Infection Analysis

Michael Sieler

16 December, 2022

Do infection counts differ between the diets?

There is a discrepancy between Mike and I's results about whether infection counts differ between the diets. They do, but it depends on how you calculation infection. After digging into the datasets Mike previously sent me on Nov 24th, 2020 and Dec 7th, 2022, I realized that we were calculating infection differently. Mike's counts are based on severity score (0 to 3), while mine are based on presence or absence of infection in the coelom, ovary, kidney or lumen. Using severity score as a metric for infection count, there are indeed signficant differences between the diets. However, when comparing presence/absence of infection between the diets the association is weak, or not significant at the 0.05 threshold. The question remains, which method (severity score or presence/absence) is best for determining differences in infection burden?

One issue I found with the severity score data is that there are 11 cases where an infection is present (e.g., in the lumen), but severity score is 0. Conversely, there are two cases (e.g., DSF 644 & 651) where infection is absent, but severity score is 1. As far as I can tell, there is no indication whether these fish were marked incorrectly as having an infection.

Import Data

Nov 24, 2020

This dataset was emailed to me on November 24th, 2020 from Mike and Colleen. It is titled `OSU Diet Study Primary Excel CA072820 28 July b.xlsx`.

```
# Import spreadsheet
data.Nov24 <- readxl::read_excel(file.path(proj.path,"Data/Raw/data_Nov24.xlsx"))
```

Dec 7, 2022

This data was emailed to me by Mike Kent on Dec 7th, 2022. His original spreadsheet (`DSF Myco Histo Results 5 and 15 wk   Dec 2022.xlsx`) contains his own calculations and annotations, but were removed from this dataset (`mikesData.xlsx`) because it caused import issues.

```
# Import spreadsheet
data.Dec7 <- readxl::read_excel( file.path(proj.path,"Data/Raw/data_Dec7.xlsx"))
```

Counts

Compare counts between Nov 24 and Dec 7 data

```
data.Nov24 %>%
  # filter(Sex == "M") %>%
  group_by(Diet, Sex) %>%
  # group_by(Diet, Pathology.Results) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Original Data, Nov 24, 2020 (male and female fish)")
```

Sex and Diet

Table 1: Original Data, Nov 24, 2020 (male and female fish)

Diet	Sex	Count
Gemma	F	3
Gemma	M	12
Watts	F	4
Watts	M	24
ZIRC	F	6
ZIRC	M	17

```
data.Dec7 %>%
  # filter(Sex == "M") %>%
  group_by(Diet, Sex) %>%
  # group_by(Diet, Pathology.Results) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Mike's Data, Dec 7, 2020 (male and female fish)")
```

Table 2: Mike's Data, Dec 7, 2020 (male and female fish)

Diet	Sex	Count
Gemma	F	3
Gemma	M	12
Watts	F	4
Watts	M	24
ZIRC	F	6
ZIRC	M	17

```
data.Nov24 %>%
  filter(Sex == "M") %>%
  group_by(Diet, Severity.Score) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Original Data, Nov 24, 2020 (male  fish)")
```

Severity score and Diet

Table 3: Original Data, Nov 24, 2020 (male fish)

Diet	Severity.Score	Count
Gemma	0	9
Gemma	2	1
Gemma	3	2
Watts	0	20
Watts	1	3
Watts	3	1
ZIRC	0	5
ZIRC	1	3
ZIRC	2	3
ZIRC	3	6

```
data.Dec7 %>%
  filter(Sex == "M") %>%
  group_by(Diet, Severity.Score) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Mike's Data, Dec 7, 2020 (male fish)")
```

Table 4: Mike's Data, Dec 7, 2020 (male fish)

Diet	Severity.Score	Count
Gemma	0	9
Gemma	2	1
Gemma	3	2
Watts	0	20
Watts	1	3
Watts	3	1

Table 4: Mike's Data, Dec 7, 2020 (male fish)

Diet	Severity.Score	Count
ZIRC	0	5
ZIRC	1	3
ZIRC	2	3
ZIRC	3	6

```
data.Nov24 %>%
  filter(Sex == "M") %>%
  group_by(Diet, Pathology.Results) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Original Data, Nov 24, 2020 (male fish)")
```

Pathology results and Diet

Table 5: Original Data, Nov 24, 2020 (male fish)

Diet	Pathology.Results	Count
Gemma	Negative	7
Gemma	Positive	5
Watts	Negative	13
Watts	Positive	11
ZIRC	Negative	4
ZIRC	Positive	13

```
data.Dec7 %>%
  filter(Sex == "M") %>%
  group_by(Diet, Infection.Mike) %>%
  count(name = "Count") %>%
  flextable() %>%
  set_caption(caption = "Mike's Data, Dec 7, 2020 (male fish)")
```

Table 6: Mike's Data, Dec 7, 2020 (male fish)

Diet	Infection.Mike	Count
Gemma	Negative	9
Gemma	Positive	3

Table 6: Mike's Data, Dec 7, 2020 (male fish)

Diet	Infection.Mike	Count
Watts	Negative	20
Watts	Positive	4
ZIRC	Negative	5
ZIRC	Positive	12

Statistical Tests

Chi-Square

Nov 24   Counts from Nov 24 spreadsheet

```
# (Gemma, Watts, ZIRC)
p <- c( 5, 11, 13)  # Positive
n <- c( 7, 13, 4 )  # Negative

d <- rbind( p, n)
d <- as.table(d)
coin::chisq_test(d)
```

```
##
##   Asymptotic Pearson Chi-Squared Test
##
## data:  Var2 by Var1 (p, n)
## chi-squared = 4.836, df = 2, p-value = 0.0891
```

Dec 7   Mike's numbers from email

```
# Mike's numbers from his email
# (Gemma, Watts, ZIRC)
p <- c( 3, 3, 14)  # Positive
n <- c( 9, 20, 4 )   # Negative

d <- rbind( p, n)
d <- as.table(d)
coin::chisq_test(d)
```

```
##
##   Asymptotic Pearson Chi-Squared Test
##
## data:  Var2 by Var1 (p, n)
## chi-squared = 19.08, df = 2, p-value = 7.192e-05
```

Counts from spreadsheet

Note: These counts differ from Mike's email, but still return a significant result

```
# (Gemma, Watts, ZIRC)
p <- c( 3, 3, 12)  # Positive
n <- c( 9, 21, 5 )  # Negative

d <- rbind( p, n)
d <- as.table(d)
coin::chisq_test(d)
```

```
##
##   Asymptotic Pearson Chi-Squared Test
##
## data:  Var2 by Var1 (p, n)
## chi-squared = 15.527, df = 2, p-value = 0.000425
```

GLM

Nov 24 Presence/Absence

```
mod.Nov.glm <- data.Nov24 %>%
  filter(Sex == "M") %>%
  mutate(Infected =
           case_when(
             Pathology.Results == "Negative" ~ 0,
             Pathology.Results == "Positive" ~ 1
           )) %>%
    glm(formula = Infected ~ Diet, family = "binomial")

mod.Nov.glm %>% summary()
```

```
##
## Call:
## glm(formula = Infected ~ Diet, family = "binomial", data = .)
##
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.7011  -1.1073   0.7325   1.2491   1.3232
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.3365     0.5855  -0.575   0.5655
## DietWatts     0.1694     0.7146   0.237   0.8126
## DietZIRC      1.5151     0.8184   1.851   0.0641 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 73.001  on 52   degrees of freedom
## Residual deviance: 67.955  on 50   degrees of freedom
## AIC: 73.955
##
## Number of Fisher Scoring iterations: 4
```

```
mod.Nov.glm %>% Anova(type = 2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Infected
##      LR Chisq Df Pr(>Chisq)
## Diet   5.0461  2    0.08021 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# mod.1 %>% plot()
```

Severity Score

```r
mod.Nov.glm <- data.Nov24 %>%
  filter(Sex == "M") %>%
  mutate(Infected =
           case_when(
             Severity.Score == 0 ~ 0,
             Severity.Score > 0 ~ 1
           )) %>%
    glm(formula = Infected ~ Diet, family = "binomial")

mod.Nov.glm %>% summary()
```

```
##
## Call:
## glm(formula = Infected ~ Diet, family = "binomial", data = .)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5645  -0.7585  -0.6039   0.8346   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0986     0.6667  -1.648   0.0994 .
## DietWatts    -0.5108     0.8628  -0.592   0.5538
## DietZIRC      1.9741     0.8531   2.314   0.0207 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 69.17  on 52  degrees of freedom
## Residual deviance: 55.72  on 50  degrees of freedom
## AIC: 61.72
##
## Number of Fisher Scoring iterations: 4
```

```r
mod.Nov.glm %>% Anova(type = 2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Infected
##       LR Chisq Df Pr(>Chisq)
## Diet     13.45  2   0.001201 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# mod.1 %>% plot()
```

Dec 7 Mike's Infection Count

```r
mod.Dec.glm <- data.Dec7 %>%
  filter(Sex == "M")  %>%
  mutate(Infected =
           case_when(
             Infection.Mike == "Negative" ~ 0,
             Infection.Mike == "Positive" ~ 1,
           )) %>%
    glm(formula = Infected ~ Diet, family = "binomial")

mod.Dec.glm %>% summary()
```

```
##
## Call:
## glm(formula = Infected ~ Diet, family = "binomial", data = .)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q      Max
## -1.5645  -0.7585  -0.6039   0.8346   1.8930
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.0986     0.6667  -1.648   0.0994 .
## DietWatts    -0.5108     0.8628  -0.592   0.5538
## DietZIRC      1.9741     0.8531   2.314   0.0207 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 69.17  on 52  degrees of freedom
## Residual deviance: 55.72  on 50  degrees of freedom
## AIC: 61.72
##
## Number of Fisher Scoring iterations: 4
```

```r
mod.Dec.glm %>% Anova(type = 2)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Infected
##      LR Chisq Df Pr(>Chisq)
## Diet    13.45  2   0.001201 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# mod.1 %>% plot()
```