Swallowed Fluticasone Propionate is an Effective Long-Term Maintenance Therapy for Children with Eosinophilic Esophagitis: Supplemental statistical analysis

Introduction

Below follows an abbreviated summary of our model building. We build a mixed linear models to test the hypothesis of a *sustained* decrease in peak esosinophil count after treatment initiation. To account for patient variability and the correlation of repeated observations we include a random effect for endoscopy follow up visit

Load EoE data

We load the data from the Rdata file *EoEdata9Feb.Rdata*

Dataset as table

Patient	Endoscopy	Date	Eos	Diet	PPI	Endoscopy.factor	treat	furtherfollowup	Followup
1	1	2012-04-30	52	1	0	1	0	0	0
1	2	2012-07-23	2	1	0	2	1	0	84
2	1	2009-08-10	82	1	0	1	0	0	0
2	2	2009-12-14	83	1	0	2	1	0	126
3	1	2010-05-10	38	0	0	1	0	0	0
3	2	2010-12-17	0	0	0	2	1	0	221

This is an observational open label study of 54 subjects, each with several (between 1 to 6) follow up endoscopies at variable intervals. The outcome is continuous (Eos), a repeated measure of eosinophil *count*, showing on average a response to treatment over time.

Diet indicated if patient was pretreated with diet PPI indicated concurrent treatment with proton inhibitor

Endoscopy codes for the follow up visit endoscopy, either as integer or as factor.

treat indicates start of treatment (after first visit) furtherfollowup codes for endoscopy visits after the second. Followup is days of follow up at which Eos was measured (in days).

Explore response

Table of mean and medians

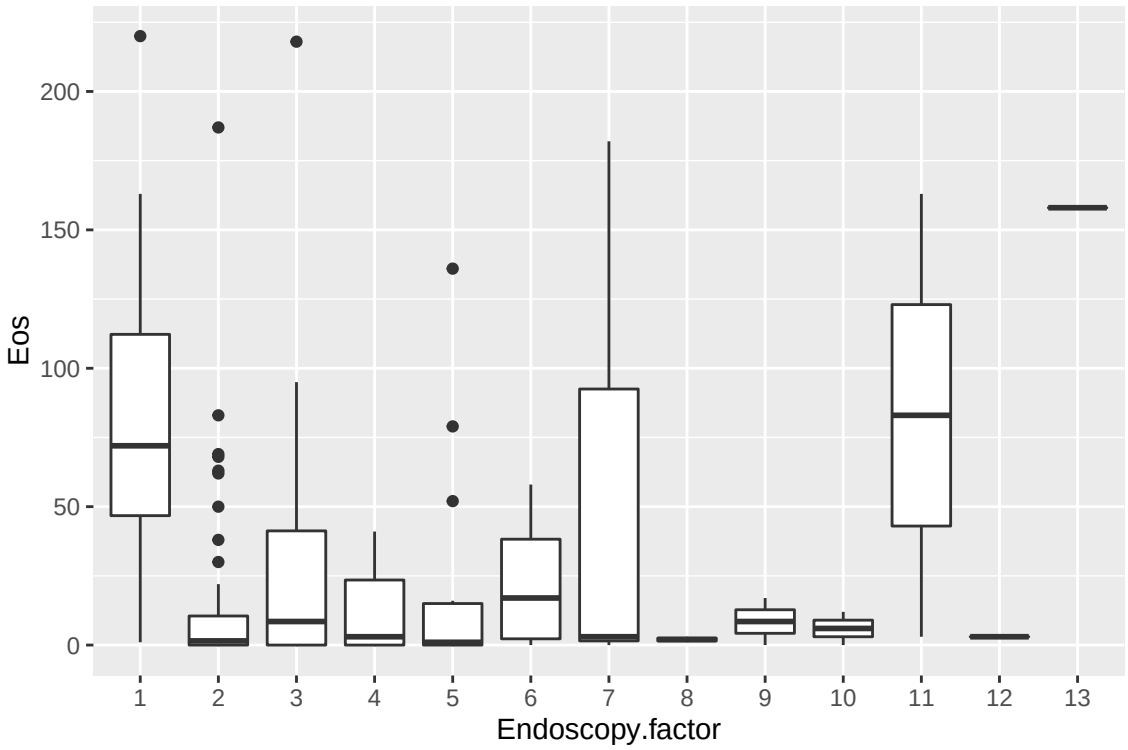
Endoscopy	Number	MedianEos	meanEos
1	54	72.0	78.46
2	54	1.5	14.65

Endoscopy	Number	MedianEos	meanEos
3	30	8.5	27.50
4	19	3.0	11.32
5	14	1.0	21.64
6	8	17.0	22.25
7	3	3.0	61.67
8	2	2.0	2.00
9	2	8.5	8.50
10	2	6.0	6.00
11	2	83.0	83.00
12	1	3.0	3.00
13	1	158.0	158.00

This table list the median Eos count by endoscopy.

