Stoeckli, Sabrina
Department Consumer Behavior, Institute for Marketing and Management, University of Bern
Engehaldenstrasse 4, 3012 Bern, Switzerland
sabrina.stoeckli@imu.unibe.ch

Rojas Mora, Alfonso
Laboratory of Ecology and Epidemiology of Parasites, Institute of Biology, University of Neuchatel,
Rue Emile-Argand 11, 2000 Neuchatel, Switzerland
alfonso.rojas@unine.ch

Data Science Project

Do parasites affect male fertility in a common bird?

Conceptual Design Report

19 October 2020

Abstract

Parasites are unavoidable natural enemies of all living organisms, and they result in several costs to their hosts. For instance, in order to fight parasites, hosts have to re-allocate resources from their somatic maintenance and reproduction investment towards their defenses against the parasite. Due to such trade-off between defense against parasites and other bodily functions, it is commonly observed that parasitism results in shortened lifespans and lowered fertility. While production of ova has been argued to be costly, sperm production is deemed to incur in small costs. This research hypothesizes that male fertility is likely affected by parasitism. Here, we use data coming from an experiment that experimentally infected a common House Sparrows (*Passer domesticus*) with the coccidian *Isospora sp.* to test whether such an infection leads to loses in male fertility. Coccidia are common avian parasites, infecting the majority of bird species. The data used in this project comes from 30 uninfected and 30 infected males, which were sperm sampled before the infection and after the 9^{th} day and 18^{th} day of infection. Ejaculates were recorded under the microscope, and the videos were analyzed using a computer assisted sperm analyzer (CASA) plug-in for ImageJ. In this “module 1 conceptual design report”, we describe the data management of the data analysis. The subsequent conduct of the statistical modelling and data visualization will follow in the “module 2 poster presentation”.

Table of Contents

1 Project Objectives

Parasites are unavoidable, and thus organisms have developed several mechanisms of defense against parasitism. However, defense against parasites is costly [1], as for it requires the re-allocation of resources that would be otherwise used for the maintenance of somatic and reproductive functioning [2]. Thus, parasitism usually leads to shortened lifespans and lowered fertility. Sperm production is likely costly, and thus parasitism can directly impact male fertility by reducing the resource pool that can be allocated to sperm production. Hence, it could be predicted that infected males would produce ejaculates of lower quality compared to males that are uninfected.

Having said this, it is the main objective of this project to test the hypothesis that male fertility is affected by parasitism. To assess our hypothesis, we use data obtained from a study that experimentally infected captive wild House Sparrows (*Passer domesticus*) with the coccidian *Isospora sp.* We expect that males infected with *Isospora* have ejaculates with relatively low proportions of motile sperm and/or ejaculates with less motile sperm. Note that ejaculate quality is a multivariate trait that refers to the fertilizing ability of an ejaculate. It can be measured as the proportion of alive sperm, the motility of sperm, and the proportion of abnormal sperm. Hence, to test our hypothesis we require ejaculate data from various males as well as information on their infection status by a parasitic disease. In an experiment, we therefore control the infection status of males by previously treating them against a parasite and experimentally reinfecting half of them. Under such a design, we can provide evidence of how parasites affect ejaculate quality through different stages of the infection (i.e. before infection, acute stage, and chronic stage). While it is more difficult to generate specific predictions on how ejaculate quality will change between infection stages, the main prediction hypothesized above should hold.

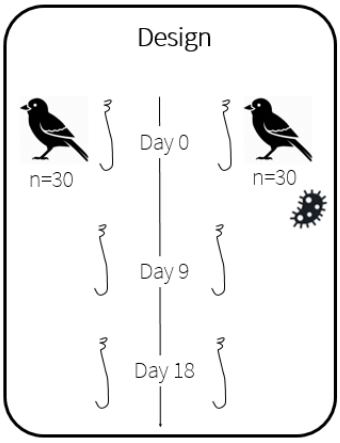


Figure 1. Illustration of the experimental design. All birds are sperm sampled on day 0, day 9 and day 18. Half of the birds (n=30) are infected on day 0 (after sperm sampling). Birds reach an acute infection stage on day 9 and a chronic infection stage on day 18.

2 Methods

The methods for this project can be discussed from the perspective of three different phases: field phase, laboratory phase, and data management. For the field phase, we need a set of aviaries to maintain in captivity the birds to be used in the project. This includes the appropriate bird food, and appropriate food containers for the species that is kept captive.

