

Unicare zero inflated admissions counts models summary

Summary remarks

- We compare inpatient admissions for Unicare - Study vs all the other members
- 270 window for inpatient admissions.
- The zero inflated models appear to be a good approximation of the data generating process.
- The various models are in agreement as to the effect of `ibis`.
- Statistical significance for coefficient for cohort-`ibis` vs control - is achieved in some cases.
- The difference in mean admissions is more that 50% reduced for the `ibis` vs control. This is borne out by the models as well.
- On the other hand the same coefficient value gives statistical signifiacne for less than 10% percent increase in the probability of zero admissions for `ibis` cohort., depending on values of covariates.
- Additional covariates were selected based on correlations and other considerations. The inclusion of these can reduce the variance of estimates, even for a randomized control study where there is complete balance across cohorts.

Models

Based on earlier model results we consider Bayes and frequentist zero inflated Poisson and negative binomials models.

For the zero inflated Poisson model we use the following model specification:

$$\begin{aligned}\text{counts}_i &\sim ZIP(\pi_i, \mu_i) \\ \log \mu_i &= \beta_0 + \beta_1 \text{cohort}_i + \beta_2 \text{chf}_i + \beta_3 \text{age}_i + \beta_4 \text{afib}_i \\ \log \frac{\pi_i}{1 - \pi_i} &= \gamma_0 + \gamma_1 \text{age}_i\end{aligned}$$

Note that the zero inflated negative binomial models model the mean the same way. The expected value for these models, which is the average inpatient count, is

$$(1 - \pi_i)\mu_i$$

We scale the age variable as

$$age \rightarrow \frac{age - 60}{10}$$

for interpretability and for numerical stability.

Effects sizes.

Presently we are not modeling the π using cohort. So we can compare cohorts with the same values of other predictors-whatever they are- simply using the ratios of the means μ_i .

We cannot do across the board comparison of the effect on the probability of zero admissions as it varies depending on the values of the other covariates as well as the zero inflation probability π . But we can compare for given values of the covariates.

We do these comparisons below.

Summary statistics

We consider admissions occurring within 270 day window of observation.

```
# A tibble: 2 x 2
  cohort count
  <chr>   <int>
1 control   881
2 ibis      82
```

Mean inpatient counts

Overall

```
# A tibble: 1 x 1
  mean_count
    <dbl>
1      0.372
```

By cohort

```
# A tibble: 2 x 2
  cohort mean_count
  <chr>    <dbl>
1 control    0.390
2 ibis       0.171
```

So this is more than 50% reduction.

Proportion if zero admissions

Overall

```
# A tibble: 1 x 1
  prop_zero_admit
    <dbl>
1      0.198
```

By cohort

```
# A tibble: 2 x 2
  cohort prop_zero_admit
  <chr>    <dbl>
1 control    0.796
2 ibis       0.866
```

Mean for patients with one or more admissions

Overall

```
# A tibble: 1 x 1
  mean_at_least_one_admit
      <dbl>
1         0.198
```

By cohort

```
# A tibble: 2 x 2
  cohort mean_at_least_one_admit
  <chr>      <dbl>
1 control    0.204
2 ibis       0.134
```

Model results

Frequentist zero inflated Poisson model

Call:

```
pscl::zeroinfl(formula = inpatient_count ~ cohort + age + chf + atrial_fibrillation |
  age, data = patients_events, dist = "poisson")
```

Pearson residuals:

	Min	1Q	Median	3Q	Max
	-0.681	-0.455	-0.402	-0.312	16.120

Count model coefficients (poisson with log link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.2698	0.1113	2.42	0.0153 *
cohortibis	-0.7187	0.3239	-2.22	0.0265 *
age	-0.2456	0.0608	-4.04	5.4e-05 ***
chf	0.4644	0.1443	3.22	0.0013 **
atrial_fibrillation	0.4300	0.1447	2.97	0.0030 **

Zero-inflation model coefficients (binomial with logit link):

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.3549	0.1482	9.14	< 2e-16 ***

```
age          -0.4835      0.0975   -4.96  7.1e-07 ***
```

```
---
```

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

```
Number of iterations in BFGS optimization: 16
```

```
Log-likelihood: -726 on 7 Df
```

Frequentist zero inflated negative binomial model

```
Call:
```

```
pscl::zeroinfl(formula = inpatient_count ~ cohort + age + chf + atrial_fibrillation |  
  age, data = patients_events, dist = "negbin")
```

```
Pearson residuals:
```

```
      Min      1Q Median      3Q      Max  
-0.579 -0.398 -0.372 -0.291 15.937
```

```
Count model coefficients (negbin with log link):
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.4663	0.2364	-1.97	0.04856 *
cohortibis	-0.6441	0.3477	-1.85	0.06398 .
age	-0.2874	0.0913	-3.15	0.00165 **
chf	0.7563	0.2015	3.75	0.00017 ***
atrial_fibrillation	0.7098	0.1990	3.57	0.00036 ***
Log(theta)	-0.3909	0.3327	-1.17	0.24010

```
Zero-inflation model coefficients (binomial with logit link):
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.308	0.397	0.77	0.44
age	-0.656	0.154	-4.25	2.1e-05 ***

```
---
```

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

```
Theta = 0.676
```

```
Number of iterations in BFGS optimization: 16
```

```
Log-likelihood: -698 on 8 Df
```

Bayesian zero inflated Poisson model

Estimate	Est.Error	Q2.5	Q97.5
----------	-----------	------	-------

Intercept	0.285	0.1094	0.0685	0.49130
zi_Intercept	1.336	0.1382	1.0700	1.61402
cohortibis	-0.511	0.2656	-1.0393	0.00693
age	-0.209	0.0601	-0.3253	-0.09031
chf	0.413	0.1372	0.1452	0.68345
atrial_fibrillation	0.376	0.1397	0.1004	0.64854
zi_age	-0.387	0.0845	-0.5544	-0.22165

Model fit and predictions

We can see that these models are in agreement, and the results do fit the data well.

Expected number of inpatient admissions; probabilities of zero admissions

If we take the mean of the `cohortibis` count coefficients for the two models as an estimate, the models suggest a mean admission count for ibis cohort is 0.54 times that of the control, assuming the same values for other predictors, which amounts to a 46 percent reduction.

zero admissions probabilities

We can compare prevalences for outcomes in the data vs those predicted by the model. This was done in the `zer0_infl_admissions_models_2025-05_short.html`. We give a graphical illustration below with the Bayes model. Presently, we will compare what the models give as probabilities of admissions count outcomes for a single patient with the same age and conditions covariates, with one being `ibis` cohort and the other `control`.

For 60 year old patient, without no `chf` or `afib`, for `ibis` vs `control`, On the other hand, for the probability of zero admissions, the zero inflated Poisson model gives,

	cohort	age	chf	atrial_fibrillation	0	1	2	3	4
1	control	0	0		0 0.850	0.0725	0.0475	0.02072	0.006785
2	ibis	0	0		0 0.903	0.0691	0.0221	0.00469	0.000749
		5							
1					1.78e-03				
2					9.56e-05				

while for patients with both chronic conditions we have

	cohort	age	chf	atrial_fibrillation	0	1	2	3	4	
1	control	1	1		0	0.763	0.0942	0.0768	0.0417	0.01700
2	ibis	1	1		0	0.838	0.1059	0.0421	0.0111	0.00221

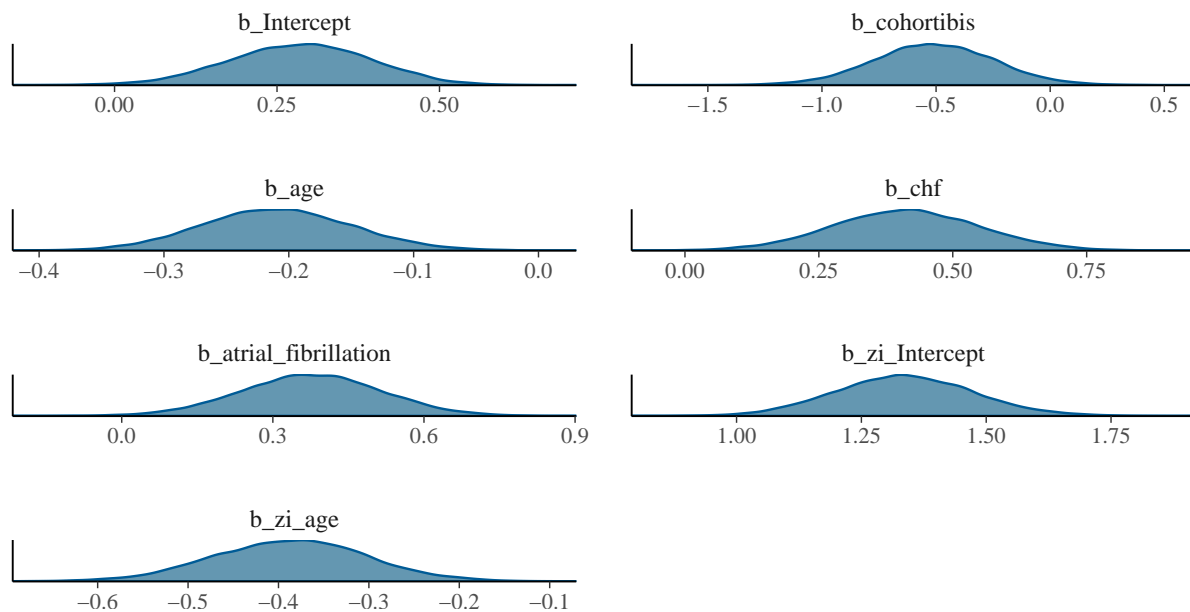
	5
1	0.005541
2	0.000351

In either case we get a little under 6-10% increase. This agrees with the difference in proportion of zero admits for ibis vs control above.

Post predictive checks

We use Bayes model here. The coefficients are similar, but we can get a more complete picture of both the distributions of the coefficients, and the how well the model reproduces the observations, as well as model variability.

Posterior distributions for zero-inflated model parameters; scaled age

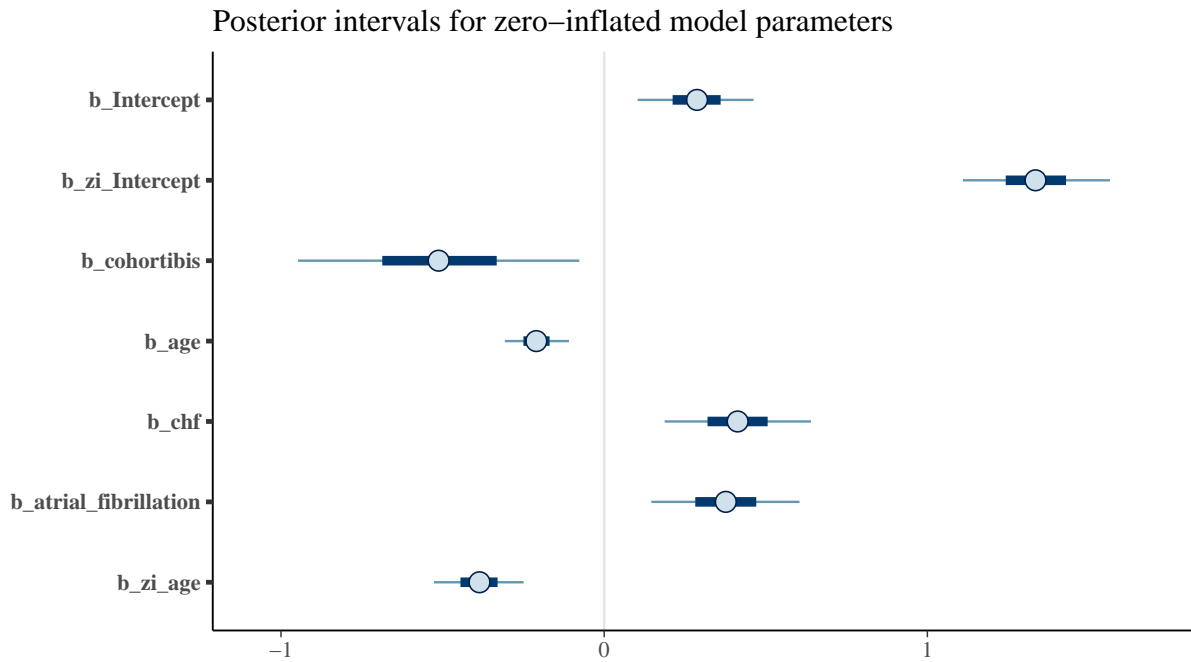


Credible intervals

These are 50% and 90% “credible intervals”.

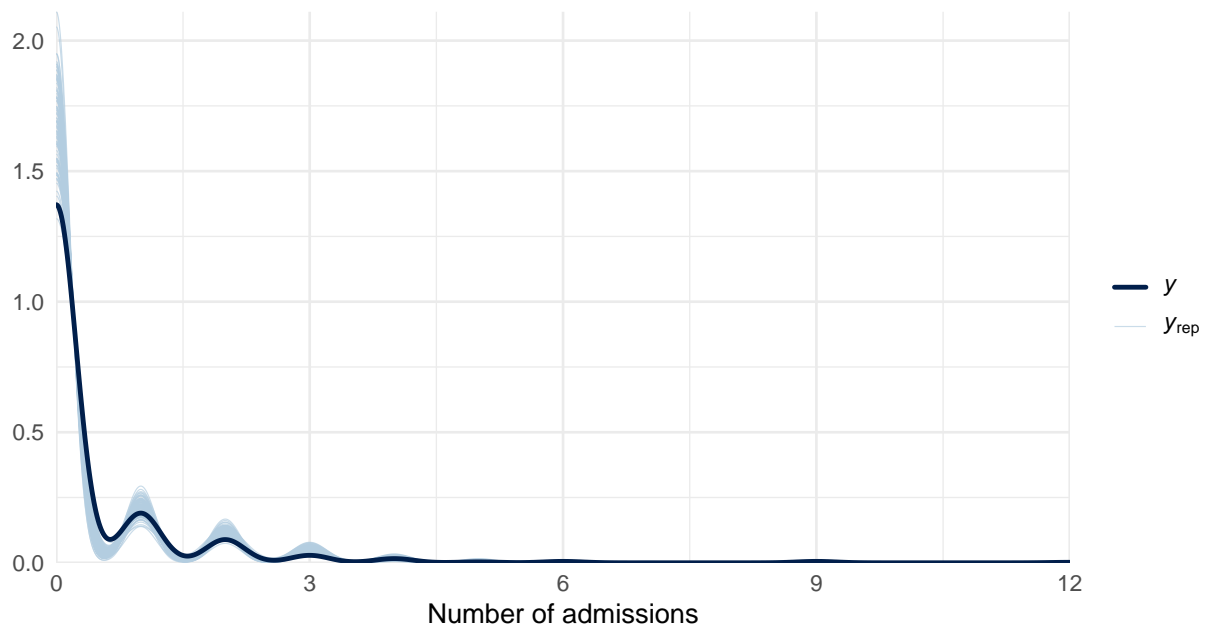
Remark These are the middle quantiles in the distribution for the various coefficients. We can interpret this as a probability the coefficient lies within the interval, according to the model. By contrast, a frequentist confidence interval does not admit a probabilistic interpretation this

way. In particular, we can say that the probability the true parameter is less than the upper bound of a 90% credible interval is 95%, because of the area in the left tail. In that sense, a 90% credible interval is comparable to the upper limit of a 95% frequentist confidence interval, if, as in this case, we are concerned with that upper limit being greater than zero or not.



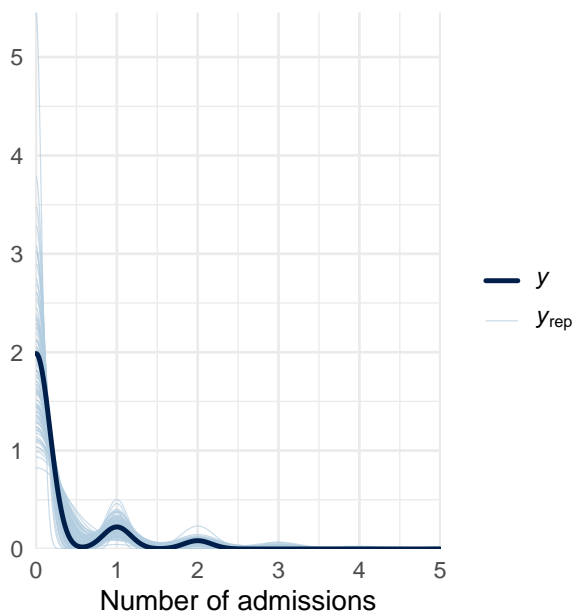
If anything, the model may inflate the zeros too much.

100 draws from the posterior

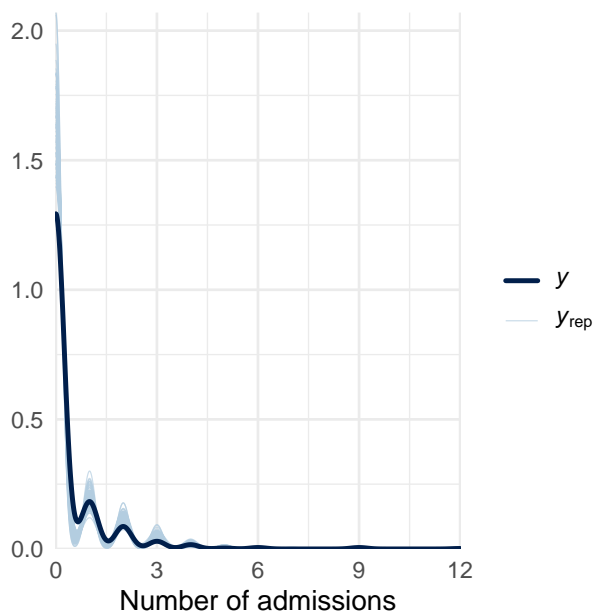


We plot posterior draws by cohort. Not surprisingly there is more uncertainty with the Ibis cohort, which has much smaller sample size.

Density overlay for Ibis cohort

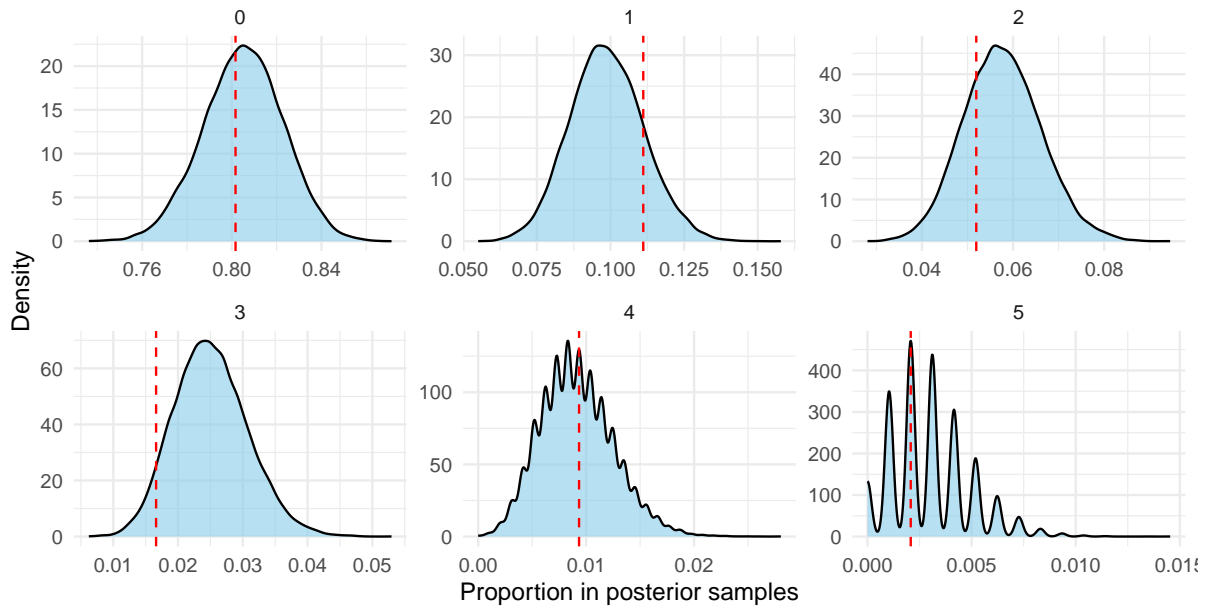


Density overlay for control cohort



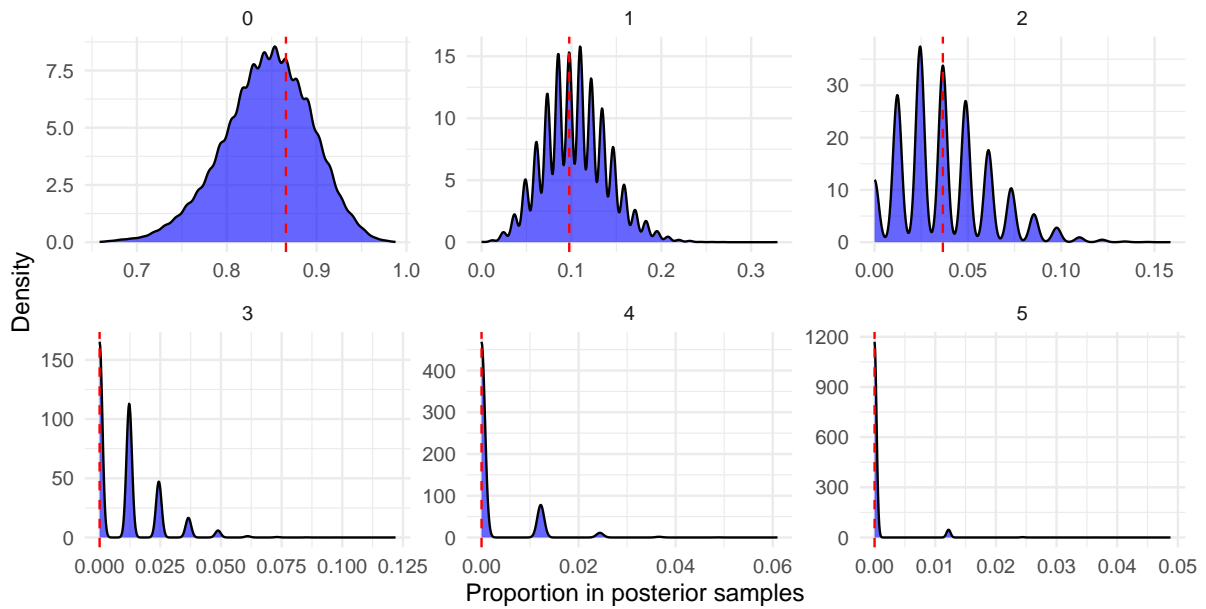
We can plot proportions of various counts in the posterior and compare with actual proportions in the data.

Posterior predictive distribution of proportion for each count (0–10)



And also do this by cohort. There is not enough data

Posterior predictive distribution by count (IBIS cohort)



Posterior predictive distribution by count (Control cohort)

