20250801_abner_emmeans.Rmd

2025-08-01

```
# First, fit your model and store it
model <- glm.nb(
    is.referenced.by.count ~ da_factor + log(age.in.months) + container.title +
        container.title*da_factor + log(age.in.months)*da_factor +
        container.title*log(age.in.months) +
        log(age.in.months) * da_factor * container.title,
    data = my_data,
    link = "log"
)
# Define the age values you want to examine (in months)
age_values <- c(12, 36, 60, 120)  # Adjust these as needed
# Get emmeans on the link scale for all combinations
emm <- emmeans(model,  ~ da_factor + age.in.months | container.title,
        at = list(age.in.months = age_values), CIs = TRUE,
        type = "response")
# Get pairwise comparisons (differences) between da_factor levels
differences <- contrast(
    emm, by = c("age.in.months", "container.title"),
    method = "revpairwise",
    ratios = TRUE, CIs = TRUE
)
# See the contrasts
summary(differences)
```

```
## age.in.months = 12, container.title = Antimicrobial Agents and Chemotherapy:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.552 0.0623 Inf    1  -5.265  <.0001
##
## age.in.months = 36, container.title = Antimicrobial Agents and Chemotherapy:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.826 0.0515 Inf    1  -3.069  0.0021
##
## age.in.months = 60, container.title = Antimicrobial Agents and Chemotherapy:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.996 0.0448 Inf    1  -0.081  0.9355
##
## age.in.months = 120, container.title = Antimicrobial Agents and Chemotherapy:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.285 0.0545 Inf    1   5.925  <.0001
##
## age.in.months = 12, container.title = Applied and Environmental Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.863 0.0593 Inf    1  -2.150  0.0316
##
```

```
## age.in.months = 36, container.title = Applied and Environmental Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.072 0.0431 Inf    1   1.727  0.0842
##
## age.in.months = 60, container.title = Applied and Environmental Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.186 0.0347 Inf    1   5.826  <.0001
##
## age.in.months = 120, container.title = Applied and Environmental Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.360 0.0311 Inf    1  13.440  <.0001
##
## age.in.months = 12, container.title = Genome Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.126 0.7950 Inf    1   0.169  0.8662
##
## age.in.months = 36, container.title = Genome Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.096 0.4030 Inf    1   0.251  0.8020
##
## age.in.months = 60, container.title = Genome Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.083 0.2290 Inf    1   0.376  0.7068
##
## age.in.months = 120, container.title = Genome Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.065 0.0567 Inf    1   1.175  0.2399
##
## age.in.months = 12, container.title = Infection and Immunity:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.901 0.1580 Inf    1  -0.599  0.5494
##
## age.in.months = 36, container.title = Infection and Immunity:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.037 0.1070 Inf    1   0.350  0.7267
##
## age.in.months = 60, container.title = Infection and Immunity:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.107 0.0856 Inf    1   1.315  0.1883
##
## age.in.months = 120, container.title = Infection and Immunity:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.210 0.0772 Inf    1   2.989  0.0028
##
## age.in.months = 12, container.title = Journal of Bacteriology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.923 0.1300 Inf    1  -0.565  0.5719
##
## age.in.months = 36, container.title = Journal of Bacteriology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.885 0.0739 Inf    1  -1.469  0.1419
##
## age.in.months = 60, container.title = Journal of Bacteriology:
##  contrast  ratio     SE  df null z.ratio p.value
```

```
##  Yes / No  0.867 0.0508 Inf    1  -2.438  0.0147
##
## age.in.months = 120, container.title = Journal of Bacteriology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.844 0.0276 Inf    1  -5.193  <.0001
##
## age.in.months = 12, container.title = Journal of Clinical Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.930 0.1230 Inf    1  -0.549  0.5831
##
## age.in.months = 36, container.title = Journal of Clinical Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.232 0.0944 Inf    1   2.721  0.0065
##
## age.in.months = 60, container.title = Journal of Clinical Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.404 0.0786 Inf    1   6.058  <.0001
##
## age.in.months = 120, container.title = Journal of Clinical Microbiology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.676 0.0783 Inf    1  11.064  <.0001
##
## age.in.months = 12, container.title = Journal of Microbiology &amp;amp; Biology Education:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No nonEst     NA  NA    1      NA      NA
##
## age.in.months = 36, container.title = Journal of Microbiology &amp;amp; Biology Education:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No nonEst     NA  NA    1      NA      NA
##
## age.in.months = 60, container.title = Journal of Microbiology &amp;amp; Biology Education:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No nonEst     NA  NA    1      NA      NA
##
## age.in.months = 120, container.title = Journal of Microbiology &amp;amp; Biology Education:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No nonEst     NA  NA    1      NA      NA
##
## age.in.months = 12, container.title = Journal of Virology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.798 0.0725 Inf    1  -2.482  0.0131
##
## age.in.months = 36, container.title = Journal of Virology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.987 0.0500 Inf    1  -0.255  0.7988
##
## age.in.months = 60, container.title = Journal of Virology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.090 0.0409 Inf    1   2.291  0.0220
##
## age.in.months = 120, container.title = Journal of Virology:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.246 0.0459 Inf    1   5.974  <.0001
##
```

```
## age.in.months = 12, container.title = mBio:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.024 0.0891 Inf    1   0.278  0.7813
##
## age.in.months = 36, container.title = mBio:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.213 0.0530 Inf    1   4.411  <.0001
##
## age.in.months = 60, container.title = mBio:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.311 0.0486 Inf    1   7.311  <.0001
##
## age.in.months = 120, container.title = mBio:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.458 0.0769 Inf    1   7.161  <.0001
##
## age.in.months = 12, container.title = Microbiology Resource Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.557 0.3020 Inf    1  -1.080  0.2800
##
## age.in.months = 36, container.title = Microbiology Resource Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.637 0.1410 Inf    1  -2.033  0.0421
##
## age.in.months = 60, container.title = Microbiology Resource Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.678 0.1380 Inf    1  -1.910  0.0562
##
## age.in.months = 120, container.title = Microbiology Resource Announcements:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.738 0.2740 Inf    1  -0.818  0.4134
##
## age.in.months = 12, container.title = Microbiology Spectrum:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.053 0.0812 Inf    1   0.664  0.5064
##
## age.in.months = 36, container.title = Microbiology Spectrum:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.332 0.0819 Inf    1   4.657  <.0001
##
## age.in.months = 60, container.title = Microbiology Spectrum:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.486 0.1550 Inf    1   3.798  0.0001
##
## age.in.months = 120, container.title = Microbiology Spectrum:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.724 0.2920 Inf    1   3.214  0.0013
##
## age.in.months = 12, container.title = mSphere:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.045 0.1650 Inf    1   0.276  0.7823
##
## age.in.months = 36, container.title = mSphere:
##  contrast  ratio     SE  df null z.ratio p.value
```