For the laboratory phase, we need a microscope with a mounted video camera in order to record the fresh ejaculates. Further, we need sperm swimming chambers with a fixed dept as well as a microscope plate heater to maintain a constant temperature for the sperm during the video recordings.

Finally, for the data management we require external drives to store all the uncompressed video recordings (ca. 1.5 Gb per minute of recording). In order to analyze the videos, we require a computer with ImageJ installed as well as the Computer Assisted Sperm Analyzer (CASA) plug-in. Given that the computational power required for this step is not high, this analysis could be performed on a normal laptop or desktop computer. For more details on the CASA plug-in see Wilson-Leedy and Ingermann (2006) [3].

To organize the data obtained from the CASA plug-in, we need the following python libraries: Pandas, NumPy and os. Specifically, we use these libraries to summarize and reshape the data obtained from each video and then merge everything into a single table containing all the data. Further, we use the libraries Matplotlib and Seaborn to visualize the data. We will also use the library scikit-learn to reduce the dimensionality of the data with a principal component analysis. Finally, we require the libraries statsmodels and scipy to make the linear mixed models to assess repeatability of measurements as well as to fit the corresponding models to the data.

3 Data

In the current project, we aim to assess whether a parasite infection can lead to loses in male fertility. The final set of variables consists of 53 variables. Besides some more general variables such as DATE, MANIPULATION DAY and ID, the data set contains various variables that are indicators for male fertility. In the following, we provide an overview of the final data frame that builds the basis of our statistical modelling and data visualization.

#	Column	Description
0	DATE	Date
1	SAMPLING_GROUP	Sampling group
2	EXP_DAY	Experimental day
3	MANIPULATION_DAY	Day of manipulation, (after 0, 9, and 18 days)
4	TREATMENT	Infected vs. not infected
5	AVIARY	Number of aviary (total of 15 aviaries)
6	ID	Identifier for bird
7	COLOR_RING	Ring on leg of every bird, for behavioral observations (not relevant)
8	SAMPLING_TIME	Time of sampling day (e.g., 10 am)
9	BODY_MASS	Body mass
10	VIDEO	Number of video
11	CLOACA_HIGH1	Cloaca height (in mm) at measurement 1
12	CLOACA_HIGH2	Cloaca height (in mm) at measurement 2
13	CLOACA_HIGH3	Cloaca height (in mm) at measurement 3
14	CLOACA_WIDTH1	Cloaca width (in mm) at measurement 1
15	CLOACA_WIDTH2	Cloaca width (in mm) at measurement 2
16	CLOACA_WIDTH3	Cloaca width (in mm) at measurement 3
17	CLOACA_LENGTH1	Cloaca length (in mm) at measurement 1
18	CLOACA_LENGTH2	Cloaca length (in mm) at measurement 2
19	CLOACA_LENGTH3	Cloaca length (in mm) at measurement 3
20	INFECTION_STAGE	Infection stage: before vs. acute vs. chronic
21	VCL	Velocity curvilinear; Point to point velocity (total distance traveled per sec)
22	VAP	Velocity average path; Point to point velocity on a path constructed using a roaming average
23	VSL	Velocity straight line; Velocity measured using the first point and the point reached that is furthest from th is origin during the measured time period
24	LIN	Linearity (=VSL/VAP), describes path curvature
25	WOB	Amplitude of Lateral Head displacement (ALH) corresponds to Wobble (WOB; =VAP/VCL), describes side to side movement of the sperm head
26	PROG	Progression; The average distance of the sperm from its origin on the average path during all frames analyzed
27	BCF	Beat cross frequency; determined by detecting the frequency at which VCL crosses VAP
28	VCL15	VCL for 15% fastest sperms

29	VAP15	VAP for 15% fastest sperms
30	VSL15	VSL for 15% fastest sperms
31	LIN15	LIN for 15% fastest sperms
32	WOB15	WOB for 15% fastest sperms
33	PROG15	PROGR for 15% fastest sperms
34	BCF15	BCF for 15% fastest sperms
35	VCL10	VCL for 10% fastest sperms
36	VAP10	VAP for 10% fastest sperms
37	VSL10	VSL for 10% fastest sperms
38	LIN10	LIN for 10% fastest sperms
39	WOB10	WOB for 10% fastest sperms
40	PROG10	PROG for 10% fastest sperms
41	BCF10	BCF for 10% fastest sperms
42	VCL5	VCL for 5% fastest sperms
43	VAP5	VAP for 5% fastest sperms
44	VSL5	VSL for 5% fastest sperms
45	LIN5	LIN for 5% fastest sperms
46	WOB5	WOB for 5% fastest sperms
47	PROG5	PROG for 5% fastest sperms
48	BCF5	BCF for 5% fastest sperms
49	MOTILITY	proportion of motile sperm
50	TRACKS	total number of sperm
51	cloaca_depth	mean score of cloaca depth of the three measurements
52	cloaca_height	mean score of cloaca heigth of the three measurements
53	cloaca_length	mean score of cloaca length of the three measurements

