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
```

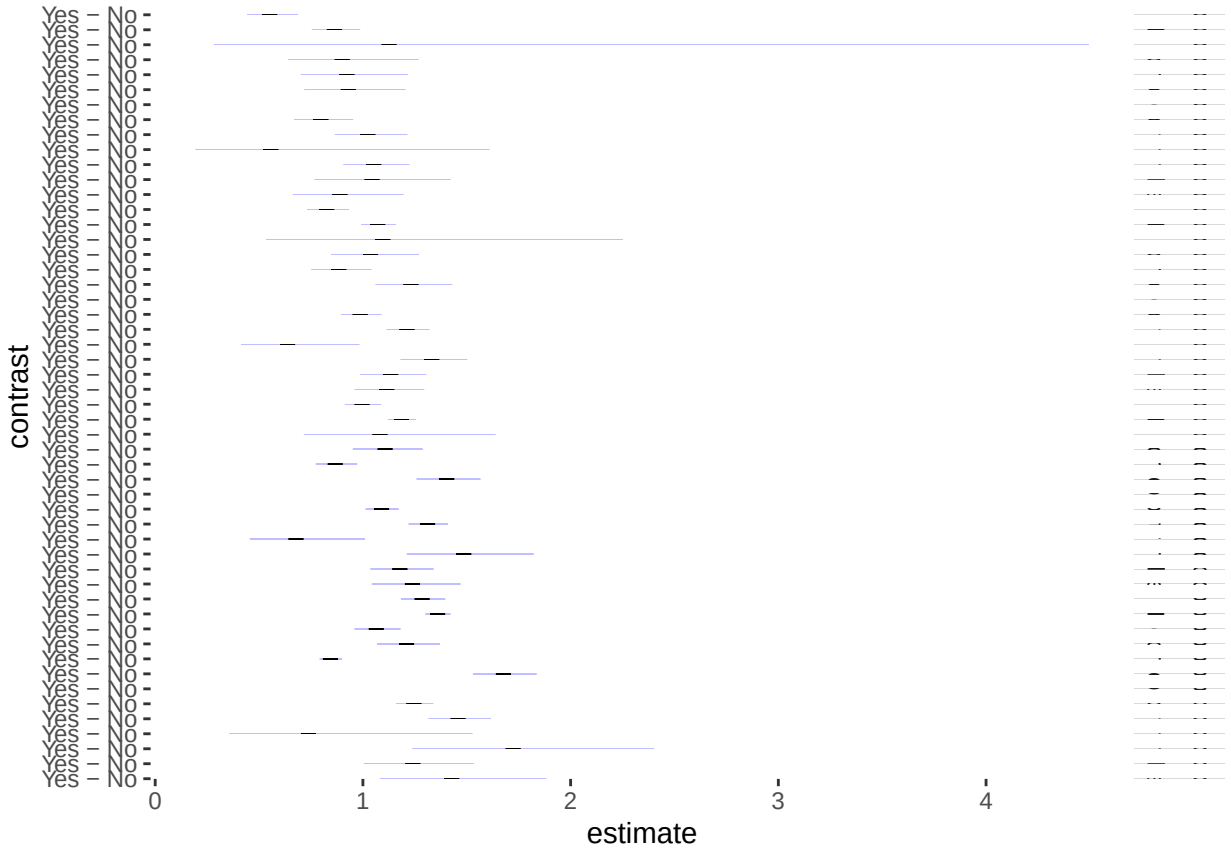
```
##  Yes / No  1.134 0.0815 Inf    1   1.753  0.0797
##
## age.in.months = 60, container.title = mSphere:
##  contrast  ratio     SE  df null z.ratio p.value
##  Yes / No  1.178 0.0776 Inf    1   2.492  0.0127
##
## age.in.months = 120, container.title = mSphere:
##  contrast  ratio     SE  df null 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()`).
```

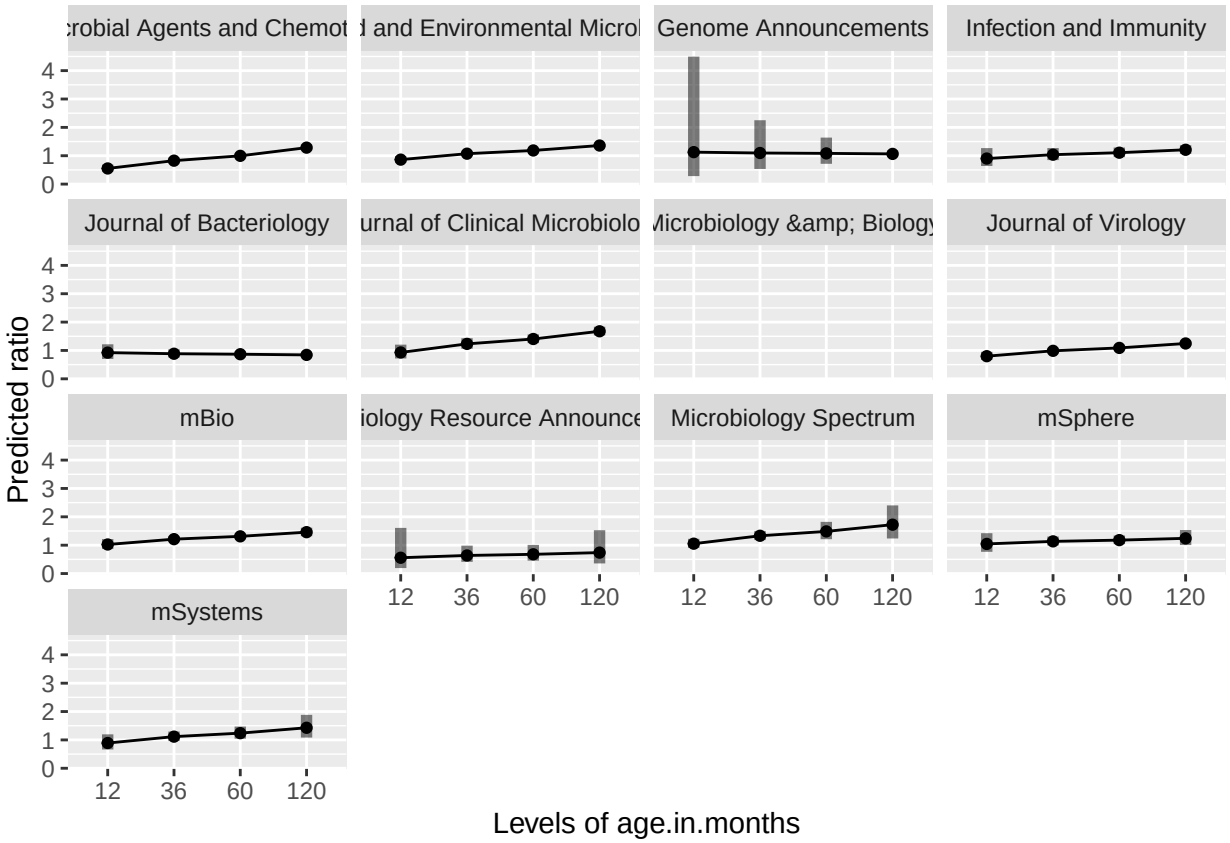
```
## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
```




```
# # working on plotting better
# emmip(model, ~ age.in.months | container.title, CIs = TRUE, type = "response",  at = list(age.in.mont
#     geom_point(aes(x = age.in.months, y = is.referenced.by.count), data = my_data, size = 1, color =

emmip(differences, ~ age.in.months | container.title, CIs = TRUE, engine = "ggplot")
```

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

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



The Other 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".

```
library(DHARMa)
```

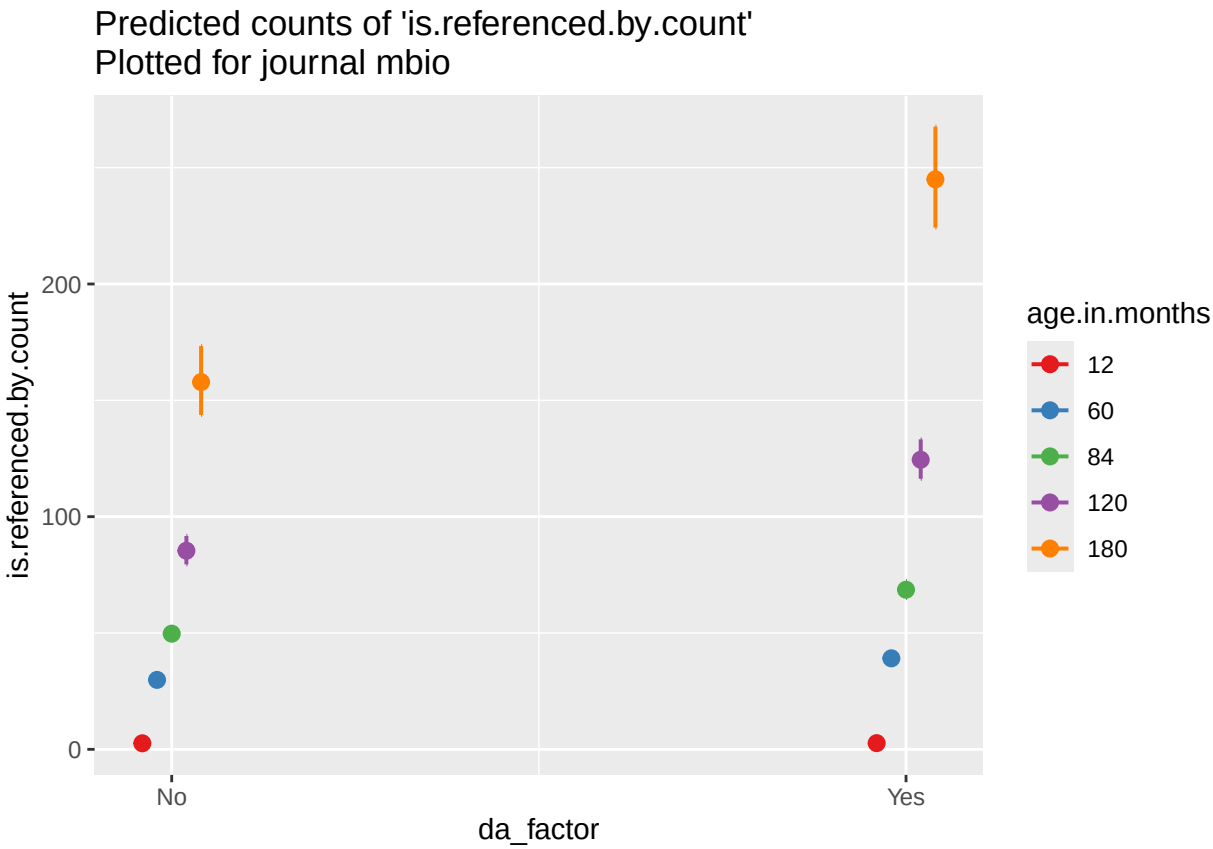
```
## This is DHARMa 0.4.7. For overview type '?DHARMa'. For recent changes, type news(package = 'DHARMa')
```