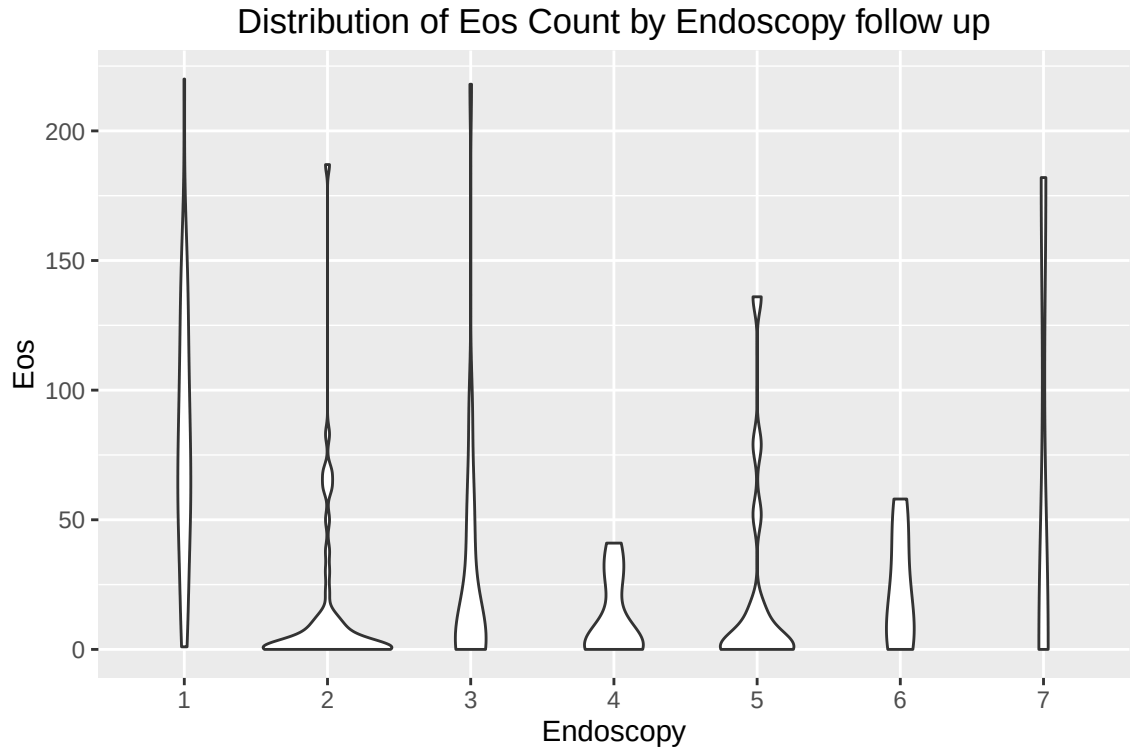
Box and violin plots

Boxplot



This boxplot shows the distribution of Eos peak counts by endoscopy.

Violin plot



We look at the distribution of peak eosinophil counts by endoscopy appointment, (limiting ourselves to the first 7 follow up endoscopies for better visualization).