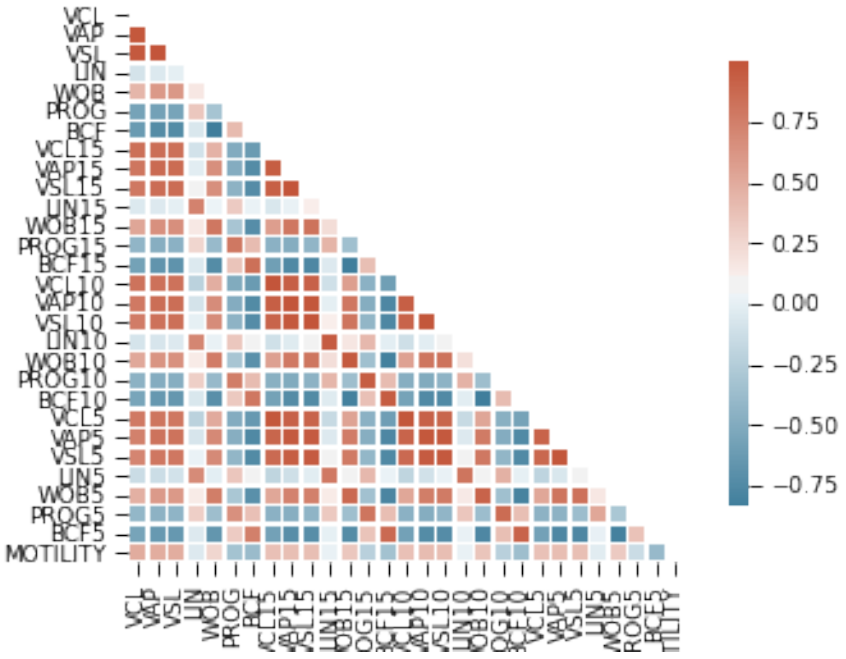


Figure 2. Graphical representation of the Pearson correlation coefficients for all the sp erm quality measurements.

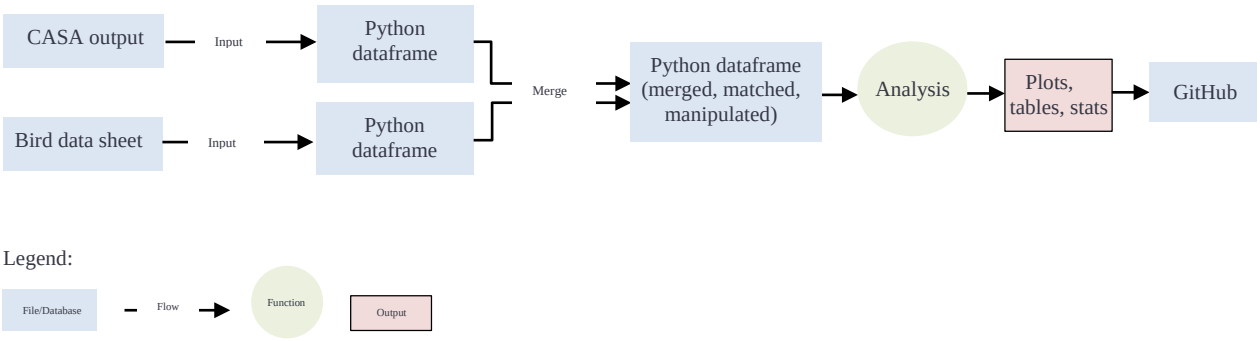
4 Metadata

All the data and the Jupiter Notebook with our code is stored on GitHub. All the metadata for this project, namely the structure of our GitHub project, is described in a README file on GitHub. For readability and comprehension reasons, we include detailed explanations of the different measures (i.e., variables) and data management steps in this CDR and the Jupiter Notebook.

