Obesity & postmenopausal breast cancer

Rochita Das, Sandipan Pramanik & Pallavi Ray

1 Introduction

Over last few decades, we are seeing a significant rise in obesity and health risks related to it. According to World Health Organization (WHO), worldwide obesity has nearly tripled since 1975. The data also suggests that the prevalence of overweight and obesity is spreading across age groups and genders. Reporting some recent WHO estimates, in 2016

- more than 1.9 billion adults aged 18 years and older were overweight. Of these over 650 million adults were obese,
- 39% of adults aged 18 years and over (39% of men and 40% of women) were overweight,
- Overall, about 13% of the world's adult population (11% of men and 15% of women) were obese.
- In 2019, an estimated 38.2 million children under the age of 5 years were overweight or obese.

Existing literature suggests that obesity is a common risk factor for diabetes and coronary artery disease. But its causation on cancer, morbidity and mortality is comparatively less known. There are evidences which suggest that obesity is associated with a high risk of developing breast cancer, particularly among postmenopausal women, and with worse diseases among women of all ages. Elevated body weight is usually associated with hormone-responsive tumors. So for analyzing the effect of weight maintainance diet on decreasing tumor volume, a pre clinical study is conducted to determine its impact on rats after menopause and also possibly quantify the effect of diet.

Our primary objective here is to determine whether dietary intervention (preventing weight gain) after menopause potentially decreases tumor growth. Focusing on this, the rest of the report is organized as follows. In Section 2 we briefly explain the experimental study design. In Section 3 we carry out some exploratory analysis on the data. This helps us to better understand the data itself and also the underlying data generating mechanism. In Section 4 we provide our statistical analysis for the study. Finally, we conclude the report in Section 5 by citing our primary findings and presenting some conclusive remarks.

2 Experimental study design

The experimental study was conducted on 66 rats and it started by monitoring them for first couple of weeks. To divide them into more homogeneous subgroups based on obesity, after the initial monitoring stage, the rats were then classified into *Lean (L)* and *Obese (OB)* weight groups according to their percentage of body fat. At this time, for each rat all the tumors with volumes greater than 1 cm^3 were identified and the rats were kept under observation for next couple of weeks. After this, an ovariectomy was performed on each of these rats and the week of overiectomy was labeled as the Week 0 or the baseline week. The rats were then randomly assigned to one of the two *diet* groups: *Ad Libitum (AdLib)* and *Weight Maintenance (WM)*. Finally to study the impact of diet, all the rats were monitored for 8 more weeks and each time their tumor(s) volumes were measured.

In addition to the above description, there are couple of things about the study that we should keep in mind. First, each rat may have multiple tumors. At the time of dividing them into two weight groups these tumor(s) were identified and then monitored for rest of study. Second, tumor volumes were monitored starting from Week 0, the time of ovariectomy, and until the following 8 weeks. Tumor volume at Week 0 is considered as baseline effect. Third, for few rats it was not possible to monitor their tumor volumes for the entire 8 weeks and they got sacked early. Some of the reasons for sacking are: (a) the tumor volume exceeds a certain size, (b) the tumor growth is more than a percentage of the rat weight, (c) the tumor is too close to their skin, etc. Fourth, a tumor volume is reported as 0 for either due to absence of tumor or due to instrumental limitation (tumor is too small to measure).

3 Exploratory data analysis

In this section we performed an exploratory analysis for a better understanding of the observed data. Our primary variable of interest in this study is the tumor volumes, as we want to study the impact of weight maintenance diet on tumor growth.

In Figure 1 we plot tumor volume for each identified tumor starting from Week 0 to Week 8. There are four different combinations of diet and weight group. Each panel corresponds to each of these combinations and the blue lines in the respective panels denote the average tumor volume for that combination of diet and weight group. We notice an overall decreasing pattern or trend in the tumor volumes as week progresses. The decrement is slightly more prominent for WM diet and most prominent for the rats in Lean group with WM diet.

In Figure 2 we specifically compare the average tumor volumes (the blue lines in Figure 1) for each of the four combinations. In connection to our primary objective, this plot has two features of interest. First, as week progresses the average tumor volume corresponding to WM diet becomes smaller than that of AdLib diet. This seems to suggest a possible impact of dietary intervention on tumor growth. Second, the average tumor volumes versus the weeks curves exhibit a rough quadratic pattern for each of the four combinations. It would be interesting and more effective if we can incorporate these traits in our analysis.

Next we focus on exploring how the volumes are distributed across diets and weight groups. For this, in Figure 3 we draw the histogram of tumor volumes for each combinations. The plot reflects two key findings. First, an overwhelming occurrence of 0 tumor volumes

have been observed (approximately 27% of the reported volumes were exactly 0). Second, the underlying distribution is *positively skewed*.

