The School of Mathematics

THE UNIVERSITY
of EDINBURGH

Understanding the Relationship between Hospital Prescription Rates & Antimicrobial Resistance

by

Ravi Patel, S1889112

July 2019

Supervised by
Dr. Gail Robertson, Dr. Meghan Perry, Dr. Serveh Sharifi Far & Dr. Vanda Inácio de Carvalho

Own Work Declaration

All work contained within is my own, except where otherwise referenced.

Contents

1 Introduction **3**
1.1 Background & Motivation . . . 3
1.2 Objective . . . 3
1.3 Data . . . 3
1.4 Software and Algorithms . . . 3
1.5 Terminology . . . 3

2 EDA **4**
2.1 Single Variable . . . 4
2.2 Multiple Variables . . . 5
2.3 EDA Summary . . . 8

3 Model **9**
3.1 Ordered Logit . . . 9
3.2 Resistance Score as the Latent Variable . . . 10
3.3 Bayesian Specification . . . 10

4 Bayesian Workflow **12**
4.1 Prior Predictive . . . 12
4.2 Model Fit and Algorithm Diagnostics . . . 13
4.3 Posterior Predictive (Hypothesis Model) . . . 14
4.4 Model Comparison . . . 17
4.5 Posterior Predictive (`mc41`) . . . 19

5 Results **23**
5.1 DDD and AMR . . . 23
5.1.1 Route . . . 24
5.2 Drug-Specific Effects . . . 25
5.3 Class Group Heterogeneity . . . 26
5.4 Comments on Other Fitted Models . . . 26

6 Model Discussion **27**
6.1 Concerns . . . 27
6.2 Alternative Model: Zero-one inflated Beta (ZOIB) . . . 27

7 Conclusion **28**

Appendices **32**

A Data Preparation **32**

B Models Compared **33**

C R Session Information **34**

D Code & Data Access **35**
D.1 Stan Code Example . . . 35

Executive summary

Antimicrobial resistance (AMR) is a big threat, with a low estimate placing the potential death toll due to AMR in 2050 at 10,000,000 (O'Neill, 2016). Studies at the individual (Costelloe et al., 2010) and population (Goossens et al., 2005) level show a positive association between drug use and AMR – our goal is to investigate this association in a hospital setting. Using a Bayesian ordered logit model with drug-varying intercepts, we conclude there is a 95% probability of a positive association between drug use and resistance – specifically for invasively administered drugs. Additionally, there is a 97% probability that this association is smaller when considering non-invasive administration. Both these effects are very small in magnitude, and poor predictive performance means these results should be looked upon sceptically. We also propose a potential idea to expand the zero-one-inflated beta model that could account for ordering, and may be a good model for similar such studies in future.

1 Introduction

1.1 Background & Motivation

Antimicrobial resistance (AMR) is similar to antibiotic resistance, also including non-bacterial microbes such as viruses, parasite, and fungi (World Health Organisation (WHO), 2018). Their use is widespread in common medical procedures such as C-sections, joint replacements, chemotherapy, and organ transplants (Roope et al. (2019), WHO (2018)), making the rise of resistance a serious threat.
Cassini et al. (2019) claim that in 2015, resistant infections accounted for between 28,480 and 38,430 deaths, and between 768,837 and 989,068 disability-adjusted life years lost (i.e. a year lost in healthy life) at 95% confidence. O'Neill (2014, 2016) estimates there will be 10,000,000 deaths from AMR in 2015, and suggests this may even be an underestimate.
Furthermore, we would expect an economic consequence due to both morbidity and mortality. The World Bank (2017) estimate in a pessimistic (but not unrealistic) scenario, that the economic impact of AMR could be similar to that of the 2008 financial crisis.

1.2 Objective

This paper aims to provide a small contribution to the literature on AMR by examining the effect of prescription rates in a hospital setting. Previous research has focused on individual level resistance, such as Costelloe et al. (2010) who conducted a meta-analysis finding bacterial resistance developed to administered antibiotics in primary care patients. Similarly, at a population level, Goossens et al. (2005) showed an association between high-consuming countries (measured in DDD per 1000 inhabitants per day), and antibiotic resistance.

1.3 Data

Data was collected over 3 months from the Western General Hospital (WGH) in Edinburgh. Each row represents an antibiotic prescription case within a ward, within a week (a ward-week). Variables of interest are the drug name (e.g. amoxicillin), the class the drug falls into (e.g. Broad penetration β-lactams), the route of drug administration, the TRAK code (a ward identifier), the average age of patients in a ward-week (henceforth just called age), the defined daily dose of drugs administered in a ward week per 1000 occupied bed days (DDD/1000 OBD), and the proportion of clinical isolates tested in a ward week that showed resistance to the prescribed antibiotic (resistance proportion). DDD/1000 OBD can be viewed as a measure of the drug strength in a ward-week for an antibiotic that accounts for the number of patients in the ward. Our interest is the association between DDD/1000 OBD and the resistance proportion. After data preparation (see appendix), we were left with 2548 rows.

1.4 Software and Algorithms

Models were fit with `brms` (Bürkner, 2018), an `R` (R Core Team, 2019)) interface to the Stan language (Stan Development Team, 2018) which implements Hamiltonian Monte Carlo (HMC) (Duane et al., 1987; Neal, 2012) and an improvement – the No U-Turn Sampler (NUTS) (Hoffman & Gelman, 2014). Plots were created with `ggplot2` (Wickham, 2016), `cowplot` (Wilke, 2019), and `bayesplot` (Gabry & Mahr, 2019). Data manipulation was performed using `dplyr` (Wickham et al., 2019) and `data.table` (Dowle & Srinivasan, 2019), while tables were created with `xtable` (Dahl et al., 2019). Session information for `R` is in the appendix.

1.5 Terminology

I avoid the terms 'fixed effects' and 'random effects' following the advice of Gelman & Hill (2007). I instead follow `brms` terminology and refer to population-level effects, and group-level effects. These are also sometimes referred to as 'constant' and 'varying' (respectively) for ease of writing, as per Gelman (2005).

2 EDA

2.1 Single Variable

Figure 1 shows the age is approximately normal (though with a second mode in the 80s), while the DDD is right-skewed due to extreme values. Both have large standard deviations, which can cause issues in sampling and complicate prior specification. Consequently, we standardise the age, and take the logarithm of the DDD/1000 OBD.

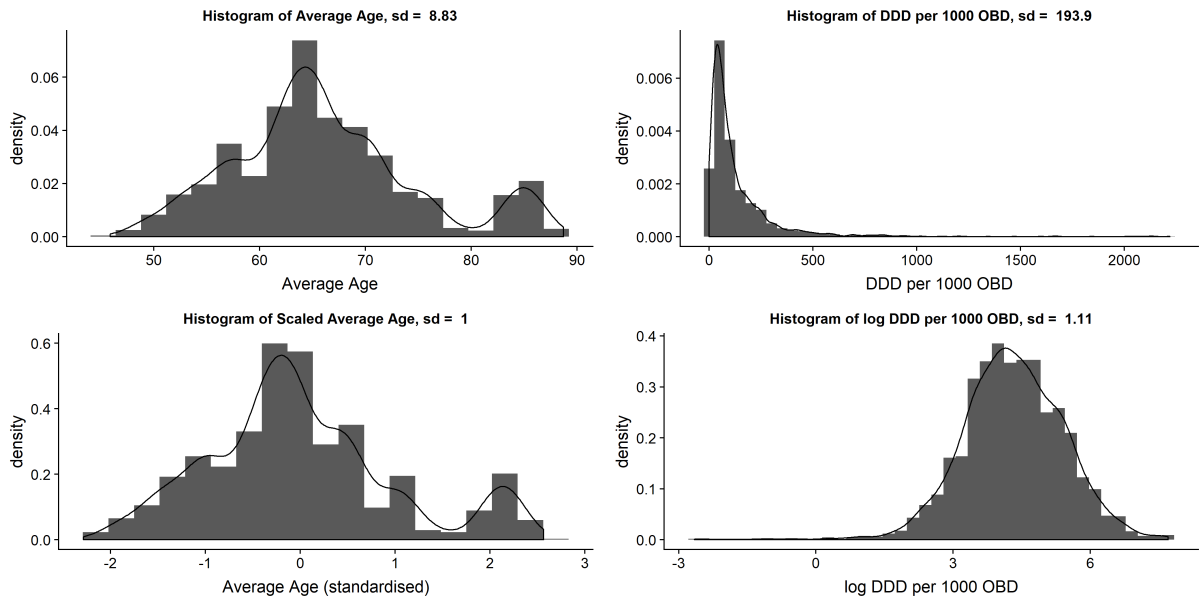

Figure 1: Continuous Covariate Histograms

Figure 2 illustrates the relationship of proportion with factor variables, and the week. Red points signify the standard deviations, and annotated numbers refer to sample sizes. We see that non-invasive administration is associated with higher resistance, but no visible time trend. Some wards appear to have less resistance than others, however there is no natural grouping in the wards as all of the error bars (1 standard error) contain adjacent wards.
Drug classes show separation into different groups based on the proportion. We group based on both observations, and proportion differences. For instance we group the top 2 classes as their means and errorbars overlap. However we do not include "BL - Broad Pen" in the group as it has enough observations to be its own group. The groupings are shown in Table 1. This enables cleaner visualisation, and will be used as a starting point for modelling.
An important feature is the volatility in the data, with standard deviations generally above 30% for subgroups. This makes it difficult to generate good predictions on the data, as there is too much noise relative to the small dataset.

Group	Drug Classes
A	Anti-TB; Chloramphenicol; Nitroimidazole; Oxalidinone
B	β-Lactam - Carbapenem; Polymixin; Glycopeptides/Lipopeptides
C	Other; Aminoglycoside
D	β-Lactams (Extended Pen; Narrow Pen; Cephalosporin); Quinolones
E	Macrolide/Lincosamide
F	β-Lactam - Broad Pen
G	Tetracycline; Trimethoprim and sulphur based drugs

Table 1: Drug Class Groupings

Figure 2: Week and Discrete Covariates vs Proportion

Figure 3 visualises the proportion in 3 ways. The left panel exhibits a large number of 0s, 1s, and possibly 0.5s. This is clearer in the middle panel, which shows a frequency barplot for each *individual* proportion. There are spikes at very specific values, such as 0, 1, 0.5, 0.25 and 0.33. As such, we can capture most of the shape in the proportion by grouping them into relevant intervals. This yields the barplot in the right panel, motivating the use of ordinal regression.

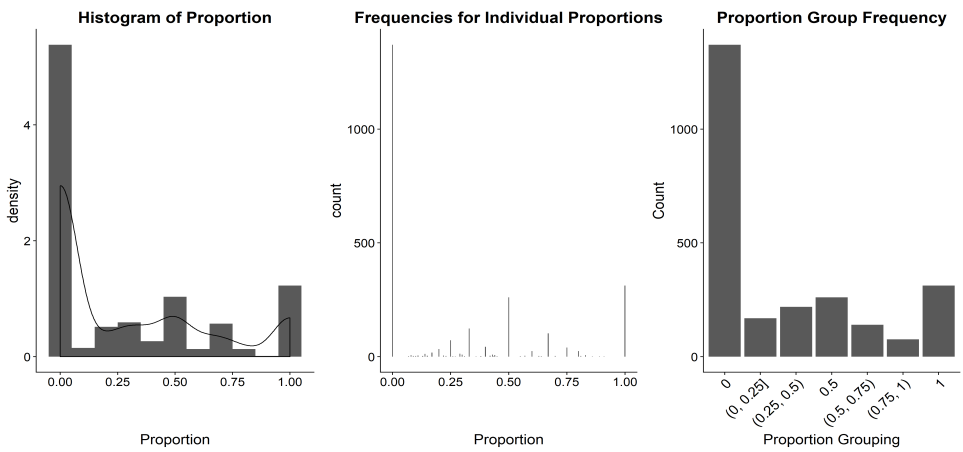

Figure 3: Proportion Distribution

2.2 Multiple Variables

Recall that our interest is the association of DDD with resistance, so we include it in our model. We will also include some sort of drug-specific effect – the grouped class, raw class, or drug name – based on the class-level proportion differences seen earlier.

Figure 4 shows the estimated linear relationship between the proportion and the DDD, segmented by route. Invasively administered drugs seemingly have a lower resistance proportion for low levels of DDD, but increases in DDD have a stronger bearing on resistance than for non-invasively administered drugs. As such, we will include the route in our model, and interact it with age to account for the different slopes.

Figure 5 suggests a negative relationship between age and DDD, which may be expected due

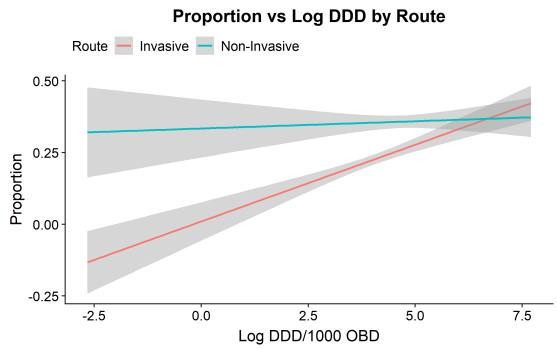

Figure 4: Proportion-DDD Relationship by Route

to the increased sensitivity of the body to certain drugs in elderly patients as suggested in Hughes (2002) and Turnheim (1998). This relationship slightly varies by class, but not enought to motivate varying slopes. Adjusted R^2 values range from 0 to 0.12 for each class, with all coefficients significant at the 5% level except for group E. All slope coefficients were negative, ranging from -0.05 to -2.99.