5 Data Quality

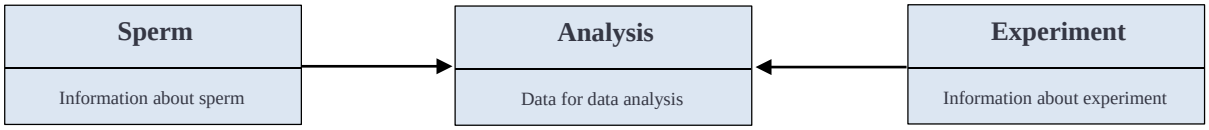
The data stems from a controlled experiment conducted in 2015 by Alfonso Rojas Mora, one of the authors of this CDR. More specifically, data comes from wild birds being captive, and thus receiving a standard diet. Further, all birds were given antiparasitic medications to assure that only the experimental individuals were infected. Thus, given the controlled character of the experiment, we can assume that—from a methodological perspective—the data is of high quality. For instance, dynamic information about the infection state (e.g., acute infection vs. chronic infection) could be recorded. Further, variables such as the cloaca dimensions were measured repeatedly each time, which increases accuracy of the cloaca dimension measures. Finally, all the sperm samples were freshly collected by a certified animal experimenter (ARM) and there was not more than five minutes between the sample collection and the video recording. Finally, all video analyses were conducted by an automatized computer system, which has been validated as a standard method for sperm analysis (CASA, see [3]).