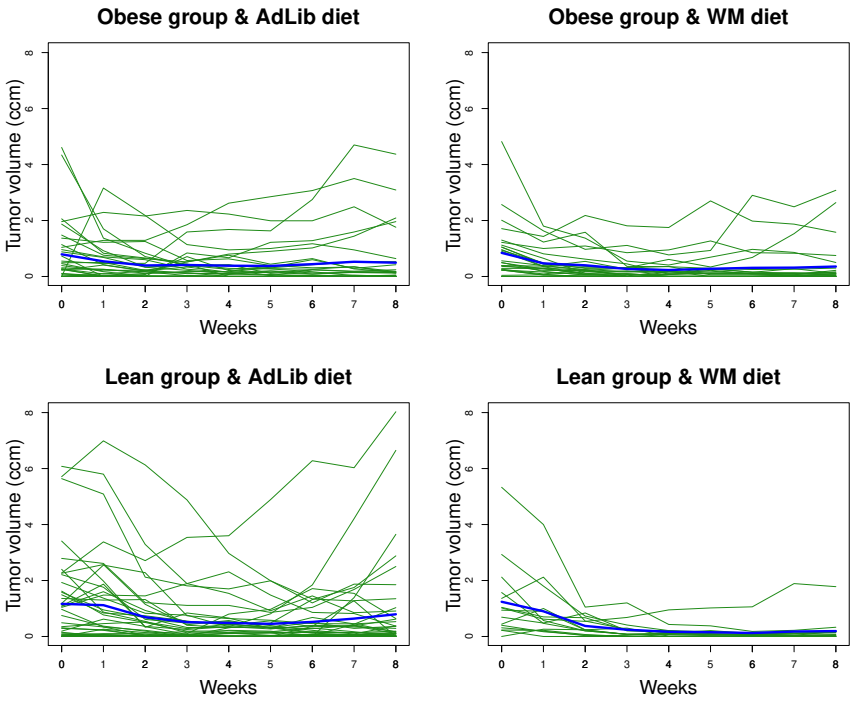

Figure 1: Group-wise comparison of tumor volumes

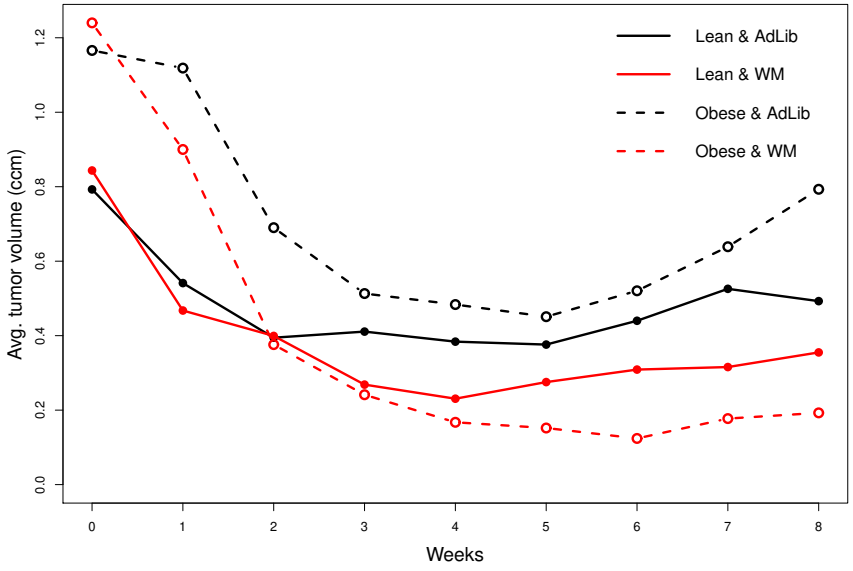

Figure 2: Group-wise comparison of average tumor volumes

Apart from all the aforementioned findings, there are some other attributes that need to be considered while modeling the data. First, we include *fixed effects* of both diet (treatment

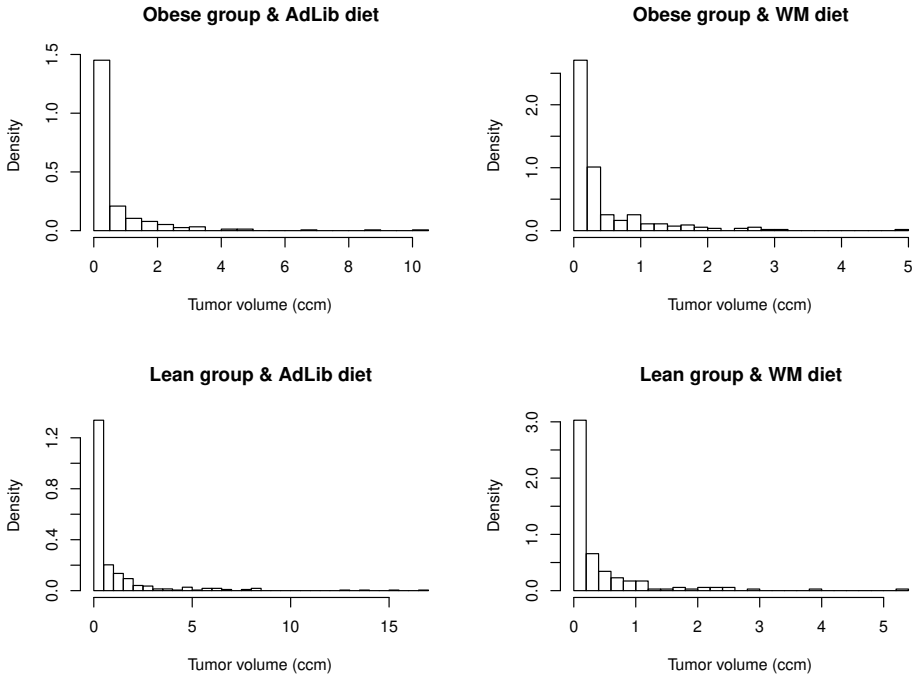

Figure 3: Group-wise histogram of tumor volumes

effect) and weight group (block effect), as we focus on the two diets and two weight groups. Second, for analyzing the impact of WM diet, we are particularly focused on the 8 weeks after the overiectomy. So in the analysis we consider fixed effects of weeks. Third, the rats are observed for 9 weeks including the baseline. So these are repeated measures on the same rat. Fourth, in this 2×2 experimental study a random sample of 66 rats were monitored. As they merely represent the large population of rats, we consider a *random effect* for each rat leading to a rat specific effect for analyzing the tumor volumes. Lastly, as it is mentioned in Section 2, there were rats with multiple tumors. Since there are multiple measurements on them over weeks, we consider a possible existence of a *nested effect* of tumors within rats.

In the following section, we aim for a sophisticated analysis for the observed data while trying to incorporate as much of the above findings as possible. Since there are both fixed and random effects that need to be implemented in the model, we resort to a *mixed effect* model with a likelihood on the tumor volumes which is both right skewed and *zero-inflated*.

4 Data Analysis

As it is mentioned in the earlier sections, the response variable, i.e. the tumor volume contains numerous zeros, most of them occurred due to measurement limitations. We also note that the tumor volume is a positive and continuous quantity and the histograms in Figure 3 reveals a very similar shape as a gamma density with zero inflation. Based on these observations, we were motivated to model the tumor volume with zero-inflated generalized gamma regression with mixed effects (combination of both fixed and random effects). The

fixed effects part of the model consists of the block effect, treatment effect and interaction between block and treatment effects while the effects due to rats considered to be random effects. We have used the statistical software **R** throughout our analysis and the zero-inflated models were implemented using **R** package `glmmTMB`.

4.1 Treatment effect over weeks

As a starting point, we modeled the tumor volumes corresponding to each week separately to see whether there is a significant treatment (diet) effect on the tumor volume as time progressed. For each week, the model contains an intercept term, coefficient for block (weight group) effect, coefficient for treatment effect and coefficient for interaction of block and treatment effects. Table 1 contains the estimated values of the coefficients and the P-values of the hypothesis test that the corresponding coefficient is zero against the alternative hypothesis that the corresponding coefficient is not zero.