Figure 1 for manuscript

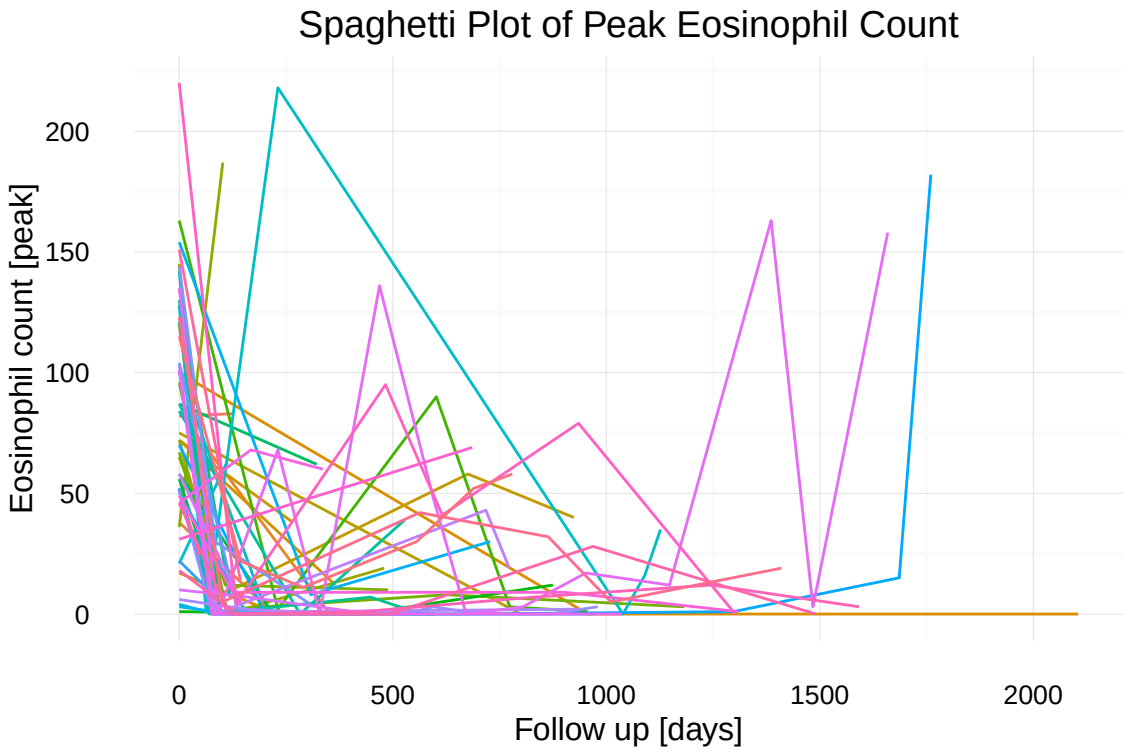


Figure 1: Above, the peak eosiophil count is plotted against patient follow up [in days] with individual patient trajectories indicated by colored lines. After treatment initiation with swallowed fluticasone, the peak eosiophil count drops and remains low, possibly with a slight increase into the second year (possibly due to either decreasing compliance or treatment efficacy) and increasing uncertainty as fewer patients were followed beyond two years.

Suplemental Table: Regression coefficients final model B

Coefficient	Rhat	lower 95%	posterior median	upper 95%
(Intercept)	1	4.02	4.45	4.92
treat	1	-2.89	-2.12	-1.39
furtherfollowup	1	-0.61	0.26	1.10
b[(Intercept) Patient:1]	1	-1.36	-0.14	1.02
b[furtherfollowup Patient:1]	1	-2.23	-0.11	2.06

Suplemental Table: Posterior median estimates (50%) for fixed effects and examplary random effects regression coeffients are shown above with their corresponding 95% upper and lower bounds, followed by an exemplary random effect for patient 1. For our mixed effects negative binomial regression model, the difference in the logs of expected peak eosinophilic counts is expected to change by the respective regression coefficient, given the other predictor variables in the model are held constant. **In a representative patient with a starting peak eosinophilic count of 80, this might mean a reduction to a peak eosinophilic count of 10, which is mostly sustained in further follow ups.** Numerically, -2.1, the reduction in log Eos count estimated by the regression coefficient for treatment initiation, equals the difference between log(80) and log(10). The consistently low Rhat values below 1.1 suggest adequate model convergence.

Model

Our mixed model was formulated as

Formula

$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 * treat_{ij} + \beta_2 * furtherfollowup_{ij} + \beta_3 * Diet_i + \beta_4 * PPI_i +$

$b_{0ij} + b_{1ij} * furtherfollowup_{ij}$

With μ_{ij} the eosinophil (peak count) for i-th patient at the j-th appointment and $treat_{ij}$ a patient specific indicator for the second follow up visit, and $furtherfollowup_{ij}$ an indicator for visits beyond the second, respectively, (both included as a fixed effect and as a random effect (grouped by $patient_i$) and accounting for the correlation among repeated observations in the same subject), $Diet_i$ is an indicator of pre-treatment with diet for the ith patient prior to initiation of oral flovent therapy and PPI_i indicating concurrent therapy with proton pump inhibitor.

Fitting advanced hierarchical models with classical software packages can be challenging, so we used a Bayesian hierarchical modelling approach with the default priors of the software package rstanarm (Gabry, J. & Goodrich, B. (2016), 'rstanarm: Bayesian Applied Regression Modeling via Stan', R package version 2.9.0-1.). The models were run with 2000 iterations and six chains in parallel and converged quickly as evidenced by the Rhat conversion diagnostic being smaller than 1.05 for all parameters, evidence for good mixing in graphical plots of the MCMC chains. Our models were robust to changes in our prior specification.

We build our model sequentially, testing the inclusion of additional fixed effects (concurrent proton pump inhibitor treatment and concurrent diet modification) by comparing the expected log predicted density for a new dataset and found that including these predictors improved the model fit. We found that the inclusion of random effects as described above improved the model fit, indicating that the aditional variabilty may be best explained by between patient variability.