6 Data Flow

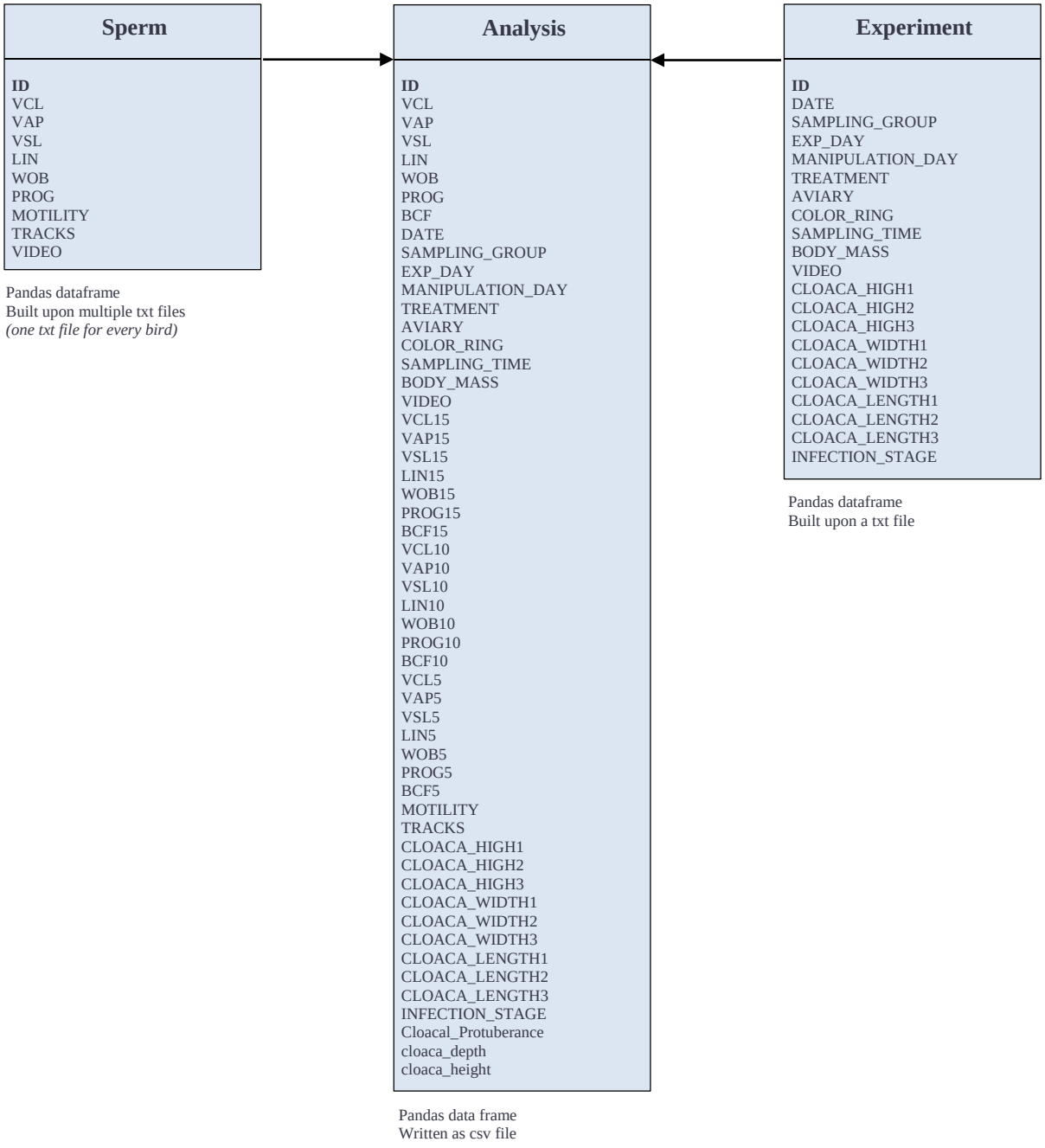


7 Data Model

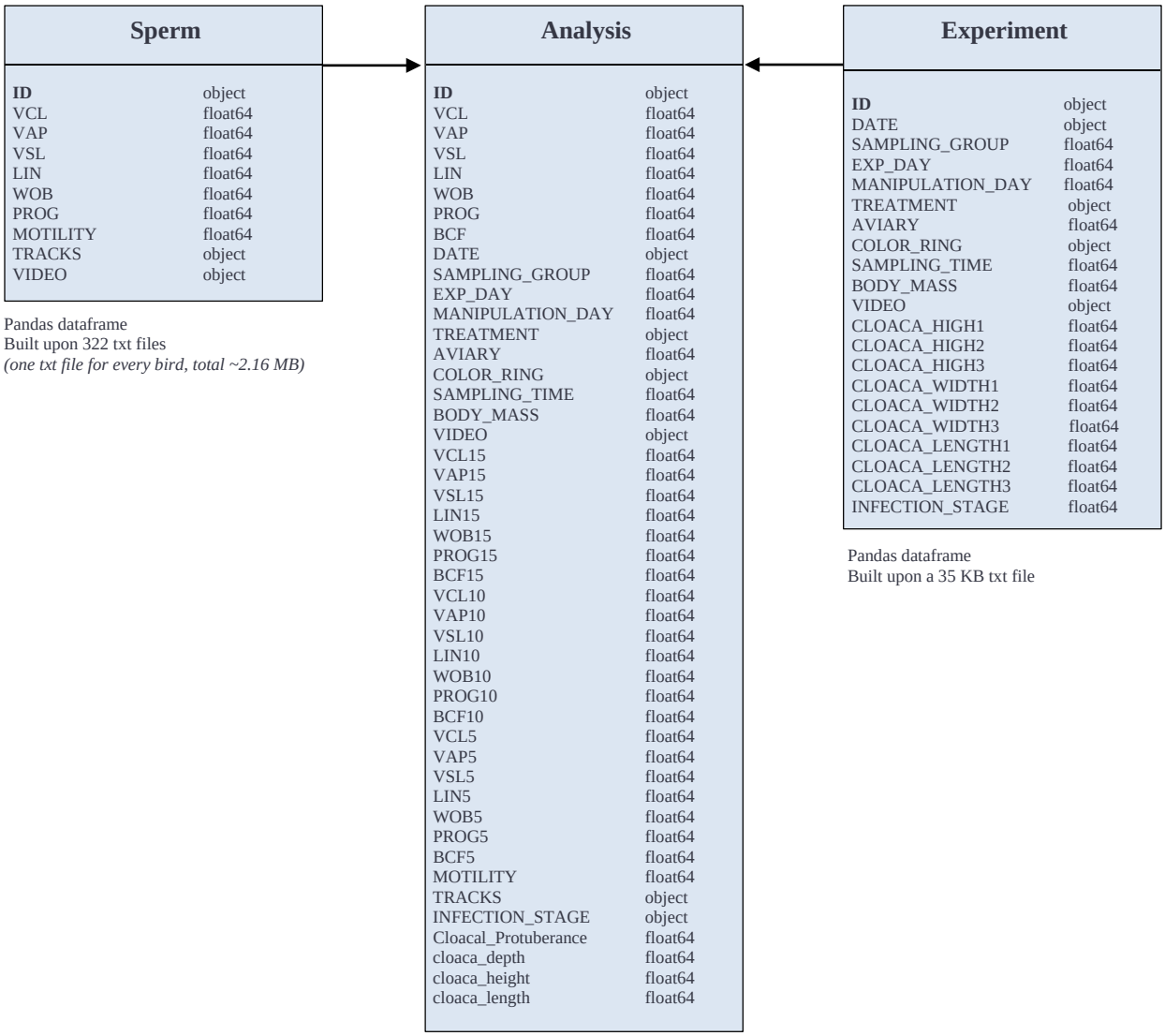
Conceptual model



Logical model



Physical model



8 Risks

There are two potential risks that need to be considered. First, there is the risk of data and code loss. The most likely reasons for this are storage failures, unintended data deletion or unauthorized access with criminal intentions. Risk diversification measures include that our raw and processed data as well as our code are stored in a cloud-based storage service (Dropbox), on an external hard drive as well an on Git.

