Designing Exposure Scores to Examine Influenza Transmission among Ferrets

Sean Tavares
Yutian Huang
Ruohao Wu
Xiaonuo Xu

36726 Statistical Practice
Prof. Valérie Ventura
March 24, 2023

Introduction

The Lakdawala Lab studies the intracellular assembly of influenza viruses and person-to-person transmission of viruses, with a specific focus on the influenza A virus. In one of their current PHIGH research projects, the lab has established a method to study the airborne transmissibility of influenza viruses using the ferret model [1]. The approach taken in this research is novel to the study of influenza virus transmission and can provide an effective way of predicting the transmission of the virus.

For this project, we are helping the Lakdawala Lab to understand the factors that can influence the transmission of the influenza virus among ferrets from a statistical perspective. We devised an automated data pipeline that can process the experiment data and generate covariates for statistical analysis. To examine the correlation between behaviors and successful transmission of the influenza virus, we applied exploratory data analysis on the covariates generated using the data pipeline. The data pipeline we devised will provide the Lakdawala Lab with tools to process their current and future experimental data for their research so that they can conduct any relevant statistical analysis they need.

Data

The data comes from BORIS (Behavioral Observational Research Interactive Software). As the name suggests, this software allows researchers to track observed behaviors from research experiments. Our client has hired coders who track ferret behaviors from each experiment using this software. They do this by watching a recorded video of the experiment, taken from multiple angles, and encoding the behaviors of interest into the BORIS file. For each behavior, there is a record of when it “START”s and when the behavior “STOP”s. The data is organized as follows:

- Time - the time at which behavior occurs or terminates (depending on Status)
- Subject - the specific ferret that is acting out the behavior of interest
- Behavior - the behavior of interest being recorded; Touching, Bite/Chew, Snuggle, etc.
- Behavioral category - classification for above behavior; either Distance, Location, or Interaction
- Modifier 1 - the target of the Subject’s Behavior; can be an inanimate object or another ferret
- Comment - any comments made by the coder for this behavior
- Status - either START (behavior begins) or STOP (behavior ends)

The durations of these behaviors are calculated by subtracting the Time when Status = START from the Time when Status = STOP. This allows us to compare the lengths of these interactions in addition to when they occurred. We developed an R markdown script to clean the experiment's data, extracting useful information and adding a variable called "Duration" which is the START Time - STOP Time. This step ensures the data is well-structured and free from inconsistencies or errors, paving the way for subsequent analyses.

Method

The main thought of our work is the design of exposure scores, which are positive variables that increase with the duration of the donor and recipient interactions. We define interactions with 3 words: the modifier the recipient and donor interact with, and the actions of the donor and recipients. For example, there are some combinations of the 3 qualities. Additionally, for a given 3 qualities, we calculated synchronous and asynchronous scores. A synchronous score is calculated for interactions where the donor's action and the recipient's action happened to the modifier at the same time. An asynchronous score is calculated for interactions where the recipient action happens after the donor action. We separate these two scores because, during synchronous interactions, there is also the possibility of direct transmission of the virus from donor to recipient. It is important to take this into account when comparing different combinations of actions, so we decided to treat these as distinct interactions.

1. **Feature Engineering**
   In this stage, we created another script to generate extra features not initially present in the dataset. These additional covariates enable us to examine the interactions between ferrets and specific objects more thoroughly. For instance, we investigated the frequency and duration of contact between ferrets and objects *immediately* following an interaction by the Donor on that same object. The duration metric is of particular interest since both the client and our team agreed that intuitively, these "first-time" interactions have the highest likelihood of successful transmission compared to subsequent interactions. We calculated a weighted sum of these "first-time" interactions based on the following factors (see example in Figure 1):
    1. The length of time the Donor interacted with the object (marked as d)
    2. The duration of contact between the other ferret and the object (marked as e)
    3. The time elapsed between the end of the Donor's interaction and the start of the subsequent ferret's connection with the same object (marked as b)

   According to the factors we created and variables we already had, we created a table to include all the information we used in this study. The table includes these variables: Donor, Donor action, object (the modifier interacted with by both Donor and Recipient), variable d (the duration that the Donor contacted with the object), recipient ferret (one of R1, R2, R3, R4), other ferrets' action, b (the duration of other ferret's contacted with the same modifier after Donor contacted), e (the duration of other ferret contacted the object). This table can be used to build models and do analyses in the future. For now, we only have the table for one experiment, so we did not build a model but focused on EDA results.

2. Rules for Assigning Weights

1. The longer the donor interacts with the object, the greater the weight assigned.
2. The longer the recipient touches the object, the greater the weight assigned.
3. The more elapsed time between the donor and recipient interacting with the object, the lower the weight assigned.

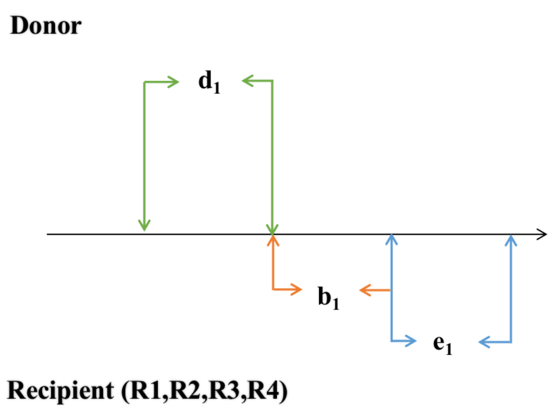

Figure 1. Simple interaction situation

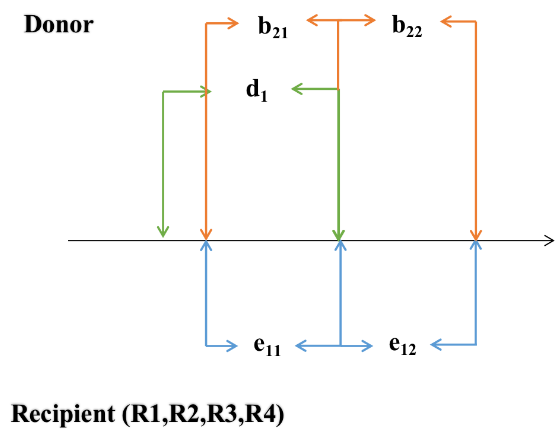

Figure 2. Recipient start during the Donor Duration

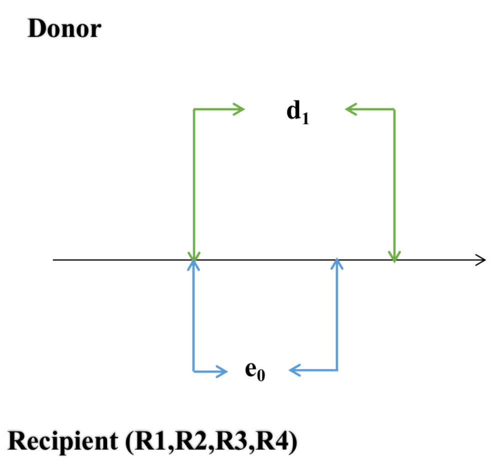

Figure 3. Recipient end time during the Donor Duration

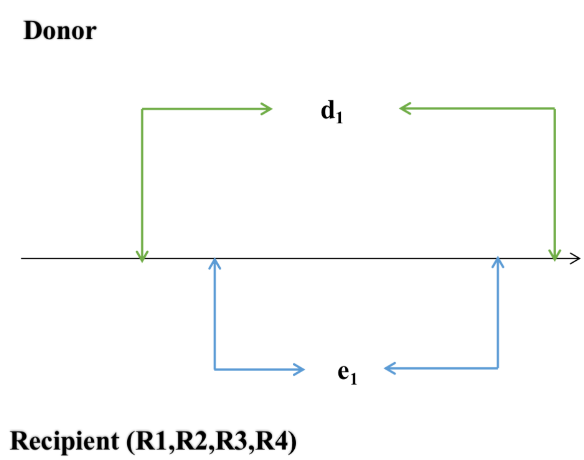

Figure 4. Recipient duration in the Donor Duration

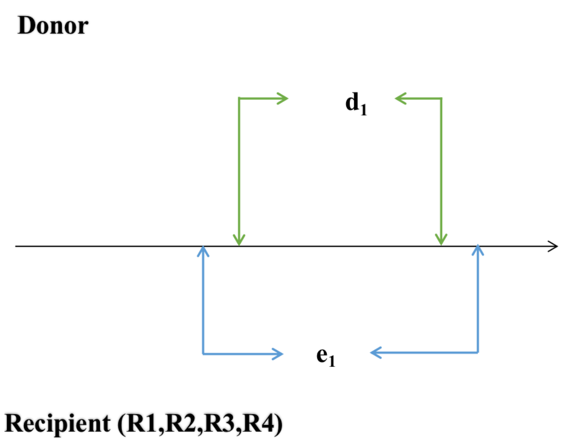

Figure 5. Donor Duration in the recipient duration

D and e are the durations of the donor's and recipient's interactions with the modifiers, respectively, and b is the time elapsed between the two interactions. The subscript 1 indicates that it is the first interaction of the donor and recipient with the modifier. Figure 1 shows the simplest situation in which the start time of the recipient contacted object is later than the Donor. Figure 2 shows a situation when the start time of the recipient is in the donor duration and the end time is out of the donor duration. We separate these two parts into Synchronous and Asynchronous interactions, at that time the elapsed time between exposures will be zero and donor duration is similar to calculating Synchronous and Asynchronous exposure scores. Figure 3 shows the end time of the recipient in the donor duration, figure 4 shows the recipient duration in the donor duration and Figure 5 shows the donor duration in the recipient duration. In these three cases, we need to calculate the exposure score by using the previous donor duration and elapsed time.

3. Weighted Functions

To compare the different behaviors of the recipient ferrets, we developed a weighted function that involves d, b, and e. The result of this weighted function is an exposure score which quantifies the interactions between the recipients and an object after the donor first interacts with it. The weighted function (*w)* is shown in *equation (1)*. It is a multiple of the functions for d, b, and e.

$$w(e,\ b,\ d)\ =\ \frac{1}{1+e^{-d}} * e^{-\alpha * b} * \frac{1}{1+e^{-\beta * e}} \quad equation\ (1)$$