We include age as a potential confounder, as it may have a direct influence on the proportion due to human-human resistance transmission (Holmes et al., 2016) in the elderly as a result of nursing homes. Age is therefore considered a potential "fork" confounder as discussed in McElreath (2019) and Pearl (2009).

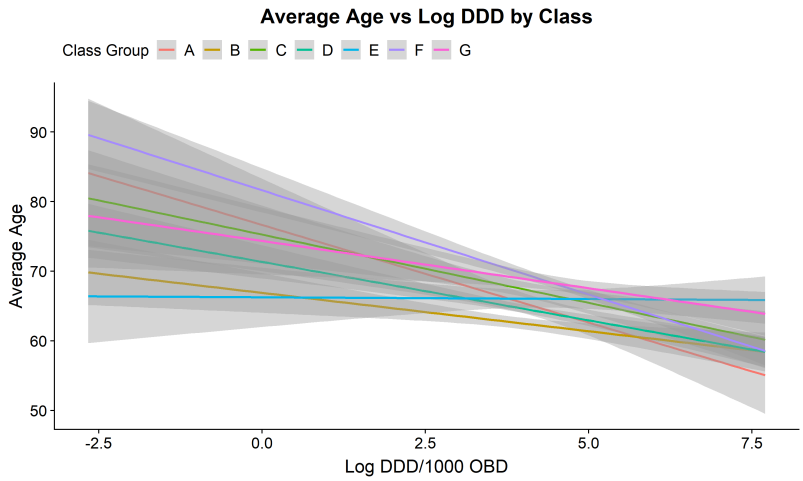

Figure 5: Age-DDD Relationship by Class Grouping

Figure 6 shows a slight positive slope for groups B, C, and D but flat slopes for E, F, and G. However due to the level of variation in the confidence intervals and the small differences, we posit population-level effects, as any group-level effects will be too subtle with our noisy data. Figure 7 suggests that the week has little influence on the proportion. There is a downward slope for group E, however there is a large degree of variation. Therefore we consider week to have no effect on the proportion.

Figure 8 points to a strong relationship between the age and the ward, with the ward seemingly a good predictor of the age due to the low age variation within most wards. A linear regression of age on TRAK yields an adjusted R^2 of 0.9329, suggesting including both TRAK and age would be redundant. We select age for theoretical reasons, and to avoid including too many parameters.

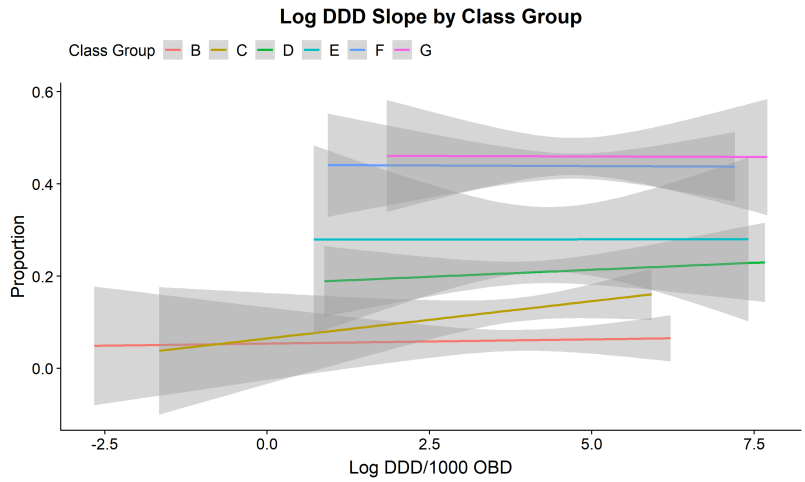

Figure 6: DDD-Proportion Relationship

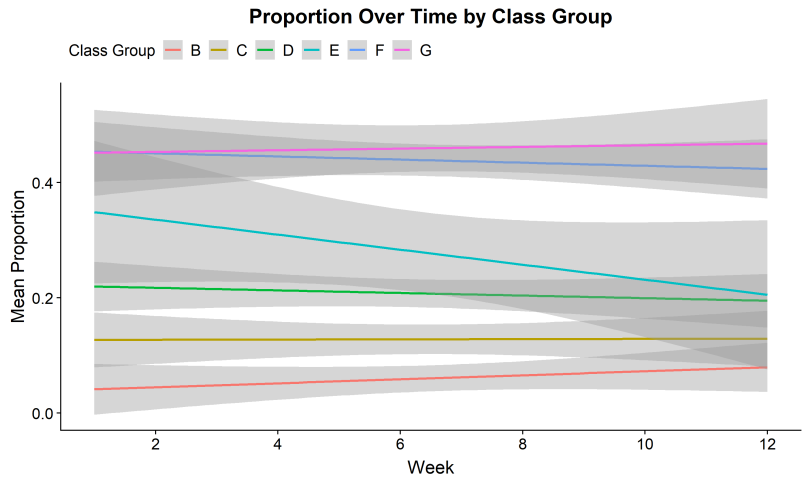

Figure 7: Proportion over time by Class Group

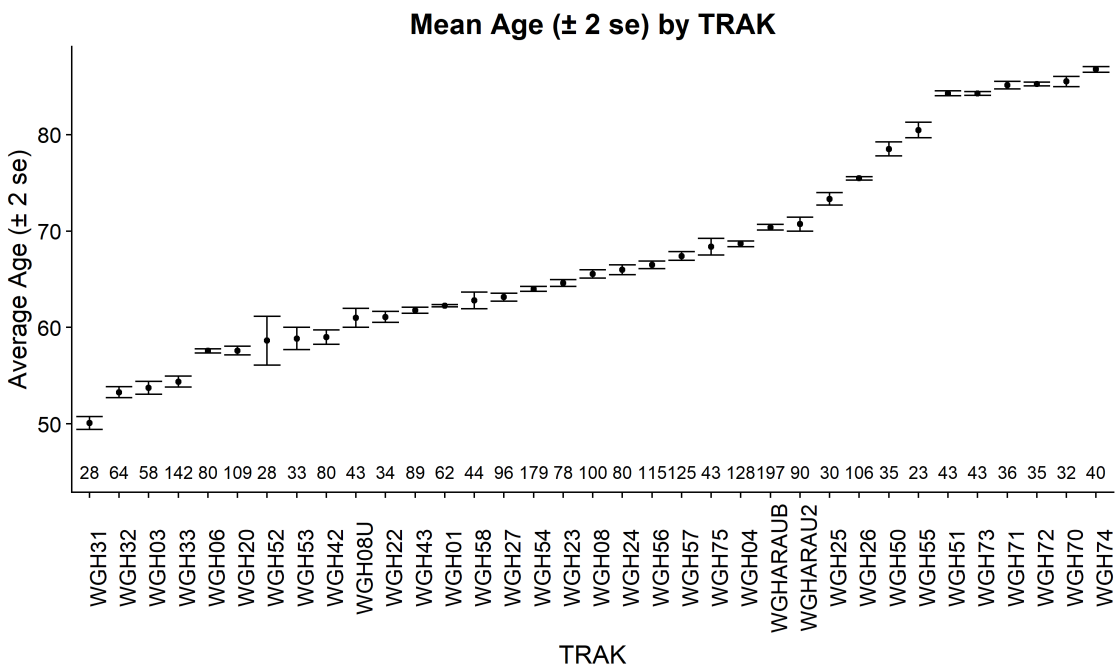

Figure 8: TRAK-Age Relationship

2.3 EDA Summary

We have

1. Transformed the proportion to a set of ordered groups.
2. Grouped drug classes based on their level of resistance for initial modelling.
3. Standardised the average age, and logarithmised the DDD/1000 OBD.
4. Decided to include log DDD as a population level effect, with an interaction on the route.
5. Included age as a population-level effect
6. Not included week due to a lack of any visible relationship in the data.
7. Not included TRAK due to its strong relationship with age.

3 Model

We use a Bayesian ordered logit with varying intercepts. This section provides an overview of ordered logit, discusses the latent variable, and specifies our priors.

3.1 Ordered Logit

This subsection gives an overview of ordered logit. Options for a more comprehensive treatment include Walker & Duncan (1967), McCullagh (1980), and Greene (2012) (ordered probit).

In ordered logit, our response variable is the binned proportion (y). Ordinal regression supposes there is a latent variable ($y^* \in \mathbb{R}$) which leads to our bins by the process below. We have thresholds $c_j, j \in \{0, 1, \ldots K = 7\}$, where $c_0 = -\infty$, and $c_7 = \infty$.

$$y = \begin{cases} 0 & \text{if } y^* \in (-\infty, c_1] \\ (0, 0.25] & \text{if } y^* \in (c_1, c_2] \\ (0.25, 0.5) & \text{if } y^* \in (c_2, c_3] \\ 0.5 & \text{if } y^* \in (c_3, c_4] \\ (0.5, 0.75] & \text{if } y^* \in (c_4, c_5] \\ (0.75, 0.99) & \text{if } y^* \in (c_5, c_6] \\ 1 & \text{if } y^* \in (c_6, \infty) \end{cases}$$

For the cumulative logit, we assume the latent variable follows a logistic distribution with scale 1 (variance $\frac{\pi^2}{3}$) and mean η (the usual linear predictor $\mathbf{x}'\beta$, excluding the intercept). We define the cumulative distribution function (CDF) of the standard logistic distribution as Λ.

$$y^* \sim Logistic(\eta, 1)$$

$$\Lambda(x) = \frac{exp(x)}{1 + exp(x)}$$

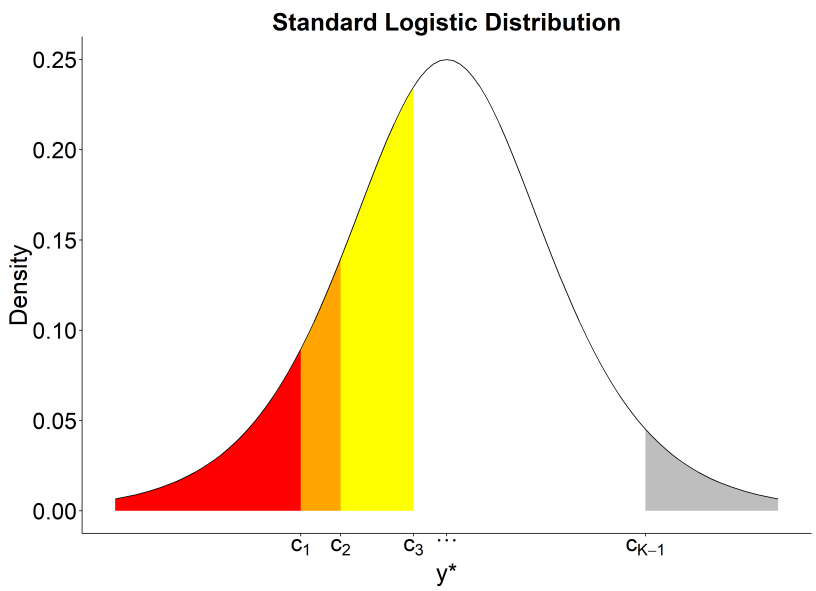

Figure 9: Latent Variable Categorisation

The probability y is in category k given η is $Pr(y = k) = Pr(y^* \in (c_{k-1}, c_k))$ $= \Lambda(c_k - \eta) - \Lambda(c_{k-1} - \eta)$. This concept is illustrated in Figure 9. For example, the yellow region is the probability the latent variable is between c_2 and c_3, i.e. the probability of the observed category being category 3. We therefore define the pdf and likelihood as:

$$f(y) = \prod_{j=1}^{K} [\Lambda(c_j - \eta) - \Lambda(c_{j-1} - \eta)]^{I(y=j)} \tag{3.1}$$

$$L = \prod_{i=1}^{n} \prod_{j=1}^{K} [\Lambda(c_j - \eta_i) - \Lambda(c_{j-1} - \eta_i)]^{I(y_i=j)} \tag{3.2}$$

Where η_i refers to covariates changing by observation.
Define $\gamma_j = Pr(y^* \leq c_j)$. We then have $\gamma_j = Pr(y^* \leq c_j) = \Lambda(c_j - \eta)$. Applying Λ^{-1}:

$$\Lambda^{-1}(\gamma_j) = logit(\gamma_j) = log\left(\frac{\gamma_j}{1-\gamma_j}\right) = c_j - \eta \tag{3.3}$$

We can interpret the coefficients in η in relation to the log odds. Suppose $\beta_1 > 0$ is the coefficient on x_1. A one unit increase in x_1 increases the log-odds of a higher category by β_1. This interpretation is independent of the category, which is where "proportional odds" comes from.

3.2 Resistance Score as the Latent Variable

It may be concerning that we know the bins were created using the proportion, which is non-logistic. The application of ordered logit therefore requires a level of abstraction. We posit an underlying "resistance score" that is impossible to measure, but can be partially captured by the proportion. Our latent variable is mapped to the proportion, which we transform to our bins. We suppose there is a function f which maps from the reals to the closed interval [0,1]. Additionally, there is a function g which maps from the interval (c_1, c_{K-1}) to the open interval (0,1). We define f below, along with a mathematical statement of the function mappings.

$$f : \mathbb{R} \mapsto [0,1]$$
$$g : (c_1, c_{K-1}) \mapsto (0,1)$$

$$f(y^*) = \begin{cases} 0 & \text{if } y^* \leq c_1 \\ g(y^*) & \text{if } y^* \in (c_1, c_{K-1}) \\ 1 & \text{if } y* \geq c_{K-1} \end{cases}$$

$g'(y^*) > 0$, to ensure higher scores lead to higher proportions.
The latent variable is 0 if below c_1, and 1 if above c_{K-1}. Otherwise, it passes through some function g to an intermediate proportion. We divide this intermediate proportion into 5 regions to get our binned observations.
There also exists a discrimination parameter equal to the inverse standard deviation, which allows heterogeneity in the latent variable. It is modelled on the log-scale, since it must be positive. We allow for heterogeneity by each class group, using the "A" group as the reference category. For more information, see Bürkner & Vuorre (2018).