```
library(sjPlot)
#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")
```

```r
j <- 8 #mbio


  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]])
summary(model)
```

```
##
## Call:
## MASS::glm.nb(formula = 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, init.theta = 1.597744281)
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    -2.81124    0.15592 -18.030  < 2e-16 ***
## da_factorYes                   -0.35970    0.20003  -1.798  0.07215 .
## log(age.in.months)              1.51608    0.03718  40.781  < 2e-16 ***
## da_factorYes:log(age.in.months)  0.15390    0.04860   3.166  0.00154 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(1.5977) family taken to be 1)
##
##     Null deviance: 6961.8  on 2437  degrees of freedom
## Residual deviance: 2603.7  on 2434  degrees of freedom
##   (60 observations deleted due to missingness)
## AIC: 20128
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  1.5977
##           Std. Err.:  0.0482
##
##  2 x log-likelihood:  -20118.0870
```

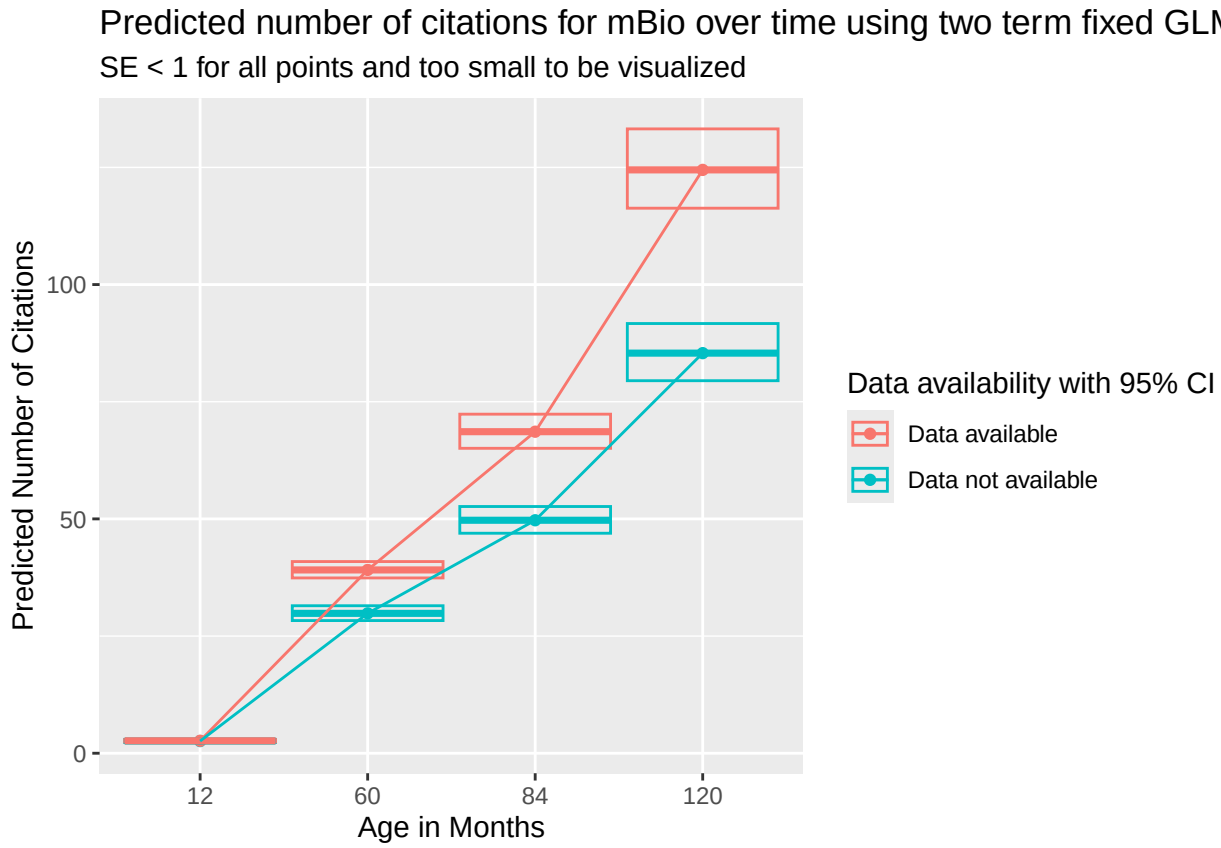
```r
  plot_model <- plot_model(model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84,120,180]

 print(plot_model)
```



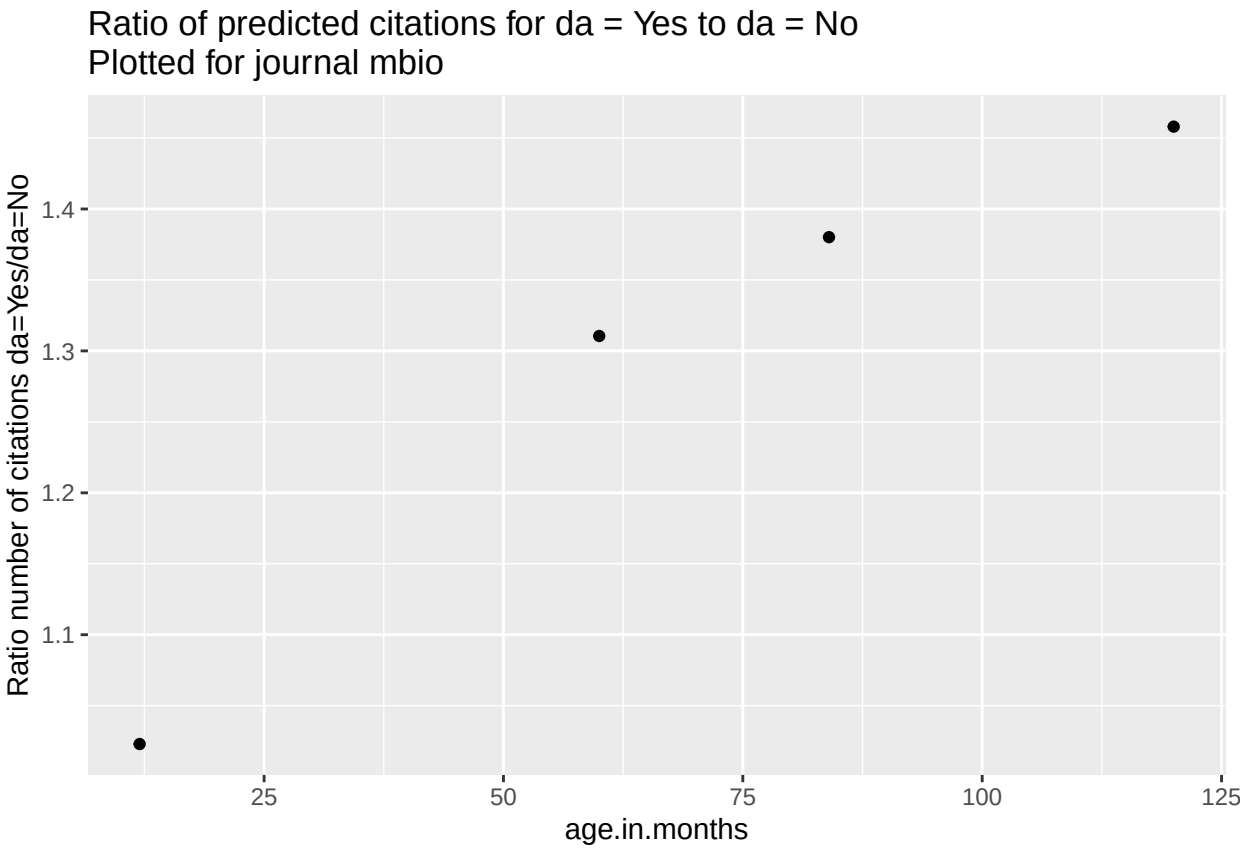

```
model_data <- get_model_data(model, type = "pred", terms = c("da_factor", "age.in.months[12,60,84,120]
  tibble(da_factor = ifelse(.$x == 1, "Data not available", "Data available"), predicted_citations = .

kableExtra::kable(model_data)
```

x	predicted	std.error	conf.low	conf.high	group	group_colda_factor	predicted_citations	age.in.months
1	2.601432	0.0668792	2.281838	2.965789	12	12	Data not available	2.601432	12
1	29.847220	0.0270486	28.306105	31.472241	60	60	Data not available	29.847220	60
1	49.710146	0.0292298	46.942326	52.641164	84	84	Data not available	49.710146	84
1	85.366628	0.0364169	79.485880	91.682461	120	120	Data not available	85.366628	120
2	2.661219	0.0506662	2.409648	2.939056	12	12	Data available	2.661219	12
2	39.114855	0.0227761	37.407152	40.900517	60	60	Data available	39.114855	60
2	68.607500	0.0270623	65.063307	72.344756	84	84	Data available	68.607500	84
2	124.466847	0.0346578	116.292821	133.215412	120	120	Data available	124.466847	120

```
ggplot(data = model_data, aes(x = age.in.months, y = predicted_citations,
                              color = da_factor)) +
  geom_point() +
  geom_crossbar(aes(ymin = conf.low,
                    ymax = conf.high)) +
  # geom_linerange(aes(ymin = predicted_citations - std.error,
  #                    ymax = predicted_citations + std.error)) +
  geom_path(aes(x = age.in.months, y = predicted_citations, group = da_factor)) +
  labs(title = "Predicted number of citations for mBio over time using two term fixed GLM",
       subtitle = "SE < 1 for all points and too small to be visualized",
       x = "Age in Months",
       y = "Predicted Number of Citations",
       color = "Data availability with 95% CI")
```




```
tibble(da = model_data$x, predicted = model_data$predicted, age.in.months = as.numeric(as.character(mod
    pivot_wider(names_from = da, values_from = predicted) %>%
    mutate(ratio = `2`/`1`) %>%
    ggplot(., aes(x = age.in.months, y = ratio)) +
    geom_point() +
     labs(title = paste0("Ratio of predicted citations for da = Yes to da = No \nPlotted for journal ",
          y = "Ratio number of citations da=Yes/da=No")
```


Ratio of predicted citations for da = Yes to da = No
Plotted for journal mbio
Ratio number of citations da=Yes/da=No
1.4
1.3
1.2
1.1
25
50
75
100
125
age.in.months