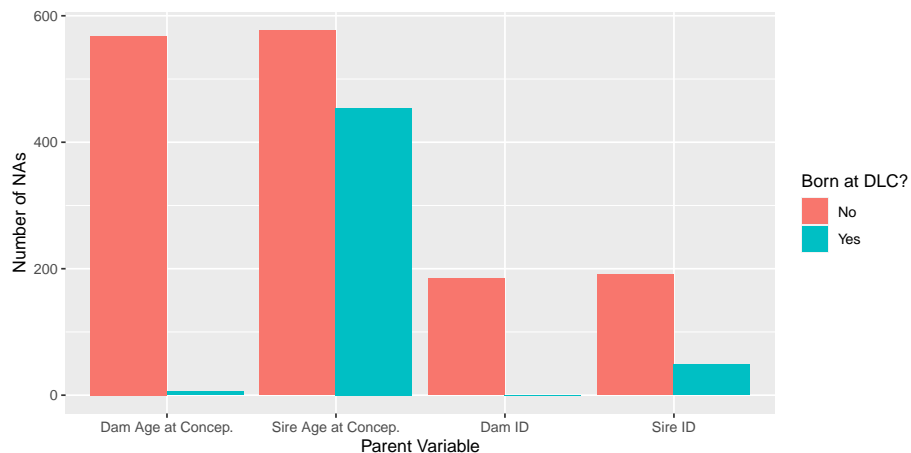


Figure 4: Missing parent variables for the cases where the animal was and was not born at the DLC.

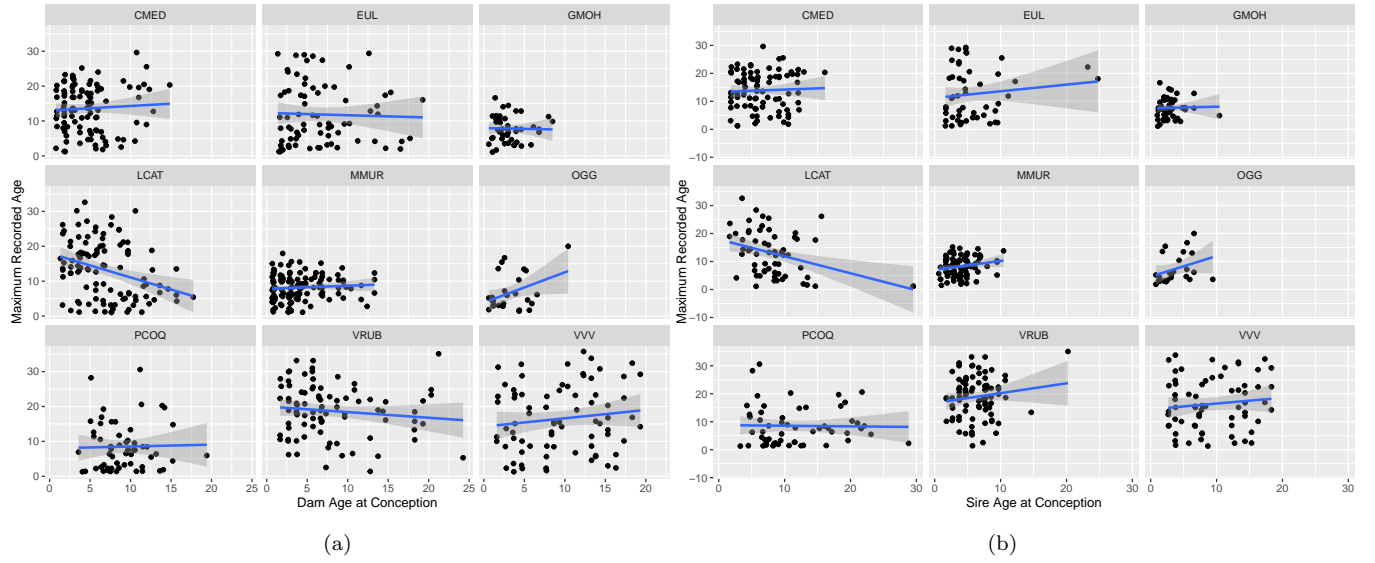


Figure 6: Dam and sire ages at conception plotted against maximum age recorded, which in this case corresponds to deaths only (animals censored or recorded as alive were ignored), for animals who survived their first year. Plots faceted by the nine Taxons with the most observations, and regression lines given in each case.