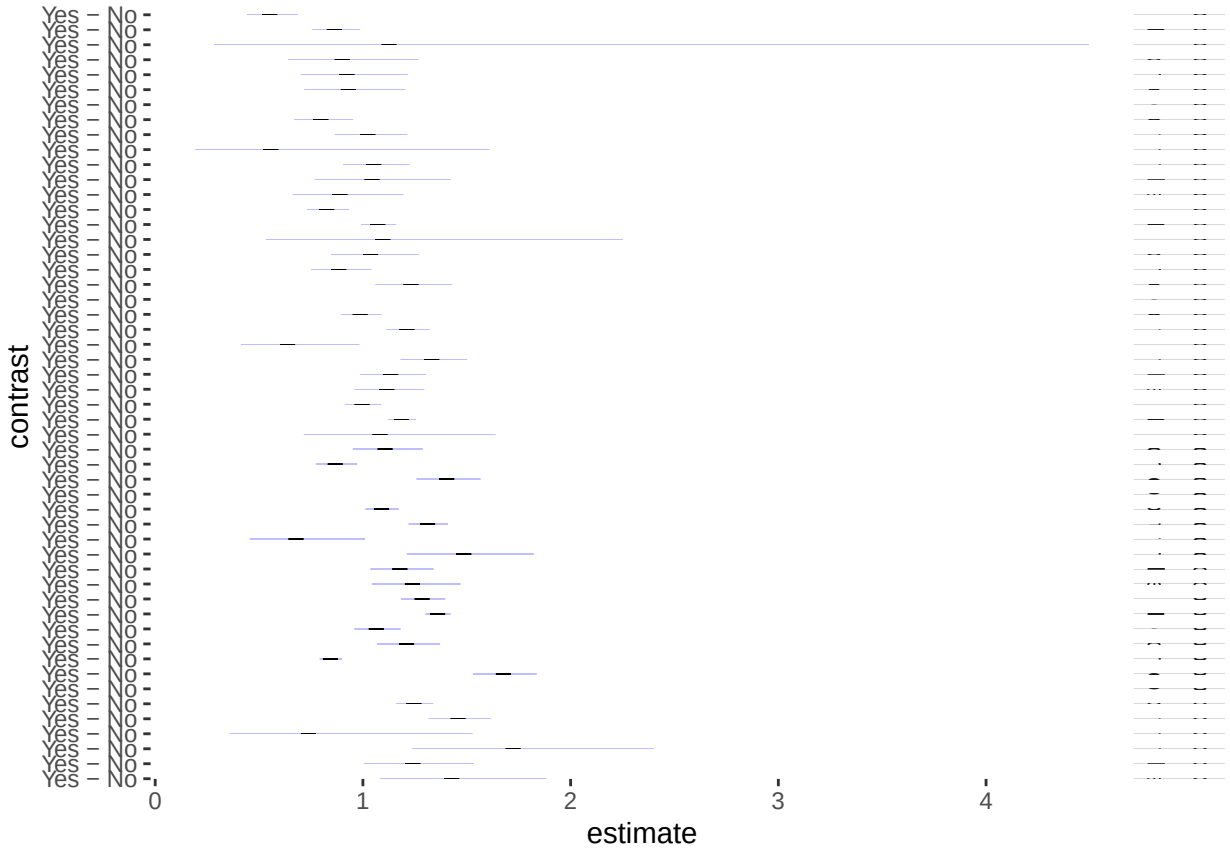
```
##  Yes / No  1.134 0.0815 Inf    1   1.753  0.0797
##
## age.in.months = 60, container.title = mSphere:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.178 0.0776 Inf    1   2.492  0.0127
##
## age.in.months = 120, container.title = mSphere:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.241 0.1340 Inf    1   1.998  0.0458
##
## age.in.months = 12, container.title = mSystems:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  0.890 0.1340 Inf    1  -0.773  0.4394
##
## age.in.months = 36, container.title = mSystems:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.115 0.0853 Inf    1   1.424  0.1546
##
## age.in.months = 60, container.title = mSystems:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.238 0.1090 Inf    1   2.439  0.0147
##
## age.in.months = 120, container.title = mSystems:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.428 0.2020 Inf    1   2.519  0.0118
##
## Tests are performed on the log scale
```

```
# Plot the contrasts
plot(differences, ratios = TRUE)
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_segment()`).
```

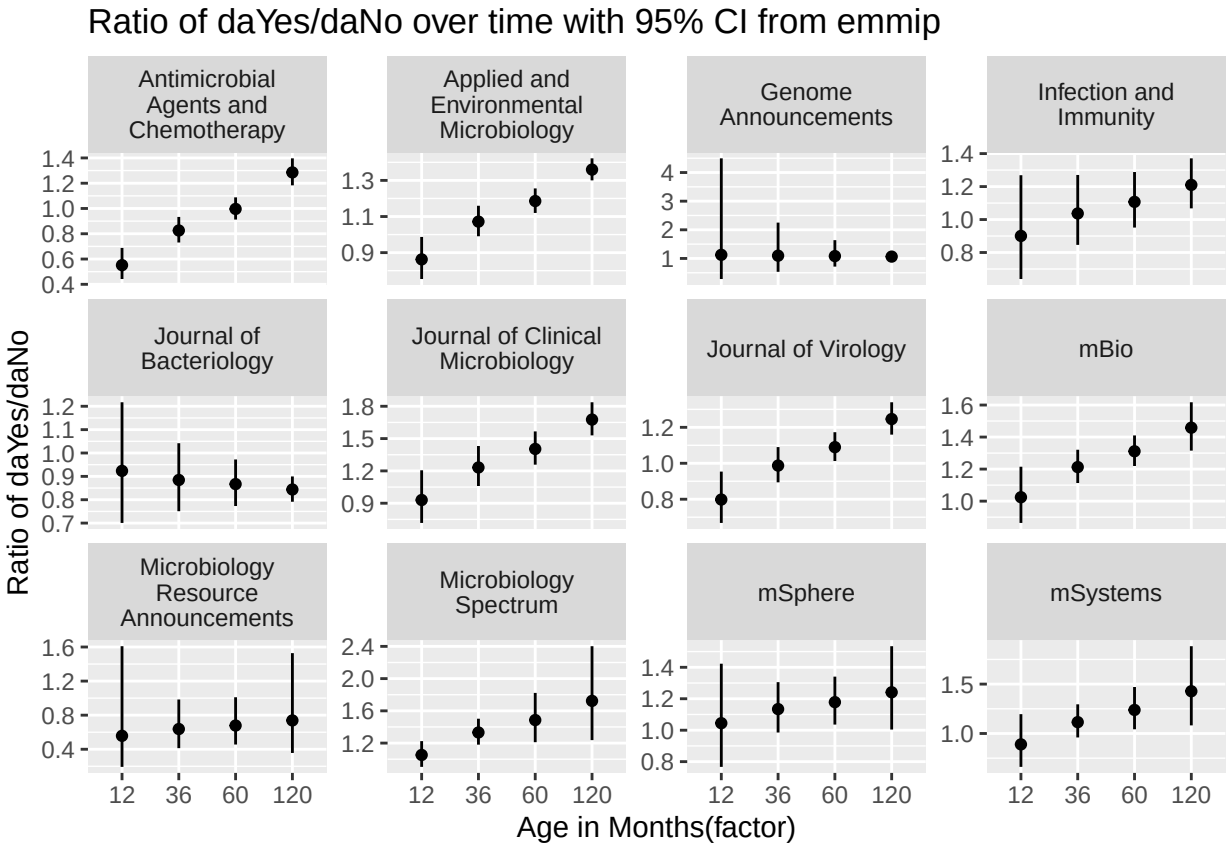
```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
# # working on plotting better
ratio_data <-  tibble(emmip(differences, ~ age.in.months | container.title, CIs = TRUE, plotit = FALSE))
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
ratio_data %>%
  filter(container.title != "Journal of Microbiology &amp;amp; Biology Education") %>%
ggplot(ratio_data, mapping = aes(x = xvar, y = yvar)) +
  geom_point() +
  geom_linerange(aes(ymin = LCL, ymax = UCL)) +
  facet_wrap(vars(container.title), scales = "free_y",
             labeller = label_wrap_gen(width = 20))  +
  labs(title = "Ratio of daYes/daNo over time with 95% CI from emmip",
       x = "Age in Months(factor)",
       y = "Ratio of daYes/daNo")
```

