Antibody Response Induced by HIV Vaccines and T-cell Suppression Treatments in Rhesus Macaques

Group 3: Kan Luo, Shih-Ni Prim, Frederick Davey, Rizwana Rehman

11/12/2020

Contents

1 Introduction

1.1 About the Study

A dominant vaccine development strategy is to induce neutralizing antibodies by immunizing humans with the virus' glycoproteins. However, HIV vaccines that adopted this strategy mostly failed due to the fact that HIV is an RNA virus, which mutates rapidly to escape the inhibition of neutralizing antibodies. By the time the body generates neutralizing antibodies against the glycoproteins of some HIV strains, the RNA virus has already mutated. Thus, the existing neutralizing antibody fails to recognize, bind with, and neutralize the HIV virus. One possible solution is to increase the number

of potential neutralizing antibodies that will cycle in the body by releasing a variety of antibodies after glycoprotein immunization.

During the experiment, 20 rhesus macaques were given glycoprotein immunization and supplemental antibody doses, as well as one of three treatments (two experimental regulatory T-cell suppression treatments and one control). For the analysis of mutation frequency and CDR3 count, each antibody within the same treatment was treated as an independent observation.

Regulatory T (Treg) cells prevent autoimmune diseases and suppress allergic reactions by inhibiting adaptive antibody immune response in the germinal center. Theoretically, this adaptive response lowers the effectiveness of vaccines. Thus the experiment used T-cell suppression treatments to investigate the effect on immunization. These drugs are widely used in post transplant immuno-suppression treatment to prevent rejection.

While we might expect different variance within subject vs between subjects, the number of potential antibodies observed is much higher than the number sampled in a blood draw.

1.2 About the Dataset

Our dataset includes measurements of antibodies measured in 20 rhesus macaques after they were given the same HIV vaccine at three different time-points and one of three randomly selected anti-Treg treatments(drugs). Blood samples were collected two weeks after vaccine dosing, and antibodies were isolated from those samples. A different number of antibodies were collected from each blood sample, limited by assay yield. Each observation contains information about the antibody isolated post the glycoprotein immunization.

A human antibody is formed by a heavy chain and light chain. For heavy chain, a human has about 51 V-gene segments, 25 D-gene segments and 6 J-gene segment. For light chain there are 71 V-gene segments and 9 J-gene segments[ref.5]. Any heavy chain V-D-J combination and light chain V-J combinations can randomly happen in germline center. Theoretically, there can be $51 * 25 * 6 * 71 * 9 = 4.88835 \times 10^6$ combinations of gene segments. Considering the frequently happened mutation and other factors, each individual can have over **10 billion** different antibodies. Thus, we decided to follow the convention of vaccine studies and treat each antibody as independent. [Kan – can you identify a reference article or journal here that uses this convention? We don't need to quote it / change the answer, it's just defending our claim that it's a standard practice]

Below is the list of variables with a brief description from our dataset. Please note that in each antibody, there are two sets of heavy chain and light chain, all of which forming a Y-shape immunoglobulin. Thus many of the variables start with H or L, indicating which chain the information comes from.

1.3 List of variables

- Monkey_id: Lists the identity of monkey
- Treatment(Drug): Treatment A is the mock control, and treatment B and C are two different kinds of Treg inhibitor treatments.
- Time_Points: 0 represents before immunization; 1 represents 2 weeks post 1st immunization; 2 represents 2 weeks post 2nd immunization; and 3 represents 2 weeks post 3rd immunization, respectively.

- Isotype: The category of antibody type; there are 5 kinds of immunoglobulin isotypes: IgG, IgA, IgM, IgE, IgD. The two most important kinds are IgG and IgM. IgM occurs in the acute stage of infection and perform an role of primary response. The secondary response IgG appears later in serum with higher binding affinity, and neutralizing potentials against toxins and virus. IgA mostly found in mucosal tissues such as Nasal mucosa. Non-dominant IgD and IgE are typically lower than 1% in blood.
- H_ID and L_ID: heavy chain and light chain IDs for the particular observation
- H_VBase: the number of nucleotide of the heavy chain variable region
- H_Substitutions: the number of relative nucleotide mutations in heany chain.
- HMuFreq: calculated by H_Substitutions / H_VBase
- H_CDR3: the number of amino acid of the heavy chain's third complementarity determining region
- L_VBase: the number of nucleotide of the light chain variable region
- L_Substitutions: the number of relative nucleotide mutations in light chain.
- LMuFreq: calculated by L_Substitutions / L_VBase
- L_CDR3: the number of amino acid of the light chain's third complementarity determining region.`H_CDR3` and `L_CDR3` indicates the length of the third complementarity-determining region on the variable heavy chain and light chain. The longer they are, the more potential there is to produce diverse antibodies. [Kan, could you check to see if this is correct?] In other words, we want the values to be higher.
- Binding: affinity of antibodies against a selected HIV glycoprotein. The larger value indicates stronger binding.`Binding` indicates the rate of neutralizing, meaning how much the antibodies bind with the virus and thus make the virus ineffective. This is the most important measure of the study.

1.4 Research questions

In the current report, we test if the different Treg treatments and number of vaccine injections(? which variables indicates this) cause changes in the antibody characteristics and if the changes are related to immunization/treatment time points.

Specifically, we evaluate:

$Q1$. Do treatments, and/or isotypes have effects on the mutation frequency and/or the amino acid count in the third complementarity determining region (CDR3)?

$Q2$. How does the binding strength of the antibodies develop in response to the number of vaccine dosages by treatment? This will be evaluated with a longitudinal analysis by test subject.[NEEDS TO BE UPDATED]

2 Methods

For exploratory data analysis we used histograms and Q-Q plots. We also computed correlations among variables.

2.1 Data Summaries

The dataset has 2465 antibodies collected from 20 rhesus monkeys. We first present our exploratory data analysis and summaries. An overview of the number of antibodies collected per treatment at each time-point is given below (Figure 1) along with tables for treatment counts per time-point and Isoptype per treatment for individual time-points.

[Figure 1 about here.]

[Table 1 about here.]

[Table 2 about here.]

Next, the histogram of `Isotype` is presented in Figure 2.

[Figure 2 about here.]

As expected, we observed that IgG and IgM occupied the biggest proportion of all antibodies in all time points. Before immunization (time point 0), there were similar weight of IgG and IgM found in blood. After the 1st immunization (time point 1), primary immune response resulted an increase of IgM, followed with IgG increase at later time point 2 and 3.

For the response variables, we began with an outlier check, and then reviewed their distributions across treatment and time-points.

[Figure 3 about here.]

[Table 3 about here.]

[Figure 4 about here.]

L_CDR3	H_CDR3	d2	Z
47	14	952.8125	30.8110233

```
## [1] 972
```

Row 972 from `Data2` is in fact an outlier, as shown in the summary and Table 3, Figure 3, and Figure 4. The value for `L_CDR3` is quite unlikely. Since we can't go back to the original data, we remove the data point and will use the new dataset `Data3`.

We examined our responses: `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq` and `Binding`.

We observed that for `H_CDR3` the distributions were roughly normal with the center around 13 at different time-points (Figure 4) without taking into account different treatments. Figure 5 represents the distribution of `H_CDR3` with respect to treatments at different time-points, and slightly centered around 9 for `L_CDR3` at different time- points (Figure 5).

[Figure 5 about here.]

[Figure 6 about here.]

Figures for L_CDR3 are shown next.

[Figure 7 about here.]

[Figure 8 about here.]

Q-Q plots of H_CDR3 and L_CDR3 are shown below.

[Figure 9 about here.]

HMuFreq and LMuFreq were calculated by dividing H_Substitution by H_VBase for heavy chain and similarly for light chain. These two variables show how much the antibodies mutate. A higher mutation rate is usually indicative of better virus neutralization. Below we present comparison of mutation rate between heavy chain and light chain.

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

Next, the Q-Q plots of HMuFreq and LMuFreq are shown.

[Figure 14 about here.]

Lastly, Histogram of Binding with respect to treatments at different time points and Q-Q plot are shown. We observed that Binding was not normally distributed. However, since our sample size is larger than 2000, we can use the Central Limit Theorem and assume normality.

[Figure 15 about here.]

[Figure 16 about here.]

We checked whether response variables could be correlated. In the plot below, we observed that none of the response variables were highly correlated.

[Figure 17 about here.]

2.2 Multivariate Data Analysis

For $Q1$ We wanted to test whether predictors Drug and Isotype had effects on the five responses: H_CDR3, HMuFreq, L_CDR3, LMuFreq, and Binding. We used $MANOVA$ due to large sample size (n = 2464) without worrying about the normality assumption. In the output, we noticed that both the main effects of Drug and Isotype are significant.

```
##             Df   Pillai approx F num Df den Df    Pr(>F)
## drug         2 0.056004   14.122     10   4902 < 2.2e-16 ***
## it           4 0.295992   39.204     20   9812 < 2.2e-16 ***
## Residuals 2454
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.2.1 Pairwise comparison

To know more details about which groups have different means, we use pairwise comparisons for each treatment group, drug, and Isotype. (We set α as 0.05, and use a Bonferroni correction.)

```
## [1] "L_CDR3  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2      0.0451 0.0450 2458  -0.0870    0.177
##  1 - 3      0.1018 0.0492 2458  -0.0429    0.246
##  2 - 3      0.0567 0.0480 2458  -0.0844    0.198
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "LMuFreq  pairwise CI's"
##  contrast  estimate      SE   df lower.CL upper.CL
##  1 - 2    -0.000176 0.00256 2458 -0.00770  0.00735
##  1 - 3     0.002320 0.00280 2458 -0.00592  0.01056
##  2 - 3     0.002496 0.00273 2458 -0.00554  0.01053
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "H_CDR3  pairwise CI's"
##  contrast estimate    SE   df lower.CL upper.CL
##  1 - 2      0.4907 0.160 2458   0.0204    0.961
##  1 - 3      0.4438 0.175 2458  -0.0712    0.959
##  2 - 3     -0.0469 0.171 2458  -0.5492    0.456
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "HMuFreq  pairwise CI's"
##  contrast estimate      SE   df lower.CL  upper.CL
##  1 - 2      -0.006 0.00199 2458 -0.01186 -0.000145
##  1 - 3       0.011 0.00218 2458  0.00461  0.017432
##  2 - 3       0.017 0.00213 2458  0.01076  0.023276
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "Binding  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2       0.118 0.0522 2458  -0.0355    0.271
##  1 - 3      -0.356 0.0571 2458  -0.5235   -0.188
##  2 - 3      -0.473 0.0557 2458  -0.6372   -0.310
```

```
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667

## [1] "L_CDR3  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  A - D      0.1136 0.1510 2456  -0.3838    0.611
##  A - E      0.0667 0.3289 2456  -1.0170    1.150
##  A - G      0.0041 0.1364 2456  -0.4451    0.453
##  A - M      0.0441 0.1361 2456  -0.4042    0.492
##  D - E     -0.0469 0.3091 2456  -1.0651    0.971
##  D - G     -0.1095 0.0769 2456  -0.3627    0.144
##  D - M     -0.0695 0.0764 2456  -0.3210    0.182
##  E - G     -0.0626 0.3022 2456  -1.0581    0.933
##  E - M     -0.0225 0.3021 2456  -1.0177    0.973
##  G - M      0.0400 0.0404 2456  -0.0931    0.173
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "LMuFreq  pairwise CI's"
##  contrast  estimate      SE   df lower.CL upper.CL
##  A - D    -0.021206 0.00858 2456 -0.04947  0.00705
##  A - E    -0.006829 0.01869 2456 -0.06840  0.05475
##  A - G    -0.023476 0.00775 2456 -0.04900  0.00205
##  A - M    -0.021669 0.00773 2456 -0.04714  0.00380
##  D - E     0.014376 0.01756 2456 -0.04348  0.07223
##  D - G    -0.002270 0.00437 2456 -0.01666  0.01212
##  D - M    -0.000463 0.00434 2456 -0.01476  0.01383
##  E - G    -0.016647 0.01717 2456 -0.07321  0.03992
##  E - M    -0.014839 0.01716 2456 -0.07138  0.04170
##  G - M     0.001807 0.00230 2456 -0.00576  0.00937
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "H_CDR3  pairwise CI's"
##  contrast estimate    SE   df lower.CL upper.CL
##  A - D       0.297 0.530 2456   -1.448    2.043
##  A - E      -2.229 1.154 2456   -6.032    1.574
##  A - G      -0.728 0.479 2456   -2.304    0.849
##  A - M       0.503 0.478 2456   -1.070    2.076
##  D - E      -2.527 1.085 2456   -6.100    1.046
##  D - G      -1.025 0.270 2456   -1.914   -0.137
```

```
##  D - M       0.206 0.268 2456   -0.677    1.089
##  E - G       1.502 1.060 2456   -1.992    4.995
##  E - M       2.733 1.060 2456   -0.760    6.225
##  G - M       1.231 0.142 2456    0.764    1.698
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "HMuFreq  pairwise CI's"
##  contrast estimate      SE   df lower.CL upper.CL
##  A - D     0.00523 0.00659 2456  -0.0165  0.02693
##  A - E     0.00848 0.01435 2456  -0.0388  0.05576
##  A - G    -0.01628 0.00595 2456  -0.0359  0.00332
##  A - M     0.00393 0.00594 2456  -0.0156  0.02349
##  D - E     0.00325 0.01349 2456  -0.0412  0.04768
##  D - G    -0.02151 0.00335 2456  -0.0326 -0.01046
##  D - M    -0.00130 0.00333 2456  -0.0123  0.00968
##  E - G    -0.02476 0.01319 2456  -0.0682  0.01868
##  E - M    -0.00455 0.01318 2456  -0.0480  0.03888
##  G - M     0.02021 0.00176 2456   0.0144  0.02602
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
## [1] "Binding  pairwise CI's"
##  contrast estimate     SE   df lower.CL upper.CL
##  A - D     0.30395 0.1610 2456   -0.227    0.834
##  A - E     0.20847 0.3508 2456   -0.947    1.364
##  A - G    -0.66527 0.1454 2456   -1.144   -0.186
##  A - M     0.30220 0.1451 2456   -0.176    0.780
##  D - E    -0.09548 0.3296 2456   -1.181    0.990
##  D - G    -0.96922 0.0820 2456   -1.239   -0.699
##  D - M    -0.00175 0.0814 2456   -0.270    0.267
##  E - G    -0.87374 0.3223 2456   -1.936    0.188
##  E - M     0.09373 0.3222 2456   -0.968    1.155
##  G - M     0.96747 0.0431 2456    0.825    1.109
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999
```

Here are the pairs that have significant differences:

- Drug
  - L_CDR3: none

  - LMuFreq: none
  - H_CDR3:
    * 1 > 2
  - HMuFreq:
    * 1 < 2
    * 1 > 3
    * 2 > 3
  - Binding:
    * 1 < 3
    * 2 < 3

- Isotype
  - L_CDR3: none
  - LMuFreq: none
  - H_CDR3:
    * IgD < IgG
  - HMuFreq:
    * IgD < IgG
    * IgG > IgM
  - Binding:
    * IgA < IgG
    * IgD < IgG
    * IgG > IgM

In short, `L_CDR3` and `LMuFreq` do not have significant paired differences.

For `HMuFreq`, drug 2 has the highest mean, followed by drug 1 and control. More specifically, treatment groups 5 and 6 (two doses in drug 2) have the highest mutation rates. IgG has higher mutation rate than IgD.

For `Binding`, drug 3 has the highest mean, but drug 1 and drug 2 do not have significant differences. IgG has higher binding rate than IgA, IgD, and IgM.

We could conclude that the drug groups do increase mutation rate; however, they do not increase binding rate. That is to say, although the treatments do help increase the diversity of antibodies, they are not specific to the HIV antigens and thus do not increase binding.

2.3 Longitudinal Analysis

For the longitudinal analysis of binding strength vs number of vaccine doses, we will be using the gls function from the nlme package[7]. (NEEDS TO UP DATED)

To answer our research question Q2, **How does the binding strength of the antibodies develop in response to the number of vaccine dosages by treatment?**, we use longitudinal data analysis, including general linear models and linear mixed models.

As seen in Figure 18 and Figure 19, the mean trend is not linear, and the different time points have different variances. This information suggests that we should use piecewise linear models and set variances as unequal over time.

[Figure 18 about here.]

[Figure 19 about here.]

We first consider a model with time point as the only covariate:

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + e_{ij}$$

Thus we will use a piecewise linear model, in which each segment has different intercepts and slopes. We use three indicator variables: $S1, S2, S3$ as the indicator variables, where

$$S1 = \begin{cases} 1 & \text{if } 0 \leq \text{Timepoint} < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$S2 = \begin{cases} 1 & \text{if } 1 \leq \text{Timepoint} < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$S3 = \begin{cases} 1 & \text{if Timepoint} \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

The new model is thus

$$Y_{ij} = S1(\beta_0 + \beta_1 Time_{ij}) + S2(\beta_2 + \beta_3 Time_{ij}) + S3(\beta_4 + \beta_5 Time_{ij}) + e_{ij}$$

We also want to make sure that the trend is continuous at timepoint = 1 and 2.

Our model is $Y_{ij} = \beta_0(S1 + 2S2 - S2Time_{ij}) + \beta_1(S1Time_{ij} + 2S2 - S2Time_{ij}) + \beta_4(-S2 + S2Time_{ij} + S3) + \beta_5(-2S2 + 2S2Time_{ij} + S3Time_{ij}) + e_{ij}$ where

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

Again, our final mean model is

$$Y_{ij} = \beta_0(S1 + 2S2 - S2Time_{ij}) + \beta_1(S1Time_{ij} + 2S2 - S2Time_{ij}) +$$

$$\beta_4(-S2 + S2Time_{ij} + S3) + \beta_5(-2S2 + 2S2Time_{ij} + S3Time_{ij}) + e_{ij}$$

which can be written as

$$Y_{ij} = S1(\beta_0) + S1Time_{ij}(\beta_1) + S2(2\beta_0 + 2\beta_1 - \beta_4 - 2\beta_5) + S2Time_{ij}(-\beta_0 - \beta_1 + \beta_4 + 2\beta_5)$$
$$+ S3(\beta_4) + S3Time_{ij}(\beta_5) + e_{ij}$$

From the model above, we can find the intercepts and slopes for all three segments of the mean trend and make a plot, as seen in Figure 20:

S1: $-0.2221651 + 0.2432183 * time$

S2: $(2*-0.2221651+2*0.2432183-0.7699600+2*0.2432756)+(0.2221651-0.2432183+0.7699600-2*0.2432756)*time = -0.2413024 + 0.2623556 * time$

S3: $0.7699600 - 0.2432756 * time$

[Figure 20 about here.]

As shown in Figure 20, the two segments S1 and S2 look linear. So now we'll refit the model with only two piecewise sections; we'll call them S4 and S5. The new model is therefore

$$Y_{ij} = S4(\beta_0 + \beta_1 Time_{ij}) + S5(\beta_2 + \beta_3 Time_{ij}) + e_{ij}$$

$$S4 = \begin{cases} 1 & \text{if Timepoint} < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$S5 = \begin{cases} 1 & \text{if Timepoint} \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

We also want to make sure that the trend is continuous at Time_Point = 2.

Our model is then $Y_{ij} = \beta_1(-2S4 + S4Time_{ij}) + \beta_2(S4 + S5) + \beta_3(2S4 + S5Time_{ij}) + e_{ij}$ where

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

Again, our model is $Y_{ij} = \beta_1(-2S4 + S4Time_{ij}) + \beta_2(S4 + S5) + \beta_3(2S4 + S5Time_{ij}) + e_{ij}$, which can be written as $Y_{ij} = S4(-2\beta_1 + \beta_2 + 2\beta_3) + S4Time_{ij}(\beta_1) + S5(\beta_2) + S5Time_{ij}(\beta_3) + e_{ij}$

We first find the mean trend for S4 and S5:

S4: $(-2 * 0.5310975 + 0.5720853 + 2 * -0.0000723) + 0.5310975 * time = -0.4902543 + 0.5310975 * time$ S5: $0.5720853 - 0.0000723 * time$

We can make the plot again to see if the model is reasonable, as shown in Figure 21. Indeed, there is a linear line between Time_Point 0 and 2 and one between Time_Point 2 and 3. The two lines are contiuous at Time_Point 2. A comparison of AIC And BIC of these two models, shown in Table 4, indicates that the second model (`fit.gls2`) is indeed a better model. We'll add random effects to it.

[Figure 21 about here.]

[Table 4 about here.]

Next we check whether adding random effects improve the model. We assume that random effects exist in the intercept and slope. Our linear mixed model is then: $Y_{ij} = \beta_1(-2S4 + S4Time_{ij}) + \beta_2(S4 + S5) + \beta_3(2S4 + S5Time_{ij}) + b_{0i} + b1iTime_{ij} + e_{ij}$ where

$$\mathbf{b}_i \sim N\left(0, \mathbf{D} = \begin{pmatrix} D_{11} & D_{12} \\ & D_{22} \end{pmatrix}\right)$$

and

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

[Table 5 about here.]

As shown in Table 5, the model `fit.a2` (random intercept and slope, AR1 correlation, unequal variances) has the lowest AIC And BIC, so it seems the best model. We now check residuals for both models.

[Figure 22 about here.]

All of the Q-Q plots in Figure 22 are reasonable, so we'll use `fit.a2` for further analysis.

Now we would like to know if the slopes between Time_Point 0 and 2 and between Time_Point 2 and 3 equal zero. H_0 : slope of $S4 = 0$ and slope of $S5 = 0$, which means $H_0 : \beta_1 = 0$ and $\beta_3 = 0$

Thus, we can check for two tests:

$$\mathbf{L_1} = 0$$

where $\mathbf{L_1} = (1, 0, 0)$ and $= (\beta_1, \beta_2, \beta_3)^T$ and

$$\mathbf{L_2} = 0$$

where $\mathbf{L_2} = (0, 0, 1)$ and $= (\beta_1, \beta_2, \beta_3)^T$

[Table 6 about here.]

As shown in Table 6, the slop of S4 has a very small p-value, while the slope of S5 is quite large, indicating that the change in Binding rate between Time_Point 0 and Time_Point 2 is significant while the change between Time_Point 2 and Time_Point 3 is not significant. We conclude that Time_Point 2, when the monkeys had received two vaccines, had the highest Binding rate, while the last vaccine shot at Time_Point 3 did not make a difference to the Binding rate.

2.4 Add drugs as a covariate

Next we add drugs as a covariate to see if it has effects on Binding. We use two indicator variables: `D2` and `D3`, where

$$D2 = \begin{cases} 1 & \text{if Drug = 2} \\ 0 & \text{otherwise} \end{cases}$$

$$D3 = \begin{cases} 1 & \text{if Drug = 3} \\ 0 & \text{otherwise} \end{cases}$$

Assuming that the random effects are the same for each drug, our full model is:

$$Y_{ij} = \beta_1(-2S4 + S4Time_{ij}) + \beta_2(S4 + S5) + \beta_3(2S4 + S5Time_{ij}) +$$
$$\beta_4 D2(-2S4 + S4Time_{ij}) + \beta_5 D2(S4 + S5) + \beta_6 D2(2S4 + S5Time_{ij}) +$$
$$\beta_7 D3(-2S4 + S4Time_{ij}) + \beta_8 D3(S4 + S5) + \beta_9 D3(2S4 + S5Time_{ij}) + b_{0i} + b1iTime_{ij} + e_{ij}$$

where

$$\mathbf{b}_i \sim N\left(0, \mathbf{D} = \begin{pmatrix} D_{11} & D_{12} \\ & D_{22} \end{pmatrix}\right)$$

and

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

```
## Linear mixed-effects model fit by REML
##  Data: dataLDA1
##        AIC      BIC    logLik
##   3207.673 3306.373 -1586.836
##
## Random effects:
##  Formula: ~time | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6424735 (Intr)
## time        0.6319522 -0.999
## Residual    0.2329293
##
## Correlation Structure: AR(1)
##  Formula: ~1 | id
##  Parameter estimate(s):
##       Phi
## 0.4172276
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | time
##  Parameter estimates:
##         1         0         2         3
## 1.0000000 0.4112624 6.0679296 5.7289228
## Fixed effects: list(meanform3)
##         Value Std.Error   DF    t-value p-value
## v1  0.3490556 0.3191170 2438  1.0938170  0.2741
## v2 -1.1765071 0.5022272 2438 -2.3425795  0.0192
## v3  0.7940787 0.3682851 2438  2.1561521  0.0312
## v4  0.2970310 0.4663132 2438  0.6369775  0.5242
## v5 -1.4458003 0.7399858   17 -1.9538217  0.0674
## v6  0.8573187 0.5404821 2438  1.5862111  0.1128
## v7 -0.1785046 0.4859598 2438 -0.3673237  0.7134
## v8  0.3263319 0.7369987   17  0.4427849  0.6635
## v9 -0.2201034 0.5500331 2438 -0.4001638  0.6891
##  Correlation:
##    v1     v2     v3     v4     v5     v6     v7     v8
## v2 -0.610
## v3  0.846 -0.938
## v4 -0.684  0.418 -0.579
## v5  0.414 -0.679  0.637 -0.605
## v6 -0.577  0.639 -0.681  0.842 -0.938
## v7 -0.657  0.401 -0.556  0.449 -0.272  0.379
## v8  0.416 -0.681  0.639 -0.285  0.462 -0.436 -0.634
## v9 -0.567  0.628 -0.670  0.388 -0.426  0.456  0.863 -0.937
```

```
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.19647370 -0.31178413 -0.10810405  0.05433803 14.06000633
##
## Number of Observations: 2464
## Number of Groups: 20
```

Now we want to find whether the drugs have any effects. To see whether Drug 1 and Drug 2 have any difference, we want to perform a hypothesis test on $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$, thus we can do the test

$$\mathbf{L_3} = 0$$

where

$$\mathbf{L_3} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

and $= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$

To see whether Drug 1 and Drug 3 have any difference, we want to perform a hypothesis test on $H_0 : \beta_7 = \beta_8 = \beta_9 = 0$, thus we can do the test

$$\mathbf{L_4} = 0$$

where

$$\mathbf{L_4} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and $= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$

To see whether Drug 2 and Drug 3 have any difference, we want to perform a hypothesis test on $H_0 : \beta_4 = \beta_7, \beta_5 = \beta_8, \beta_6 = \beta_9$, thus we can do the test

$$\mathbf{L_5} = 0$$

where

$$\mathbf{L_5} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

and $= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8, \beta_9)^T$

We found that, as shown in Table 7, Drug 1 and Drug 2 do not have significantly different effects on Binding rates. As shown in Table 8, Drug 1 and Drug 3 do not have significantly different effects on Binding rates. Also, as shown in Table 9, Drug 2 and Drug 3 do not have significantly different effects on Binding rates. In other words, drug groups do not have signifant effects on our longitudinal model. Thus we will retain `fit.a2` as our best model.

[Table 7 about here.]

[Table 8 about here.]

[Table 9 about here.]

3 Results

4 Discussion

Test Assumptions Results Implications (Interpretation for practical significance)

5 Conclusions

6 References

The dataset, which can be found here, was provided by Kan Luo, as he was one of authors for the following four publications that used the dataset:

1. Luo K, Liao HX, Zhang R, et al. Tissue memory B cell repertoire analysis after ALVAC/AIDSVAX B/E gp120 immunization of rhesus macaques. *JCI Insight*. 2016;1(20):e88522. Published 2016 Dec 8. doi:10.1172/jci.insight.88522
2. Bradley, T., Kuraoka, M., Yeh, C.-H., Tian, M., Chen, H., Cain, D. W., . . . Haynes, B. F. (2020). Immune checkpoint modulation enhances HIV-1 antibody induction. *Nature Communications*, 11(1), 948. doi:10.1038/s41467-020-14670-w
3. Easterhoff, D., Pollara, J., Luo, K., Tolbert, W. D., Young, B., Mielke, D., . . . Ferrari, G. (2020). Boosting with AIDSVAX B/E Enhances Env Constant Region 1 and 2 Antibody-Dependent Cellular Cytotoxicity Breadth and Potency. *Journal of Virology*, 94(4), e01120-01119. doi:10.1128/jvi.01120-19
4. Wiehe, K., Easterhoff, D., Luo, K., Nicely, N. I., Bradley, T., Jaeger, F. H., Dennison, S. M., Zhang, R., Lloyd, K. E., Stolarchuk, C., Parks, R., Sutherland, L. L., Scearce, R. M., Morris, L., Kaewkungwal, J., Nitayaphan, S., Pitisuttithum, P., Rerks-Ngarm, S., Sinangil, F., Phogat, S., . Haynes, B. F. (2014). Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity*, 41(6), 909-918. https://doi.org/10.1016/j.immuni.2014.11.014

5. Lefranc MP, Giudicelli V, Ginestoux C, Bodmer J, Muller W, Bontrop R, Lemaitre M, Malik A, Barbie V, Chaume D. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res*. 1999;27:209-212. doi: 10.1093/nar/27.1.209.

6. Jenny M Woof , Dennis R Burton,Human antibody-Fc receptor interactions illuminated by crystal structures.*Nat Rev Immunol*. 2004 Feb;4(2):89-99. doi: 10.1038/nri1266.

7. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team (2020). *nlme: Linear and Nonlinear Mixed Effects Models*. R package

List of Figures

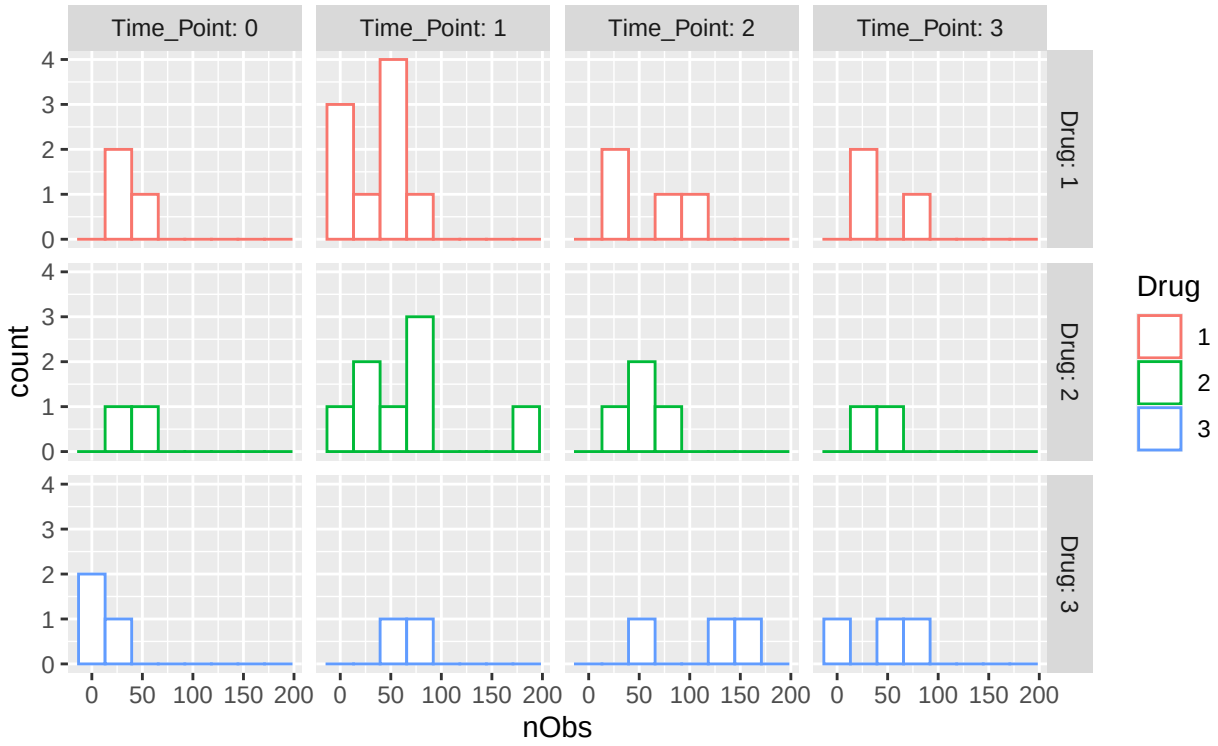

Figure 1: Histograms of Antibodies

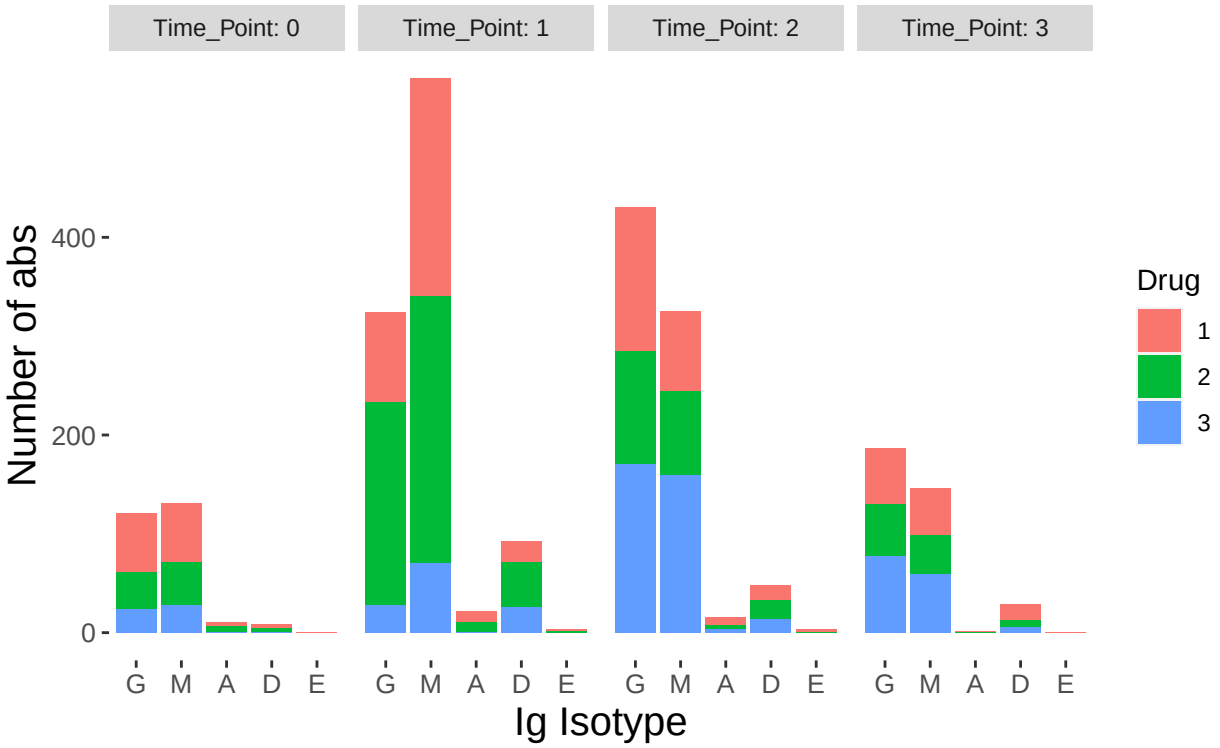

Figure 2: Histograms of Isotypes

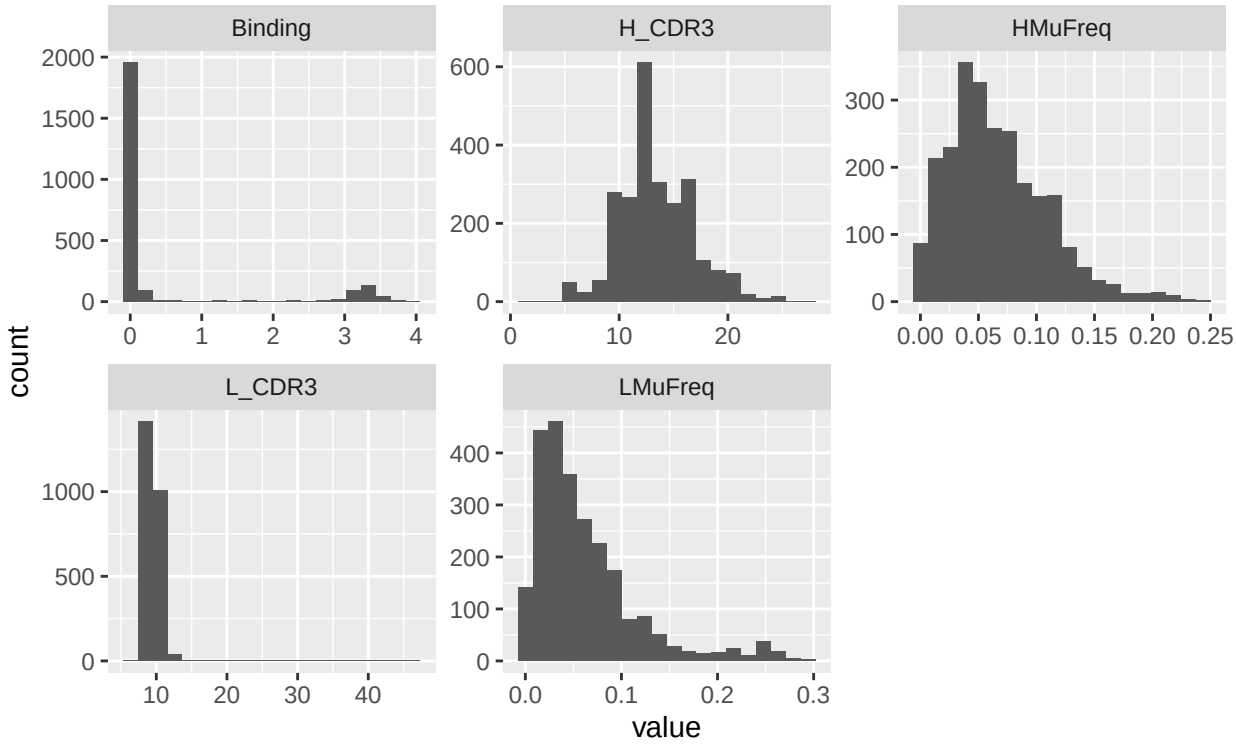

Figure 3: Histogram of Response Variables

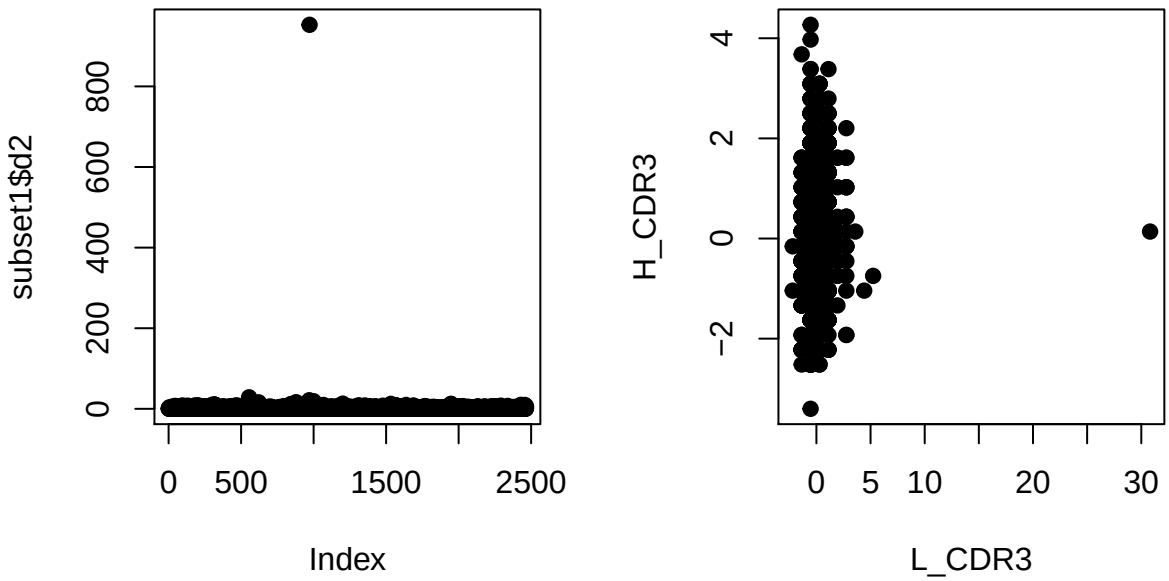

Figure 4: Mahalanobis distances and Z scores

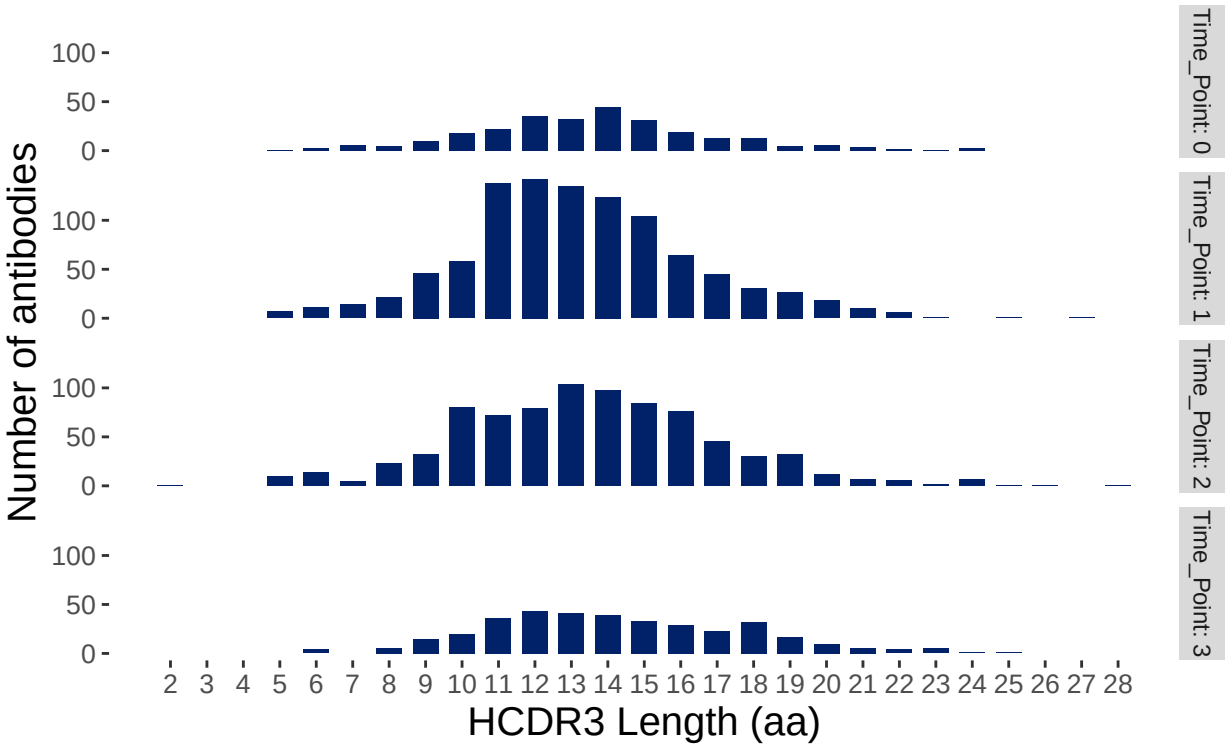

Figure 5: Histogram H_CDR3

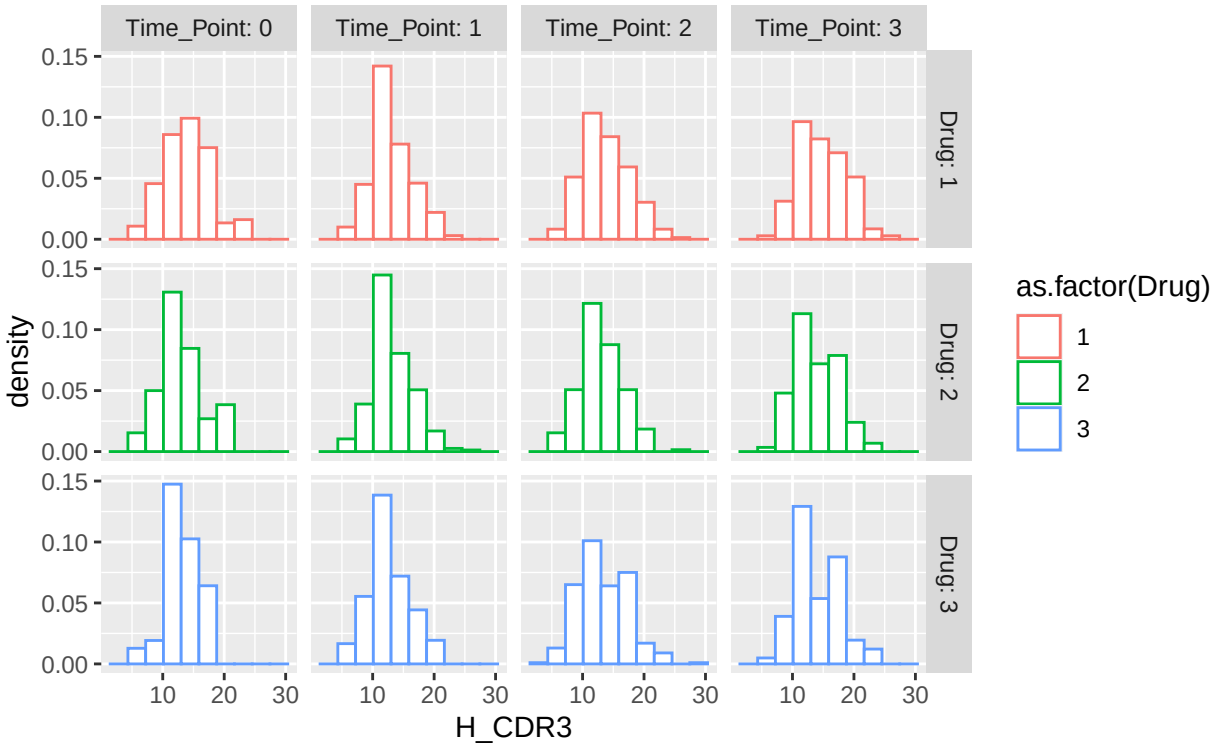

Figure 6: Histograms of H_CDR3 vs Treatment and Timepoint

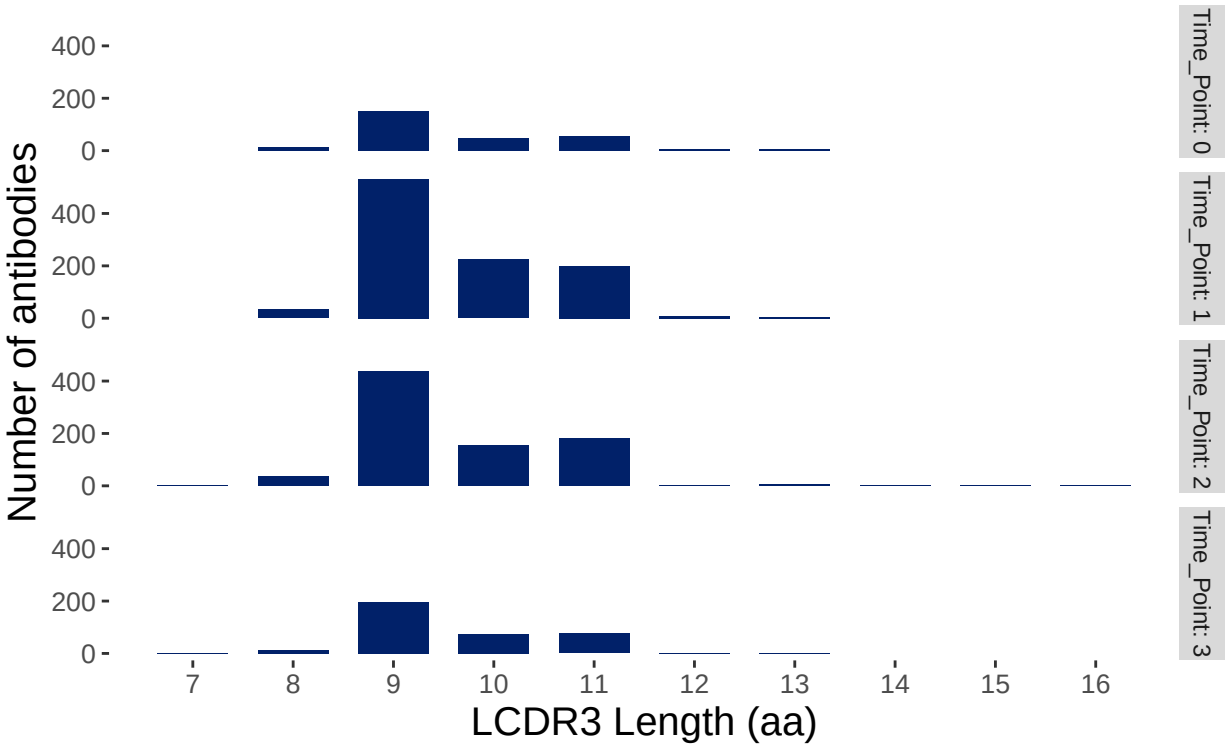

Figure 7: Histogram L_CDR3

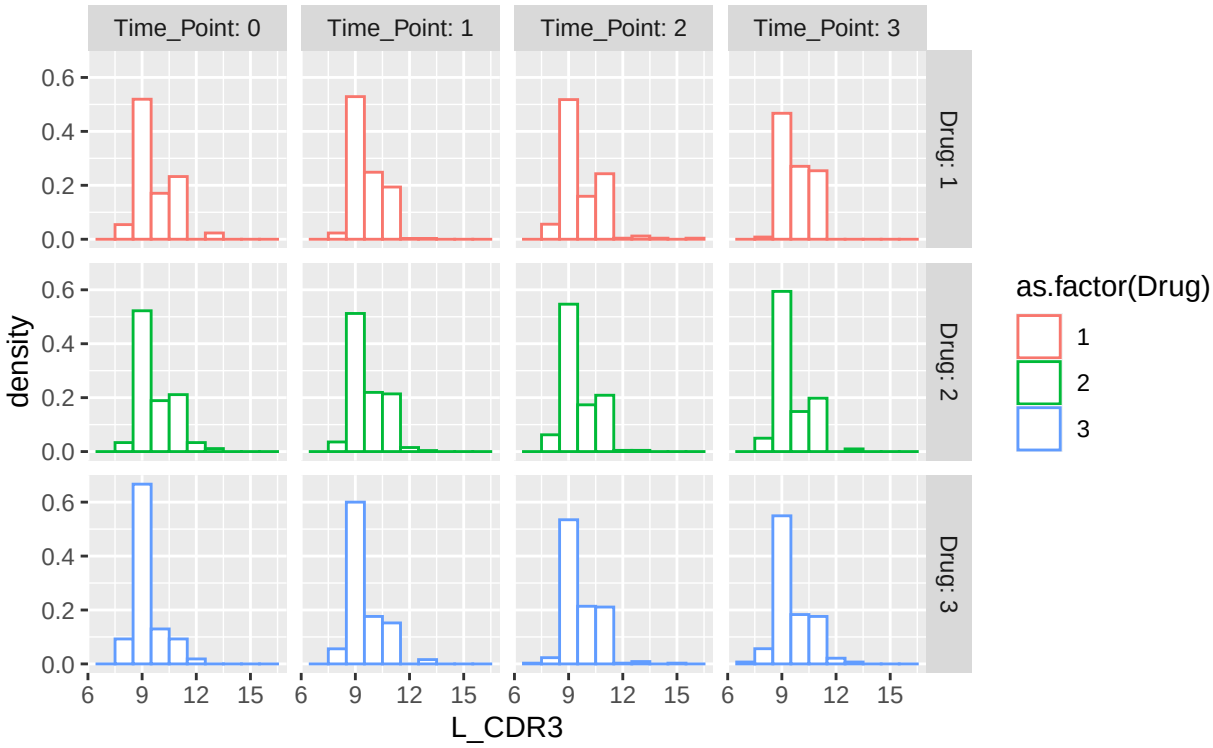

Figure 8: Histograms of L_CDR3 vs Treatment and Timepoint

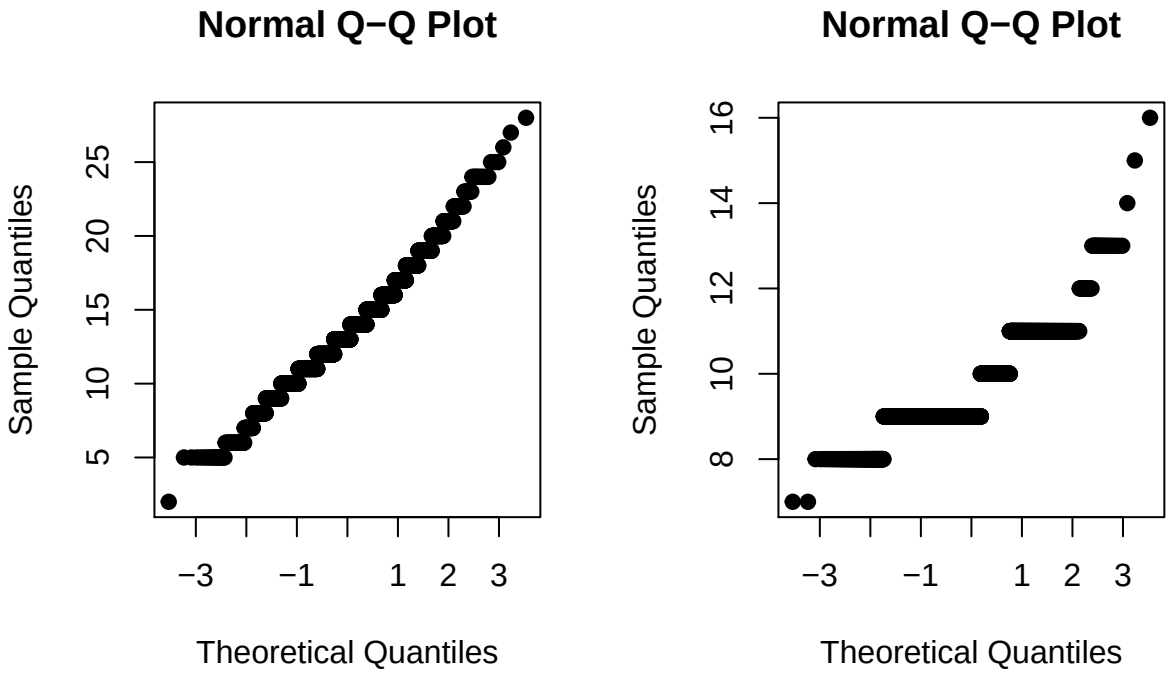

Figure 9: Q-Q Plots of H_CDR3 and L_CDR3

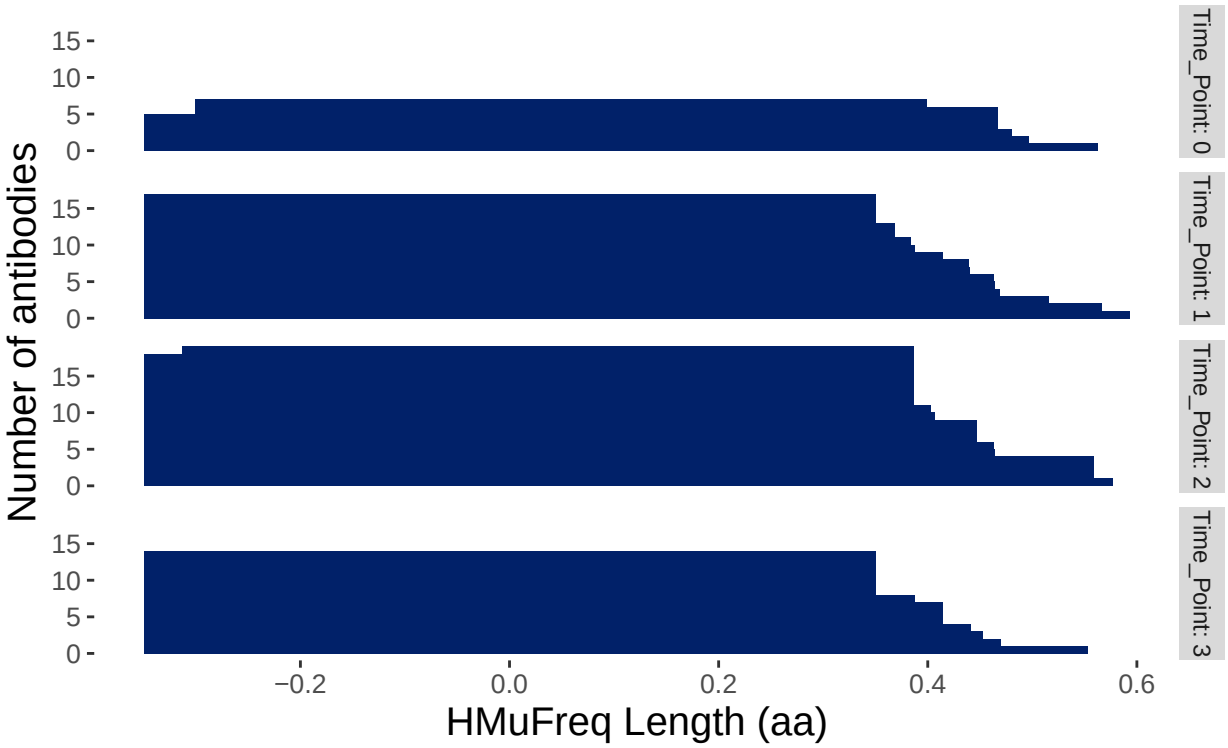

Figure 10: Histogram HMuFreq

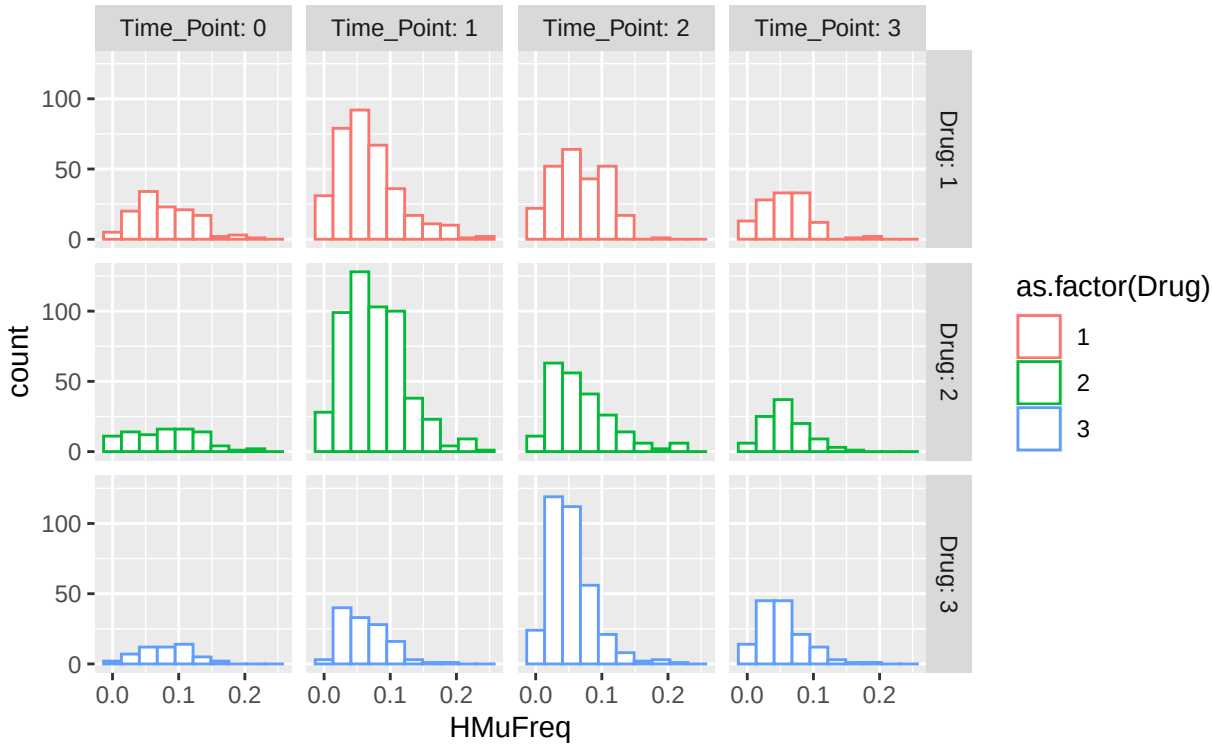

Figure 11: Histograms of HMuFreq vs Treatment and Timepoint

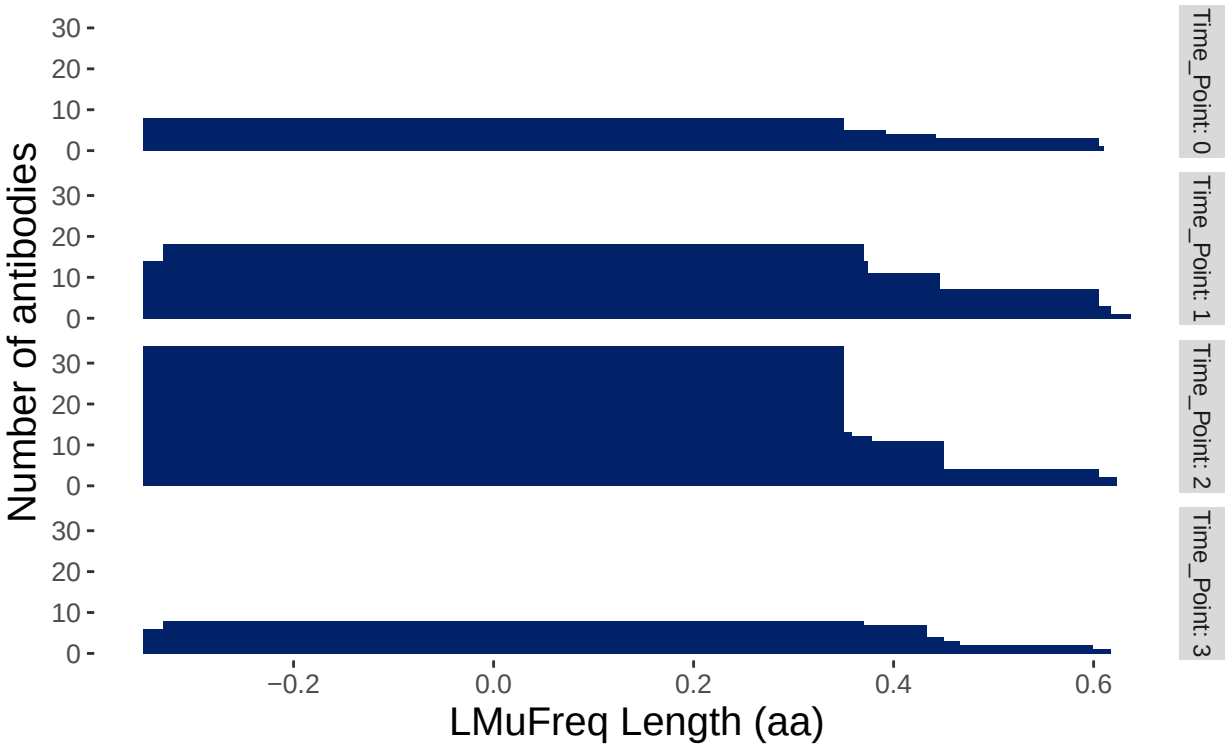

Figure 12: Histogram LMuFreq

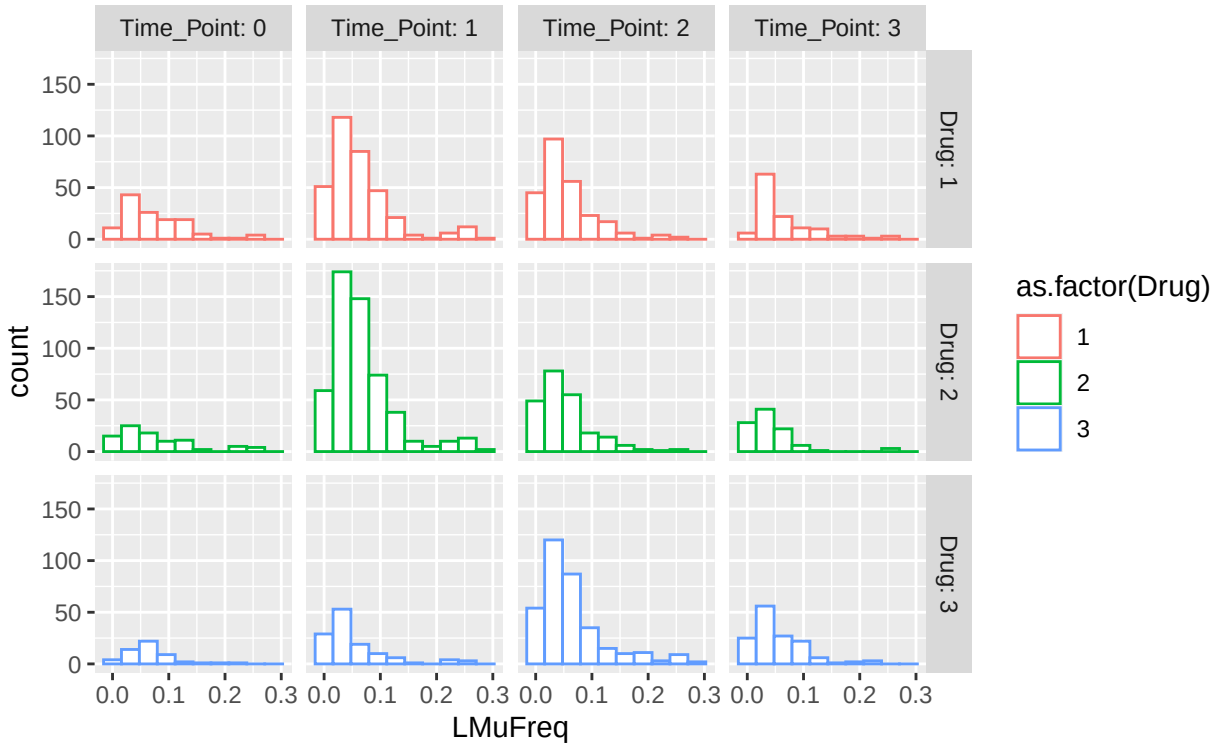

Figure 13: Histograms of LMuFreq vs Treatment and Timepoint

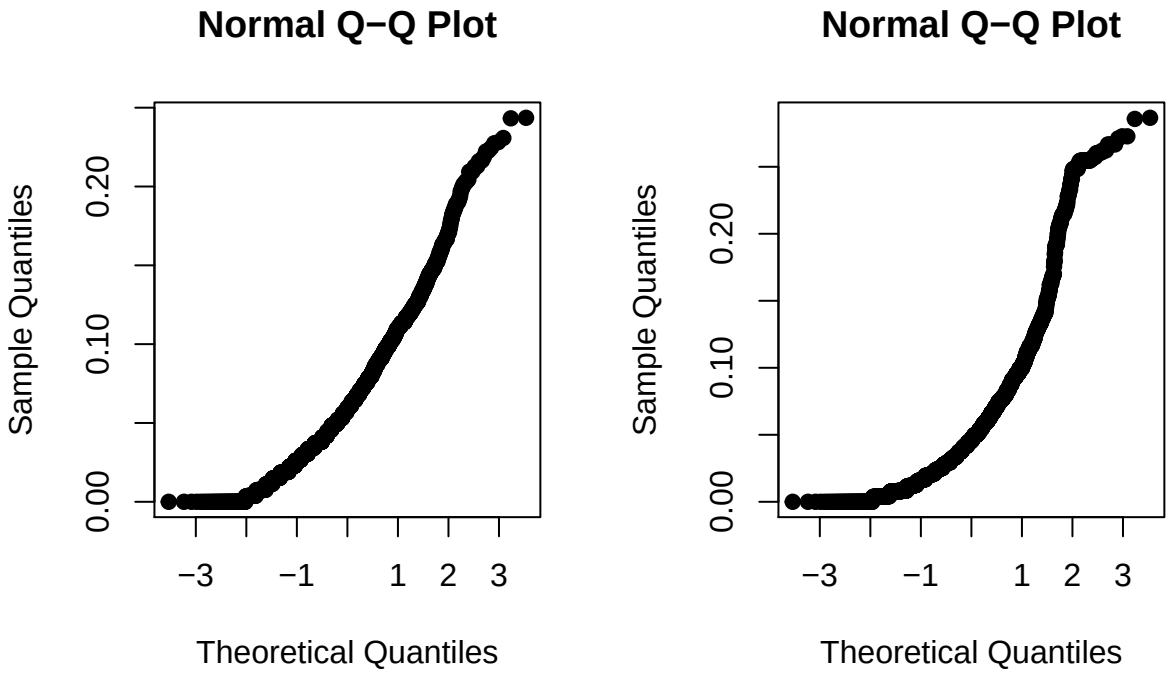

Figure 14: Q-Q Plot of HMuFreq and LMuFreq

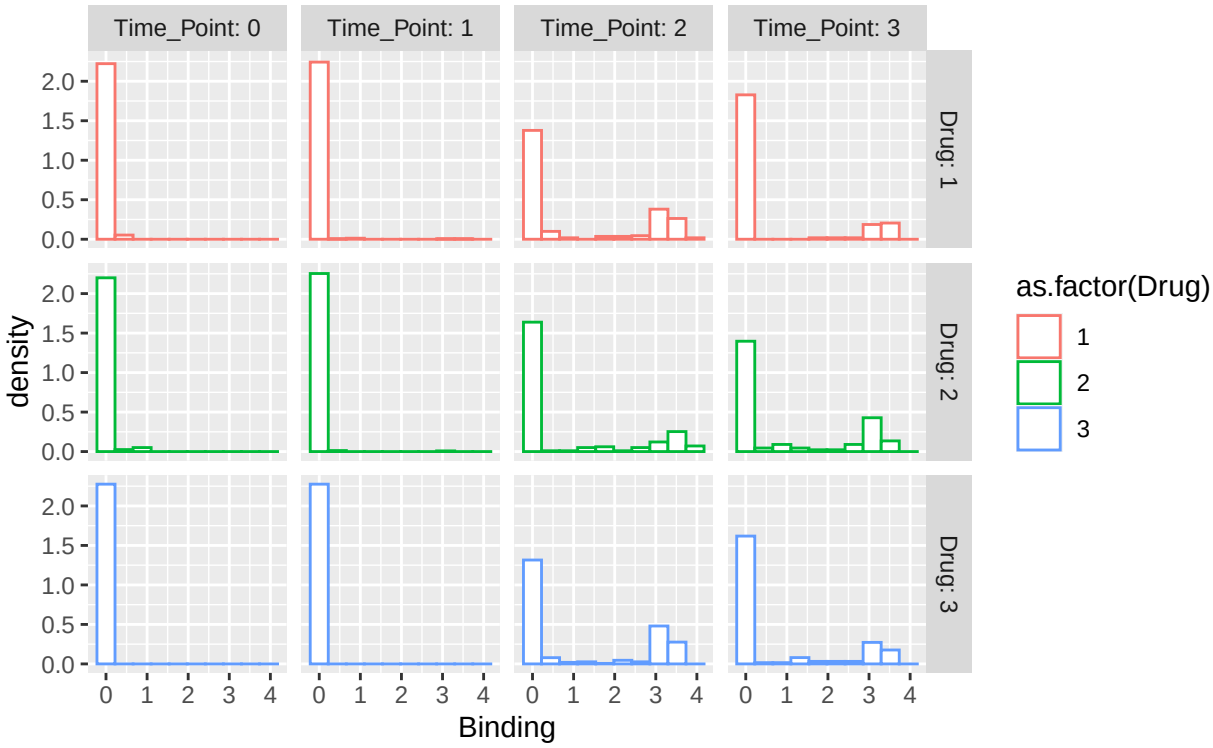

Figure 15: Histograms of Binding Strength vs Treatment and Timepoint

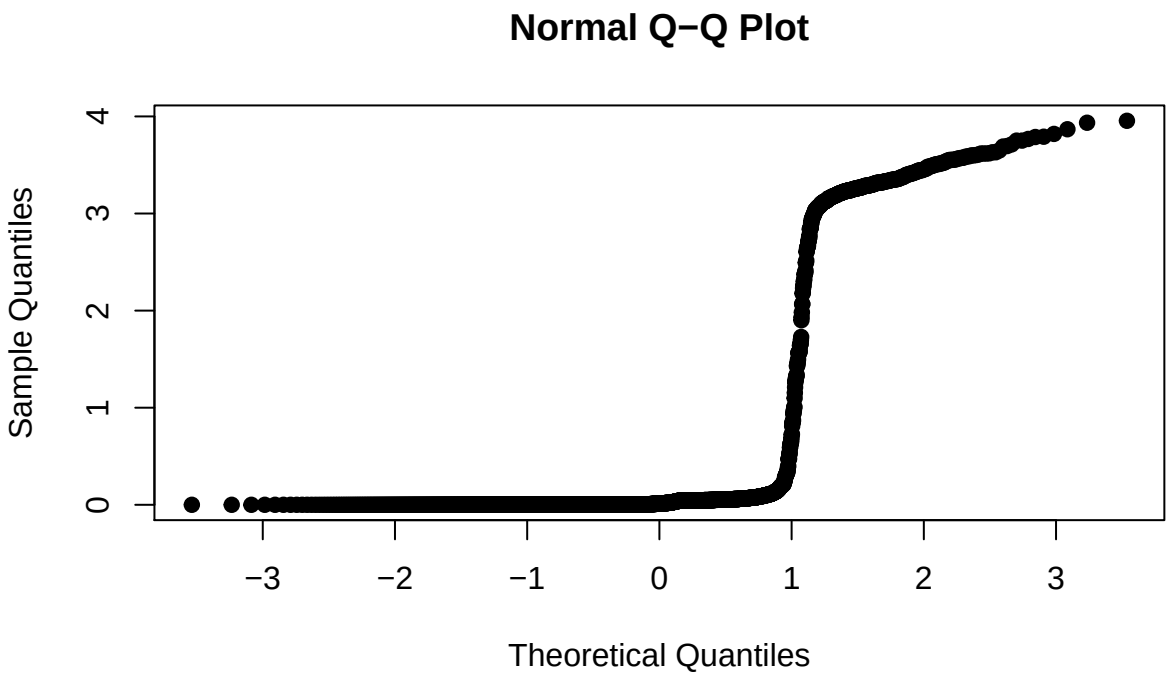

Figure 16: Q-Q Plot of Binding

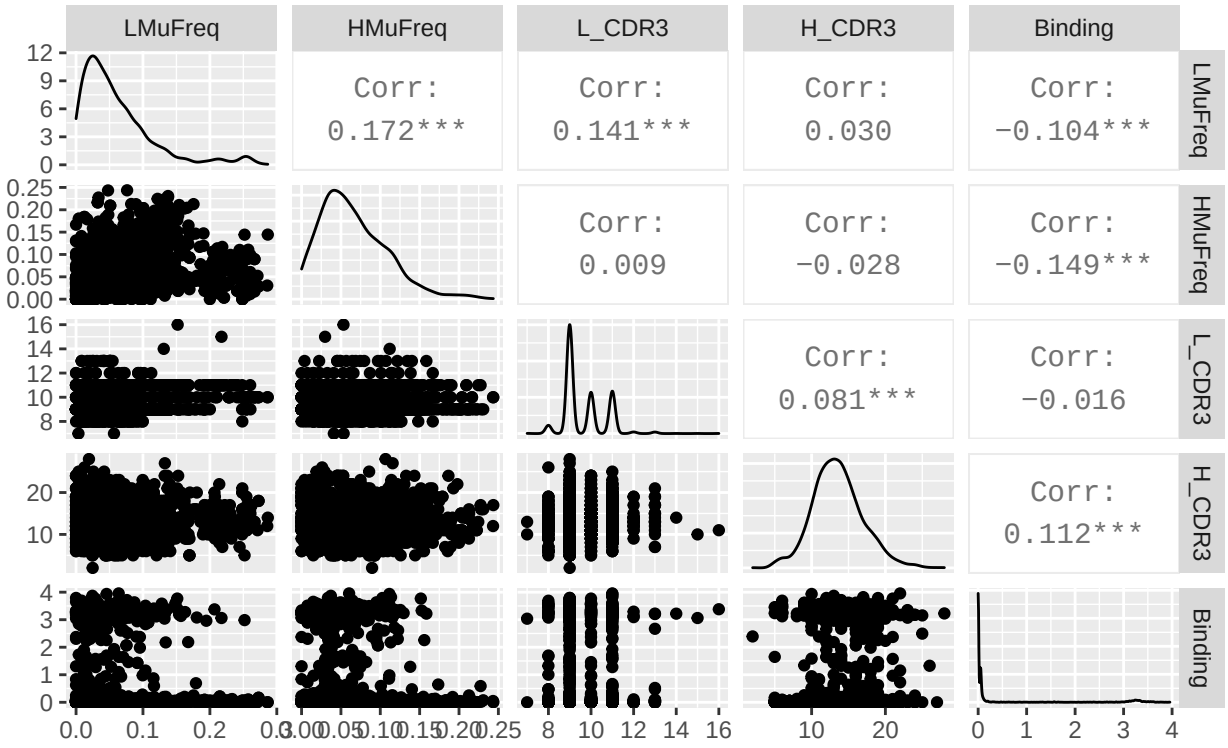

Figure 17: Plots of response variables

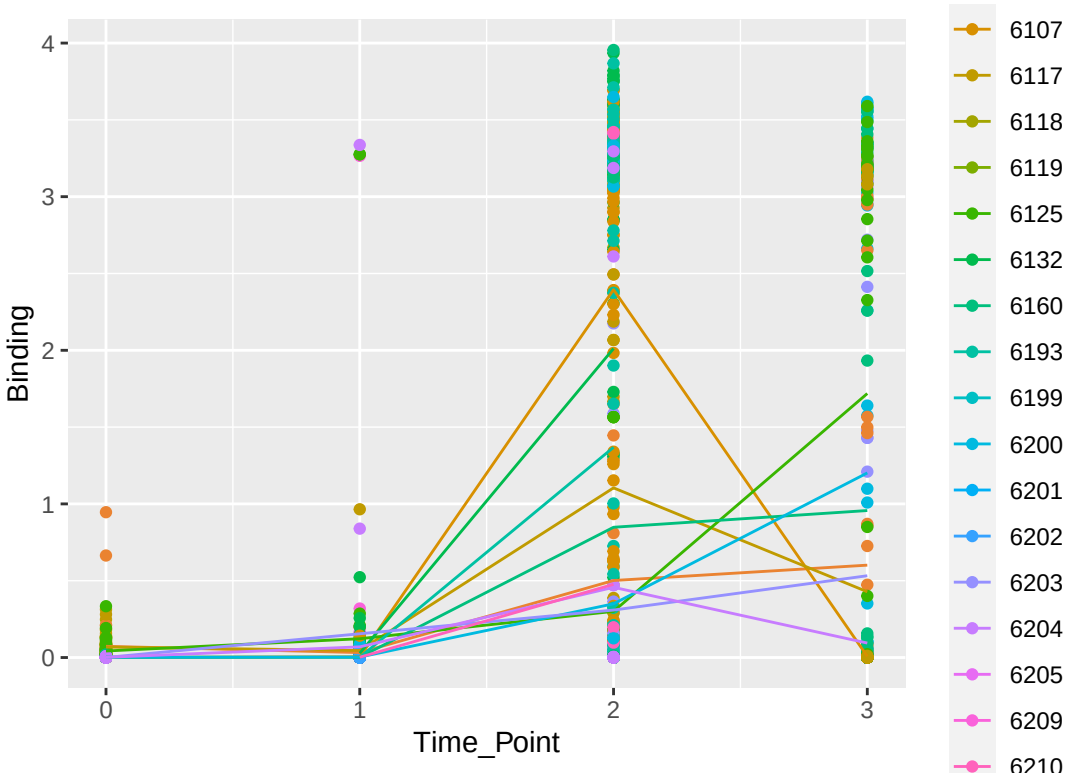

Figure 18: Mean trend by monkey

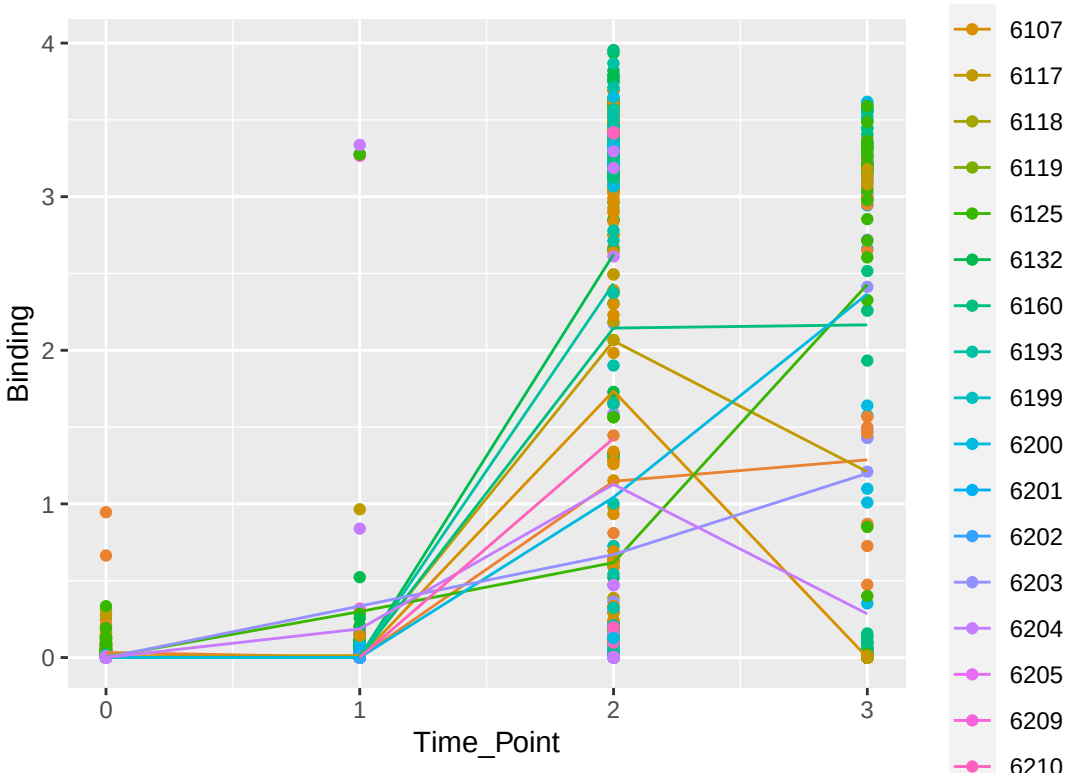

Figure 19: Variances over time by monkey

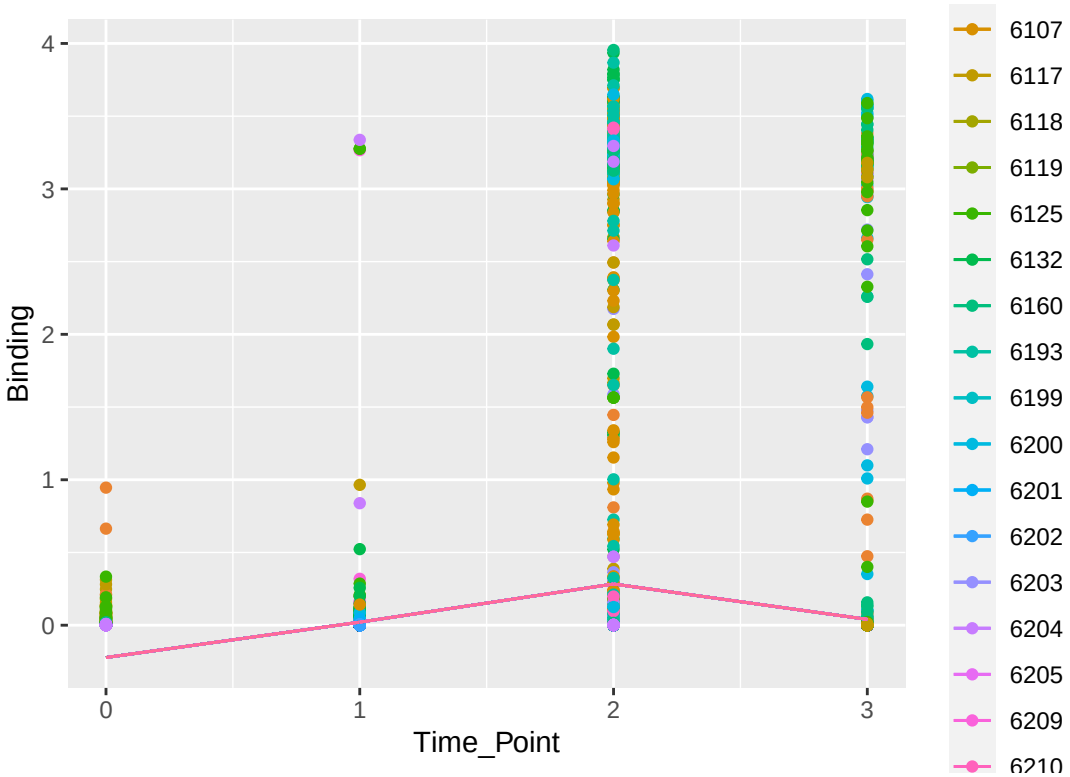

Figure 20: Piecewise Linear Function–three segments

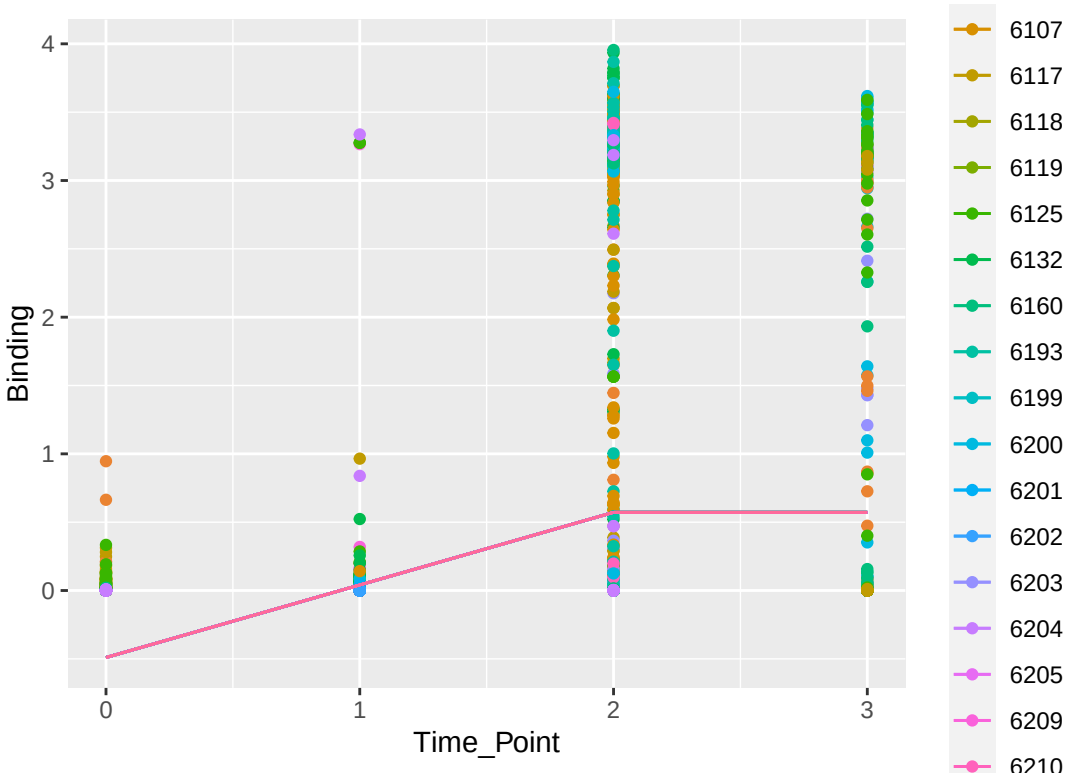

Figure 21: Piecewise Linear Function–two segments

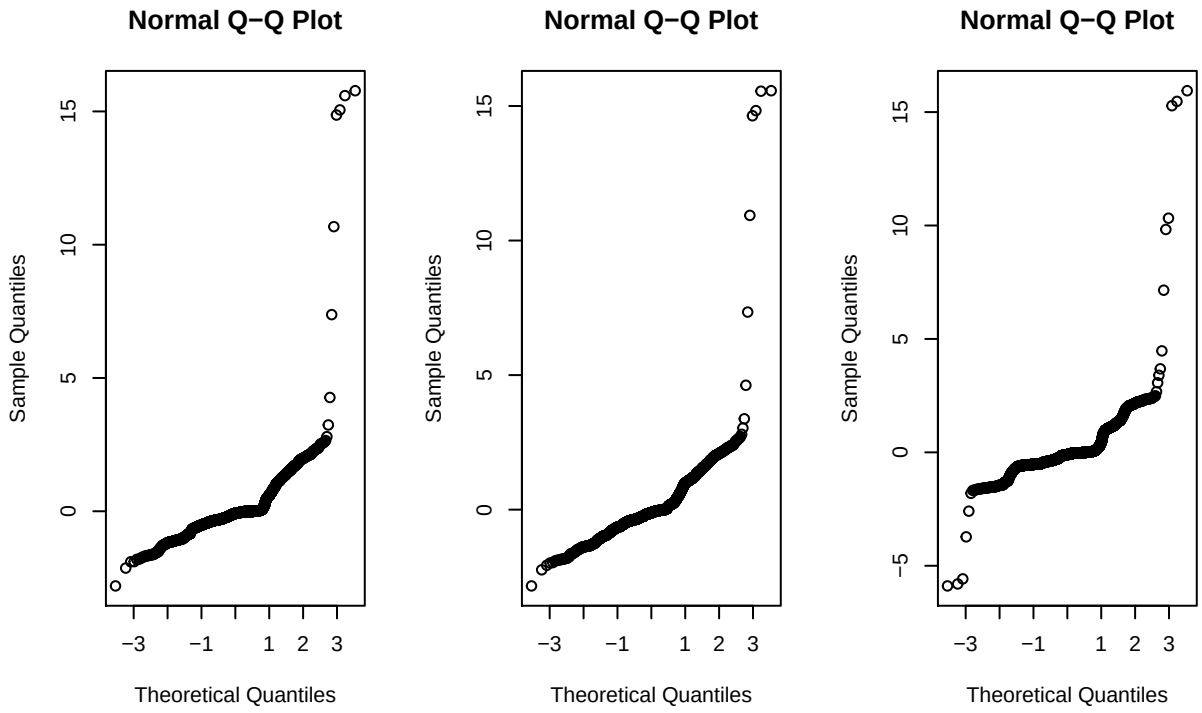

Figure 22: Q-Q plots of models: GLS, compound symmetry, AR1

List of Tables

Table 1: Frequency tables of drug vs. timepoints

0	1	2	3
129	346	251	122
90	533	225	101
54	125	347	142

Table 2: Frequency tables of drug vs. timepoints vs. isotypes

Drug	TimePoint	Isotype	Freq
1	0	A	4
2	0	A	6
3	0	A	1
1	1	A	11
2	1	A	10
3	1	A	1
1	2	A	8
2	2	A	4
3	2	A	4
1	3	A	1
2	3	A	1
3	3	A	0
1	0	D	4
2	0	D	4
3	0	D	1
1	1	D	22
2	1	D	45
3	1	D	26
1	2	D	15
2	2	D	19
3	2	D	14
1	3	D	16
2	3	D	7
3	3	D	6
1	0	E	1
2	0	E	0
3	0	E	0
1	1	E	2
2	1	E	2
3	1	E	0
1	2	E	3
2	2	E	1
3	2	E	0
1	3	E	1
2	3	E	0
3	3	E	0
1	0	G	60
2	0	G	37
3	0	G	24
1	1	G	91
2	1	G	205
3	1	G	28
1	2	G	145
2	2	G	115
3	2	G	170
1	3	G	57

Table 3: Summaries of standardized LCDR3

	V1
	Min. :-2.1860
	1st Qu.:-0.5361
	Median :-0.5361
	Mean : 0.0000
	3rd Qu.: 0.2888
	Max. :30.8110

Table 4: AIC and BIC between two gls models

	df	AIC	df.1	BIC
fit.gls	9	3323.050	9	3375.322
fit.gls2	8	3315.264	8	3361.730

Table 5: AIC and BIC for three models

	df	AIC	df.1	BIC
fit.gls2	8	3315.264	8	3361.730
fit.a1	11	3234.628	11	3298.520
fit.a2	11	3063.290	11	3127.182

Table 6: Inference of S4 ad S5 slopes

numDF	denDF	F.value	p.value
1	2442	244.2324506	0.0000000
1	2442	0.0317192	0.8586602

Table 7: Test whether drug 1 = drug 2

Fstat	p_value
1.065151	0.3626666

Table 8: Test whether drug 1 = drug 3

Fstat	p_value
1.231968	0.2964737

Table 9: Test whether drug 2 = drug 3

Fstat	p_value
1.255448	0.288075