Second, there is the risk of data analysis mistakes. The most likely reason for this is the incorrect application of statistical methods and python libraries. To mitigate this risk, our analysis was peer reviewed.

Note that an additional risk might occur due to poor data quality. Typically, the reason for this is related to a poor methodological procedure. Our measures to prevent this risk are described in section 5 (Data Quality).

9 Preliminary Studies

In the following, we report some stats and depict some plots from the analysis that we run for our “module 2 poster presentation”. Note that the full analysis that is conducted to test the hypothesis whether costs of parasitism affect male fertility can be found on GitHub (see Jupiter Notebook and our module 2 poster).

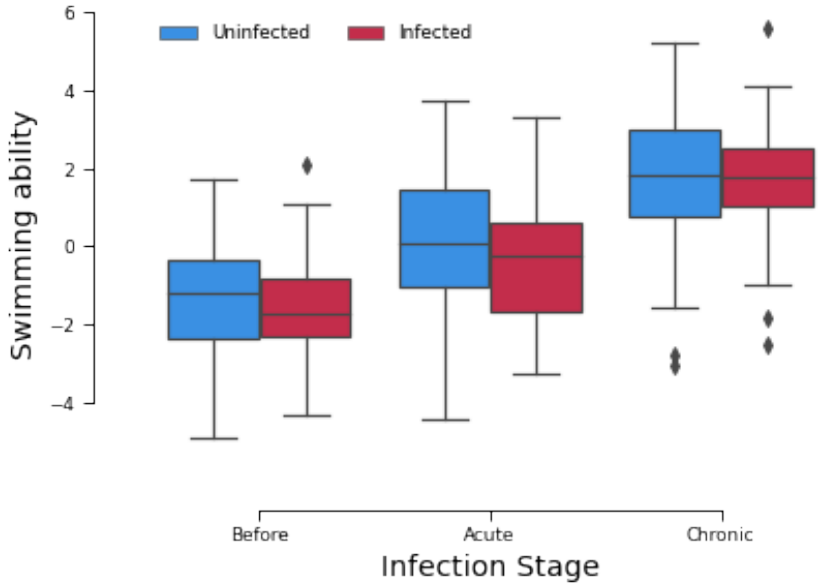
Overall, we compute six linear mixed models in the Jupiter Notebook. Therefore, we separately use our six fertility indicators (MOTILITY, PC100%, PC15%, PC10%, PC5%, Cloacal_Protuberance) as response variables for the linear mixed models. In all models, we enter TREATMENT, INFECTION_STATE and their interaction as predictors. Further, we model a random slope for all birds (i.e. ID as random effect).

Model for swimming ability of all motile sperm (PC100%)

```
mod_swimm100 = smf.mixedlm("PC1 ~ C(INFECTION_STAGE) * C(TREATMENT)",
                df100,
                groups = df100['ID'])
mod_swimm100_fit = mod_swimm100.fit()
```

```
                          Mixed Linear Model Regression Results
=============================================================================================================
Model:                              MixedLM                 Dependent Variable:                 PC1
No. Observations:                   319                     Method:                             REML
No. Groups:                         55                      Scale:                              1.5668
Min. group size:                    3                       Log-Likelihood:                     -569.5957
Max. group size:                    6                       Converged:                          Yes
Mean group size:                    5.8
-------------------------------------------------------------------------------------------------------------
                                                          Coef.  Std.Err.   z    P>|z| [0.025 0.975]
-------------------------------------------------------------------------------------------------------------
Intercept                                                 -0.383    0.266 -1.442 0.149 -0.903  0.138
C(INFECTION_STAGE)[T.Before]                              -1.238    0.241 -5.138 0.000 -1.710 -0.766
C(INFECTION_STAGE)[T.Chronic]                              2.064    0.241  8.568 0.000  1.592  2.536
C(TREATMENT)[T.Uninfected]                                 0.419    0.376  1.115 0.265 -0.317  1.156
C(INFECTION_STAGE)[T.Before]:C(TREATMENT)[T.Uninfected]   -0.252    0.344 -0.732 0.464 -0.926  0.422
C(INFECTION_STAGE)[T.Chronic]:C(TREATMENT)[T.Uninfected]  -0.345    0.347 -0.995 0.320 -1.024  0.334
Group Var                                                  1.121    0.236
=============================================================================================================

Wald test
                                    chi2        P>chi2  df constraint
Intercept                       2.079246  1.493139e-01              1
C(INFECTION_STAGE)            191.796403  2.248745e-42              2
C(TREATMENT)                    1.243775  2.647448e-01              1
C(INFECTION_STAGE):C(TREATMENT)    1.063985  5.874334e-01              2
```