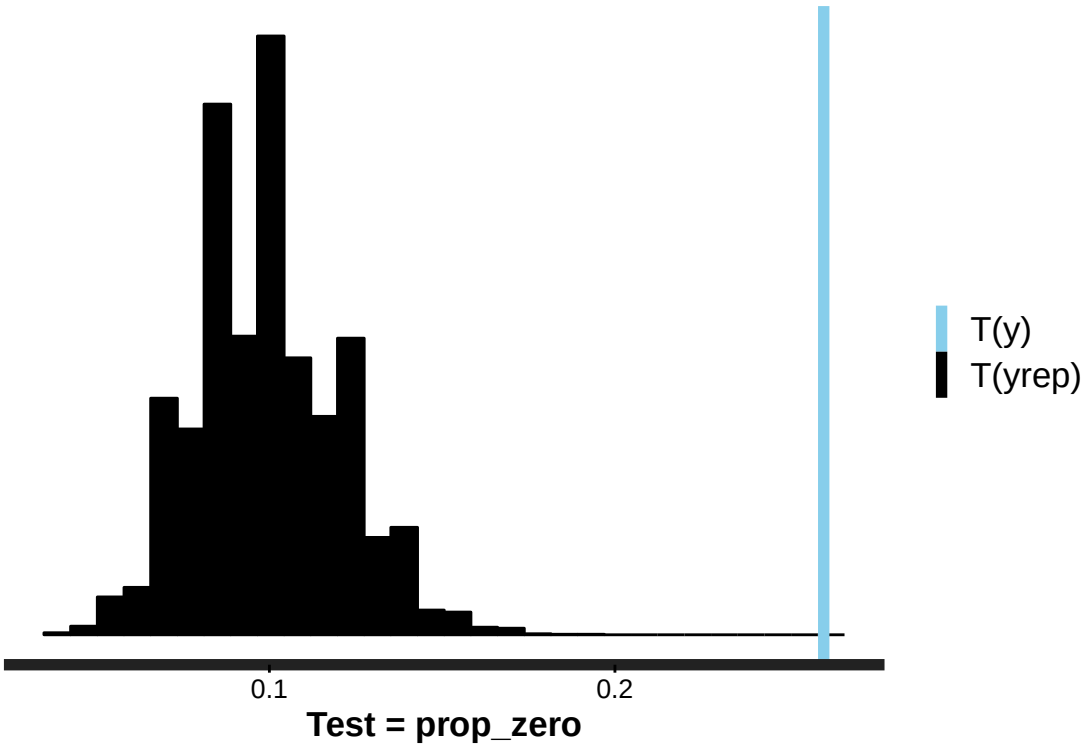
We explored our model assumptions, recognizing that our inferences of the mean effect of treatment will likely be robust to minor violations of normality and homoscedascity in our mixed effects model; Comparing the fitted and predicted values to the observed values, we observed shrinkage (a pulling of the fitted and predicted values towards the mean values of the parameter), as is typical in hierarchical models, leading the improved convergence and model stability by regularization (Gelman, Andrew, et al. Bayesian data analysis. Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC, 2014.).

Poisson model

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + Diet + PPI +
##     (1 + furtherfollowup | Patient), data = EoEdata9Feb2016,
##     family = "poisson", iter = myiter, chains = mycores - 2,
##     cores = mycores - 2)
##
## Estimates:
##                 Median MAD_SD
## (Intercept)      4.3    0.3
## treat           -1.7    0.0
## furtherfollowup -0.6    0.4
## Diet            -0.3    0.3
## PPI              0.1    0.6
##
## Error terms:
##  Groups  Name            Std.Dev. Corr
```

```
##  Patient (Intercept)     0.97
##          furtherfollowup 1.87     -0.01
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 36.9    0.6
```

Posterior predictive checks



The proportion of zeros in the peak Eos counts in the observed data is around 30% (vertical blue line). In the replicated datasets for the poisson model, zeros occur much less frequently. This may point to overdispersion and we explored alternative model (negative binomial) to account better for the large number of zeros in the observed data. The negative binomial distribution allows the mean and variance to differ, (unlike in the poisson model where they are constraint to be the same).

Negative binomial models

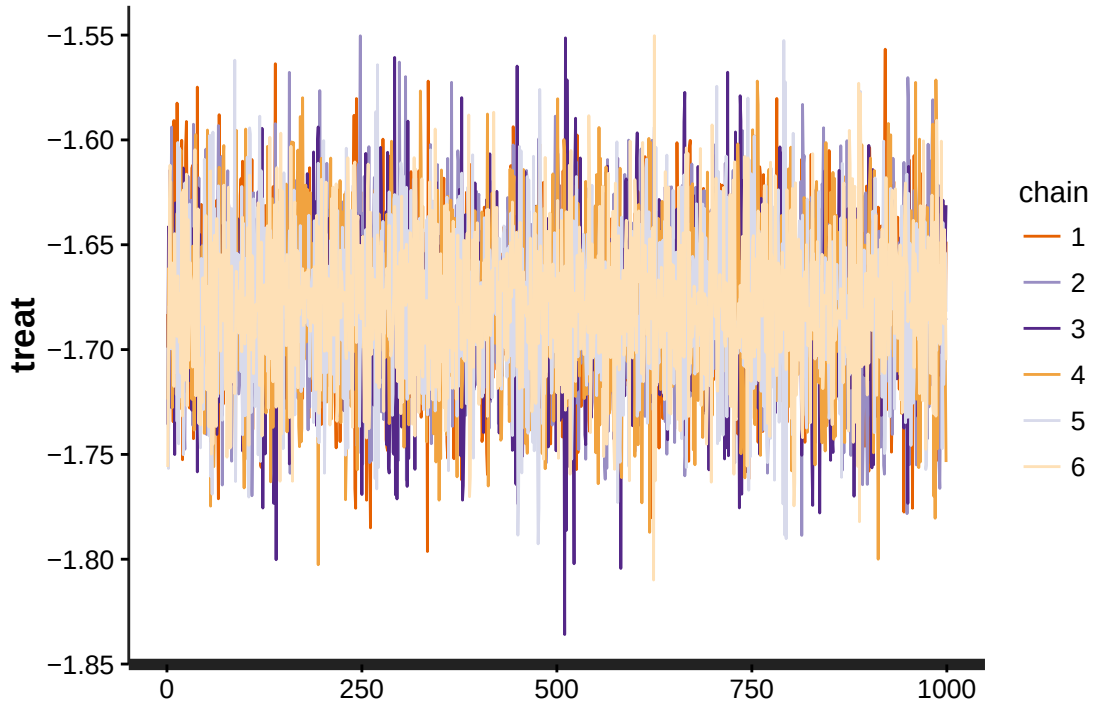
Model A

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + Diet + PPI +
##     (1 + furtherfollowup | Patient), data = EoEdata9Feb2016,
##     family = neg_binomial_2, iter = 2000, chains = 4, cores = 4)
##
## Estimates:
##                 Median MAD_SD
```

```
## (Intercept)      4.7    0.4
## treat           -2.1    0.4
## furtherfollowup  0.3    0.4
## Diet            -0.3    0.4
## PPI              0.3    0.8
## overdispersion   0.5    0.1
##
## Error terms:
##  Groups  Name            Std.Dev. Corr
##  Patient (Intercept)     0.63
##          furtherfollowup 0.83     0.61
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 46.7   11.6
```

Model B

Model convergence



The traceplot indicates good mixing of the chains

Explore in shinystan

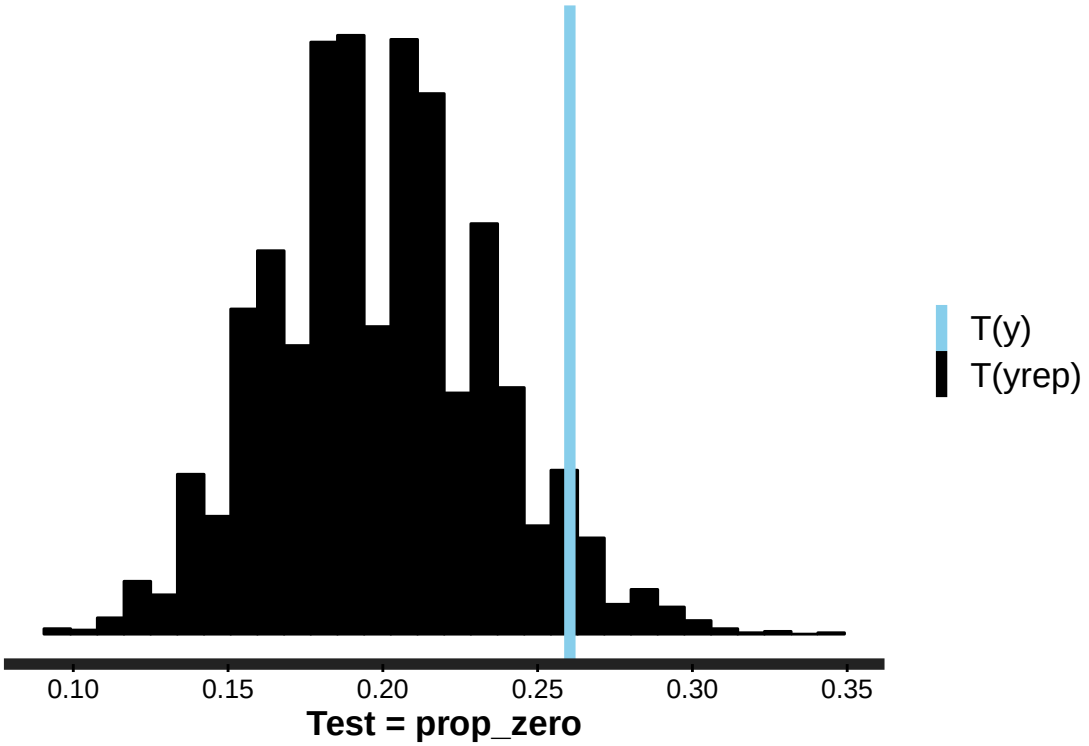
```
require(shinystan)
fit <- fit3
launch_shinystan(fit)
```

We explored the model fits in the interactive R package **shinystan** and found no evidence to suggest non-convergence, with consistently low Rhat values below 1.1.

Model B results

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + (1 + furtherfollowup |
##     Patient), data = EoEdata9Feb2016, family = neg_binomial_2,
##     iter = 2000, chains = 4, cores = 4)
##
## Estimates:
##                 Median MAD_SD
## (Intercept)      4.4    0.2
## treat           -2.1    0.4
## furtherfollowup  0.3    0.4
## overdispersion   0.5    0.1
##
## Error terms:
##  Groups  Name            Std.Dev. Corr
##  Patient (Intercept)     0.63
##          furtherfollowup 0.89     0.60
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 45.2   11.0
```

Posterior predicitve check for negative binomial model B

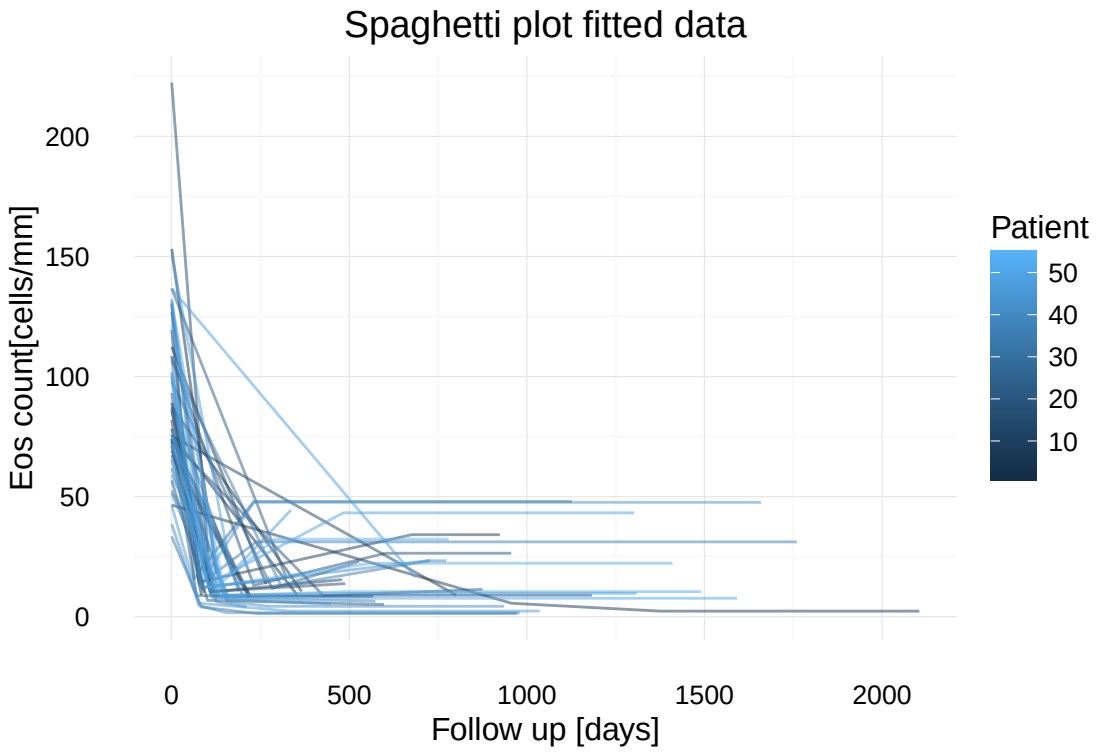


We again compare the propostion of zero counts occuring in the observed data (blue vertical line) and in the replicated datasets for the new negative binomial model. Ideally the blue line falls in the center of the black histogram, inidcating that the propostion of zero counts in the replicated sets falls around the observed value.

The proportion of zeros in the observed data falls obviously not in the center of the histogram, but it is closer, suggestion a better fit. Some difference is expected in this type of model.

Plot of fitted values for negative bionomial model B

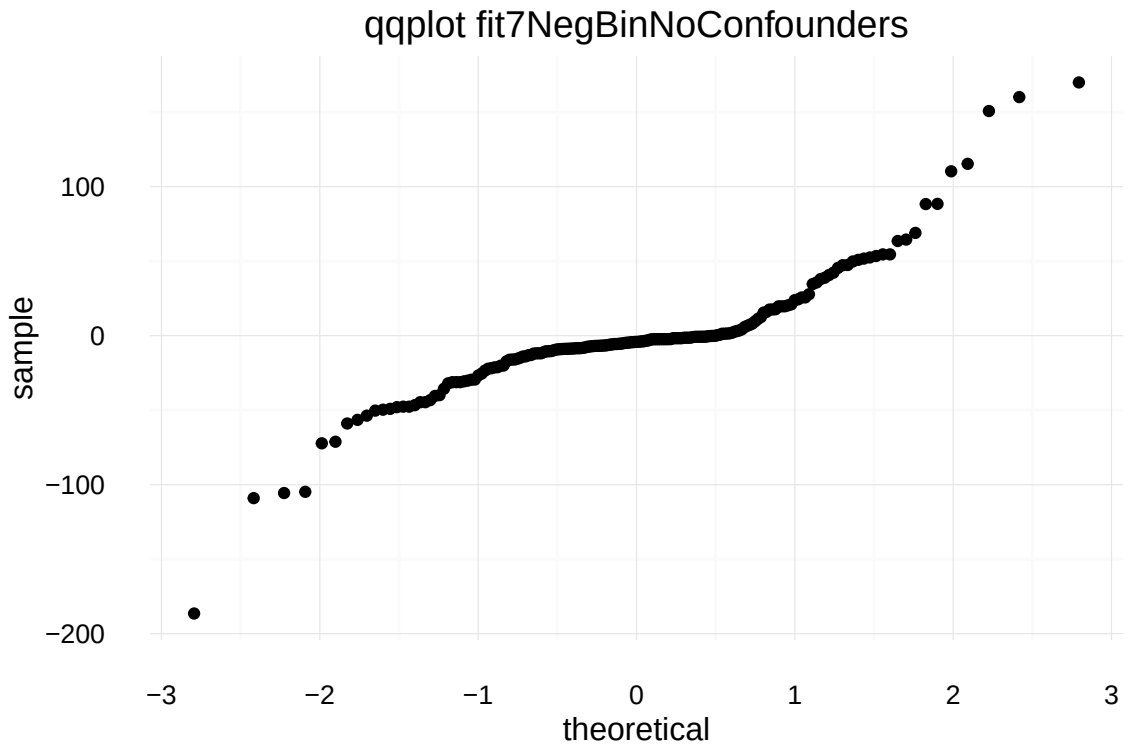
```
## [1] "SpaghettiFittedPlot.md"
```



This plot shows a resonalbe fit of the model by comparing the observed with the fitted Eos count.

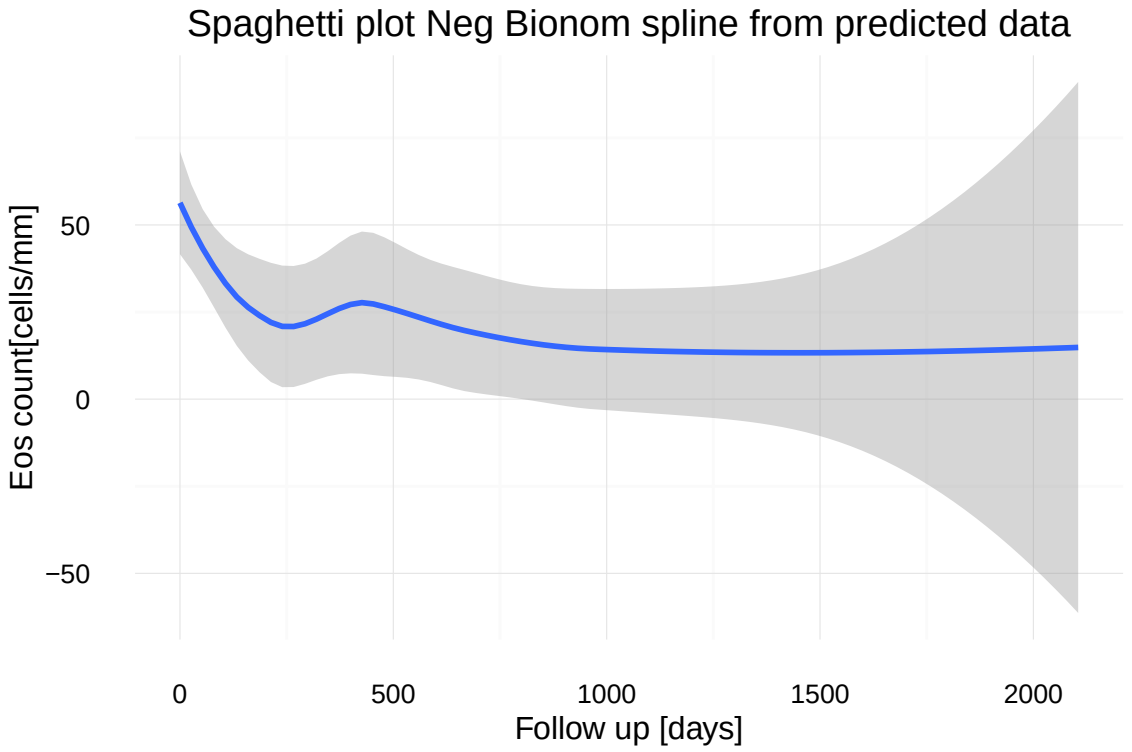
Normal distribution of residuals