Coefficients	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Intercept	0.524 (0.003)	0.492 (0.009)	−0.020 (0.937)	−0.298 (0.273)	−0.307 (0.211)	−0.521 (0.047)	−0.545 (0.071)	−0.252 (0.441)	−0.520 (0.142)
Block	−0.261 (0.362)	−0.516 (0.081)	−0.291 (0.438)	−0.287 (0.473)	−0.220 (0.567)	−0.152 (0.717)	−0.122 (0.801)	−0.286 (0.596)	−0.169 (0.764)
Treatment	−0.292 (0.358)	−0.573 (0.101)	−0.905 (**0.037**)	−0.921 (**0.042**)	−1.190 (**0.005**)	−1.020 (**0.029**)	−1.277 (**0.014**)	−1.287 (**0.029**)	−1.067 (0.092)
Block × Treatment	0.026 (0.953)	0.026 (0.957)	0.499 (0.393)	0.590 (0.339)	0.664 (0.267)	0.745 (0.259)	0.999 (0.181)	0.943 (0.265)	0.787 (0.374)

Table 1: Estimated values of the coefficients and the P-values (within parenthesis) corresponding to zero-inflated gamma regression models over 8 different weeks. The significant P-values (i.e. P-value < 0.05) are in bold.

Table 1 reveals that starting from the second week until the seventh week the treatment effects turn out to be significant at 5% level of significance. Figure 4 exhibits the same phenomenon. It is also of importance to note that the coefficients corresponding to the treatment effects for all the 9 weeks are negative. This can serve as a surrogate measure for the negative impact of diet on tumor growth, which is the main objective of the study.

From Table 1 and Figure 4 it is clear that time plays an important role along with the other external factors such as block or treatment. Thus, it is meaningful to include week effects directly in the model instead of modeling the data for each week separately. This idea motivates the subsequent section.

4.2 Fixed week effects

Since we are only interested in the 8 consecutive weeks post-overiectomy, it is logical to consider the week effects as fixed effects and are included in the fixed effects part of the model. We continue to treat rat effects as random effects. It is to be noted that one rat can have multiple tumors. The new model incorporates this nested structure of the number of tumors within each rat and this nested structure is combined with the random effects part of the model. This is another subtle difference of the new model from the previous ones. To determine the effectiveness of the weeks, we carried out a likelihood ratio test (LRT).

In this part of the analysis, we are primarily interested in model selection, meaning which of the fixed effects terms should be included in the model to explain the maximum variability in the tumor volumes. To that end, we conducted two hypotheses tests, one for the time effects where the null hypothesis (H_0) is all the coefficients corresponding to the week effects

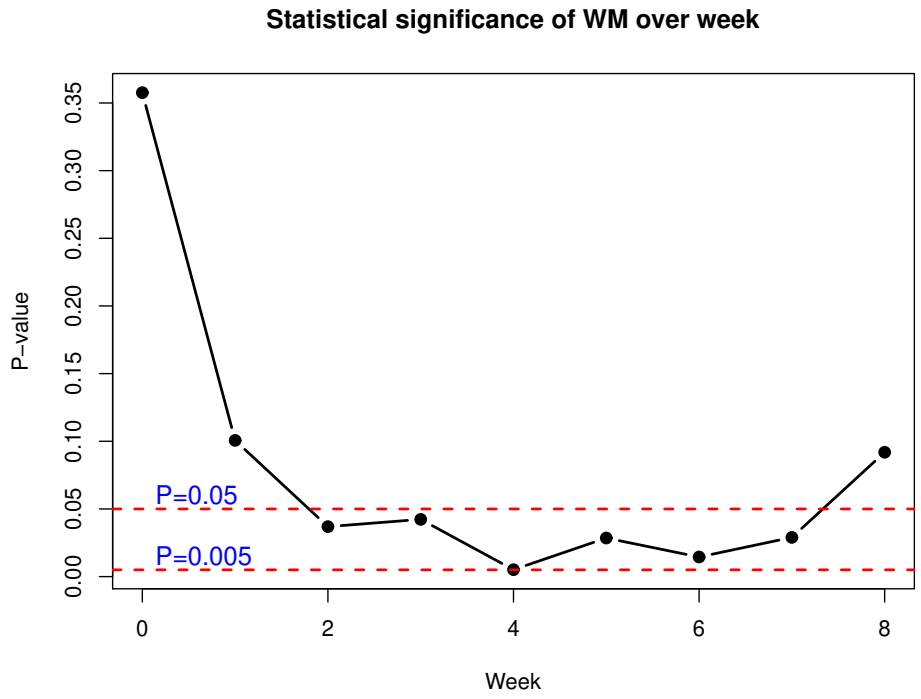

Figure 4: The P-values corresponding to the test that the coefficient for the treatment effect is zero over 9 consecutive weeks.

is zero versus the alternative hypothesis is at least of one of the coefficients is non-zero, and the second test is to determine whether or not including the block, treatment and their interaction effects in the fixed effects part of the model is beneficial. For the second test, the null hypothesis is all the coefficients corresponding to the block, treatment and their interaction effects are zero while the alternative hypothesis is at least one those coefficient is non-zero. We conducted LRTs for both the tests and the findings are tabulated in Table 2.

Tests	Null Hypothesis	P-value	Decision	Conclusion
Test-1	No fixed week effects	4.16×10^{-81}	Reject H_0	Significant week effects
Test-2	No fixed block or treatment effects	0.54	Fail to reject H_0	Not significant block or treatment effects

Table 2: Test results for tests 1 and 2 based on LRT

Based on the likelihood ratio tests, it is clear that the fixed week effects explain more variability in the tumor volumes than the effects due to block and treatment combined. Figure 2 exhibits a group-wise (moderate) quadratic pattern of the average tumor volumes over weeks and we also considered a model where the week is no longer handled as fixed effect but studied as a numeric covariate and a polynomial of degree 2 along with other fixed effects (block, treatment and their interaction) are incorporated in the model. The results for model selection based on this quadratic week model are very similar to that of Table 2, so we refrained from including the results in the main report. The details of this model selection procedure is deferred to the Appendix.

4.3 Quadratic week effects

As it is mentioned in the previous section, the tumor volumes over different combinations of weight and diet groups show a quadratic pattern over weeks and incorporating quadratic week effects did not make any difference as far as significant treatment effects are concerned. So, our intuition was to check whether removing the "quadratic trend" from the tumor volumes make any difference or not. To remove the quadratic trend it is enough to take second order difference of the tumor volumes. Figure 5 shows the boxplots for the zero to second order difference of the tumor volumes over 9 weeks for both diet groups.

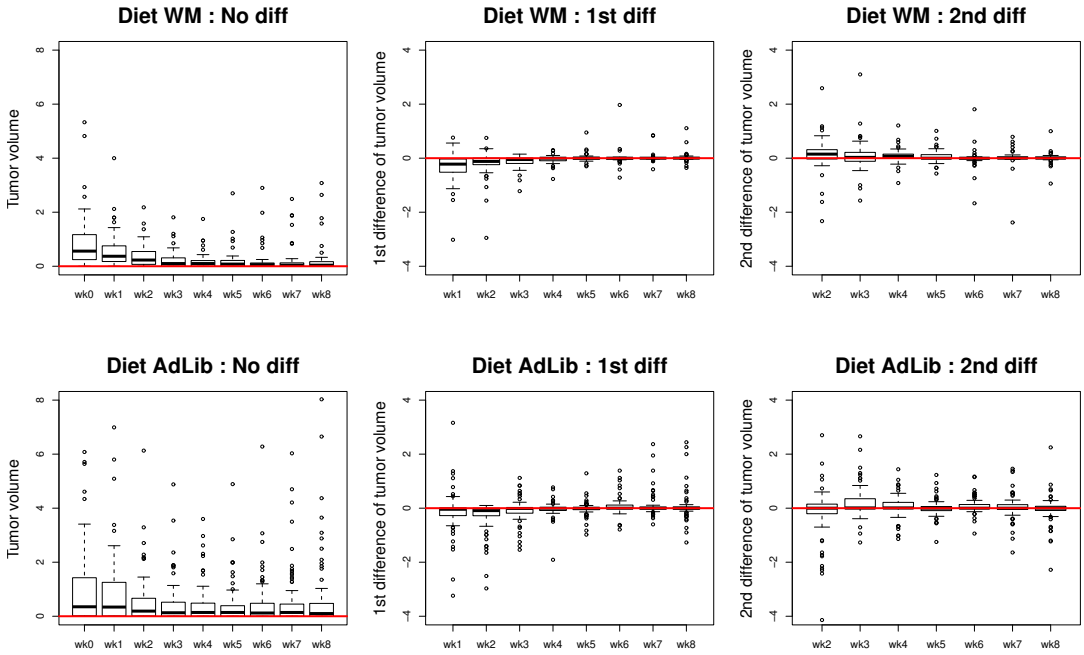

Figure 5: Boxplots for the zero to second order difference of the tumor volumes over 9 weeks.

Intuitively, the effect of weeks should be removed due to second order difference and from the boxplots, it is evident that the median second order difference of the tumor volumes stabilize around the value 0 for both the diet groups over the 9 weeks. Based on this observation it seems that treatment effect may not be significant. To verify this conjecture we performed a zero-inflated Gaussian mixed model to the second difference of the tumor volumes with block, treatment and their interaction as fixed effects and number of tumors within rats as random effects. It turns out that after removing the week effects from the tumor volumes, the treatment effect is not significant. The details of this analysis is deferred to the Appendix. Based on the preceding analyses, we gather that the main effects due to treatment or block are significant or explain a bulk amount of the variability in the tumor volumes only through the dependence on time (weeks) and we shall focus on capturing this dependence for the remainder of section 4. Thus, the study of quadratic week effects leads us to the final version of our analysis and we discuss that in the following section.

4.4 The final model : A zero-inflated Gamma regression

From previous analyses we observe that the main effect of week significantly dominates over all the other effects. So to test for the efficacy of WM diet, we consider the interaction effect between diet and week. In other words, we can imagine a hypothetical treatment with 18 levels; namely, AdLib (Week 0), AdLib (Week 1), $\cdots$, AdLib (Week 8), WM (Week 0), WM (Week 1), $\cdots$, WM (Week 8). Each of these levels correspond to the joint effect of a specific diet and a specific week. Following this, we perform a zero-inflated Gamma regression with fixed effect of block, fixed interaction effect between diet and week, random effect of rats and nested random effect of tumor within rats. This is the full model. We perform the desired hypothesis testing and obtain parameter estimates from this model.

4.4.1 Testing for the efficacy of Weight Maintenance

For testing the significance of WM diet under the above model, we test the null hypothesis that there is no interaction effect between WM diet and week. So under the null hypothesis, the effects of WM (Week 0), WM (Week 1), $\cdots$, WM (Week 8) in the hypothetical treatment are hypothesized to be 0. To conduct the testing, we again carry out the Likelihood ratio test and we conclusively reject the null hypothesis at 0.005 level of significance (Chi-square test statistic value: 246.676, df: 9, P-value: 5.02e-48). This implies that preventing weight gain after menopause and the next 8 weeks after the surgery has a significant joint effect in decreasing tumor growth.

4.4.2 Superiority of Weight Maintenance over Ad Libitum diet

In this section we compare effects of the two diets for possibly preventing tumor growth. For this we fit the full model discussed at the beginning of Section 4.4. In particular we look for the effects of different levels in the hypothetical treatment. For convenience in statistical

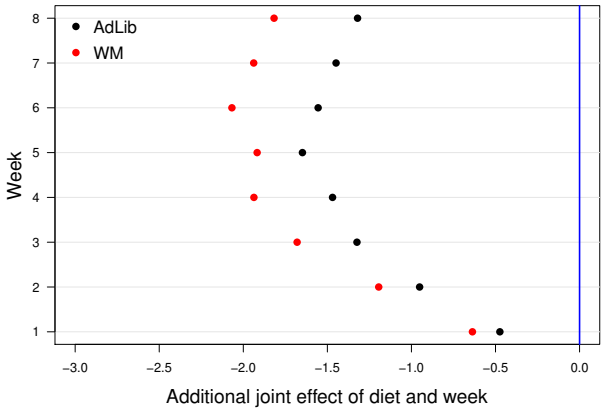

Figure 6: Plot showing the joint effect diet and week

analysis, following the common practice we include the joint effect of AdLib and Week 0 in the base effect. So for other levels in that treatment, we obtain the *additional effect* of a

specific diet and specific week. In Figure 6 we plot the additional joint effect of diet and week from Week 1 to Week 8.

There are two primary findings from the figure: (1) All the estimated effect sizes are negative. Considering all the weeks, this suggests that both the diets, AdLib and WM, are effective in decreasing tumor growth. This supports the findings from existing nutrition literature that diet prevents tumor growth. (2) But more importantly, the effect of WM diet is lower than that of AdLib diet for all the 8 weeks. This suggests that, jointly with week the Weight Maintenance is uniformly more effective than the Ad Libitum when considered over the 8 weeks.

5 Conclusion

Overall we have observed a statistically significant effect of treatment over the time as week progressed. But when we try to fit a model considering the week effects as fixed effects, we observe that the maximum variability of the response variable, the tumor volume, is being solely explained by the week, leaving all other predictors insignificant. This phenomena motivated us to incorporate the treatment effect as an interaction with time. Hence we define a set of new effects for treatment-time interaction and fit a model with these newly defined effects. The effect size estimates are negative and indicates a negative influence of dietary restrictions on tumor volumes as we may expect. Moreover, the significance of the weight maintaining diet over time is supported by the likelihood ratio test.

Based on some Biology literature, we intuitively also expect a negative effect of lean group on response variable tumor volume. But in all the models discussed in Section 4, we observe that the effect of Lean group is statistically insignificant. Here we try to guess a reasonable explanation for that. At the beginning of the study, the rats were divided into Lean or Obese group according to their body fat percentage at the time of surgery. After OVX, rats from each weight group were randomly assigned into AdLib or Weight Maintenance diet groups. As being on diet affects body fat, it is possible that a lean rat gains weight after the Adlib diet intake. Similarly, an obese rat may loose weight after the Weight Maintenance diet intake. So the weight groups, those were fixed at the start of an experiment, starts to change over the course of time. This change is neither available from the data, nor is considered in the statistical models. If it is possible to incorporate this change, this will not only make the analyses more exact, but we may also start seeing a statistically significant effect of weight groups on tumor volume.

Another limitation of this study can be the sample size. The study was conducted with only 66 rats, which is not a large sample to consider the nested effect of multiple tumors within rats. Moreover, some rats got sacked early in the study which lead to even less than 66 observations for some of the later weeks of the study. In the analyses as described in Section 4, we have also assumed that all the tumor volumes combined over weeks are uncorrelated. Since each rat was repeatedly measured over weeks, an improved analysis may be considered by assuming the tumor volumes observed from each rat to be correlated. Lastly, an abundance of zero as the observed tumor volumes made it particularly harder to test for the treatment effect. Based on our findings, future endeavors can be made to handle the problem with comparatively larger sample size and some other relevant covariates.

Appendix

Relevant **R** codes in Section 4 and their detailed outputs are provided here. For more detailed results, readers are encouraged to refer to corresponding sections here. For ease of reference, the section numbers start with A followed by the section number for which the code and output are provided.

A.4.1 Treatment effect over weeks

```
> library(glmmTMB)
> # calling the data:
> dat1 <- read.csv("tumor_data_mixed_model.csv",header = TRUE)
> dat1 <- dat1[,-1]
> head(dat1)
     ID rat_ID tumor_nos week tumor_vol block treat
1 401-1    401         1    0      1.05    OB    WM
2 401-1    401         1    1      0.44    OB    WM
3 401-1    401         1    2      0.31    OB    WM
4 401-1    401         1    3      0.31    OB    WM
5 401-1    401         1    4      0.36    OB    WM
6 401-1    401         1    5      0.06    OB    WM

> dp=dz=integer()
> for(i in 0:8){
+     glm.w <- glmmTMB(tumor_vol~block*treat+(1|rat_ID),
+                      data=subset(dat1,week==i),
+                      ziformula = ~1,family = ziGamma(link = "log"))
+     dp=rbind(dp,summary(glm.w)$coefficients$cond[,4])
+     dz=rbind(dz,summary(glm.w)$coefficients$cond[,1])
+ }

> dp <- data.frame(dp)
> names(dp) <- c("Intercept","blockOB","treatWM","blockOB.treatWM")
> dz <- data.frame(dz)
> names(dz) <- c("Intercept","blockOB","treatWM","blockOB.treatWM")

> dp #table containing p-values
    Intercept   blockOB     treatWM blockOB.treatWM
1 0.003018599 0.3624303 0.357641564       0.9526754
2 0.008970899 0.0805993 0.100653875       0.9567471
3 0.937247967 0.4380721 0.036929678       0.3926733
4 0.272966585 0.4733228 0.042287859       0.3391708
5 0.211256481 0.5666973 0.005098861       0.2670417
6 0.046831743 0.7173799 0.028508629       0.2591653
7 0.071021916 0.8007327 0.014477254       0.1810222
```

```
8 0.441467576 0.5957738 0.028966674       0.2647233
9 0.142234264 0.7640738 0.091883138       0.3742111

> dz #table containing coefficient estimates
    Intercept    blockOB    treatWM blockOB.treatWM
1  0.52380249 -0.2611078 -0.2918268      0.02601282
2  0.49150379 -0.5161381 -0.5730754      0.02586429
3 -0.01973164 -0.2909037 -0.9045592      0.49867055
4 -0.29757383 -0.2872938 -0.9206395      0.58990965
5 -0.30672175 -0.2201365 -1.1903230      0.66394006
6 -0.52055030 -0.1523380 -1.0203482      0.74509821
7 -0.54479526 -0.1220266 -1.2765827      0.99893595
8 -0.25188604 -0.2862250 -1.2872807      0.94344709
9 -0.52006959 -0.1686288 -1.0666312      0.78725484
```

A.4.2 Fixed week effects

```
> #Likelihood Ratio Tests (LRT) for fixed effects:
> glm0 <- glmmTMB(tumor_vol~block*treat+as.factor(week)+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm0)
 Family: Gamma  ( log )
Formula:
tumor_vol ~ block * treat + as.factor(week) + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  1761.8   1843.3   -864.9   1729.8     1186

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 2.1660   1.4717
 rat_ID           (Intercept) 0.2633   0.5132
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.498

Conditional model:
                Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept)       0.14784    0.25081   0.589    0.556
blockOB          -0.35309    0.38047  -0.928    0.353
treatWM          -0.60381    0.45477  -1.328    0.184
as.factor(week)1 -0.51736    0.09455  -5.472 4.45e-08 ***
as.factor(week)2 -1.02602    0.09714 -10.563  < 2e-16 ***
as.factor(week)3 -1.44260    0.10263 -14.056  < 2e-16 ***
as.factor(week)4 -1.63161    0.10639 -15.337  < 2e-16 ***
as.factor(week)5 -1.73552    0.10757 -16.134  < 2e-16 ***
as.factor(week)6 -1.73428    0.10852 -15.981  < 2e-16 ***
as.factor(week)7 -1.62045    0.10995 -14.737  < 2e-16 ***
as.factor(week)8 -1.48706    0.11487 -12.946  < 2e-16 ***
blockOB:treatWM   0.58704    0.62389   0.941    0.347
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> glm1 <- glmmTMB(tumor_vol~block*treat+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm1)
 Family: Gamma  ( log )
Formula:          tumor_vol ~ block * treat + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  2144.2   2184.9  -1064.1   2128.2     1194

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 1.4869   1.219
 rat_ID           (Intercept) 0.1656   0.407
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.81

Conditional model:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     -0.7932     0.2073  -3.826  0.00013 ***
```

```
blockOB          -0.1673     0.3251  -0.515  0.60684
treatWM          -0.3494     0.3841  -0.910  0.36299
blockOB:treatWM   0.1834     0.5284   0.347  0.72854
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #p-value for testing H0: No fixed week effects:
> pchisq(as.numeric(-2*(as.vector(logLik(glm1))-as.vector(logLik(glm0)))),
+        df.residual(glm1)-df.residual(glm0),lower.tail = FALSE)
[1] 4.165333e-81

> glm2 <- glmmTMB(tumor_vol~as.factor(week)+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm2)
 Family: Gamma  ( log )
Formula:          tumor_vol ~ as.factor(week) + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  1758.0   1824.2   -866.0   1732.0     1189

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 2.1743   1.4746
 rat_ID           (Intercept) 0.3028   0.5503
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.498

Conditional model:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.11020    0.16555  -0.666    0.506
as.factor(week)1 -0.51548    0.09452  -5.454 4.93e-08 ***
as.factor(week)2 -1.02449    0.09711 -10.550  < 2e-16 ***
as.factor(week)3 -1.44072    0.10260 -14.042  < 2e-16 ***
as.factor(week)4 -1.62933    0.10636 -15.319  < 2e-16 ***
```

```
as.factor(week)5 -1.73380    0.10753 -16.123  < 2e-16 ***
as.factor(week)6 -1.73213    0.10849 -15.965  < 2e-16 ***
as.factor(week)7 -1.61790    0.10994 -14.716  < 2e-16 ***
as.factor(week)8 -1.48426    0.11484 -12.924  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #p-value for testing H0: No fixed effects of block and treatment:
> pchisq(as.numeric(-2*(as.vector(logLik(glm2))-as.vector(logLik(glm0)))),
+        df.residual(glm2)-df.residual(glm0),lower.tail = FALSE)
[1] 0.5402127
```

A.4.3 Quadratic week effects

```
> #LRT for Quadratic time effects:
> glm0 <- glmmTMB(tumor_vol~block*treat+week+I(week^2)+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm0)
 Family: Gamma  ( log )
Formula:
tumor_vol ~ block * treat + week + I(week^2) + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  1751.7   1802.6   -865.9   1731.7     1192

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 2.1647   1.4713
 rat_ID           (Intercept) 0.2592   0.5091
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.499

Conditional model:
```

```
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.16654    0.24736   0.673    0.501
blockOB         -0.35329    0.37975  -0.930    0.352
treatWM         -0.60343    0.45391  -1.329    0.184
week            -0.63583    0.03655 -17.394   <2e-16 ***
I(week^2)        0.05662    0.00449  12.611   <2e-16 ***
blockOB:treatWM  0.58453    0.62270   0.939    0.348
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> glm1 <- glmmTMB(tumor_vol~block*treat+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm1)
 Family: Gamma  ( log )
Formula:          tumor_vol ~ block * treat + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  2144.2   2184.9  -1064.1   2128.2     1194

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 1.4869   1.219
 rat_ID           (Intercept) 0.1656   0.407
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.81

Conditional model:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.7932     0.2073  -3.826  0.00013 ***
blockOB          -0.1673     0.3251  -0.515  0.60684
treatWM          -0.3494     0.3841  -0.910  0.36299
blockOB:treatWM   0.1834     0.5284   0.347  0.72854
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #p-value for testing coefficients of linear and quadratic week terms are 0:
> pchisq(as.numeric(-2*(as.vector(logLik(glm1))-as.vector(logLik(glm0)))),
+        df.residual(glm1)-df.residual(glm0),lower.tail = FALSE)
[1] 7.921655e-87

> glm2 <- glmmTMB(tumor_vol~week+I(week^2)+(1|rat_ID/tumor_nos),
+                 data = dat1,ziformula = ~1,family = ziGamma(link = "log"))
> summary(glm2)
 Family: Gamma  ( log )
Formula:          tumor_vol ~ week + I(week^2) + (1 | rat_ID/tumor_nos)
Zero inflation:             ~1
Data: dat1

     AIC      BIC   logLik deviance df.resid
  1747.9   1783.5   -866.9   1733.9     1195

Random effects:

Conditional model:
 Groups           Name        Variance Std.Dev.
 tumor_nos:rat_ID (Intercept) 2.1731   1.4741
 rat_ID           (Intercept) 0.2985   0.5463
Number of obs: 1202, groups:  tumor_nos:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.499

Conditional model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.09139    0.16093  -0.568     0.57
week        -0.63531    0.03655 -17.381   <2e-16 ***
I(week^2)    0.05658    0.00449  12.604   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> #p-value for testing H0: No fixed effects of block and treatment:
> pchisq(as.numeric(-2*(as.vector(logLik(glm2))-as.vector(logLik(glm0)))),
+        df.residual(glm2)-df.residual(glm0),lower.tail = FALSE)
[1] 0.5377101
>
>
>
>
> ##### zero-inflated Gaussian model on the second difference #####
> library(readr)
> tumordata = read_csv("tumordata.csv")
>
> tumor.dat = as.data.frame(tumordata[,c("Animal # - Tumor #", 'ID', 'Tumor_No',
                                         "2018_08_Adipo_by_pFatOVX", "Group",
+                                        "VOL_OVX_closest_week", "Vol_OVX+1",
+                                        "Vol_OVX+2", "Vol_OVX+3", "Vol_OVX+4",
+                                        "Vol_OVX+5", "Vol_OVX+6", "Vol_OVX+7",
+                                        "Vol_OVX+8")])
> colnames(tumor.dat) = c('ID','rat_ID', 'tumor_ID', 'block', 'treatment',
+                         'wk0', 'wk1', 'wk2', 'wk3', 'wk4', 'wk5', 'wk6', 'wk7',
+                         'wk8')
> head(tumor.dat)
     ID rat_ID tumor_ID block treatment  wk0  wk1  wk2  wk3  wk4  wk5  wk6
1 401-1    401        1    OB        WM 1.05 0.44 0.31 0.31 0.36 0.06 0.06
2 402-3    402        3    OB     AdLib 1.48 0.76 0.64 0.52 0.76 0.44 0.64
3 414-1    414        1    OB        WM 4.82 1.80 1.37 0.55 0.41 0.69 0.97
4 414-2    414        2    OB        WM 0.00 0.00 0.00 0.00 0.00 0.00 0.00
5 416-1    416        1     L        WM 5.33 4.00 1.05 1.20 0.43 0.38 0.18
6 416-2    416        2     L        WM 0.44 1.00 0.24 0.30 0.01 0.00 0.00
   wk7  wk8
1 0.03 0.02
2 0.27 0.16
3 0.86 0.50
4 0.00 0.00
5 0.22 0.33
6 0.00 0.00
>
> tumor.dat.complete = tumor.dat[complete.cases(tumor.dat),]
> head(tumor.dat.complete)
     ID rat_ID tumor_ID block treatment  wk0  wk1  wk2  wk3  wk4  wk5  wk6
1 401-1    401        1    OB        WM 1.05 0.44 0.31 0.31 0.36 0.06 0.06
2 402-3    402        3    OB     AdLib 1.48 0.76 0.64 0.52 0.76 0.44 0.64
3 414-1    414        1    OB        WM 4.82 1.80 1.37 0.55 0.41 0.69 0.97
```

```
4 414-2    414        2    OB        WM 0.00 0.00 0.00 0.00 0.00 0.00 0.00
5 416-1    416        1     L        WM 5.33 4.00 1.05 1.20 0.43 0.38 0.18
6 416-2    416        2     L        WM 0.44 1.00 0.24 0.30 0.01 0.00 0.00
   wk7  wk8
1 0.03 0.02
2 0.27 0.16
3 0.86 0.50
4 0.00 0.00
5 0.22 0.33
6 0.00 0.00
>
> vol.mat = as.matrix(tumor.dat.complete[,c('wk0', 'wk1', 'wk2', 'wk3', 'wk4',
+                                           'wk5', 'wk6', 'wk7', 'wk8')])
> sum(is.na.data.frame(vol.mat))
[1] 0
>
> ## 1st difference
> voldiff1.mat = matrix(NA, nrow(vol.mat), ncol(vol.mat)-1)
> for(i in 1:nrow(voldiff1.mat)) voldiff1.mat[i,] = diff(vol.mat[i,])
> head(voldiff1.mat)
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
[1,] -0.61 -0.13  0.00  0.05 -0.30  0.00 -0.03 -0.01
[2,] -0.72 -0.12 -0.12  0.24 -0.32  0.20 -0.37 -0.11
[3,] -3.02 -0.43 -0.82 -0.14  0.28  0.28 -0.11 -0.36
[4,]  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
[5,] -1.33 -2.95  0.15 -0.77 -0.05 -0.20  0.04  0.11
[6,]  0.56 -0.76  0.06 -0.29 -0.01  0.00  0.00  0.00
> colnames(voldiff1.mat) = c('wk1', 'wk2', 'wk3', 'wk4',
+                            'wk5', 'wk6', 'wk7', 'wk8')
> boxplot(voldiff1.mat)
> abline(h = 0, col = 2, lwd = 2)
>
> ## 2nd difference
> voldiff2.mat = matrix(NA, nrow(vol.mat), ncol(vol.mat)-2)
> for(i in 1:nrow(voldiff2.mat)) voldiff2.mat[i,] = diff(voldiff1.mat[i,])
> head(voldiff2.mat)
      [,1]  [,2]  [,3]  [,4]          [,5]  [,6]  [,7]
[1,]  0.48  0.13  0.05 -0.35  3.000000e-01 -0.03  0.02
[2,]  0.60  0.00  0.36 -0.56  5.200000e-01 -0.57  0.26
[3,]  2.59 -0.39  0.68  0.42  5.551115e-17 -0.39 -0.25
[4,]  0.00  0.00  0.00  0.00  0.000000e+00  0.00  0.00
[5,] -1.62  3.10 -0.92  0.72 -1.500000e-01  0.24  0.07
[6,] -1.32  0.82 -0.35  0.28  1.000000e-02  0.00  0.00
> colnames(voldiff2.mat) = c('wk2', 'wk3', 'wk4', 'wk5', 'wk6', 'wk7', 'wk8')
> boxplot(voldiff2.mat)
```

```
> abline(h = 0, col = 2, lwd = 2)
> ## fitting 0 inflated normal on 2nd difference
> library(glmmTMB)
>
> # data for ZINormal
> tumor.ZIN = data.frame()
> for (i in 1:nrow(voldiff2.mat)) {
+
+   tumor.ZIN =
+     rbind.data.frame(tumor.ZIN,
+       cbind.data.frame(
+        'rat_ID' = rep(tumor.dat.complete$rat_ID[i], ncol(voldiff2.mat)),
+        'tumor_ID' = rep(tumor.dat.complete$tumor_ID[i], ncol(voldiff2.mat)),
+        'block' = rep(tumor.dat.complete$block[i], ncol(voldiff2.mat)),
+        'treatment' = rep(tumor.dat.complete$treatment[i], ncol(voldiff2.mat)),
+        'wk' = paste('wk', 1 + (1:ncol(voldiff2.mat)), sep = ''),
+        'vol' = voldiff2.mat[i,]))
+ }
> head(tumor.ZIN)
    rat_ID tumor_ID block treatment  wk   vol
wk2    401        1    OB        WM wk2  0.48
wk3    401        1    OB        WM wk3  0.13
wk4    401        1    OB        WM wk4  0.05
wk5    401        1    OB        WM wk5 -0.35
wk6    401        1    OB        WM wk6  0.30
wk7    401        1    OB        WM wk7 -0.03
>
> tumor.ZIN$treatment = factor(tumor.ZIN$treatment, levels = c('AdLib', 'WM'))
>
> ZIN = glmmTMB(vol~block + treatment + (1|rat_ID/tumor_ID),
+               ziformula = ~1,
+               data = tumor.ZIN, family = gaussian())
>
> summary(ZIN)
 Family: gaussian  ( identity )
Formula:          vol ~ block + treatment + (1 | rat_ID/tumor_ID)
Zero inflation:       ~1
Data: tumor.ZIN

     AIC      BIC   logLik deviance df.resid
    1284     1317     -635     1270      819

Random effects:

Conditional model:
```

```
 Groups          Name        Variance  Std.Dev.
 tumor_ID:rat_ID (Intercept) 1.232e-11 3.510e-06
 rat_ID          (Intercept) 7.496e-12 2.738e-06
 Residual                    2.724e-01 5.219e-01
Number of obs: 826, groups:  tumor_ID:rat_ID, 118; rat_ID, 57

Dispersion estimate for gaussian family (sigma^2): 0.272

Conditional model:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.032407   0.030838   1.051    0.293
blockL      -0.004537   0.037118  -0.122    0.903
treatmentWM  0.025521   0.038541   0.662    0.508

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -23.53    5333.77  -0.004    0.996
```

A.4.4 The final model : A zero-inflated Gamma regression

```
> library(readr)
> tumordata = read_csv("tumordata.csv")
>
> tumor.dat = as.data.frame(tumordata[,c("Animal # - Tumor #", 'ID', 'Tumor_No',
                                         "2018_08_Adipo_by_pFatOVX", "Group",
+                                        "VOL_OVX_closest_week", "Vol_OVX+1",
+                                        "Vol_OVX+2", "Vol_OVX+3", "Vol_OVX+4",
+                                        "Vol_OVX+5", "Vol_OVX+6", "Vol_OVX+7",
+                                        "Vol_OVX+8")])
> colnames(tumor.dat) = c('ID','rat_ID', 'tumor_ID', 'block', 'treatment',
+                         'wk0', 'wk1', 'wk2', 'wk3', 'wk4', 'wk5', 'wk6', 'wk7',
+                         'wk8')
>
> tumor.ZIG = data.frame()
> for (i in 1:nrow(tumor.dat)) {
+
+   y.vec = as.numeric(tumor.dat[i, c('wk0', 'wk1', 'wk2', 'wk3', 'wk4',
+                                     'wk5', 'wk6', 'wk7', 'wk8')])
+
+   if(sum(!is.na(y.vec))>0){
+
+     nTime = max(which(!is.na(y.vec)))
+     tumor.ZIG =
+     rbind.data.frame(tumor.ZIG,
```

```
+       cbind.data.frame('rat_ID' = rep(tumor.dat$rat_ID[i], nTime),
+                        'tumor_ID' = rep(tumor.dat$tumor_ID[i], nTime),
+                        'block' = rep(tumor.dat$block[i], nTime),
+                        'treatment' = rep(tumor.dat$treatment[i], nTime),
+                        'wk' = paste('wk', 0:(nTime-1), sep = ''),
+                        'vol' = y.vec[1:nTime]))
+   }
+ }
> # fitting the ZIG model
> library(glmmTMB)
> ZIG = glmmTMB(vol~block + TimeTreatment + (1|rat_ID/tumor_ID),
+               ziformula = ~1,
+               data = tumor.ZIG, family = ziGamma(link = "log"))
> summary(ZIG)
 Family: Gamma  ( log )
Formula:          vol ~ block + TimeTreatment + (1 | rat_ID/tumor_ID)
Zero inflation:       ~1
Data: tumor.ZIG

     AIC      BIC   logLik deviance df.resid
  1769.2   1886.3   -861.6   1723.2     1179

Random effects:

Conditional model:
 Groups          Name        Variance Std.Dev.
 tumor_ID:rat_ID (Intercept) 2.1591   1.4694
 rat_ID          (Intercept) 0.2642   0.5141
Number of obs: 1202, groups:  tumor_ID:rat_ID, 144; rat_ID, 66

Dispersion estimate for Gamma family (sigma^2): 0.495

Conditional model:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            -0.15953    0.27330  -0.584 0.559430
blockL                  0.11561    0.30138   0.384 0.701264
TimeTreatmentAdLib-wk1 -0.47398    0.12195  -3.887 0.000102 ***
TimeTreatmentAdLib-wk2 -0.95130    0.12651  -7.519 5.50e-14 ***
TimeTreatmentAdLib-wk3 -1.32398    0.13105 -10.103  < 2e-16 ***
TimeTreatmentAdLib-wk4 -1.46918    0.13632 -10.778  < 2e-16 ***
TimeTreatmentAdLib-wk5 -1.64824    0.13670 -12.058  < 2e-16 ***
TimeTreatmentAdLib-wk6 -1.55495    0.13906 -11.182  < 2e-16 ***
TimeTreatmentAdLib-wk7 -1.44832    0.14107 -10.267  < 2e-16 ***
TimeTreatmentAdLib-wk8 -1.32044    0.14389  -9.176  < 2e-16 ***
TimeTreatmentWM-wk0    -0.07168    0.33543  -0.214 0.830775
```

```
TimeTreatmentWM-wk1    -0.63725    0.33619  -1.895 0.058029 .
TimeTreatmentWM-wk2    -1.19449    0.33621  -3.553 0.000381 ***
TimeTreatmentWM-wk3    -1.67993    0.34180  -4.915 8.88e-07 ***
TimeTreatmentWM-wk4    -1.93680    0.34357  -5.637 1.73e-08 ***
TimeTreatmentWM-wk5    -1.91769    0.34516  -5.556 2.76e-08 ***
TimeTreatmentWM-wk6    -2.06769    0.34448  -6.002 1.95e-09 ***
TimeTreatmentWM-wk7    -1.93803    0.34504  -5.617 1.95e-08 ***
TimeTreatmentWM-wk8    -1.81701    0.35162  -5.167 2.37e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Zero-inflation model:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.9969     0.0650  -15.34   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> # testing effect of WM
> modelmat = model.matrix(ZIG)
> modelmat.H0.ZIG = modelmat[,1:10]
> ZIG.H0 = glmmTMB(tumor.ZIG$vol ~ 0 + modelmat.H0.ZIG + (1|rat_ID/tumor_ID),
+                  ziformula = ~1,
+                  data = tumor.ZIG, family = ziGamma(link = "log"))
> pchisq(as.numeric(-2*(logLik(ZIG.H0)-logLik(ZIG))),
+        df.residual(ZIG.H0) - df.residual(ZIG),
+        lower.tail = FALSE) #significant p-value
[1] 5.019307e-48
```