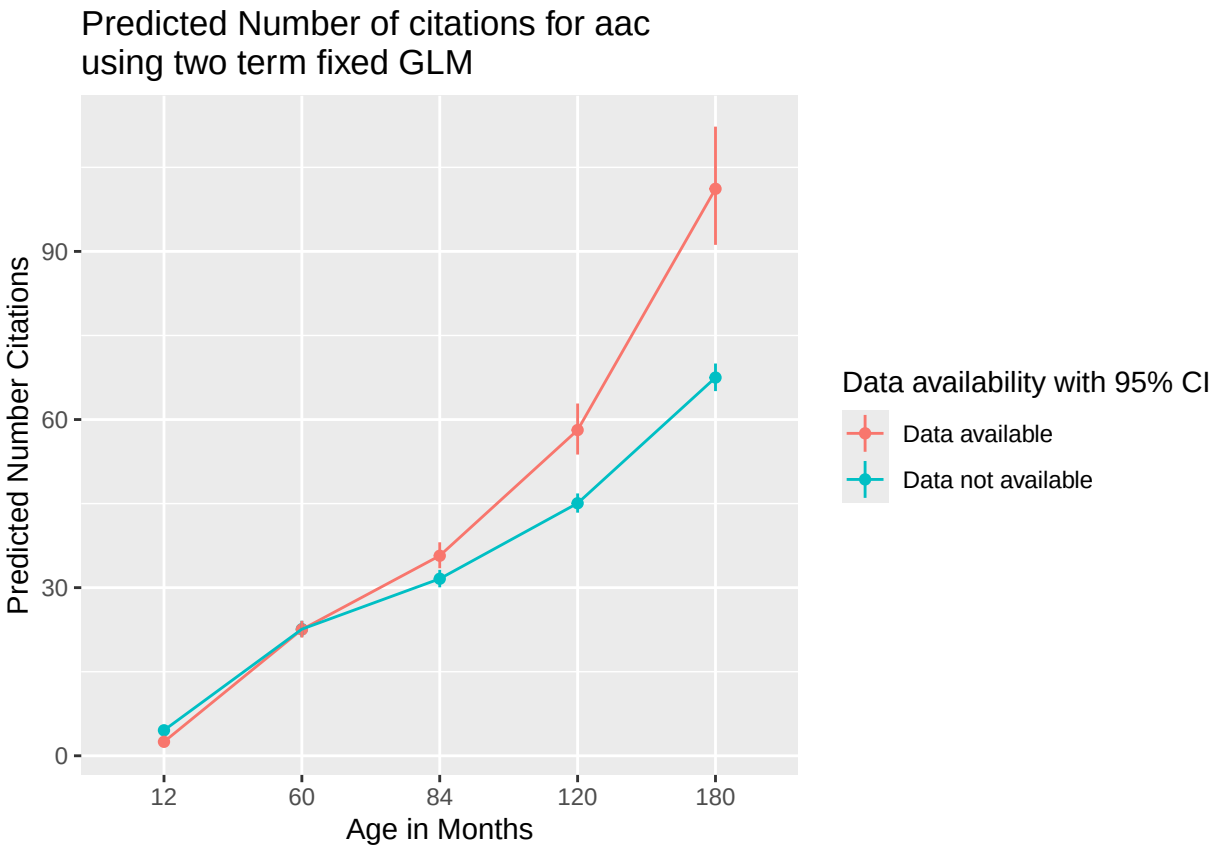


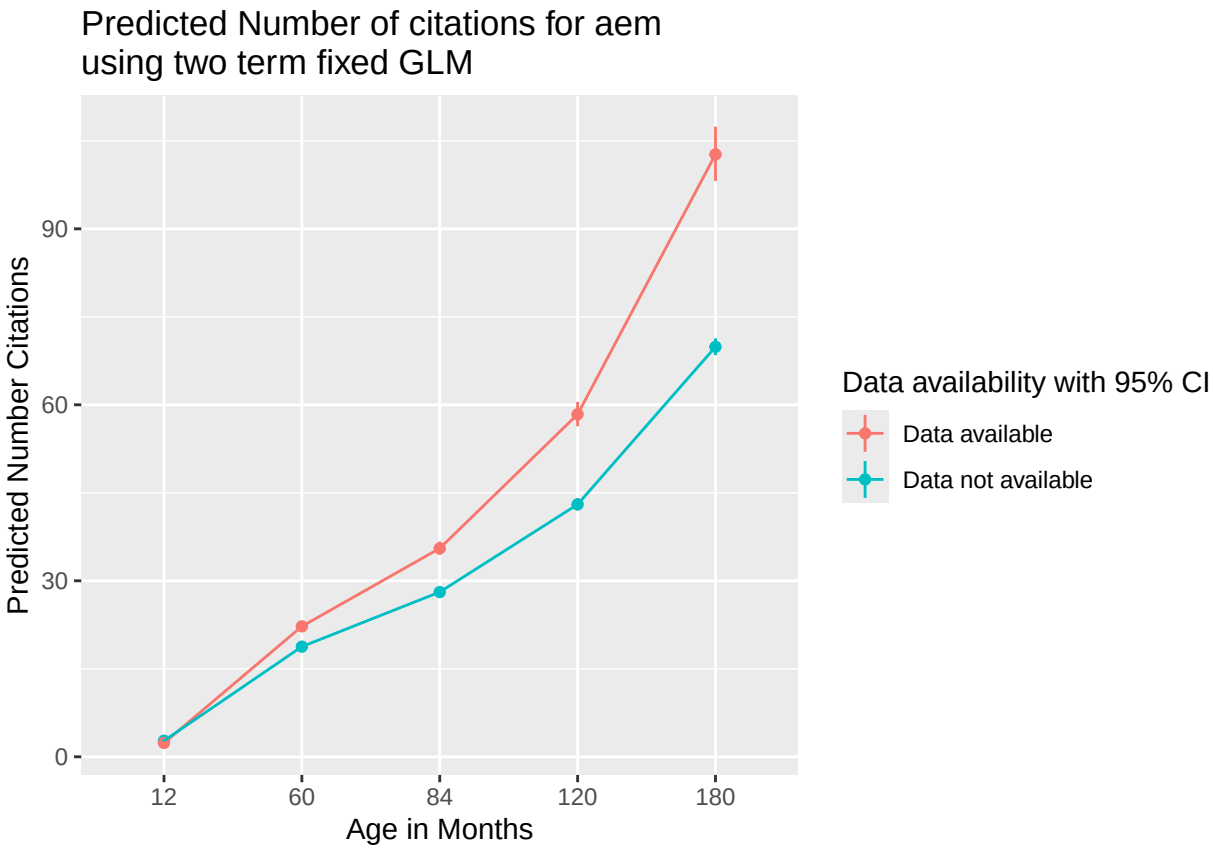
Create Plot from Abner

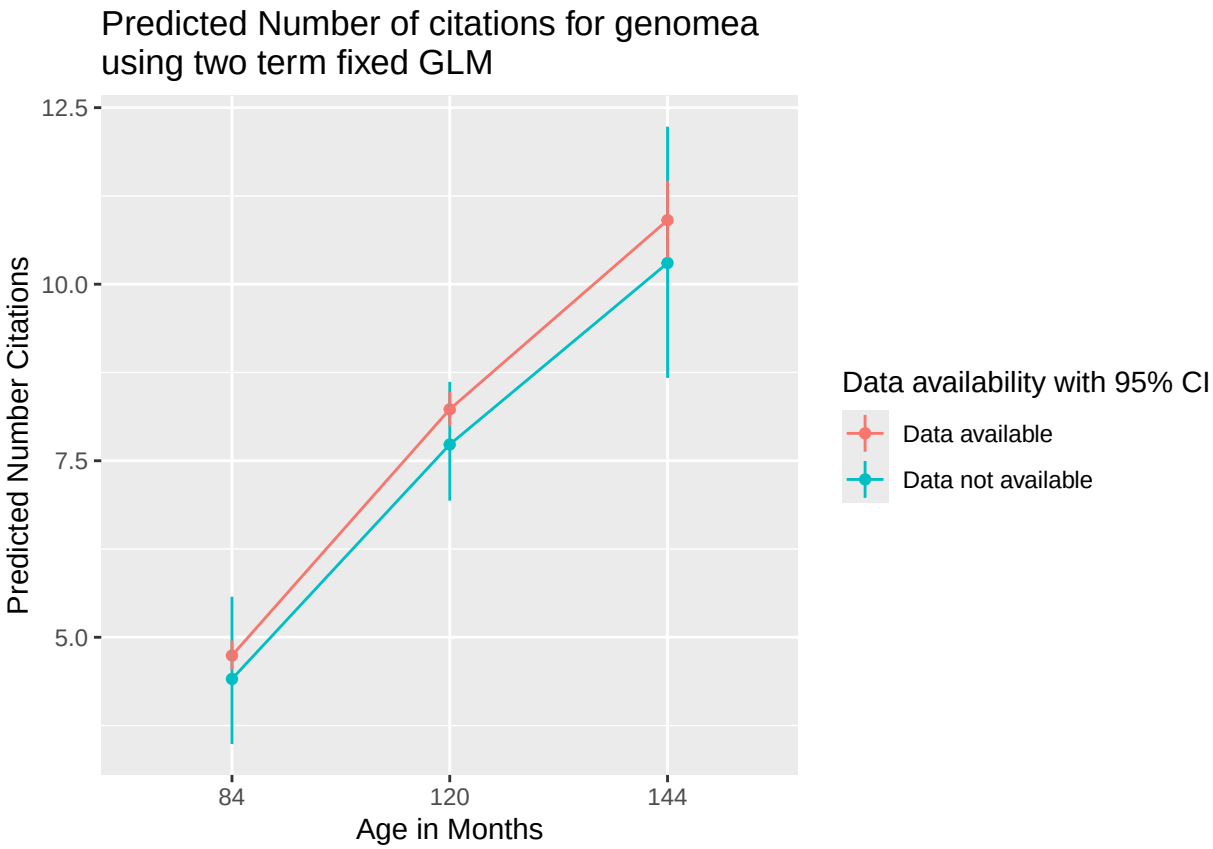
- "Also, I think this result would be even clearer if you made a plot with"age" in the horizontal axis, "predicted citations" in the vertical axis, and lines colored by "da_factor"." - AHB
- These plots are each made with a different glm.nb model (one for each journal), which is why they are not combined into a faceted plot.

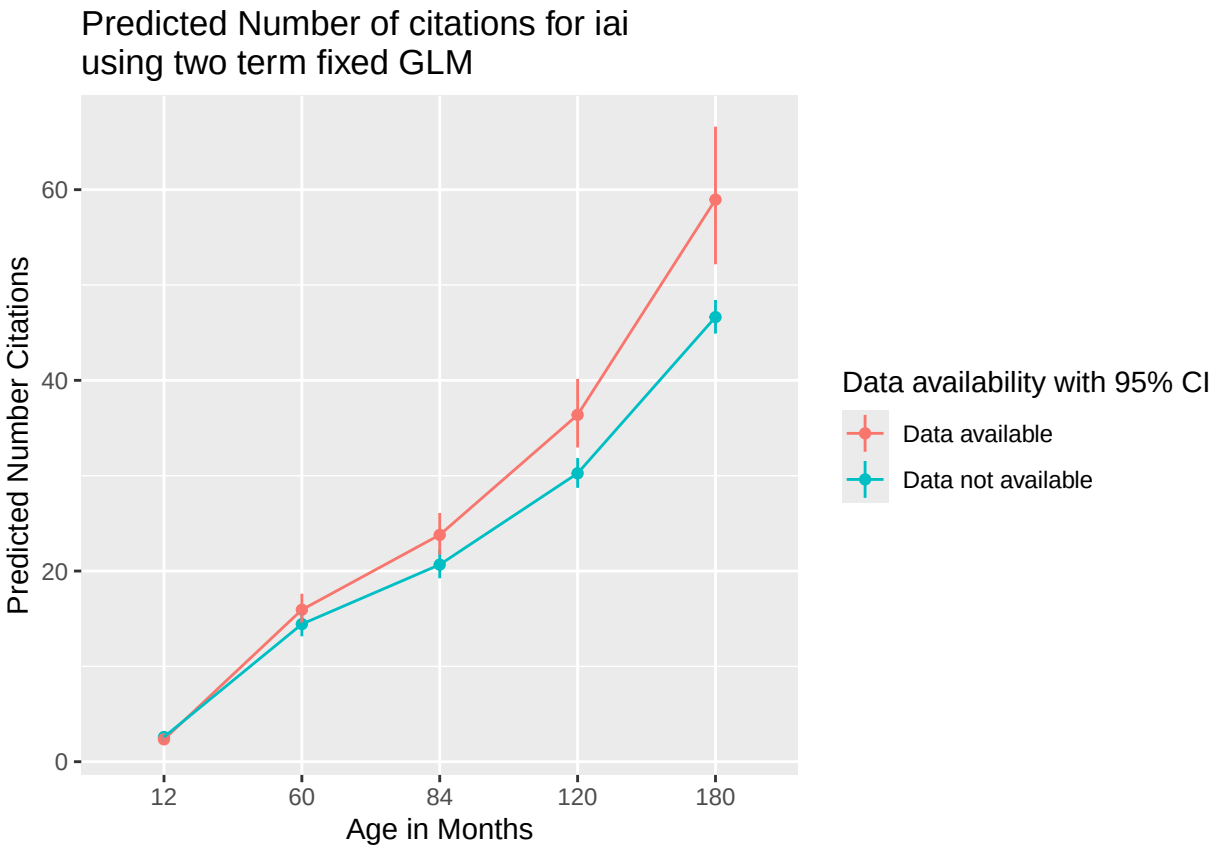
```
#smaller model with 2 terms
two_term_glmnb <-function(model_data, model_name) {

  total_model <-MASS::glm.nb(is.referenced.by.count~ da_factor + log(age.in.months) +
       + log(age.in.months)*da_factor + log(age.in.months)*da_factor, data = model_data, link = log)

  return(total_model)

}

journals <-
  nsd_yes_metadata %>%
  count(journal_abrev) %>%
  filter(journal_abrev != "jmbe")

j <- 3

for(j in 1:nrow(journals)) {
```

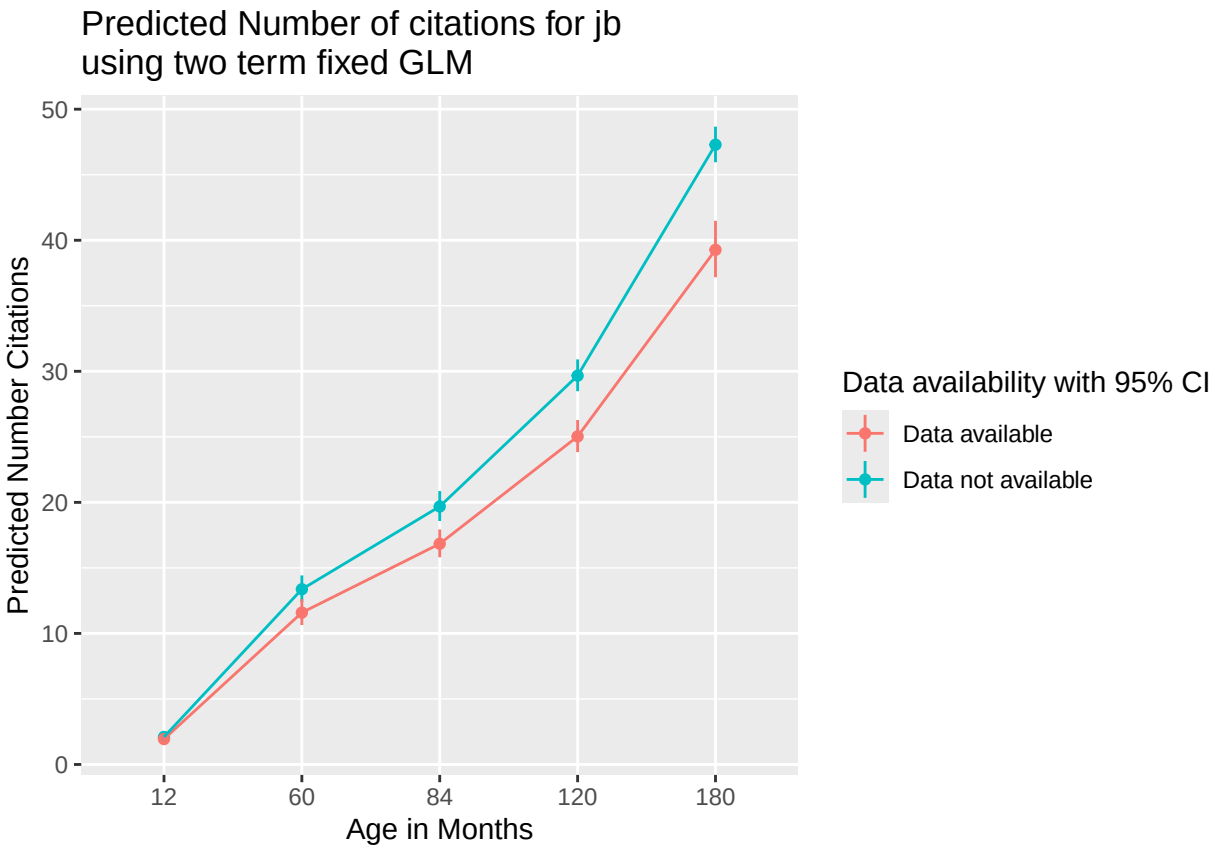
```
  journal_data <-
  nsd_yes_metadata %>%
    filter(journal_abrev == journals[[j,1]]) %>%
    mutate(da_factor = factor(da))


  model <- two_term_glmnb(journal_data, journals[[j,1]])




  if(j == 3) { #genomea = 5
  p <-  get_model_data(model = model, type = "pred", terms = c("da_factor", "age.in.months[84,120,144]")
  }
  else if (j == 9) { ##Mra - 7
    p <-  get_model_data(model = model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84]")

  }
  else if (j == 10 | j == 11 ) { #msphere, msystems 9 years
  p <-  get_model_data(model = model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84,108]
  }
  else if (j == 12 ) {
    p <-  get_model_data(model = model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84,1
  }
  else {
     p <-  get_model_data(model = model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84,
  }

 model_graph <-
   p %>%
   tibble(da_factor = ifelse(.$x == 1, "Data not available", "Data available"), predicted_citations = .$
  ggplot(mapping = aes(x = age.in.months, y = predicted_citations,
                               color = da_factor)) +
   geom_point() +
   geom_linerange(aes(ymin = conf.low,
                      ymax = conf.high)) +
   geom_path(aes(x = age.in.months, y = predicted_citations, group = da_factor)) +
   labs(title = paste0("Predicted Number of citations for ", journals[[j,1]], "\nusing two term fixed GL
        x = "Age in Months",
        y = "Predicted Number Citations",
        color = "Data availability with 95% CI")

 assign(paste0(journals[[j,1]], "_plot"), model_graph)
 print(model_graph)

}
```

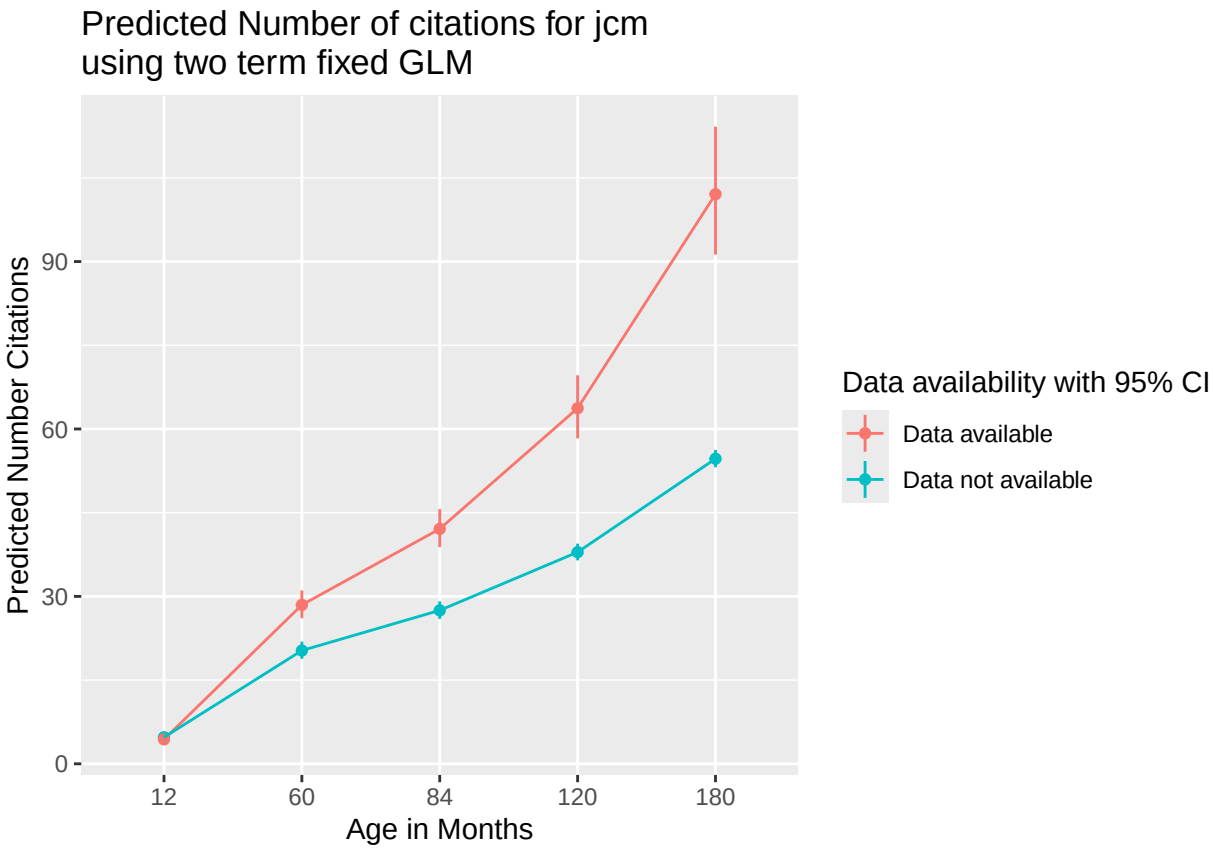

Predicted Number of citations for aac
using two term fixed GLM
Predicted Number Citations
90
60
30
0
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available