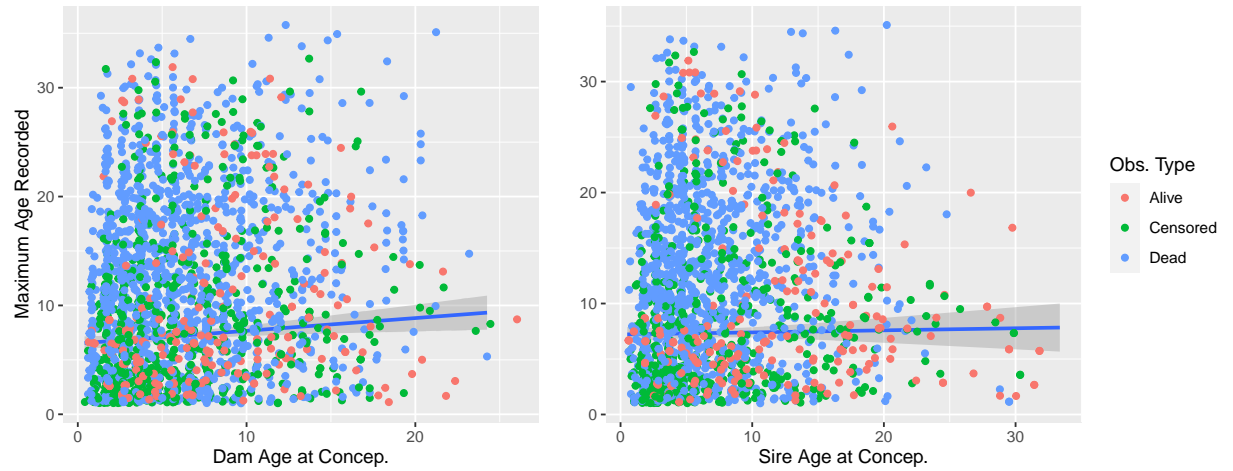


Figure 5: Dam and sire ages at conception plotted versus the maximum age recorded (which may be a censored value) for animals who survived their first year. Regression lines are given for animals recorded as “Dead” only.

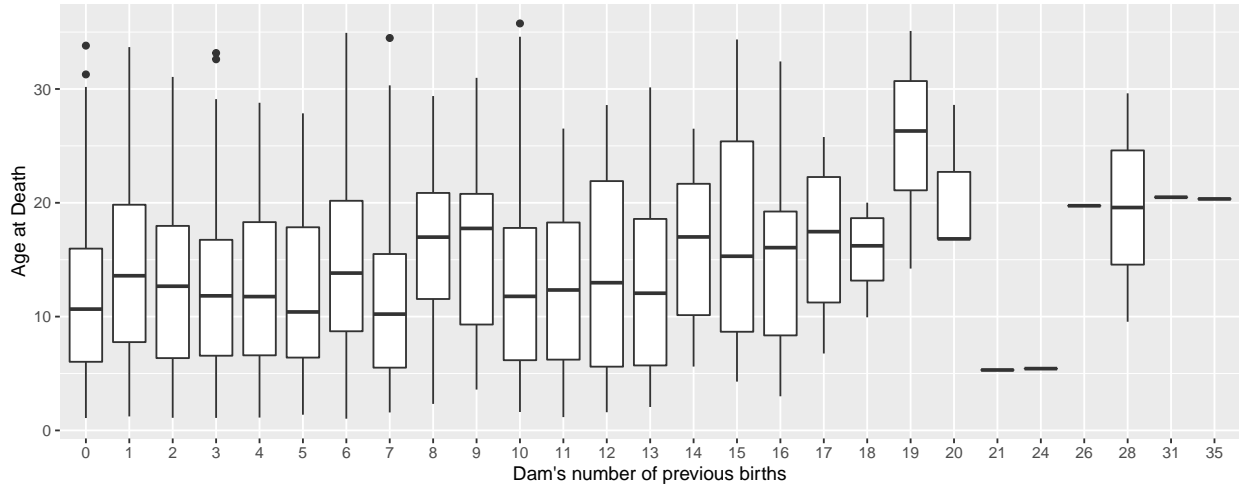


Figure 7: Distributions of recorded ages at death for individuals who survived their first year, plotted by their dam's number of previous births.