3.3 Bayesian Specification

An alternative parameterisation of the latent variable equation accounting for group-level effects is:

$$y^* = \mathbf{x}'\beta + \mathbf{z}'u + \epsilon \qquad \epsilon \sim Logistic(0,1) \tag{3.4}$$

Hence, we can redefine (3.3) as:

$$logit(\gamma_j) = c_j - \mathbf{x}'\beta - \mathbf{z}'u \tag{3.5}$$

Where $\mathbf{z}$ defines the groups, and u defines the group-level parameters. Our priors and hyperpriors are as follows:

1. Population-level parameters β_j are $\mathrm{N}(0, 3^2)$ priors to be weakly informative. Recall that the standard deviation of the standard logistic is $\frac{\pi}{\sqrt{3}} \approx 1.8$, we increase this value to 3 to add some extra uncertainty to reduce the chance of underfitting.

2. The thresholds c_j have $N(0, 3^2)$ priors for the same reason.

3. The standard deviation of group-level parameters were modelled as $Exponential(1)$, following the usage by McElreath (2019).

4. `brms` uses the non-centered parameterisation, so mean hyperpriors were not explicitly set.

5. For the discrimination, we allow heterogeneity over each class grouping, and place a N(0,3) prior on each class group intercept.

6. On the occasions we model varying slopes and intercepts, we use the LKJ(1) (Lewandowski et al., 2009) prior on the correlations.

4 Bayesian Workflow

The Bayesian workflow will be presented based on our model hypothesised following section 2. In general, we follow the steps of analysis from Gabry et al. (2019). This model is presented in `brms` syntax for conciseness.

```
Proportion_Gr ~ Age_Standardised + log_DDDp1kOBD*Route + (1|Class_Gr)
disc ~ 0 + Class_Gr
```

The first equation can be thought of in latent variable form as:

$$y^* = \beta_1 Age + \beta_2 lDDD + \beta_3 NonInvasive + \beta_4 lDDD * NonInvasive + \alpha_{[\text{Class Group}]} + \epsilon \quad (4.1)$$

We have population level effects on the standardised age and the log DDD, with a group-level intercept on the grouped class. We allow heterogeneity by each class grouping. The group intercept on the grouped class is subject to change to the ungrouped class and the drug name. Howevver `disc` (discrimination) will always be dependent on the grouped class due to sparsity issues with the ungrouped class and drug name.

4.1 Prior Predictive

Using proper priors allows us to generate samples from the prior predictive. Our goal is to ensure that the observed data is feasible under our prior.
Figure 10 shows our prior is not urealistic given our key test statistics (number of 0s, 1s, 0.5s), with the sampled data (y_{rep}) capturing the test statistic within the bounds of the histogram. Similarly, the empirical CDF (ECDF) shows the data is unlikely but not impossible under the prior.

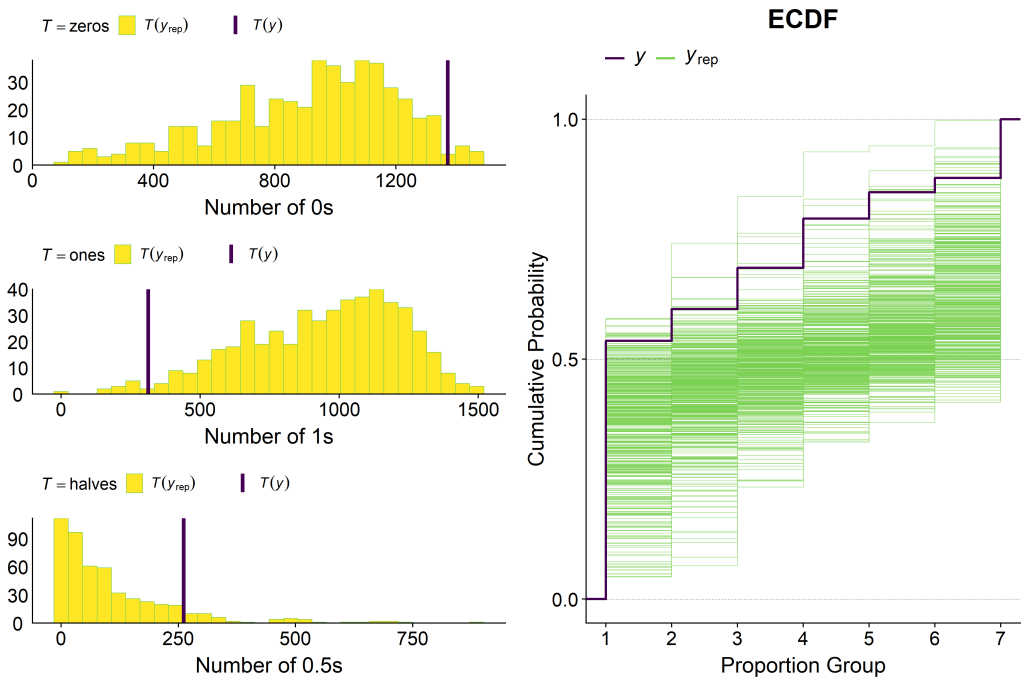

Figure 10: Prior Predictive Plots

4.2 Model Fit and Algorithm Diagnostics

We ran 3 chains, with warmup of 2000, and 5000 total iterations per chain, meaning 9000 post warmup samples with combined chains. This may seem small for those familiar with Gibbs sampling, however the NUTS (Hoffman & Gelman, 2014) allows for convergence with fewer samples, as the effective size per iteration is much higher (Bürkner, 2017).
The model had 0 divergent transitions[1]. The split $\hat{R}$ should be below 1.1 for all parameters (Gelman et al., 2013), and our model is below this threshold as per Figure 11. The minimum effective size ratio is above 0.15, meaning our effective sizes are all above 1000 which is regarded as usually being enough for stable estimates (Bürkner & Vuorre, 2018). Trace plots display no mixing issues.

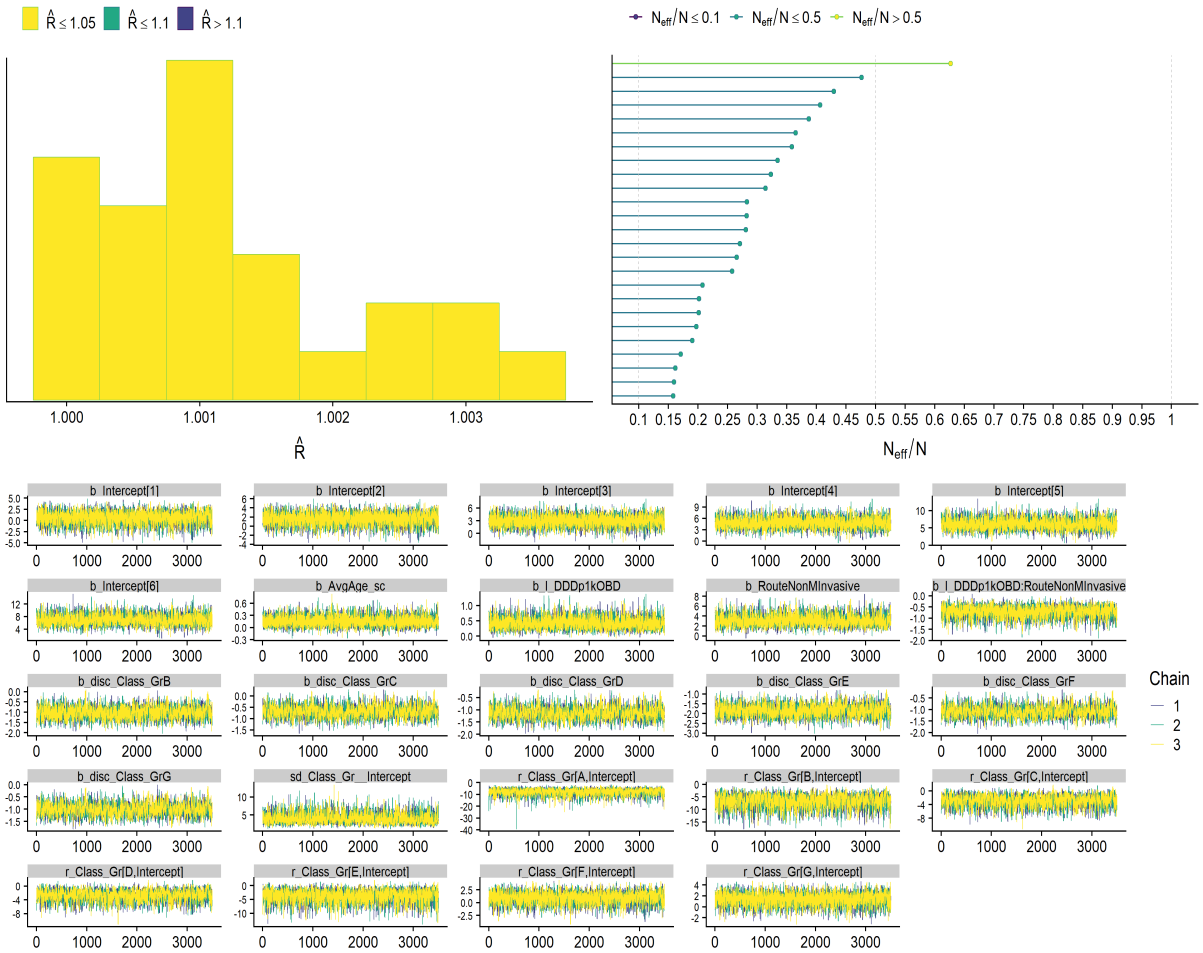

Figure 11: Convergence Diagnostics and ESS

[1]Divergent transitions are an indicator of potential bias in the HMC algorithm, and are a more sensitive diagnostic than the usual $\hat{R}$ measure. For more information, see the online case study by Betancourt (2017).

4.3 Posterior Predictive (Hypothesis Model)

This section explores whether the fitted model would be able to generate a dataset similar to what we observed in reality. We do this primarily by computing test statistics, and checking these against the true data. We examine the test statistics split by the class group, to get a more precise look at the data. Posterior predictive checks can be thought of as the Bayesian analogue to goodness-of-fit tests.

Figure 12 shows the count of each proportion group segmented by the class group. We see that the counts within each class largely match up to the observed values. The 2nd class in group F however looks suspect, with a relatively low level of uncertainty, yet being above the true count indicated by the bar. There is a substantial amount of variation in groups E to G, suggesting there is a lot of uncertainty around these posterior frequencies.

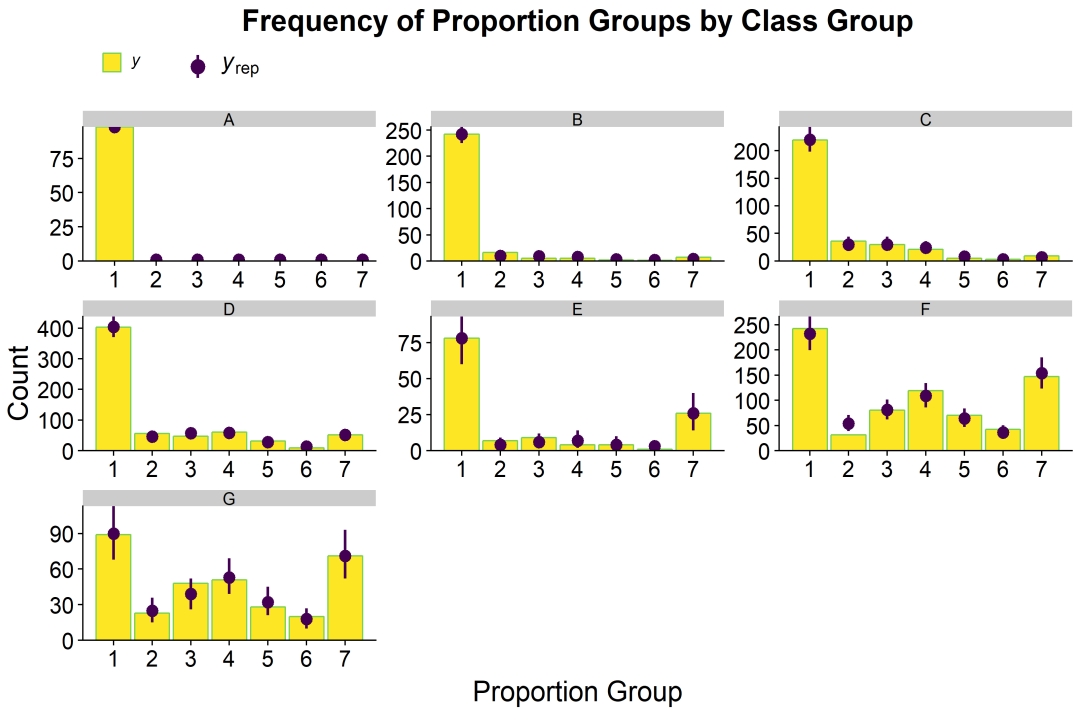

Figure 12: Frequency Plot by Grouped Class

Figure 13 shows the density of the overall dataset appears to be well fitted by the simulations from the posterior predictive.