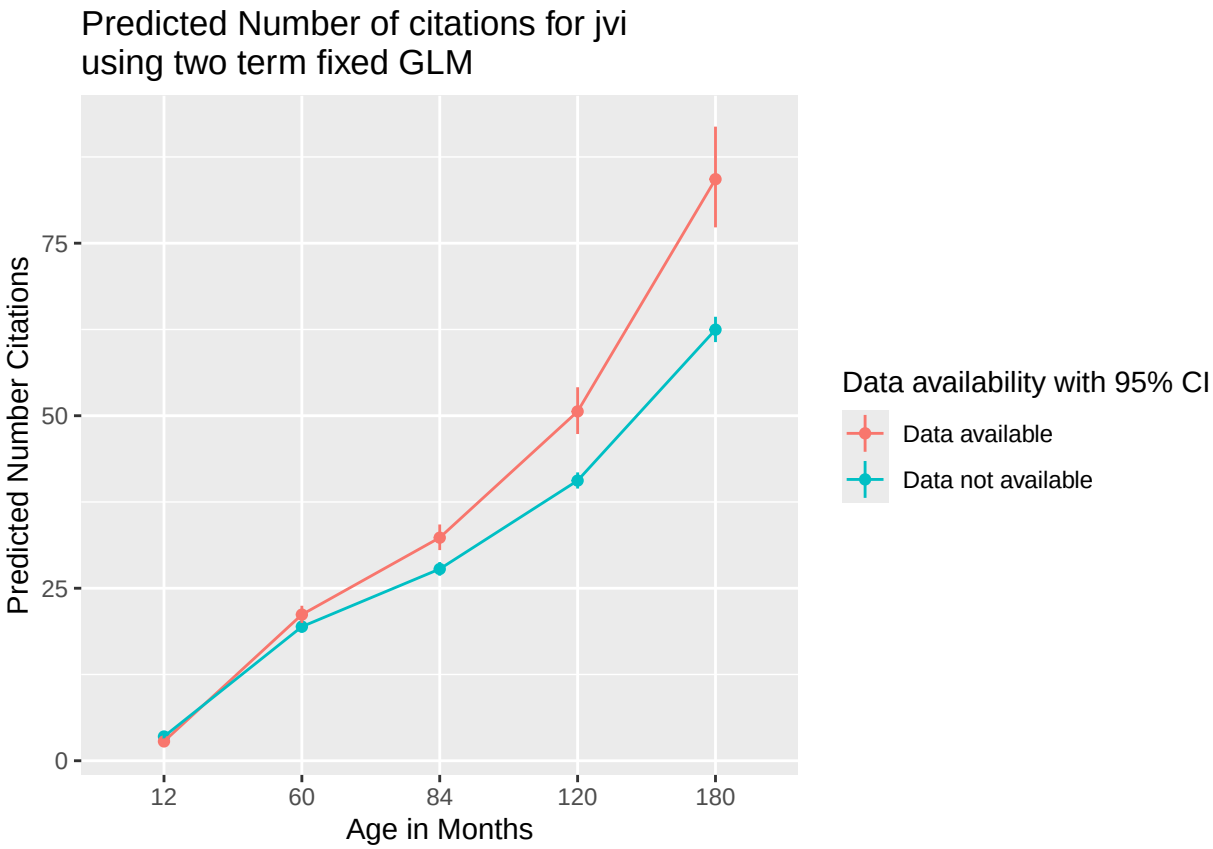

Predicted Number of citations for aem
using two term fixed GLM
Predicted Number Citations
90
60
30
0
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available

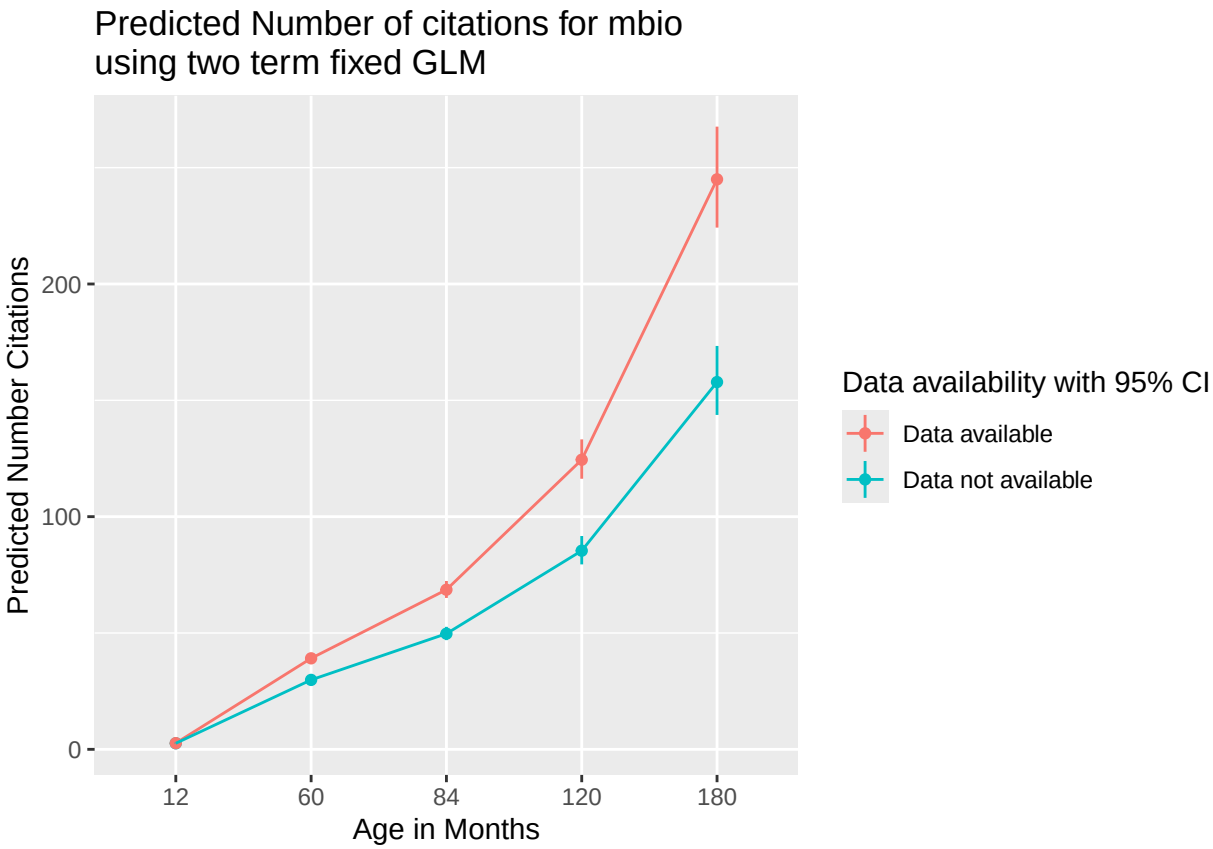

Predicted Number of citations for genomea using two term fixed GLM
Predicted Number Citations
12.5
10.0
7.5
5.0
84
120
144
Age in Months
Data availability with 95% CI
Data available
Data not available