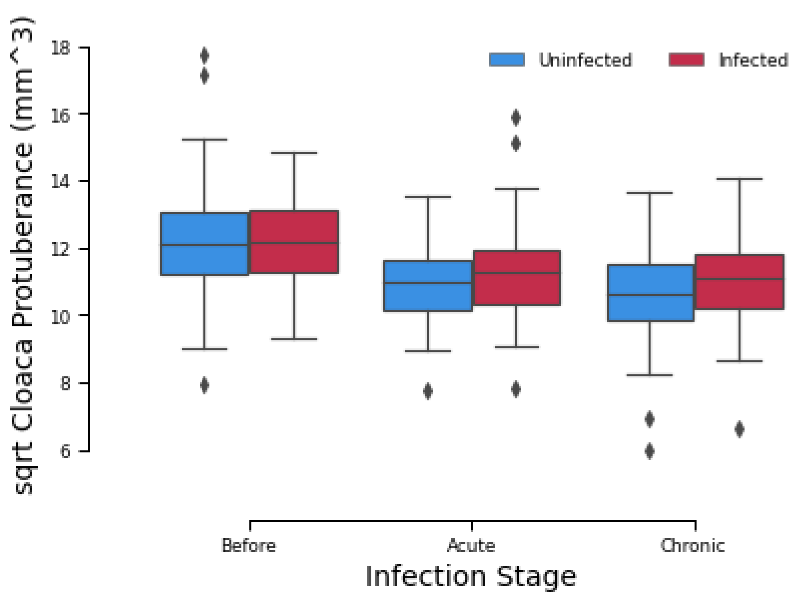


Model for cloaca protuberance

```
mod_cp = smf.mixedlm("Cloacal_Protuberance ~ C(INFECTION_STAGE) * C(TREATMENT)",
                df_cp,
                groups = df_cp["ID"])
mod_cp_fit = mod_cp.fit()
```

```
                         Mixed Linear Model Regression Results
========================================================================================
Model:                     MixedLM          Dependent Variable:          Cloacal_Protuberance
No. Observations:          336              Method:                      REML
No. Groups:                56               Scale:                       612.4549
Min. group size:           6                Log-Likelihood:              -1589.1339
Max. group size:           6                Converged:                   Yes
Mean group size:           6.0
----------------------------------------------------------------------------------------
                                                        Coef.  Std.Err.   z    P>|z|  [0.025  0.975]
----------------------------------------------------------------------------------------
Intercept                                             127.503    5.519 23.104 0.000 116.686 138.319
C(INFECTION_STAGE)[T.Before]                           21.034    4.677  4.498 0.000  11.868  30.201
C(INFECTION_STAGE)[T.Chronic]                          -3.010    4.677 -0.644 0.520 -12.176   6.157
C(TREATMENT)[T.Uninfected]                             -6.313    7.804 -0.809 0.419 -21.610   8.983
C(INFECTION_STAGE)[T.Before]:C(TREATMENT)[T.Uninfected]  9.307    6.614  1.407 0.159  -3.657  22.270
C(INFECTION_STAGE)[T.Chronic]:C(TREATMENT)[T.Uninfected] -2.341  6.614 -0.354 0.723 -15.305  10.622
Group Var                                             546.501    5.515
========================================================================================

Wald test
                                  chi2          P>chi2  df constraint
Intercept                   533.808790  4.191134e-118              1
C(INFECTION_STAGE)           31.381702   1.533028e-07              2
C(TREATMENT)                  0.654374   4.185532e-01              1
C(INFECTION_STAGE):C(TREATMENT)  3.471275   1.762878e-01              2
```

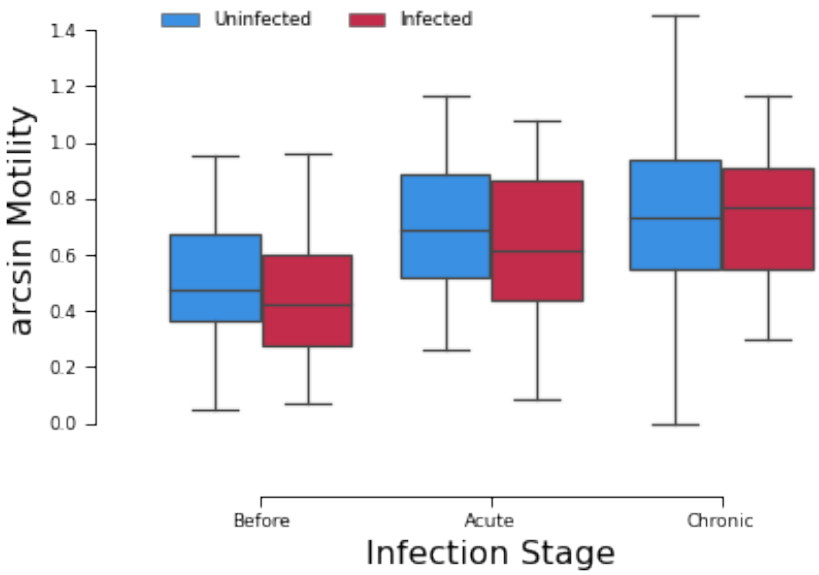


Model for motility (or proportion of motile sperm)

```
mod_motility = smf.mixedlm("MOTILITY ~ C(INFECTION_STAGE) * C(TREATMENT)",
                df_motility,
                groups = df_motility["ID"])
mod_motility_fit = mod_motility.fit()
```

```
                          Mixed Linear Model Regression Results
=============================================================================================================
Model:                          MixedLM              Dependent Variable:                  MOTILITY
No. Observations:               320                  Method:                              REML
No. Groups:                     55                   Scale:                               0.0295
Min. group size:                3                    Log-Likelihood:                      64.3926
Max. group size:                6                    Converged:                           Yes
Mean group size:                5.8
-------------------------------------------------------------------------------------------------------------
                                                      Coef.  Std.Err.    z    P>|z| [0.025 0.975]
-------------------------------------------------------------------------------------------------------------
Intercept                                              0.572    0.031 18.304 0.000  0.510  0.633
C(INFECTION_STAGE)[T.Before]                          -0.161    0.033 -4.886 0.000 -0.226 -0.097
C(INFECTION_STAGE)[T.Chronic]                          0.077    0.033  2.333 0.020  0.012  0.142
C(TREATMENT)[T.Uninfected]                             0.051    0.044  1.158 0.247 -0.036  0.138
C(INFECTION_STAGE)[T.Before]:C(TREATMENT)[T.Uninfected]  0.001    0.047  0.014 0.989 -0.092  0.093
C(INFECTION_STAGE)[T.Chronic]:C(TREATMENT)[T.Uninfected] -0.082    0.047 -1.725 0.084 -0.175  0.011
Group Var                                              0.012    0.021
=============================================================================================================

Wald test
                                    chi2        P>chi2  df constraint
Intercept                     335.023733  7.738132e-75              1
C(INFECTION_STAGE)             54.286930  1.628332e-12              2
C(TREATMENT)                    1.340169  2.470042e-01              1
C(INFECTION_STAGE):C(TREATMENT)    4.009706  1.346801e-01              2
```



10 Conclusions

Overall, we found that the parasite infection did not affect any of the fertility indicators. Yet, as the experiment progressed, all individuals increased their ejaculate swimming ability and motility. Further, sperm production was slightly decrease with time, as evidenced by the smaller cloacal protuberances.

References and Bibliography

[1] Lochmiller, R. L., & Deerenberg, C. (2000). Trade-offs in evolutionary immunology: just what is the cost of immunity?. *Oikos*, *88*(1), 87-98.

[2] Sheldon, B. C., & Verhulst, S. (1996). Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. *Trends in Ecology & Evolution*, *11*(8), 317-321.

[3] Wilson-Leedy, J. G., & Ingermann, R. L. (2007). Development of a novel CASA system based on open source software for characterization of zebrafish sperm motility parameters. *Theriogenology*, *67*(3), 661-672.