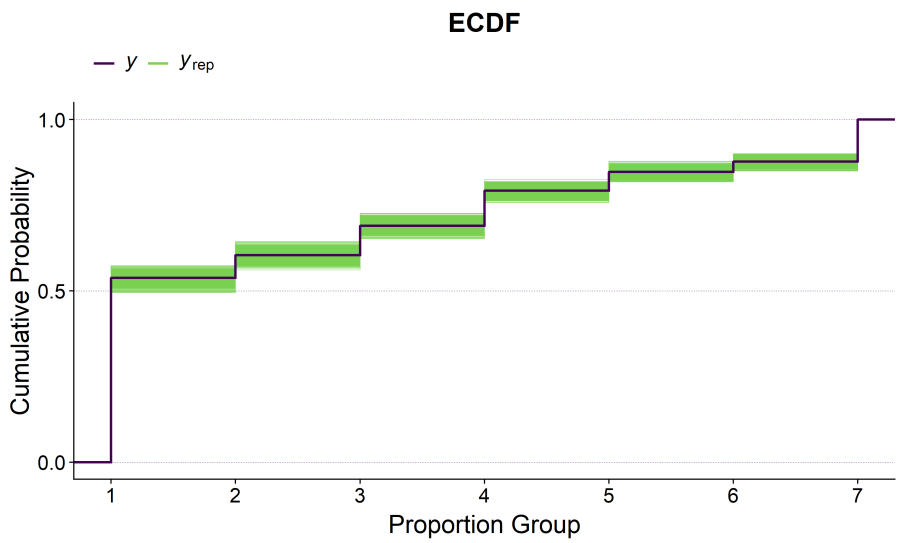

Figure 13: Posterior ECDF

Figure 14 shows the distribution of the counts of 0s, 0.5s and 1s within each class group. In general, the model captures the test statistics well. However the number of 1s in B is slightly underestimated. The number of 0.5s is not unreasonable for any groups, however B, E, and F are slightly off center. On the whole, these test statistics show an adequate fit to the data.

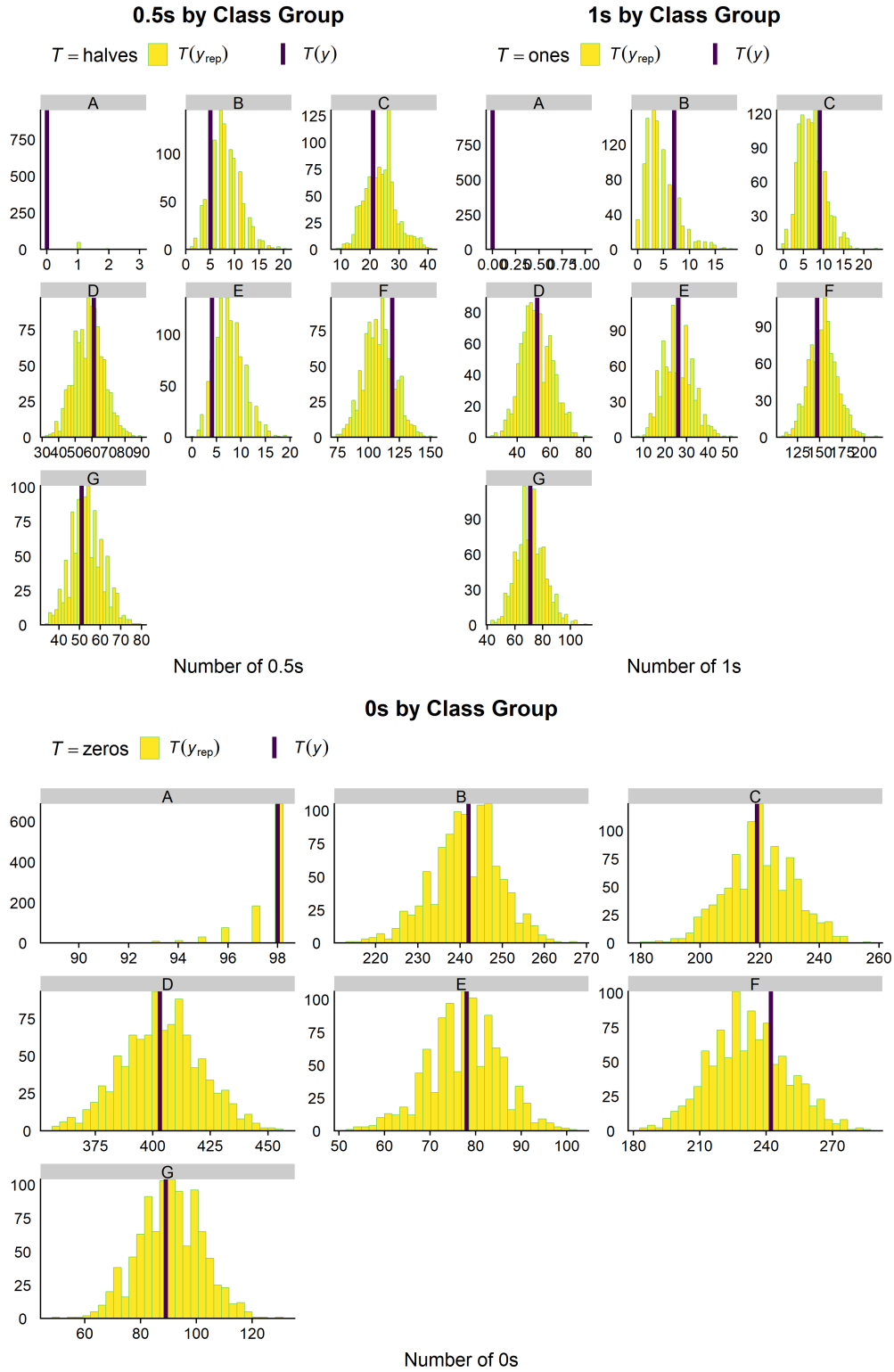

Figure 14: Test Statistics

Figure 15 shows a similar frequency plot to Figure 12, but now the groups are the true observations. In terms of direct predictions the model does poorly, proposing similar patterns within each bin. That is each bin has a peak at 0 and 1, with a wide normal shape on (0,1). Our goal may be inference, however this lack of predictive ability is concerning as it suggests that the data is too variable to predict, meaning we should be sceptical of our coefficients. Examining the predicted linear component (i.e. the values on the latent scale) in Figure 16 we see that, as desired, higher bins have a higher value of the latent variable. This breaks down for the 7th group (proportions of 1), but the general trend is promising. This suggests there is *some* ability to distinguish between proportion groups, but that there is too much variance for specific predictions to be made.

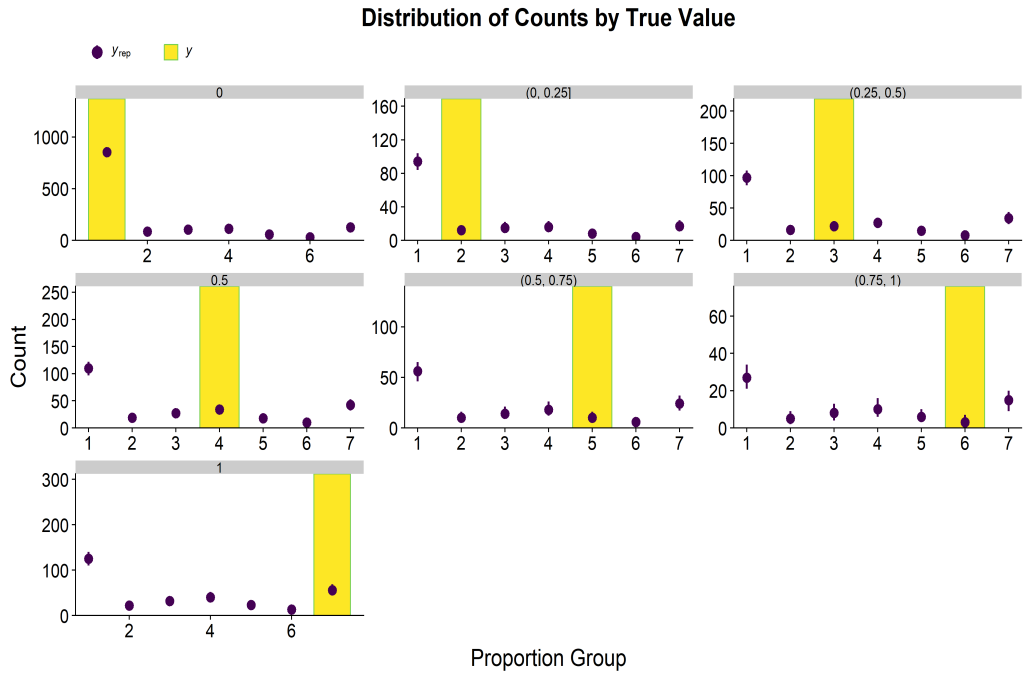

Figure 15: Frequencies vs True Values

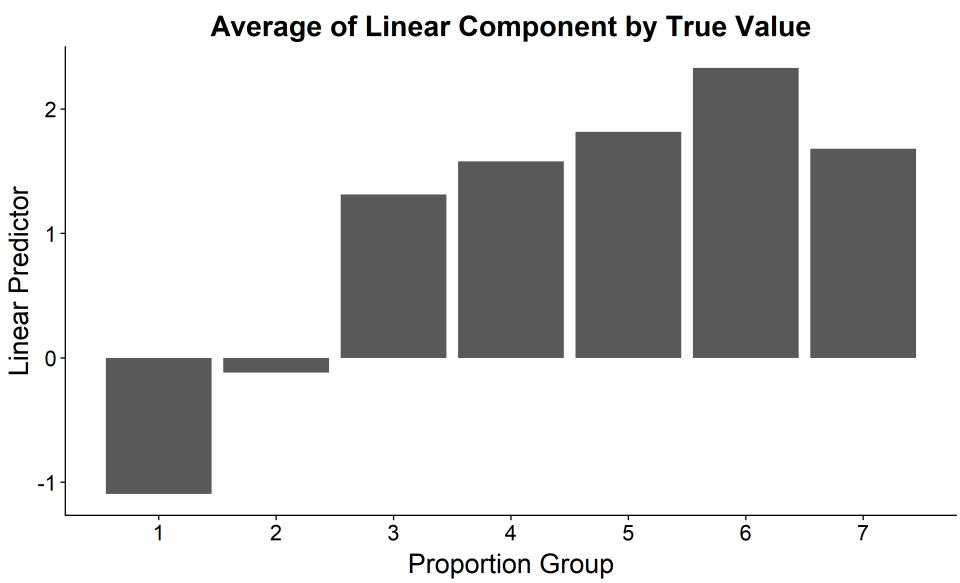

Figure 16: Linear Predictor vs. True Group

4.4 Model Comparison

We compare our hypothesised model to alternative formulations to establish room for improvement, using Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO) (Vehtari et al., 2017*a*). PSIS-LOO has been established as preferable to the widely applicable information criterion (WAIC) developed by Watanabe (2013) – another common model comparison tool. Both aim to approximate the expected log pointwise predictive density (elpd), however WAIC has a larger bias, and the bias-variance tradeoff is harder to control (Vehtari et al., 2017*b*). In general, we interpret a higher elpd as better, and a model as clearly better if the difference in elpd is greater than 2 standard errors. However this is only a general heuristic, since standard error estimates have high variance (Piironen & Vehtari, 2017).
Details of the models compared are appended. The idea is to use a different group for the group-level effect (drug class, drug name), allow varying slopes, and remove the interaction between DDD and the route. We also use an overfit model to get a rough estimate of the maximum predictive potential on the data under the ordered logit, by plotting the equivalent of Figure 15 in Figure 17.
Our hypothesised model is `mp1`, while comparison models begin with `mc`. The overfit model is `mc1`, varying slopes is `mc2`, `mc3x` models use the ungrouped class as the varying intercept, and `mc4x` models use the drug name.
No divergent transitions were reported, $\hat{R}$ was below 1.1 for all models, effective sample sizes were all above 1000, and $\hat{k}$ values (a diagnostic for PSIS-LOO) were all less than 0.7 except the overfit model, which had 6 "bad" values (in (0.7, 1]) and 1 "very bad" value (above 1).

Model	elpd_diff	se_diff	elpd_loo	se_elpd_loo	p_loo	se_p_loo	looic	se_looic
mc41	0.00	0.00	-3438.12	51.90	38.24	1.75	6876.24	103.80
mc43	-0.64	1.71	-3438.76	51.92	37.79	1.72	6877.51	103.84
mc44	-0.89	1.64	-3439.01	51.93	37.33	1.74	6878.02	103.86
mc42	-1.23	1.89	-3439.34	51.92	39.16	1.76	6878.69	103.85
mc1	-9.90	5.08	-3448.02	52.11	59.48	3.81	6896.05	104.22
mp1	-23.73	9.80	-3461.85	50.91	20.36	0.68	6923.69	101.82
mc2	-29.85	10.30	-3467.97	50.94	22.37	0.70	6935.93	101.87
mc31	-30.03	9.45	-3468.15	51.02	27.24	0.91	6936.31	102.04
mc34	-36.42	10.07	-3474.54	51.03	26.27	0.88	6949.08	102.06
mc33	-36.70	10.15	-3474.82	51.01	26.25	0.88	6949.64	102.02
mc32	-37.17	9.98	-3475.29	51.03	28.94	0.92	6950.58	102.06

Table 2: Model Comparison

Table 2 implies a varying intercept for the class group is poor relative to a name-varying intercept, but better than the raw class. In general, population-level effects for the age with an interaction between the route and DDD performed better than alternatives. This can be seen as `mc41` and `mc31` are better than others within their group, and `mp1` is better than `mc2`. The overfit model performs better than our hypothesised model, however this elpd approximation is likely very biased due to the high $\hat{k}$ values.
The prediction barplot for the overfit model can be seen in Figure 17. Again, the raw predictions themselves do poorly, suggesting there is too much variance in the dataset to generate good predictions. Similarly, Figure 18 has the same pattern as previously.
Given the similarity of the "best" model (`mc41`) to our hypothesis (the only difference is the group-level intercept), we briefly review the posterior predictive for `mc41`, and proceed to results.