The choice of the functions for d, b, and e are based on the rules for assigning weights as described in section 3 and were verified with our clients. The parameters α and β are decay parameters that control how much weight should be given to an exposure score based on the elapsed time between the Donor's interaction and the subsequent recipient's interaction with the same object as well as the duration of the recipient's interaction with the object. Since the amount of data we have are not enough for us to calibrate the α and β, we set both of them to 1 when computing the scores for the results of this project.

The function for d indicates the weight for d on the score will increase exponentially as d increases and approximates to 1 after a certain threshold value. An expit instead of a linear function is chosen for d because, intuitively, the longer the donor interacts with an object, the more virus is transmitted to the object. However, since the time unit in the experiment is second, the difference between the amount of virus transmitted for one second and two seconds will be very trivial. Hence, the change in the weight of d on the score should be trivial as well. If a linear function is used, the score for d=2 will be twice the score for d=1, which does not make sense. Therefore, an expit function is more appropriate.

The function for b is a negative exponential function, which indicates the exponential decay of the weight b has on the score. The time elapsed between the Donor's interaction and the subsequent recipient's interaction with the contaminated object can be also considered as the time the viruses are "hostless". As a result, the longer the time the viruses are "hostless", the less infectious they can be. Therefore, we decide to use the negative exponential function to model how much weight b should have on the score, because this function well resembles the decay of virus infectability in relation to time.

The function for e is also an expit function. The reason we have an expit function for e is that when a recipient first interacts with an object that is contaminated by the donor, it is assumed that the recipient will be getting viruses that are still infectious, hence, the score should be weighted more. As the interaction time becomes longer, the recipient might be getting more influenza virus. Nevertheless, the infectability of the virus might be lower since they've left their host, the donor, longer. As a result, the change in the weight of e on the score should be more stationary. The behavior of the expit function resembles our assumption for e.

Results

The main objective of our study is to generate flu exposure scores that we believe are correlated with flu transmission. Ultimately, when more data will be collected, we will calculate correlations between each exposure score and infection status(0/1) to determine which ferret correlates most strongly.

As previously mentioned, we have designed b, d, and e features based on the data we collected, and subsequently applied a weighted function to each of these components. To determine the impact of these features, we propose calculating the sum of b, d, and e, and applying the weighted function to each. The resulting summation score represents the degree of engagement between the first-time recipient and the modifier after the donor has interacted with it. A higher summation score indicates a stronger engagement between the recipient and the donor-modifier interaction.

The table below, derived from our output file, showcases the descending scores for different recipients. In the specific experiment we analyzed, R1 was the only ferret that was affected. The table reveals that R1 has 13 high scores among the 20 highest scores. This observation suggests that R1 had a greater degree of engagement with the donor and modifier interactions as compared to other ferrets. As a result, R1 may have been more susceptible to the influenza virus.

This output supports our final finding that R1 was infected, while the other ferrets remained unaffected. Furthermore, the generation of these features not only allowed us to identify the infected ferret but also provided a more extensive set of features for model building. This comprehensive feature set can be utilized in future studies to enhance our understanding of experimental outcomes and improve predictive models.

Subject	Donor Behavior	Recipient Behavior	Modifier	Score	Max_Score	Frequency
R4	Touching	Touching	Pen Walls	29.36	1	67
R3	Touching	Touching	Pen Walls	25.79	0.99	63
R1	Touching	Touching	Pen Walls	22.31	1	56
R2	Touching	Touching	Pen Walls	19.02	0.99	49
R1	Touching	Snuggle	R3	16.5	0.98	32
R1	Chase	Snuggle	R3	14.46	0.96	27
R1	Snuggle	Snuggle	R3	13.93	0.98	25
R1	Snuggle	Snuggle	R4	13.03	0.97	21
R1	Touching	Touching	R2	11.65	1	34
R1	Snuggle	Snuggle	R2	11.58	1	21
R1	Touching	Touching	Pen Walls	10.97	1	37
R1	Chase	Snuggle	R2	10.88	1	21
R4	Touching	Touching	Plush	10.79	1	39
R4	Touching	Touching	Pen Walls	10.66	0.96	32
R1	Touching	Snuggle	R4	9.64	0.9	19
R1	Chase	Snuggle	R4	9.46	0.78	17
R2	Snuggle	Touching	R4	9.25	1	32
R1	Touching	Snuggle	R2	9.21	0.99	18
R1	Chase	Touching	R2	8.99	0.95	46
R2	Touching	Touching	Cat Tower	8.85	0.96	36
R4	Touching	Touching	R1	8.14	1	17

Table 1.1 Part of final output in descending order of score

The summation table is one of our final outputs. The first column shows the recipient ferret which interacted with the modifier after/during the Donor's interaction with the same modifier. The second column shows the behavior of the Donor on the modifier The third column shows the behavior of the recipient on the modifier. The modifier column shows which object or ferret they contacted. The last three columns contain the total scores, max score, and frequency.

Discussion

If the exposure scores truly measure transmissibility as intended, then the ferrets which are not infected will have lower exposure scores compared to the scores of the ferrets which are infected. As shown in the results, the exposure scores for the ferrets in this first experiment are higher for the ferret R1 which was infected by the donor. In repeated experiments, the expectation is that these exposure scores will continue to be larger for the ferrets which are infected. These exposure scores can then be used as covariates from repeated experiments to determine which objects and behaviors are more correlated with the successful transmission of influenza than others. One approach that seems suitable would be to use logistic regression where the exposure scores would serve as predictors, and the outcome would be a binary variable denoting whether or not the particular ferret was infected. It is important to emphasize that a single set of exposure scores from any one experiment cannot be used to determine which behaviors lead to successful virus transmission, but they can indicate which behaviors may be more contagious than others. With multiple sets of exposure scores across multiple experiments, we are confident that these scores will serve as valuable and information-rich covariates in future statistical models.

References

[1] *Lakdawala Lab*. EMORY UNIVERSITY SCHOOL OF MEDICINE. (n.d.). Retrieved March 23, 2023, from https://med.emory.edu/departments/microbiology-immunology/research/labs/lakdawala/index.html