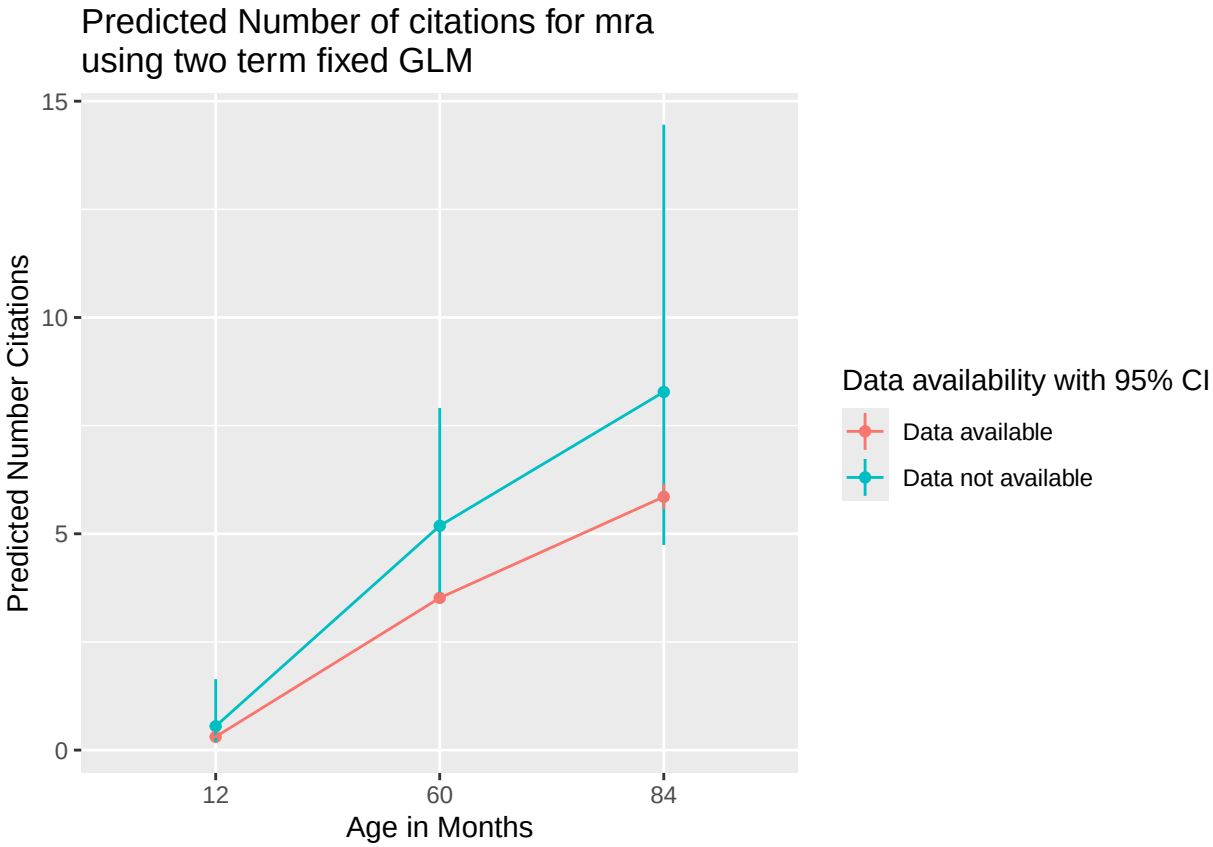

Predicted Number of citations for iai
using two term fixed GLM
Predicted Number Citations
0
20
40
60
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available

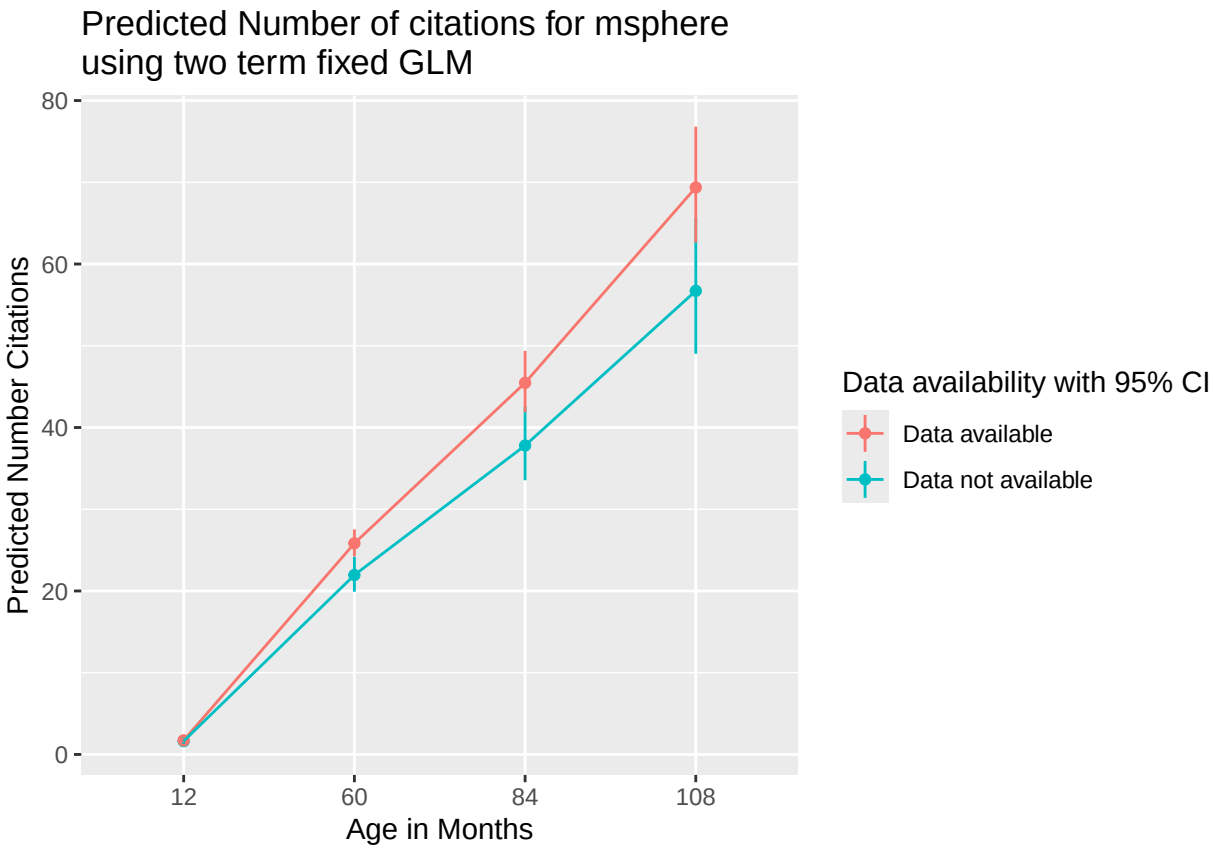

Predicted Number of citations for jb
using two term fixed GLM
Predicted Number Citations
50
40
30
20
10
0
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available

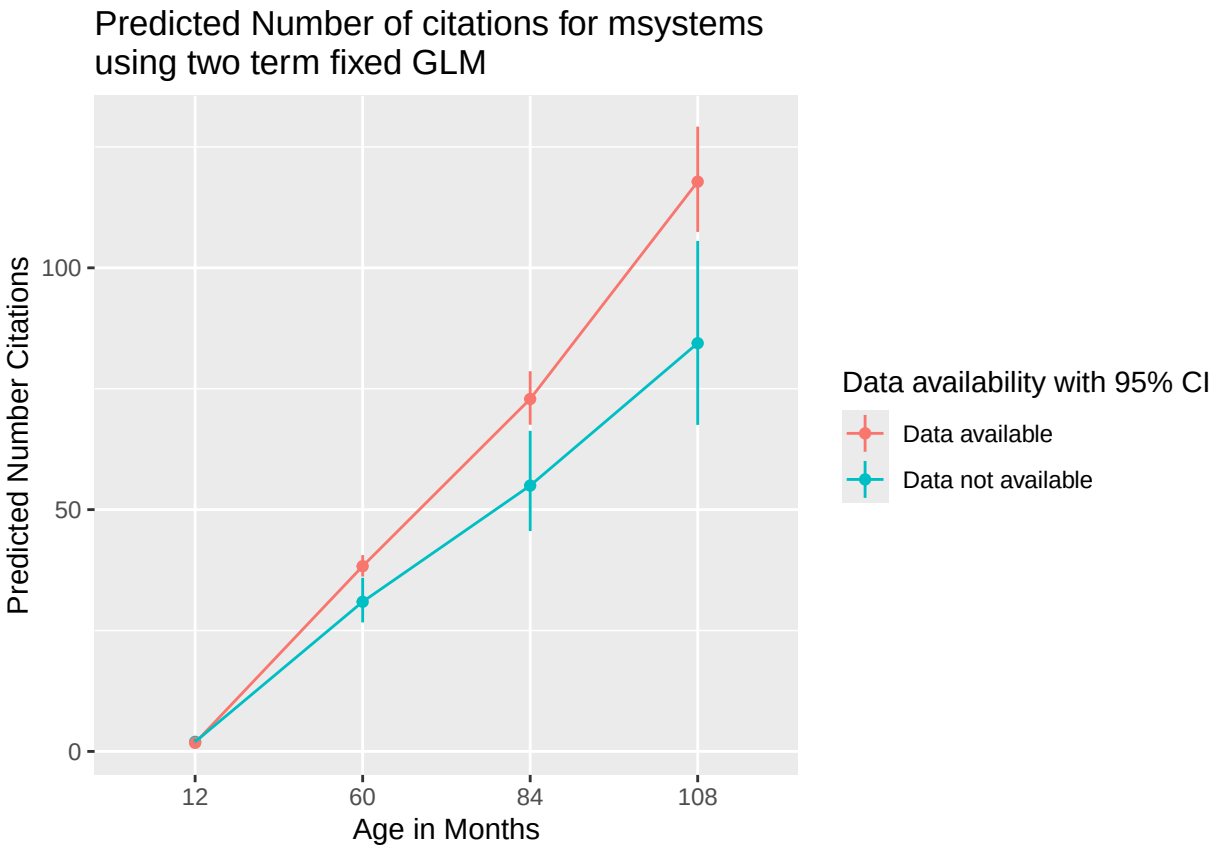

Predicted Number of citations for jcm
using two term fixed GLM
Predicted Number Citations
0
30
60
90
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available

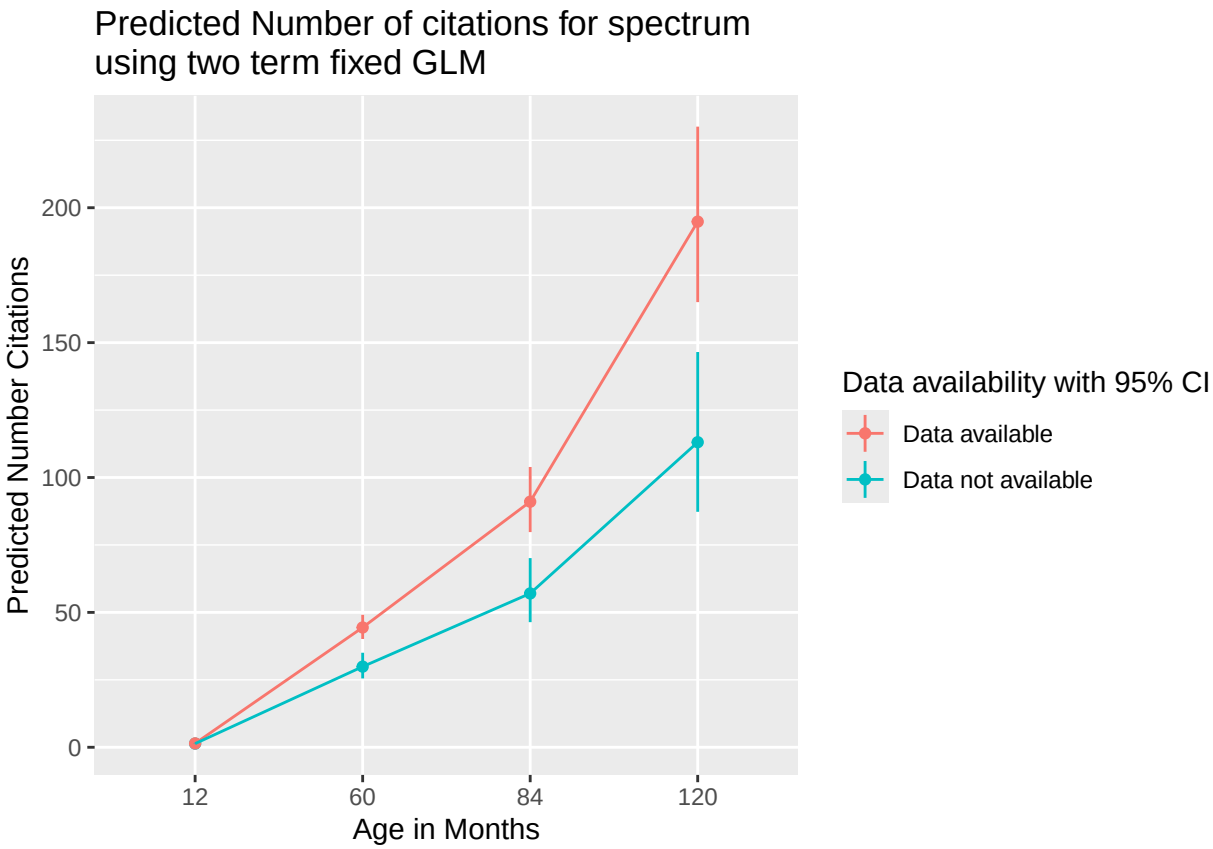

Predicted Number of citations for jvi
using two term fixed GLM
Predicted Number Citations
75
50
25
0
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available


Predicted Number of citations for mbio
using two term fixed GLM
Predicted Number Citations
200
100
0
12
60
84
120
180
Age in Months
Data availability with 95% CI
Data available
Data not available


Predicted Number of citations for mra
using two term fixed GLM
Predicted Number Citations
15
10
5
0
12
60
84
Age in Months
Data availability with 95% CI
Data available
Data not available


Predicted Number of citations for msphere using two term fixed GLM
Predicted Number Citations
80
60
40
20
0
12
60
84
108
Age in Months
Data availability with 95% CI
Data available
Data not available


Predicted Number of citations for msystems using two term fixed GLM
Predicted Number Citations
100
50
0
12
60
84
108
Age in Months
Data availability with 95% CI
Data available
Data not available




```
# plot_names <-
# journals %>%
#   mutate(plot_name = paste0(journal_abrev, "_plot")) %>%
#   select(plot_name)
#
#
# plot_grid(plotlist = mget(paste0(journals$journal_abrev,"_plot")))
```