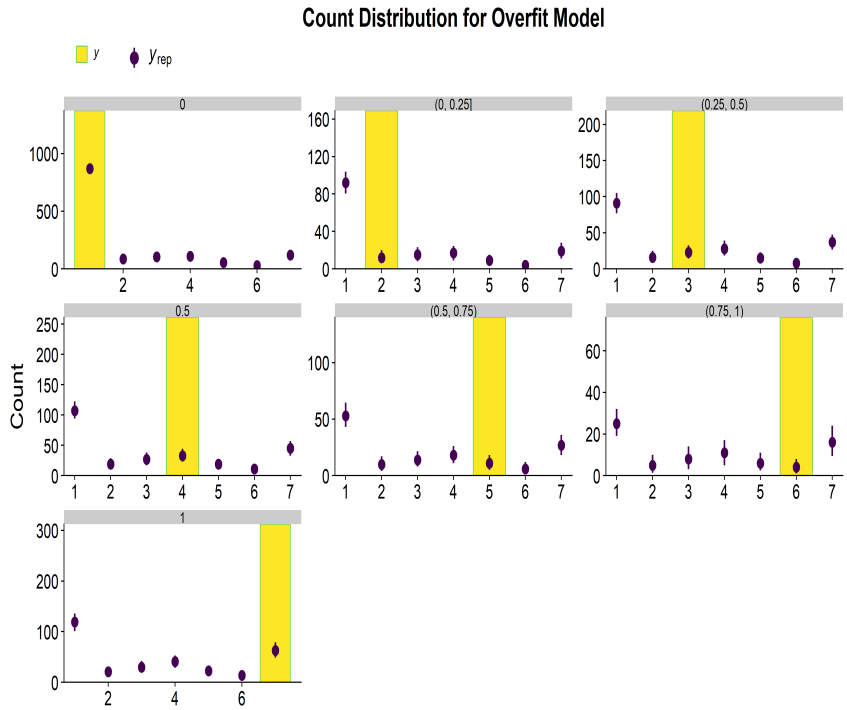

Figure 17: Frequencies vs True Values (Overfit)

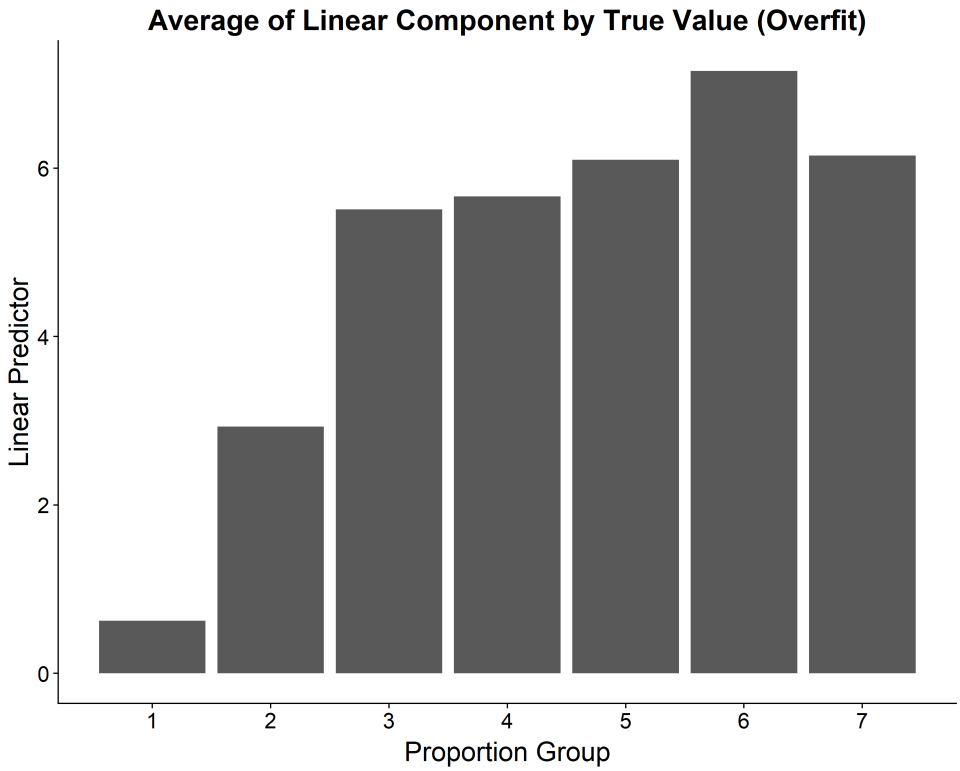

Figure 18: Linear Predictor vs True (Overfit)

4.5 Posterior Predictive (`mc41`)

First, consider Figure 19. Though at first glance there are many "wrong" estimates (e.g. the top left panel), this is just a result of shrinkage to the mean that comes with using group-level effects. Since there are not many obserations in certain groups, their intercept is pushed towards the mean as a result of the common distribution shared by each group. This is how varying intercepts/slopes prevent overfitting in hierarchical models. McElreath (2019) refers to this as adaptive regularisation. See also Gelman et al. (2013), or Gelman & Hill (2007).
The ECDF in Figure 20 looks to fit the data well.
The test statistics for 0s, 1s, and 0.5s in Figures 21, 22, and 22 largely seem appropriate. Some exceptions are the number of 1s for vancomycin, and possibly flucloxacillin, and the numbers of 0.5s for vancomycin, cefalexin, ceftriaxone, and ciprofloxacin.
The barplot of frequencies against true values (Figure 24), and the latent variable plot (Figure 25) are the same story as previously. Poor actual predictions, but a mostly desirable pattern in the latent variable.

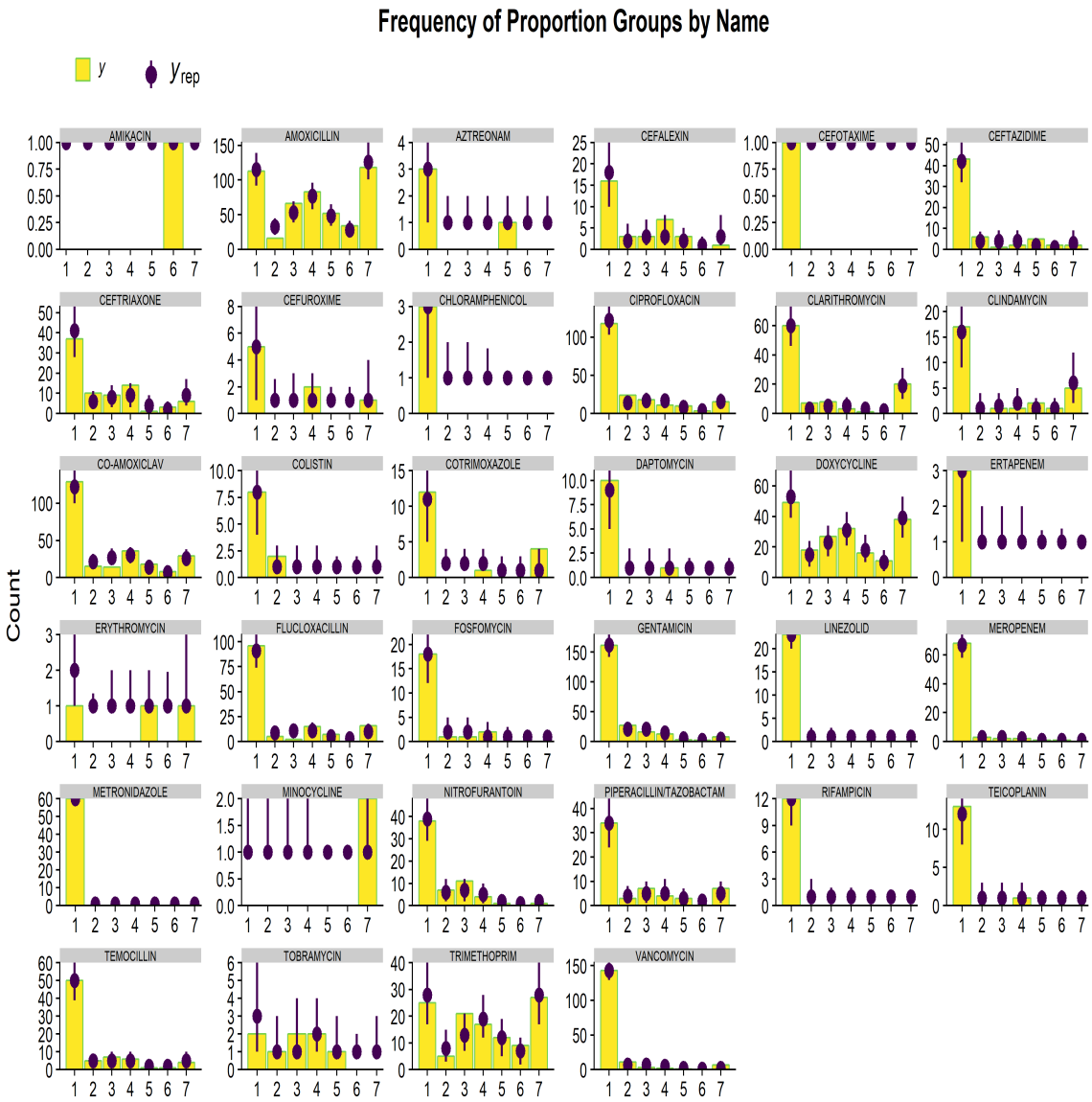

Figure 19: Frequency by Name

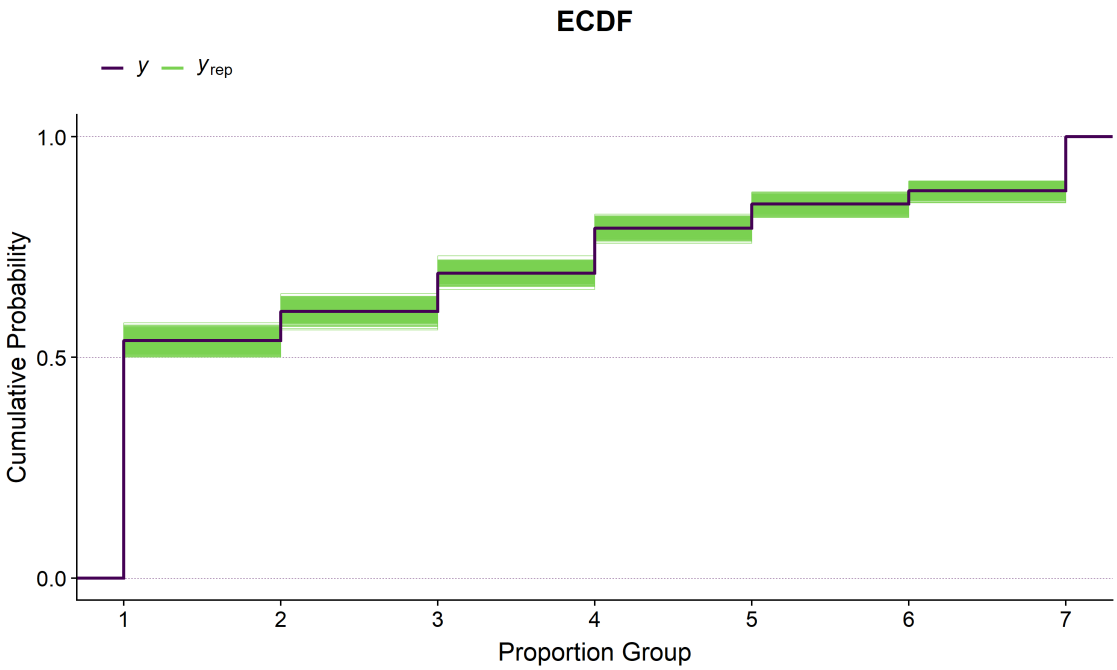

Figure 20: ECDF

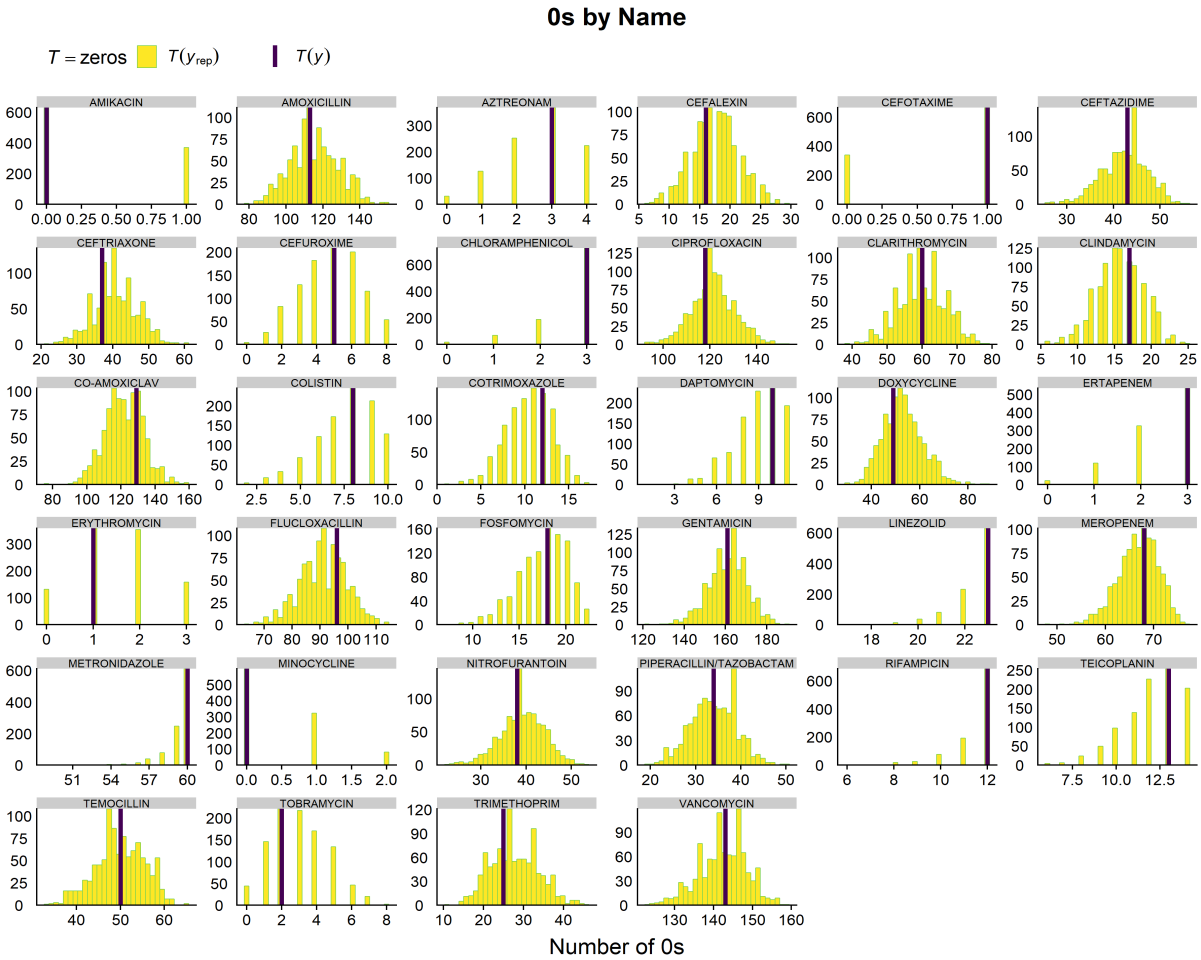

Figure 21: Test Statistic by Name (Number of 0s)

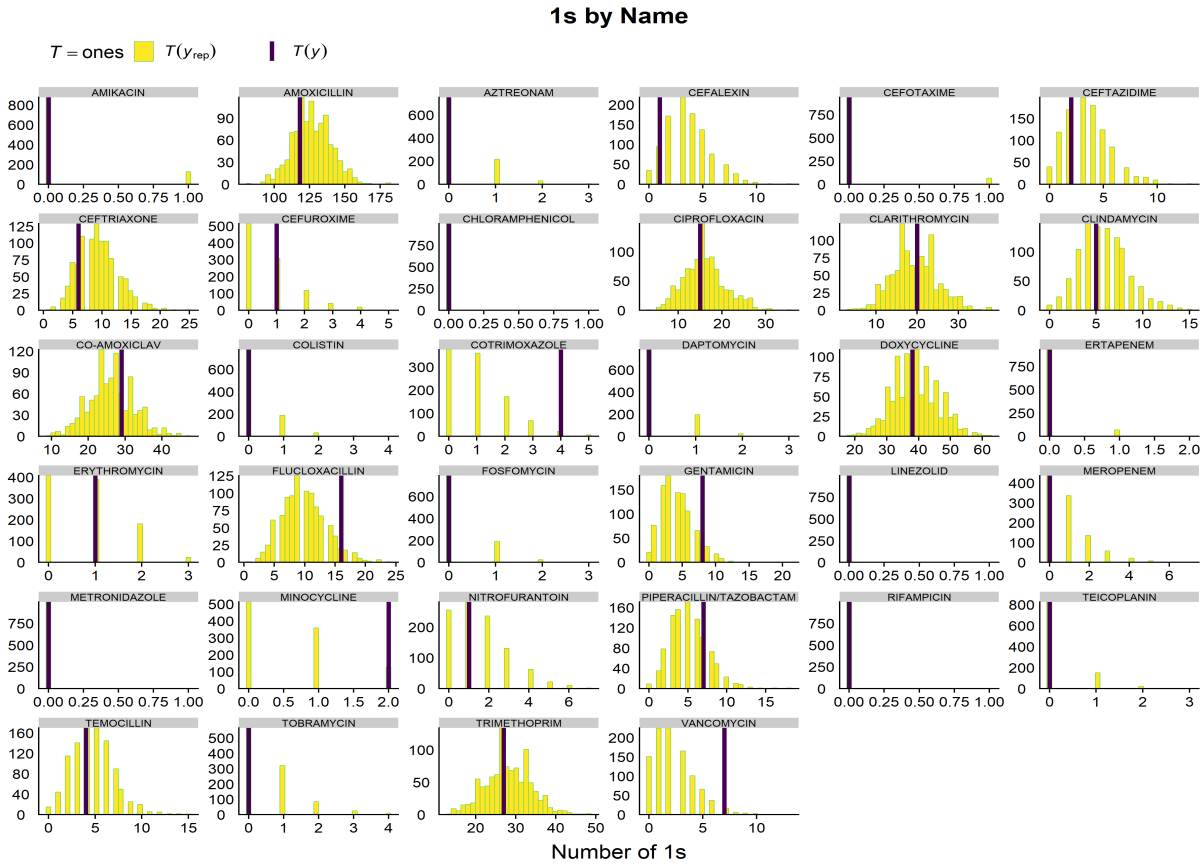

Figure 22: Test Statistic by Name (Number of 1s)

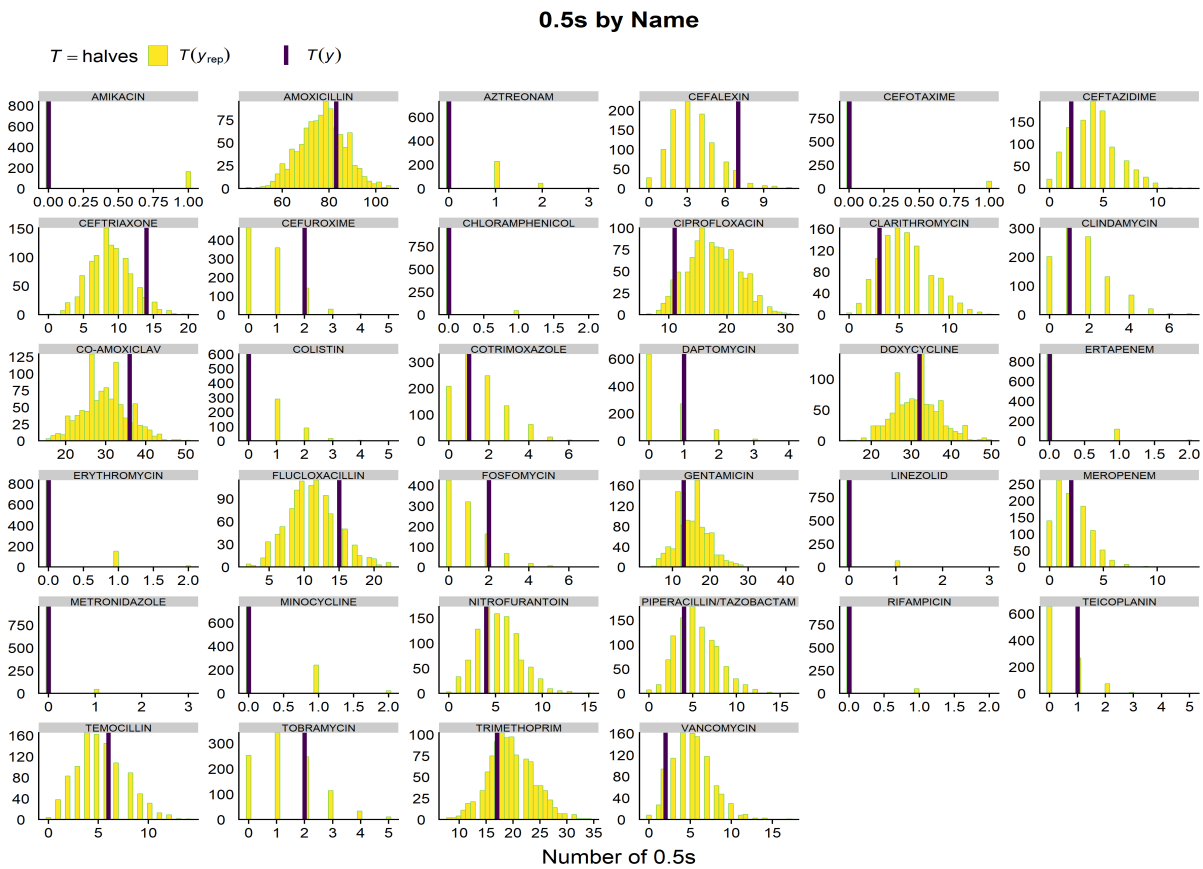

Figure 23: Test Statistic by Name (Number of 0.5s)

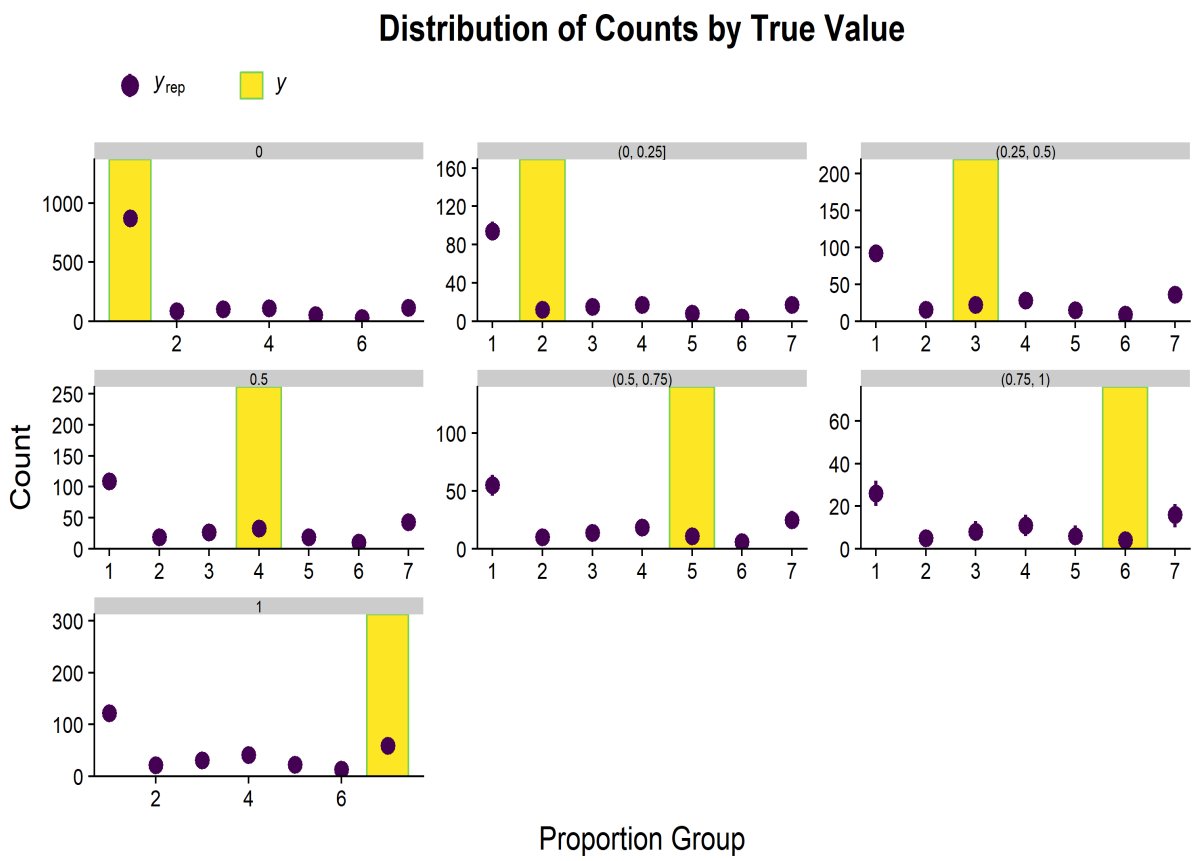

Figure 24: Predictions vs Truth

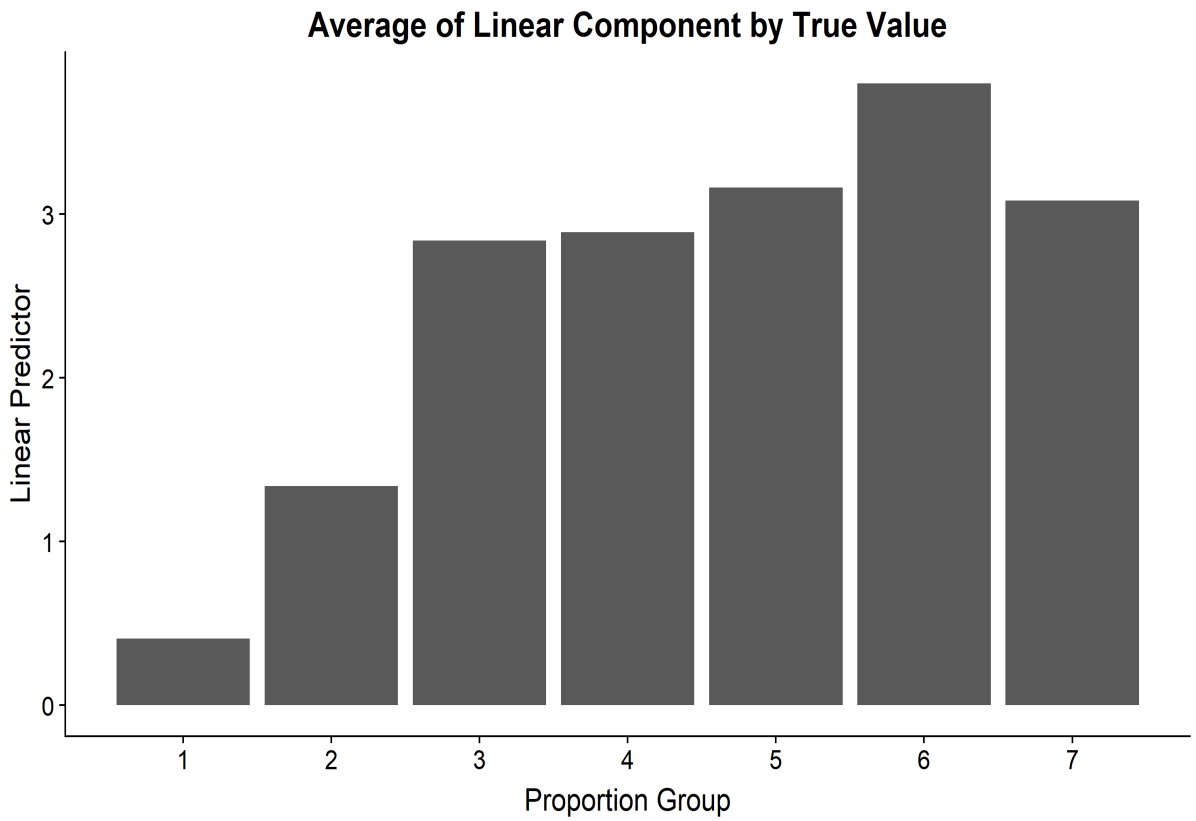

Figure 25: Linear Predictor Value by True Value

5 Results

5.1 DDD and AMR

The results for the population-level effects are in Table 3. The distributions for the key coefficients (i.e. involving DDD and route) are shown in Figure 26. Under our model, there is a 95% probability of a positive association between resistance and the DDD for invasively administered drugs. Specifically, we estimate a unit increase in the log DDD for invasive treatments increases the log-odds of a higher proportion group by 0.24.

It may be more intuitive to look at the latent scale, where Table 4 will assist us. We are still looking at *invasively administered* drugs. Looking at the 2nd column, we can say a unit increase in the log DDD closes the distance between the first and second threshold by 26%. Inverting this number, we see that to transition the whole gap, the log DDD needs to increase by 3.81. From this, we can calculate transition ratios – the multiplier needed on raw DDD to push the linear predictor from the bottom of one group to the bottom of the next group up. This is the exponent of the values in the `Difference/`β_{lDDD} column. For example, to transition from the bottom of group 2 to the bottom of group 3 given a specific value of DDD x_1, the difference $x_2 - x_1$ needs to be such that $\frac{x_2}{x_1}$ is 45.[2] This suggests that although the effect of DDD on resistance is likely positive, it is not overly important materially – particularly considering the middle transition ratios. Recalling Figure 1 – we can see that there is a low proportion of the higher values of raw DDD. The proportion of DDD values larger than 400 is only 6%, suggesting that DDD changes are unlikely to be able to discriminate between groups, since such large values are so rare.

However, it appears that transitioning from group 6 to 7 is more strongly determined by DDD (relatively), with a multiplicative effect of 24 for a bottom-to-bottom transition. In general, the *importance* of DDD in group transitions appears to be higher at more extreme proportions, in terms of the ability to change the proportion group. We might hypothesise that there is some medical tipping-point phenomenon near the more extreme proportions, however this would require more domain expertise.

Parameter	Mean	SD	2.5%	5%	95%	97.5%	Pr(>0)	Pr(<0)
c_1	1.94	0.81	0.42	0.66	3.30	3.58	0.99	0.01
c_2	2.85	0.85	1.32	1.55	4.33	4.64	1.00	0.00
c_3	4.11	0.96	2.45	2.66	5.79	6.18	1.00	0.00
c_4	5.82	1.17	3.76	4.05	7.86	8.35	1.00	0.00
c_5	6.96	1.34	4.60	4.93	9.32	9.88	1.00	0.00
c_6	7.72	1.46	5.14	5.50	10.28	10.89	1.00	0.00
β_{Age}	0.14	0.11	-0.07	-0.03	0.32	0.37	0.90	0.10
β_{lDDD}	0.24	0.15	-0.04	0.00	0.51	0.57	0.95	0.05
$\beta_{NonInvasive}$	1.75	0.97	-0.01	0.25	3.45	3.84	0.97	0.03
$\beta_{lDDD*NonInvasive}$	-0.36	0.21	-0.81	-0.72	-0.04	0.02	0.03	0.97
$\beta_{lDDD} + \beta_{lDDD*NonInvasive}$	-0.12	0.15	-0.42	-0.37	0.12	0.17	0.20	0.80

Table 3: Results Table (Population Effects)

	Difference	β_{lDDD}/Difference	Difference/β_{lDDD}	Transition Ratios
$c_2 - c_1$	0.91	0.26	3.81	45.04
$c_3 - c_2$	1.26	0.19	5.27	193.99
$c_4 - c_3$	1.71	0.14	7.15	1269.58
$c_5 - c_4$	1.15	0.21	4.79	120.90
$c_6 - c_5$	0.76	0.32	3.16	23.65

Table 4: Group Transition Table

[2] $log(x_2) - log(x_1) = 3.81 \iff \frac{x_2}{x_1} = exp(3.81) \approx 45$

For *non-invasive* administration, there is a 97% probability the association between DDD and proportion is smaller than for *invasive* administration ($\beta_{lDDD*NonInvasive}$). The *overall* ceteris paribus[3] effect of an increase in log DDD for a non-invasively administered drug is, on average, a reduction of 0.12 on the latent scale, with an 80% probability the association is negative, hence a 20% probability it is positive. Therefore, there is lots of uncertainty around the association of DDD for non-invasive drugs, so we should not commit to any conclusions. Summarising, we can note that the association between DDD and the proportion:

1. Is positive for invasively administered drugs (95%, β_{lDDD})
2. Is smaller for non-invasively administered drugs (97%, $\beta_{lDDD*NonInvasive}$)
3. *May* be negative for non-invasive drugs (80%, $\beta_{lDDD} + \beta_{lDDD*NonInvasive}$)

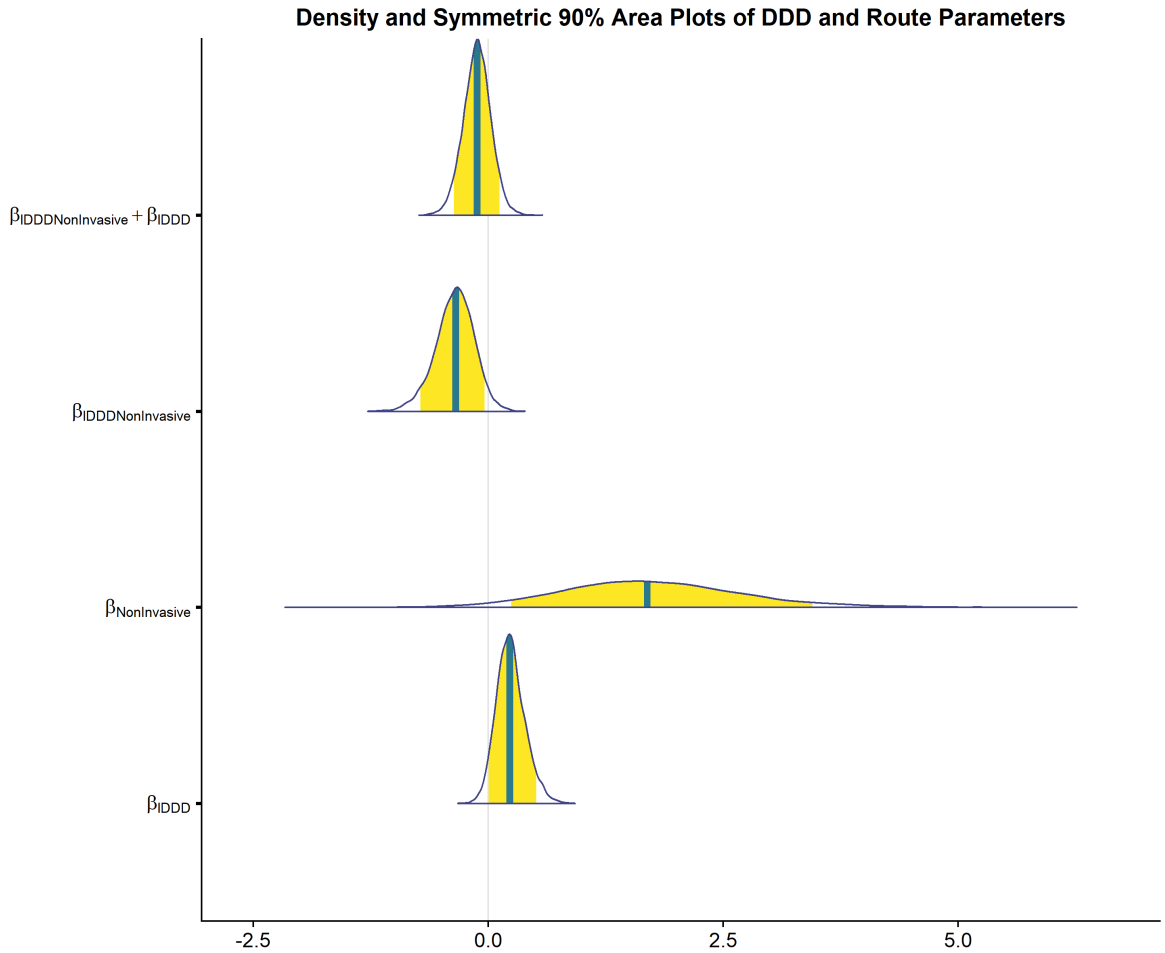

Figure 26: Posterior Area Plots (DDD and Route)

5.1.1 Route

Ceteris paribus, non-invasive administration has a much higher chance of falling into a higher bin, with a mean of 1.75, and a 97% probability of non-invasive administration being associated with higher proportions. Reviewing the thresholds, it appears that the administration route being non-invasive is enough, on average, to achieve a bottom to bottom transition for each group – suggesting it is an important predictor of AMR.

[3] All else being equal

5.2 Drug-Specific Effects

Posterior interval plots of the varying intercepts are shown in Figure 27, with lines at -3 and 3. It appears a large part of the hospital resistance is being driven by a few drugs, with only 5 drugs having 90% symmetric credible intervals not containing 0. The highest magnitudes come from amoxicillin, doxycycline, and trimethoprim (minocycline is excluded due to its wide intervals). On the other side, linezolid, meropenem, and metronidazole all have a strong negative relationship with resistance.

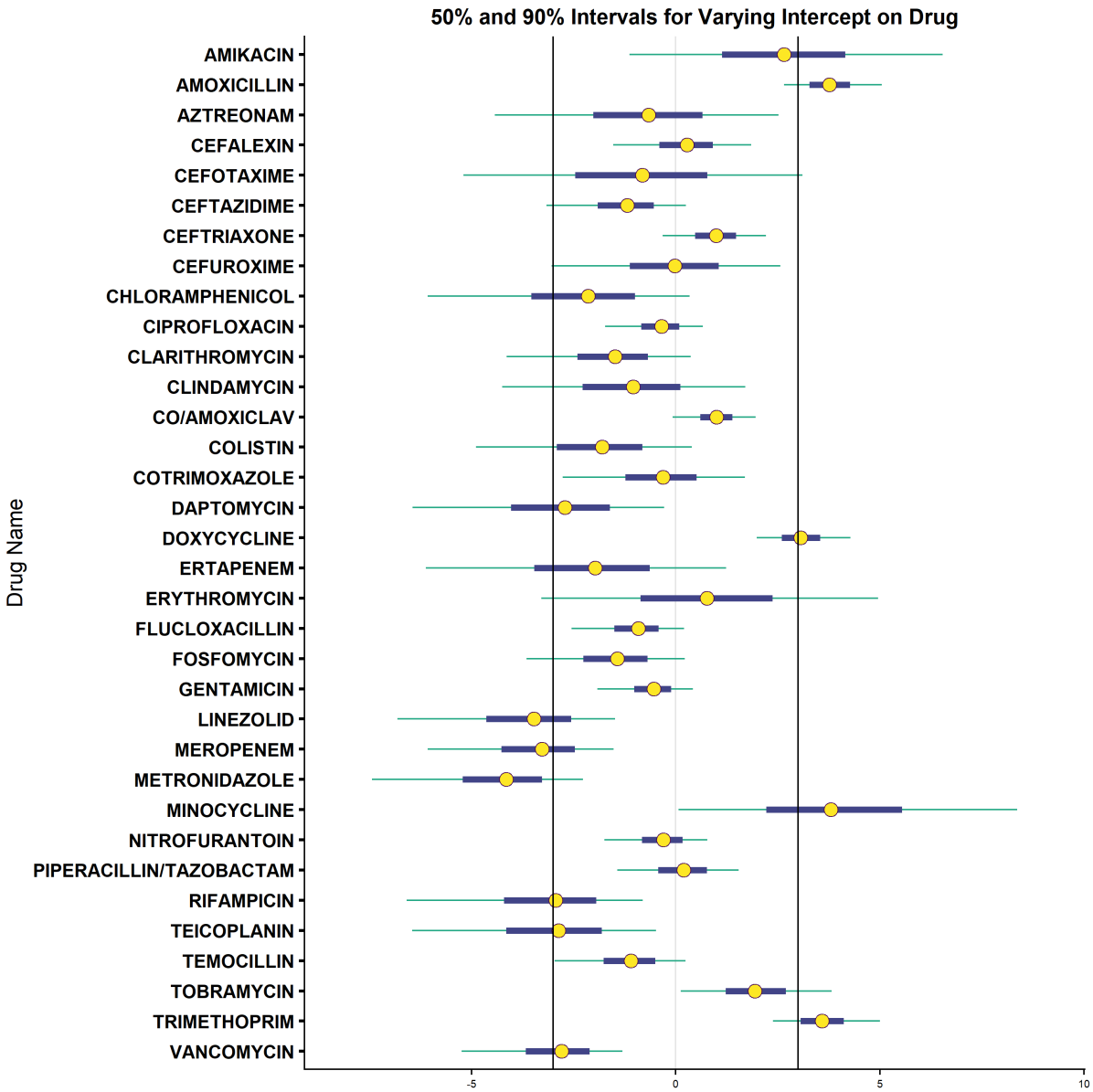

Figure 27: Posterior Interval Plots (Drug Name)

5.3 Class Group Heterogeneity

We transform the discrimination parameter to the standard deviation[4] for interpretation. σ_j refers to the standard deviation of class group j, relative to group A.
All groups are more variable than group A (as expected, since there was no resistance in A), and most variability (relative to A) is between 2 and 3. However, E shows much more variability in the latent variable, with the lower bound of its 95% credible interval being 3.53. That is 3.5 times higher than the standard deviation of the group A, and almost twice as variable as group C.

	Mean	SD	2.5%	97.5%
σ_B	2.28	0.55	1.42	3.52
σ_C	1.84	0.40	1.18	2.71
σ_D	2.90	0.59	1.91	4.19
σ_E	6.01	1.56	3.53	9.70
σ_F	2.65	0.53	1.76	3.82
σ_G	2.50	0.51	1.63	3.64

Table 5: Class Group Heterogeneity

5.4 Comments on Other Fitted Models

Our hypothesised model (`mp1`), and the model with a class-varying intercept (`mc31`) both also had probabilities of the log DDD parameter being positive for invasive administration, this time with above 99% probability. For models with log DDD as a population effect, the minimum probability of a positive relationship was 71% (excluding the overfit model due to the lack of theoretical properties, and the varying slopes models due to no concise measure).

Varying slopes were found to have near-zero standard deviations on all group-level slopes, and wide intervals for the correlations - suggesting there is not enough[5] data to determine correlations between the varying intercepts and slopes. For example `mc32` suggests only a 75% probability the standard deviation between the varying slopes of log DDD was greater than 0.05. The smallest width of a 95% credible interval for estimated correlations was 1.56 – 78% of the whole range of values. Furthermore, the highest probability a correlation was more or less than 0 was only 68%. The combination of the small standard deviation estimates, and the wide intervals for correlations suggest a group-level slope is a needless complication given our limited data.

[4] By transforming each individual sample using s = 1/exp(log(disc))). The output of `brms` already contains the discrimination on the log scale.

[5] "Not enough" refers to data *quality*, not the number of observations. It can be thought of as saying the data does not contain strong evidence

6 Model Discussion

6.1 Concerns

As noted, the predictive performance of the model was poor, mostly due to the high variability in the data, the limited covariates, and the small data (2548 rows). Additionally, though the latent variable should be higher with each group, there was a dip from group 6 to group 7. Furthermore, we should be sceptical of generalising results from this data as it is based solely on the WGH. There may be location-specific effects, or hospital-specific effects that we cannot pick up on due to examining a niche subset of the population.
A weakness of the cumulative logit is that it assumes a unit increase in the covariates has the same impact on the log-odds regardless of the group in question. Alternative methods that can account for so-called category-specific effects are the sequential model, and the adjacent category model (Bürkner & Vuorre, 2018). However the theoretical intuition is not cohesive with our problem structure, with the adjacent category model in particular being difficult to interpret.

6.2 Alternative Model: Zero-one inflated Beta (ZOIB)

Previously, a ZOIB (Ospina & Ferrari, 2008) model was fitted to the data, however even posterior histograms and density plots fit the data relatively poorly, specifically the intermediate proportions. In addition, the complexity of interpretation due to the separate logistic models internally fitted meant the model was an unattractive proposition, despite the raw proportion seemingly following such a distribution.
Though not implemented due to time constraints, it may be of interest to combine the ordered logit with the ZOIB, by recalling section 3.2. We can attempt to determine whether the latent variable implies a 0, a 1, or if it passes through the function g. In the latter case, we could use a beta regression, as all responses constructed by g are in (0,1). Currently, ZOIB ignores the ordering of data, and models the 0s and 1s as a separate process. This contributes to its difficult interpretation. For instance, a parameter suggesting an increase in a covariate x is associated with the response y being in $\{0,1\}$ cannot be interpreted directly, since we cannot see the directional effect (i.e. lower or higher values of y) of a covariate, because 0 and 1 are opposed. As such, this covariate must be interpreted in multiple equations, and its parameters do not have a clean form suggesting whether its overall impact is positive or negative.
Another approach to finding the contribution to the overall mean is marginalisation of the ZOIB. Marginalisation of a two-part model is discussed in Smith et al. (2017), but the concepts should extend to ZOIB.

7 Conclusion

The objective of this paper was to establish if there exists an association between drug use (measured in DDD/1000 OBD) and AMR (measured by proportion of clinical isolates exhibiting resistance) at the hospital level, using a Bayesian orderd logit model.
We find a 95% probability of a positive association between DDD and AMR for invasively administered drugs, with a 97% probability that non-invasive administration leads to a smaller association between DDD and AMR. This latter association may even be negative, however the probability of this is only 80% which is too small to draw conclusions from. The relative importance of invasively administered DDD in determining the proportion group is small, with a minimum transition ratio of 23.65 for a bottom to bottom transition. These ratios were smallest at more extreme proportions, perhaps implying a tipping-point phenomenon.
In general, resistance mostly arises from a few drugs - namely amoxicillin, doxycycline, and trimethoprim. The drugs with the most negative association were linzolid, meropenem, and metronidazole.
We have also suggested an alternative model (ZOIB), and how it might be improved by accounting for the natural ordering in the data, which is currently ignored. Marginalisation is also an attractive proposition.
Concerns involve the generalisability of the model due to the subpopulation of a single hospital which could mean ignored hospital-specific or location-specific effects. The average age of patients is also very high, with the minimum average age on a ward-week being 45. Additionally, there may be selection bias due to the study being in a hospital setting where people are already ill. There is also the concern of the model's poor predictive performance, even when considering the purposefully overfit model.

References

Betancourt, M. (2017), 'Diagnosing biased inference with divergences'. `https://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html`.

Bürkner, P.-C. (2017), '**brms:** an r package for bayesian multilevel models using stan', *Journal of Statistical Software* **80**(1).
URL: *https://doi.org/10.18637/jss.v080.i01*

Bürkner, P.-C. (2018), 'Advanced Bayesian multilevel modeling with the R package brms', *The R Journal* **10**(1), 395–411.
URL: *https://doi.org/10.32614/RJ-2018-017*

Bürkner, P.-C. & Vuorre, M. (2018), 'Ordinal regression models in psychology: A tutorial'.
URL: *https://doi.org/10.31234/osf.io/x8swp*

Cassini, A., Högberg, L., Plachouras, D., Quattrocchi, A., Hoxha, A., Skov Simonsen, G., Colomb-Cotinat, M., Kretzschmar, M., Devleesschauwer, B., Cecchini, M., Ait Ouakrim, D., Cravo Oliveira, T., Struelens, M., Suetens, C., L Monnet, D., Strauss, R., Mertens, K., Struyf, T., Catry, B. & Hopkins, S. (2019), 'Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the eu and the european economic area in 2015: a population-level modelling analysis', *The Lancet Infectious Diseases* **19**, 56–66.

Costelloe, C., Metcalfe, C., Lovering, A., Mant, D. & Hay, A. D. (2010), 'Effect of antibiotic prescribing in primary care on antimicrobial resistance in individual patients: systematic review and meta-analysis', *BMJ* **340**(may18 2), c2096–c2096.
URL: *https://doi.org/10.1136/bmj.c2096*

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A. & Swinton, J. (2019), *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.
URL: *https://CRAN.R-project.org/package=xtable*

Dowle, M. & Srinivasan, A. (2019), *data.table: Extension of 'data.frame'*. R package version 1.12.2.
URL: *https://CRAN.R-project.org/package=data.table*

Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987), 'Hybrid monte carlo', *Physics Letters B* **195**(2), 216 – 222.
URL: *http://www.sciencedirect.com/science/article/pii/037026938791197X*

Gabry, J. & Mahr, T. (2019), 'bayesplot: Plotting for bayesian models'. R package version 1.7.0.
URL: *mc-stan.org/bayesplot*

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M. & Gelman, A. (2019), 'Visualization in bayesian workflow', *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **182**(2), 389–402.
URL: *https://doi.org/10.1111/rssa.12378*

Gelman, A. (2005), 'Analysis of variancewhy it is more important than ever', *Ann. Statist.* **33**(1), 1–53.
URL: *https://doi.org/10.1214/009053604000001048*

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis, Third Edition*, Chapman and Hall/CRC, Boca Raton.

Gelman, A. & Hill, J. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.
URL: *https://www.xarg.org/ref/a/052168689X/*

Goossens, H., Ferech, M., Stichele, R. V. & Elseviers, M. (2005), 'Outpatient antibiotic use in europe and association with resistance: a cross-national database study', *The Lancet* **365**(9459), 579–587.
URL: *https://doi.org/10.1016/s0140-6736(05)70799-6*

Greene, W. H. (2012), *Econometric Analysis*, Pearson, Boston London.

Hoffman, M. & Gelman, A. (2014), 'The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo', *Journal of Machine Learning Research* **15**, 1593–1623.

Holmes, A. H., Moore, L. S. P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J. & Piddock, L. J. V. (2016), 'Understanding the mechanisms and drivers of antimicrobial resistance', *The Lancet* **387**(10014), 176–187.
URL: *https://doi.org/10.1016/s0140-6736(15)00473-0*

Hughes, S. (2002), 'Understanding the mechanisms and drivers of antimicrobial resistance', *British Journal of Clinical Pharmacology* **46**(6), 531–533.
URL: *https://doi.org/10.1046/j.1365-2125.1998.00842.x*

Kowarik, A. & Templ, M. (2016), 'Imputation with the R package VIM', *Journal of Statistical Software* **74**(7), 1–16.

Lewandowski, D., Kurowicka, D. & Joe, H. (2009), 'Generating random correlation matrices based on vines and extended onion method', *Journal of Multivariate Analysis* **100**(9), 1989 – 2001.
URL: *http://www.sciencedirect.com/science/article/pii/S0047259X09000876*

McCullagh, P. (1980), 'Regression models for ordinal data', *Journal of the Royal Statistical Society. Series B (Methodological)* **42**(2), 109–142.
URL: *http://www.jstor.org/stable/2984952*

McElreath, R. (2019), *Statistical Rethinking.*
URL: *https://xcelab.net/rm/sr2/*

Neal, R. M. (2012), 'MCMC using Hamiltonian dynamics'.

O'Neill, J. (2014), 'Antimicrobial resistance: Tackling a crisis for the health and wealth of nations'. London, `https://amr-review.org/Publications.html`.

O'Neill, J. (2016), 'Antimicrobial resistance: Tackling drug-resistant infections globally: final report and recommendations'. London, `https://amr-review.org/Publications.html`.

Ospina, R. & Ferrari, S. L. P. (2008), 'Inflated beta distributions', *Statistical Papers* **51**(1), 111.
URL: *https://doi.org/10.1007/s00362-008-0125-4*

Pearl, J. (2009), *Causality: Models, Reasoning and Inference*, Cambridge University Press.
URL: *https://www.xarg.org/ref/a/052189560X/*

Piironen, J. & Vehtari, A. (2017), 'Comparison of bayesian predictive methods for model selection', *Statistics and Computing* **27**(3), 711–735.
URL: *https://doi.org/10.1007/s11222-016-9649-y*

R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Roope, L. S. J., Smith, R. D., Pouwels, K. B., Buchanan, J., Abel, L., Eibich, P., Butler, C. C., Tan, P. S., Walker, A. S., Robotham, J. V. & Wordsworth, S. (2019), 'The challenge of antimicrobial resistance: What economics can contribute', *Science* **364**(6435).
URL: *https://science.sciencemag.org/content/364/6435/eaau4679*

Smith, V. A., