**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 9th day and 18th 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

[Abstract 1](#_Toc52952538)

[Table of Contents 2](#_Toc52952539)

[1 Project Objectives 3](#_Toc52952540)

[2 Methods 3](#_Toc52952541)

[3 Data 4](#_Toc52952542)

[4 Metadata 5](#_Toc52952543)

[5 Data Quality 6](#_Toc52952544)

[6 Data Flow 6](#_Toc52952545)

[7 Data Model 7](#_Toc52952546)

[8 Risks 9](#_Toc52952547)

[9 Preliminary Studies 9](#_Toc52952548)

[10 Conclusions 13](#_Toc52952549)

[References and Bibliography 13](#_Toc52952550)

# 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.

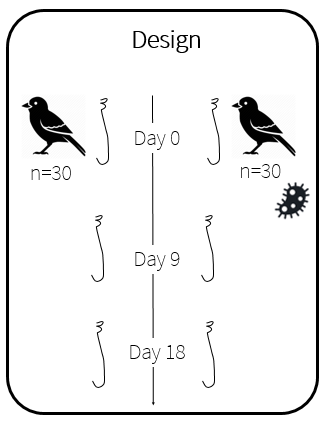
Having said this, it is the main objective of this project to test the hypothesis that male fertility is affected by parasitism. To assess our hypothesis, we use data obtained from a study that experimentally infected captive wild House Sparrows (*Passer domesticus*) with the coccidian *Isospora sp.* 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](https://imagej.nih.gov/ij/download.html) installed as well as the [Computer Assisted Sperm Analyzer](https://imagej.nih.gov/ij/plugins/docs/CASAinstructions.pdf) (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)](https://pubmed.ncbi.nlm.nih.gov/17137620/) [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 this 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

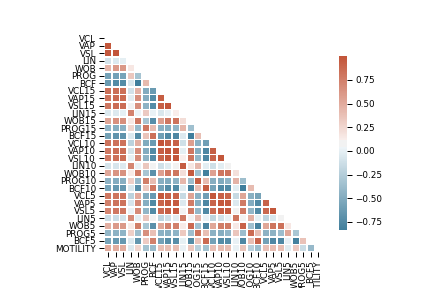


Figure 2. Graphical representation of the Pearson correlation coefficients for all the sperm quality measurements.

# 4 Metadata

All the data and the Jupiter Notebook with our code is stored on [GitHub](https://gitlab.imucb.ch/stoecklis/Project_module1-2). 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

Python dataframe

CASA output

Input

Input

Merge

GitHub

Plots,

tables, stats

Python dataframe

(merged, matched, manipulated)

Python dataframe

Bird data sheet

Legend:

Output

Flow

File/Database

# 7 Data Model

**Conceptual model**

**Analysis**

Data for data analysis

**Sperm**

Information about sperm

**Experiment**

Information about experiment

**Logical model**

**Experiment**

**ID**

DATE

SAMPLING\_GROUP

EXP\_DAY

MANIPULATION\_DAY

TREATMENT

AVIARY

COLOR\_RING

SAMPLING\_TIME

BODY\_MASS

VIDEO

CLOACA\_HIGH1

CLOACA\_HIGH2

CLOACA\_HIGH3

CLOACA\_WIDTH1

CLOACA\_WIDTH2

CLOACA\_WIDTH3

CLOACA\_LENGTH1

CLOACA\_LENGTH2

CLOACA\_LENGTH3

INFECTION\_STAGE

**Analysis**

**ID**

VCL

VAP

VSL

LIN

WOB

PROG

BCF

DATE

SAMPLING\_GROUP

EXP\_DAY

MANIPULATION\_DAY

TREATMENT

AVIARY

COLOR\_RING

SAMPLING\_TIME

BODY\_MASS

VIDEO

VCL15

VAP15

VSL15

LIN15

WOB15

PROG15

BCF15

VCL10

VAP10

VSL10

LIN10

WOB10

PROG10

BCF10

VCL5

VAP5

VSL5

LIN5

WOB5

PROG5

BCF5

MOTILITY

TRACKS

CLOACA\_HIGH1

CLOACA\_HIGH2

CLOACA\_HIGH3

CLOACA\_WIDTH1

CLOACA\_WIDTH2

CLOACA\_WIDTH3

CLOACA\_LENGTH1

CLOACA\_LENGTH2

CLOACA\_LENGTH3

INFECTION\_STAGE

Cloacal\_Protuberance

cloaca\_depth

cloaca\_height

cloaca\_length

**Sperm**

**ID**

VCL

VAP

VSL

LIN

WOB

PROG

MOTILITY

TRACKS

VIDEO

Pandas dataframe

Built upon multiple txt files

*(one txt file for every bird)*

Pandas dataframe

Built upon a txt file

Pandas data frame

Written as csv file

**Physical model**

**Sperm**

**ID** object

VCL float64

VAP float64

VSL float64

LIN float64

WOB float64

PROG float64

MOTILITY float64

TRACKS object

VIDEO object

**Analysis**

**ID** object

VCL float64

VAP float64

VSL float64

LIN float64

WOB float64

PROG float64

BCF float64

DATE object

SAMPLING\_GROUP float64

EXP\_DAY float64

MANIPULATION\_DAY float64

TREATMENT object

AVIARY float64

COLOR\_RING object

SAMPLING\_TIME float64

BODY\_MASS float64

VIDEO object

VCL15 float64

VAP15 float64

VSL15 float64

LIN15 float64

WOB15 float64

PROG15 float64

BCF15 float64

VCL10 float64

VAP10 float64

VSL10 float64

LIN10 float64

WOB10 float64

PROG10 float64

BCF10 float64

VCL5 float64

VAP5 float64

VSL5 float64

LIN5 float64

WOB5 float64

PROG5 float64

BCF5 float64

MOTILITY float64

TRACKS object

INFECTION\_STAGE object

Cloacal\_Protuberance float64

cloaca\_depth float64

cloaca\_height float64

cloaca\_length float64

**Experiment**

**ID** object

DATE object

SAMPLING\_GROUP float64

EXP\_DAY float64

MANIPULATION\_DAY float64

TREATMENT object

AVIARY float64

COLOR\_RING object

SAMPLING\_TIME float64

BODY\_MASS float64

VIDEO object

CLOACA\_HIGH1 float64

CLOACA\_HIGH2 float64

CLOACA\_HIGH3 float64

CLOACA\_WIDTH1 float64

CLOACA\_WIDTH2 float64

CLOACA\_WIDTH3 float64

CLOACA\_LENGTH1 float64

CLOACA\_LENGTH2 float64

CLOACA\_LENGTH3 float64

INFECTION\_STAGE float64

Pandas dataframe

Built upon 322 txt files

*(one txt file for every bird, total ~2.16 MB)*

Pandas dataframe

Built upon a 35 KB txt file

Pandas dataframe

Written as csv file

Analyses with Python libraries (Pandas, NumPy, os, matplotlib, seaborn, scikit-learn, statsmodels)

# 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](https://gitlab.imucb.ch/stoecklis/Project_module1-2) (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).

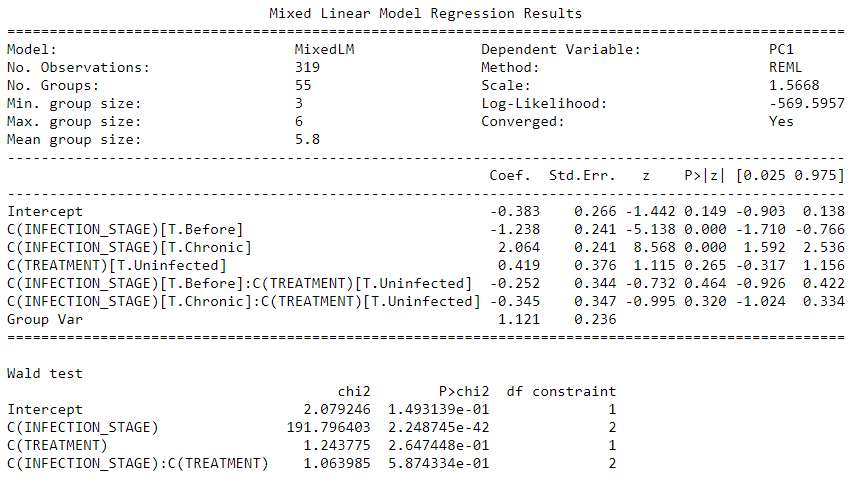
# Model for swimming ability of all motile sperm (PC100%)

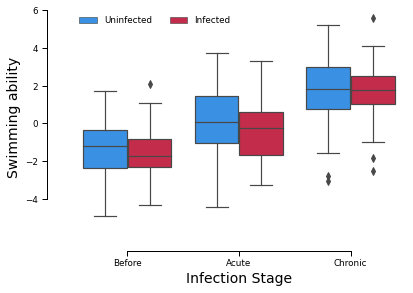
mod\_swimm100 = smf.mixedlm("PC1 ~ C(INFECTION\_STAGE) \* C(TREATMENT)",

df100,

groups = df100['ID'])

mod\_swimm100\_fit = mod\_swimm100.fit()





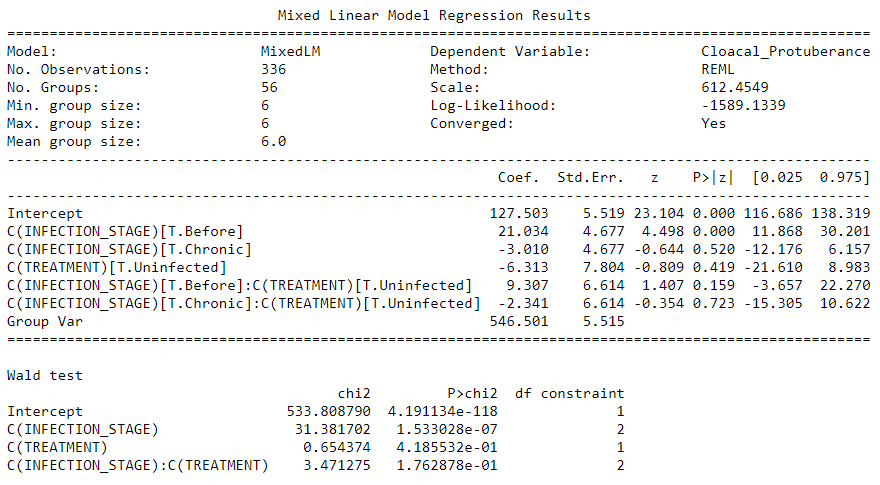
# Model for cloaca protuberance

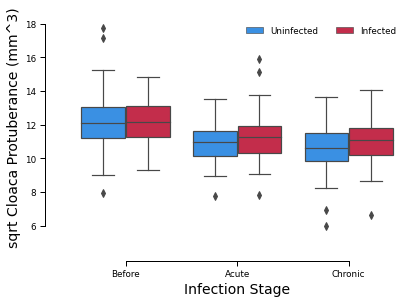
mod\_cp = smf.mixedlm("Cloacal\_Protuberance ~ C(INFECTION\_STAGE) \* C(TREATMENT)",

df\_cp,

groups = df\_cp["ID"])

mod\_cp\_fit = mod\_cp.fit()





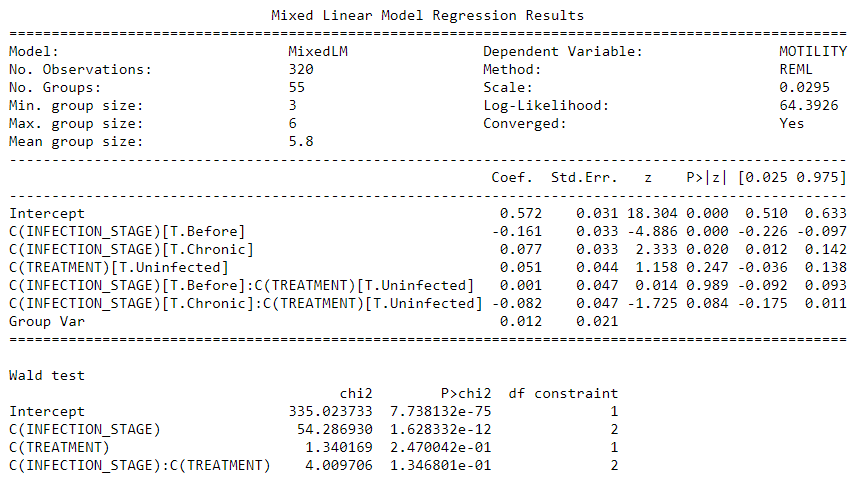
# Model for motility (or proportion of motile sperm)

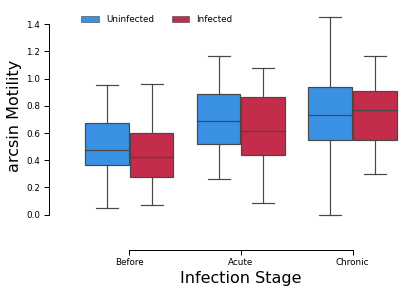
mod\_motility = smf.mixedlm("MOTILITY ~ C(INFECTION\_STAGE) \* C(TREATMENT)",

df\_motility,

groups = df\_motility["ID"])

mod\_motility\_fit = mod\_motility.fit()





# 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.