```
## [1] "myqqplot.md"
```



This plot indicates violations of some assumptions (e.g. normality), but these are expected and their absence would also not be reassuring.

Spline Model B



Model C

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + Diet + (1 +
##     furtherfollowup | Patient), data = EoEdata9Feb2016, family = neg_binomial_2,
##     iter = 2000, chains = 4, cores = 4)
##
## Estimates:
##                  Median MAD_SD
## (Intercept)       4.7    0.4
## treat            -2.1    0.4
## furtherfollowup   0.3    0.4
## Diet             -0.3    0.4
## overdispersion    0.5    0.1
##
## Error terms:
##  Groups  Name            Std.Dev. Corr
##  Patient (Intercept)     0.63
##          furtherfollowup 0.87     0.62
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 45.7   11.2
```

Model D

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + PPI + (1 +
##     furtherfollowup | Patient), data = EoEdata9Feb2016, family = neg_binomial_2,
##     iter = 2000, chains = 4, cores = 4)
##
## Estimates:
##                 Median MAD_SD
## (Intercept)      4.4    0.2
## treat           -2.1    0.4
## furtherfollowup  0.3    0.4
## PPI              0.3    0.7
## overdispersion   0.5    0.1
##
## Error terms:
##  Groups  Name            Std.Dev. Corr
##  Patient (Intercept)     0.63
##          furtherfollowup 0.83     0.59
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 46.1   11.6
```

Model E

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + Diet + PPI +
##     (1 | Patient), data = EoEdata9Feb2016, family = neg_binomial_2,
##     iter = 2000, chains = 4, cores = 4)
##
## Estimates:
##                 Median MAD_SD
## (Intercept)      4.7    0.4
## treat           -2.2    0.4
## furtherfollowup  0.5    0.4
## Diet            -0.3    0.4
## PPI              0.7    0.6
## overdispersion   0.4    0.1
##
## Error terms:
##  Groups  Name        Std.Dev.
##  Patient (Intercept) 0.72
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 48.3   13.7
```

Model F

```
## stan_glmer(formula = Eos ~ treat + furtherfollowup + (1 | Patient),
```

```
##     data = EoEdata9Feb2016, family = neg_binomial_2, iter = 2000,
##     chains = 4, cores = 4)
##
## Estimates:
##                 Median MAD_SD
## (Intercept)      4.5    0.3
## treat           -2.2    0.4
## furtherfollowup  0.6    0.4
## overdispersion   0.4    0.1
##
## Error terms:
##  Groups  Name        Std.Dev.
##  Patient (Intercept) 0.8
## Num. levels: Patient 54
##
## Sample avg. posterior predictive
## distribution of y (X = xbar):
##          Median MAD_SD
## mean_PPD 47.8   13.7
```

Compare Models

The LOO Information Criterion (LOOIC) has the same purpose as the Aikaike Information Criterion (AIC) that is used by frequentists. Both are intended to estimate the expected log predicted density (ELPD) for a new dataset.

```
##        looic  se_looic elpd_loo se_elpd_loo p_loo  se_p_loo
## modelB 1550.9   58.7   -775.4     29.4        27.3    3.9
## modelC 1552.5   58.7   -776.2     29.4        27.3    3.7
## modelA 1556.0   59.0   -778.0     29.5        27.7    4.0
## modelF 1564.1   59.1   -782.0     29.5        23.3    3.5
## modelE 1566.8   59.0   -783.4     29.5        22.8    3.4
```

The model fit is very similar and the inferences are the same.