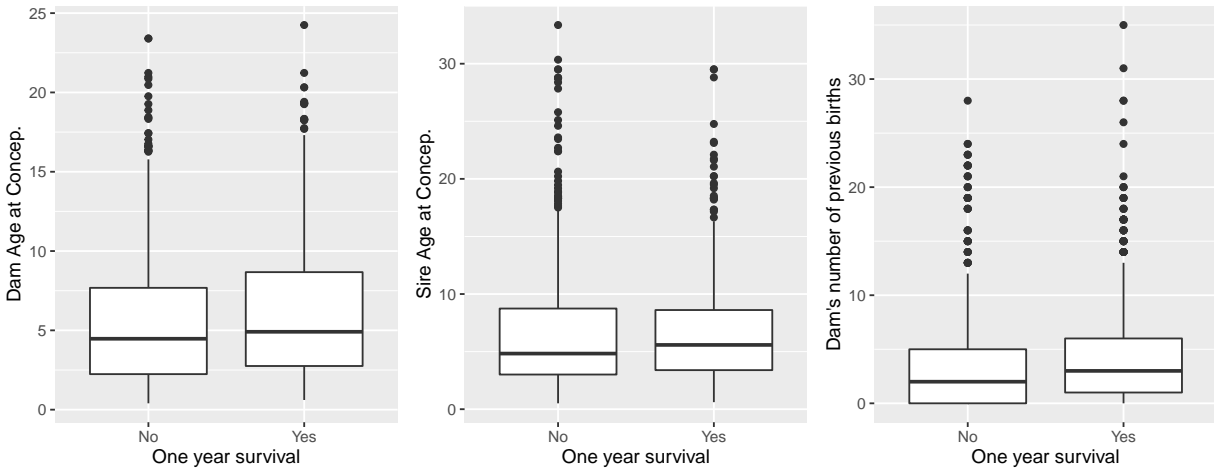


Figure 8: Distribution of Dam age at conception (left), sire age at conception (middle), and dam's number of previous births (right) for animals who did and did not survive their first year.

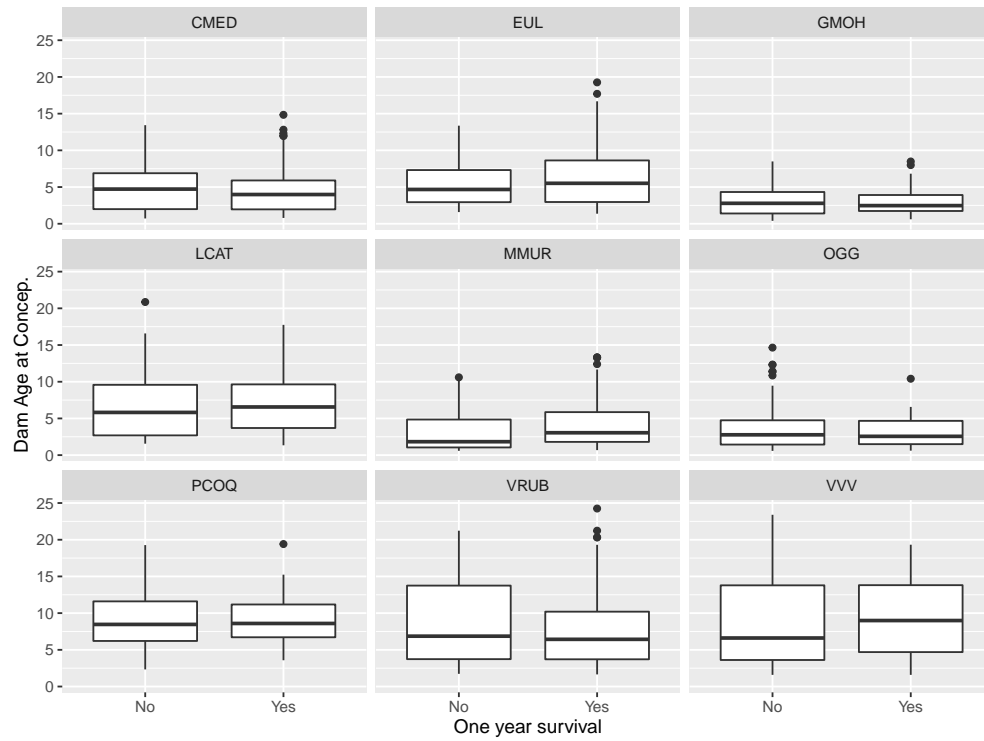


Figure 9: Distribution of dam age at conception for animals who did and did not survive their first year, faceted by the nine Taxons with the most observations.

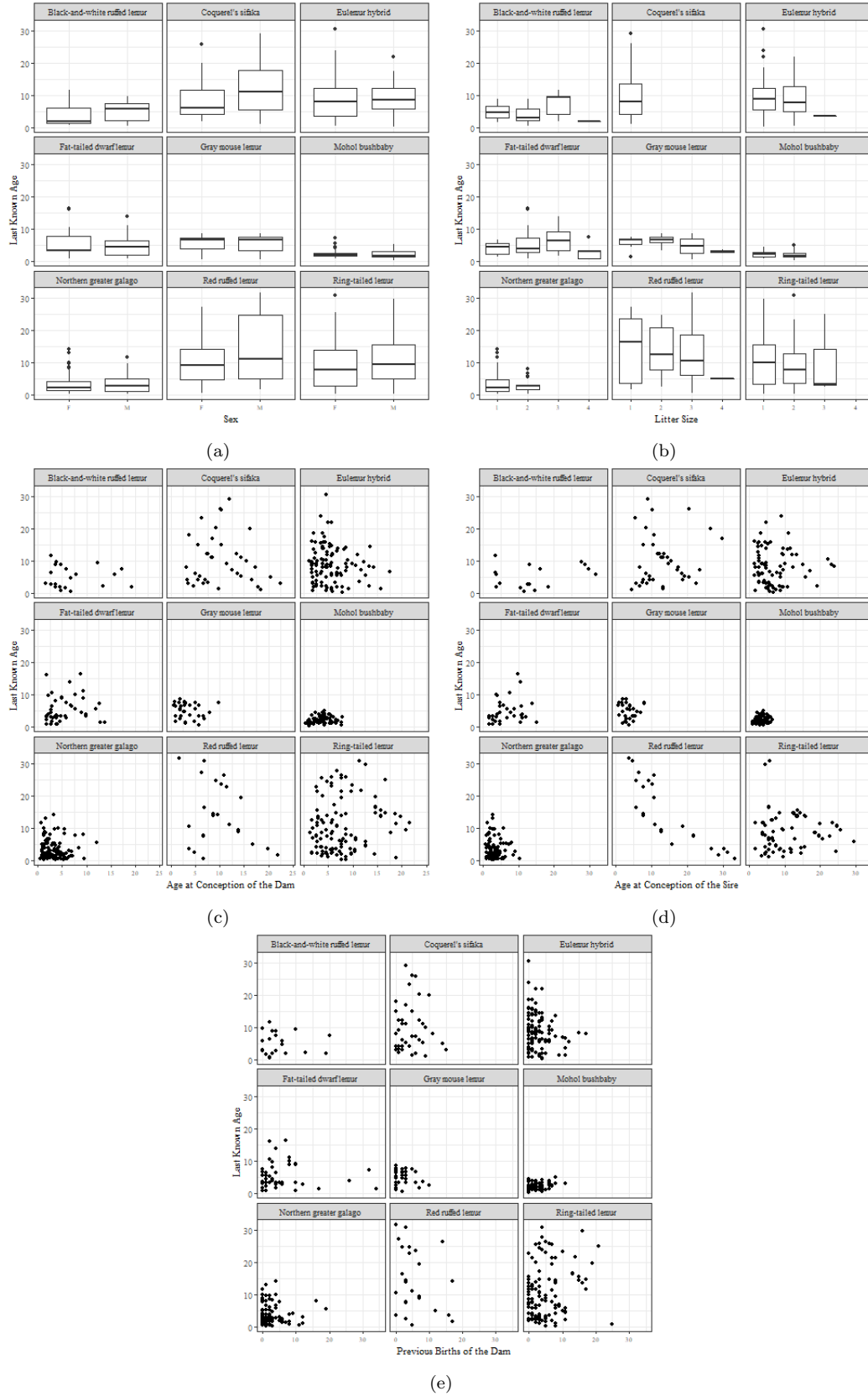


Figure 10: Distribution of age at death (in years) based on (a) sex, (b) litter size, (c) age of the dam, (d) age of the sire, and (e) previous dam births for lemurs who survived past 30 days and do not have a confirmed age of death, faceted by the nine species with the most observations.

6.2 Missing Data

Variable	Missing-Data Mechanism	Imputation Technique
Mother age	This data is missing not at random (and correlated with missing 'Father age').	We used kNN with the following variables: DOB, Mother age, Mother taxon, Father ID, Father taxon, Father age, Previous birth.
Father age	This data is missing not at random (and correlated with missing 'Mother age').	We used kNN with the following variables: DOB, Mother ID, Mother taxon, Father ID, Father taxon, Mother age, Previous birth.
Litter size		We use the mean value of the same Taxon and Sex .
Sex	This data is missing at random. While there are 0 null values, there are 220 values with no determined sex (and correlated with when the maximum age recorded is 0).	

6.3 Model Assumptions

6.3.1 Neonatal Mortality

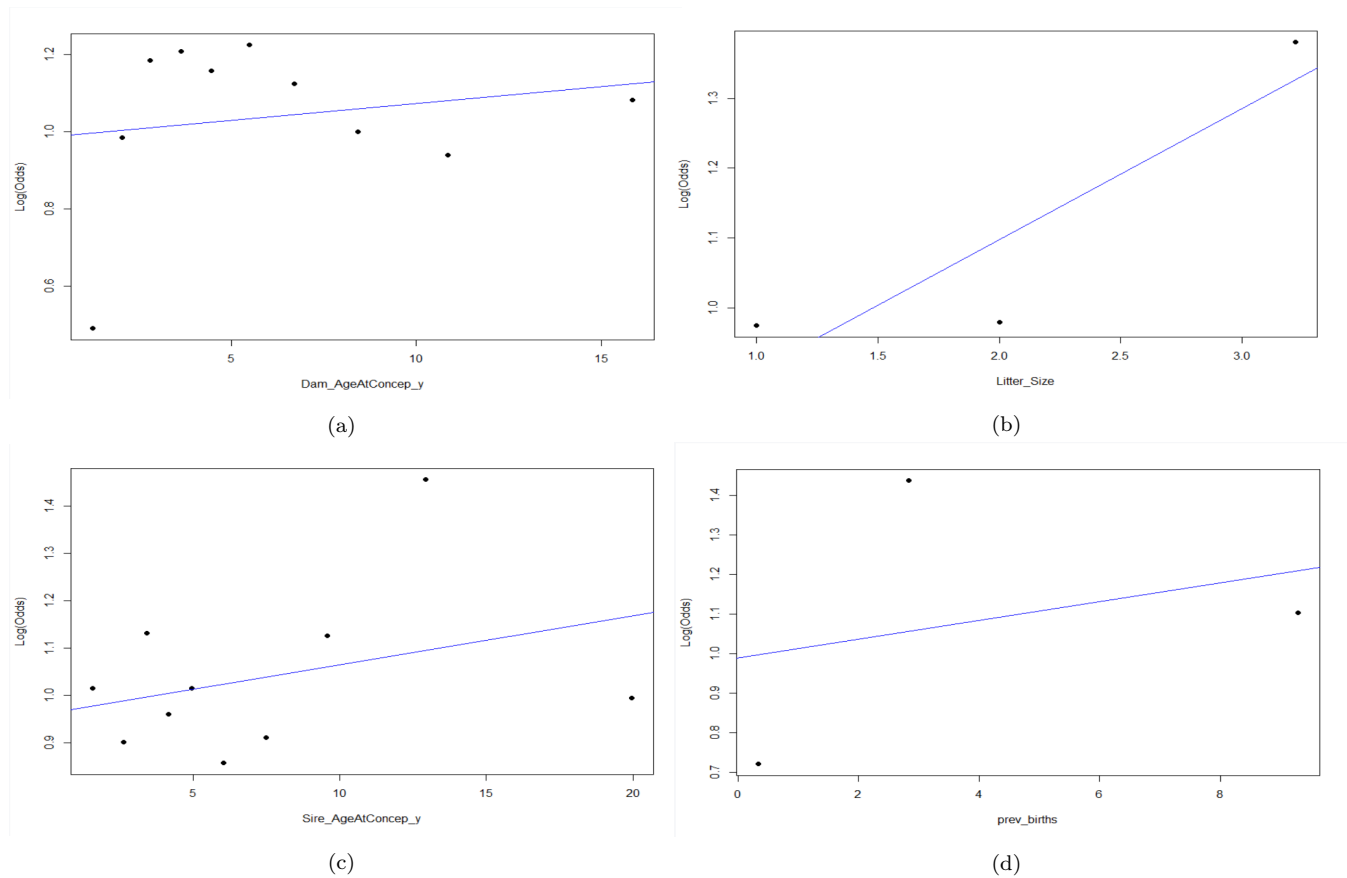


Figure 11: The empirical logit and (a) age of the dam, (b) litter size, (c) age of the sire, and (d) previous dam births.

The three conditions for logistic regression are linearity, randomness, and independence. Taking a look at the empirical logit and each numerical predictor variable (figure 11), we can see that while there is somewhat linear

relationship between the empirical logit and the variable referring to litter size, the same cannot be said for the other predictor variables. This means that the linearity condition is not met. In terms of randomness, the sample was not randomly selected, but there is no reason to suggest that our sample of lemurs from the Duke Lemur Center differs from the general population of captive-born lemurs (that is, those born at different centers). Therefore, the randomness condition is satisfied. While it is likely that the lemurs share characteristics based on species, this is accounted for in our model via interaction terms. As such, the independence condition is met.

6.3.2 Long-term Longevity

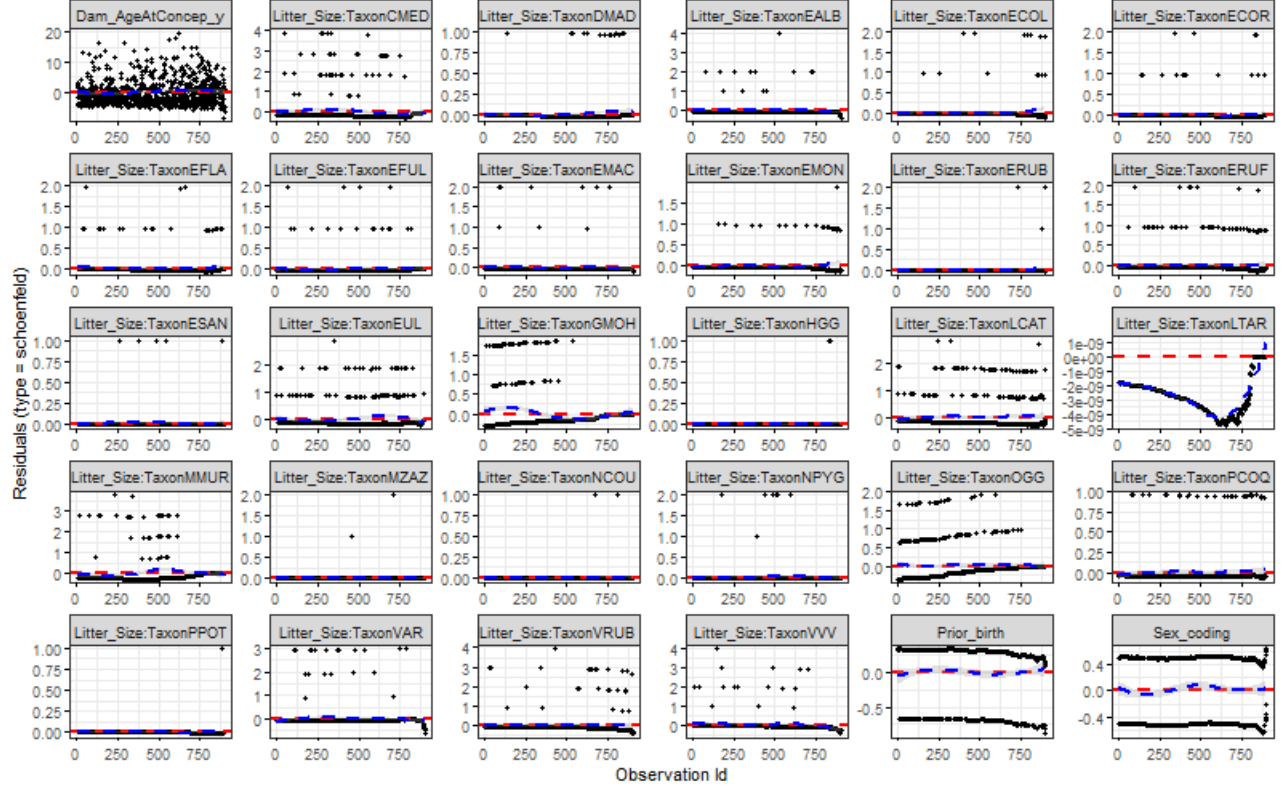


Figure 12: Scaled Schoenfeld residuals against the transformed time for the long-term longevity model.

In order to test the proportional hazards assumption, we took a look at the scaled Schoenfeld residuals for each of the terms in our model. Ideally, the Schoenfeld residuals are independent of time. Generally, there seems to be no pattern with time, except with the interaction between litter size and the TAR species. The assumption of proportional hazards seem to be supported for all covariates except the one mentioned previously.

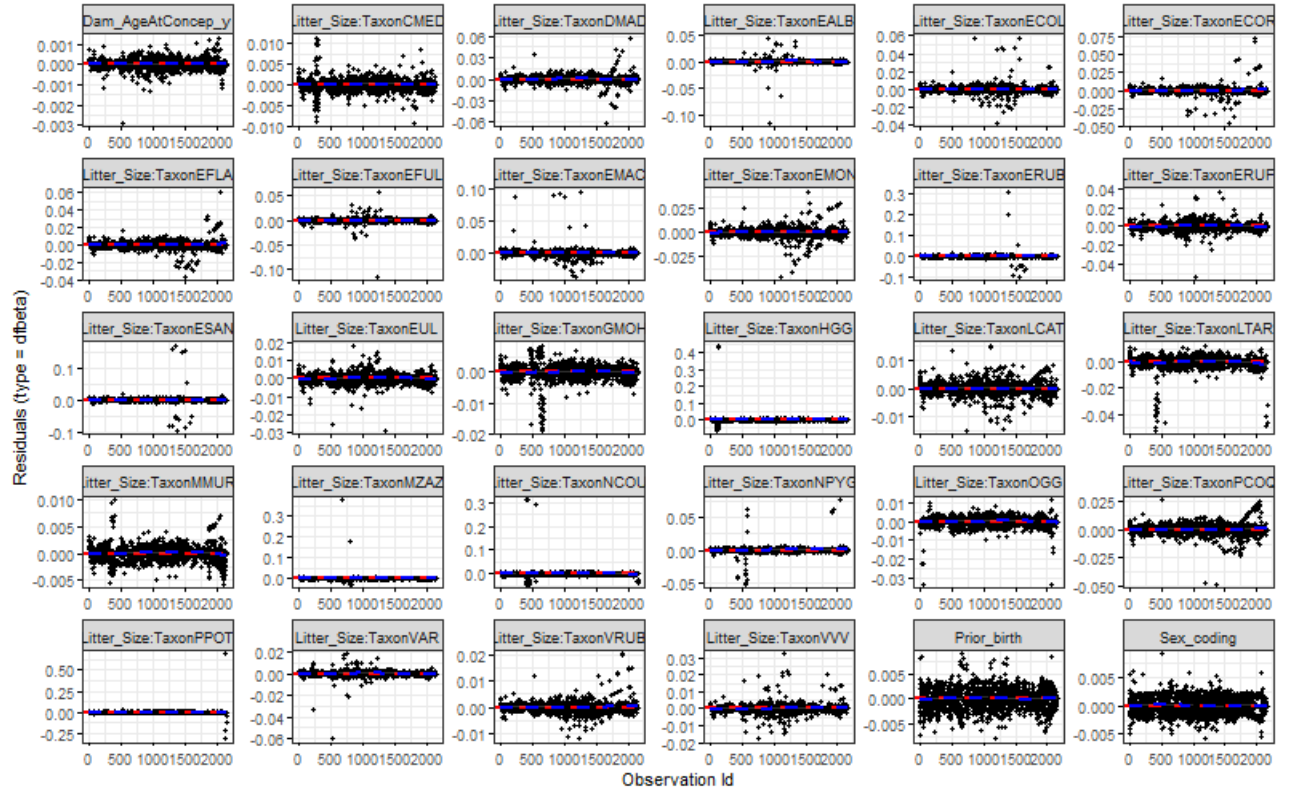


Figure 13: Estimated changes in the regression coefficients of the long-term longevity model when deleting each observation in turn.

In order to test for influential observations, we also looked at a plot of the dfbeta values. The above index plots show that comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that there are quite a few individual influential points.

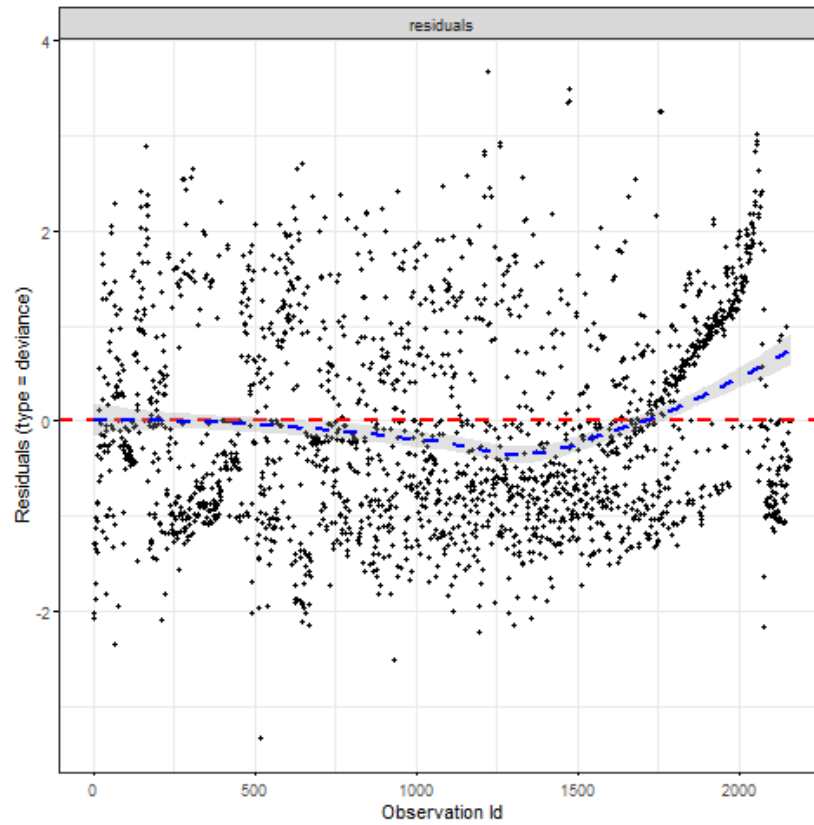


Figure 14: Deviance residuals of the long-term longevity model.

We can also check for outliers using deviance residuals. The pattern is ideally symmetric around 0, which is not the case for our model.

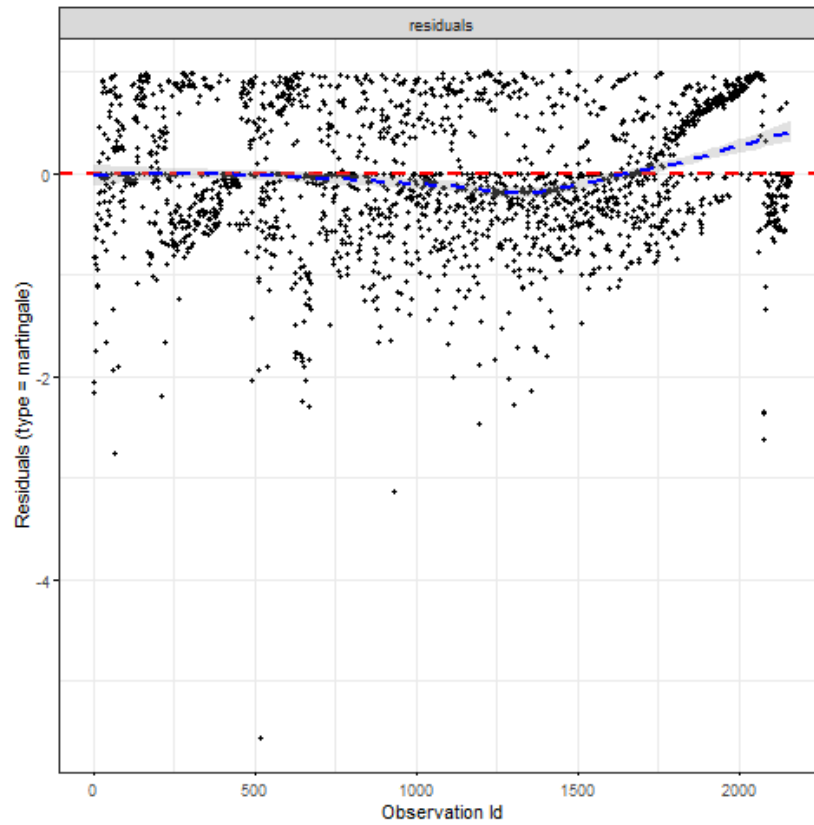


Figure 15: Martingale residuals of the long-term longevity model.

We can use Martingale residuals to detect nonlinearity. Looking at the figure above, we see that there is evidence to suggest that nonlinearity might present as seen by the fact that the blue line deviates from the red line the larger the observation ID becomes.

References

- [1] Melissa J Azur, Elizabeth A Stuart, Constantine Frangakis, and Philip J Leaf. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research*, 20(1):40–49, 2011.
- [2] Will Badr. 6 different ways to compensate for missing data (data imputation with examples), 2019.
- [3] K. Benirschke and C. J. Miller. Weights and neonatal growth of ring-tailed lemurs (*lemur catta*) and ruffed lemurs (*lemur variegatus*). *The Journal of Zoo Animal Medicine*, 12(4):107–111, 1981.
- [4] MB Blanco and SM Zehr. Striking longevity in a hibernating lemur. *Journal of Zoology*, 296(3):177–188, 2015.
- [5] N Bons, F Rieger, D Prudhomme, A Fisher, and K-H Krause. *Microcebus murinus*: a useful primate model for human cerebral aging and alzheimer’s disease? *Genes, Brain and Behavior*, 5(2):120–130, 2006.
- [6] Ricki J Colman. Non-human primates as a model for aging. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(9):2733–2741, 2018.
- [7] Klodian Dhana. Handling missing data with mice package; a simple approach — r-bloggers, 2016.
- [8] DLC. Duke lemur center. 2022.
- [9] Martine Perret and Aude Anzereay. Parental age at conception on mouse lemur’s offspring longevity: sex-specific maternal effects. *bioRxiv*, 2022.
- [10] Fabien Pifferi, Jérémy Terrien, Julia Marchal, Alexandre Dal-Pan, Fathia Djelti, Isabelle Hardy, Sabine Chahory, Nathalie Cordonnier, Loïc Desquilbet, Murielle Hurion, et al. Caloric restriction increases lifespan but affects brain integrity in grey mouse lemur primates. *Communications biology*, 1(1):1–8, 2018.
- [11] Christoph Schwitzer and Werner Kaumanns. Litter size, infant mortality and female body weight in captive black-and-white ruffed lemurs *varecia variegata*. *Endangered Species Research*, 8(3):201–209, 2009.
- [12] Sarah M Zehr, Richard G Roach, David Haring, Julie Taylor, Freda H Cameron, and Anne D Yoder. Life history profiles for 27 strepsirrhine primate taxa generated using captive data from the duke lemur center. *Scientific data*, 1(1):1–11, 2014.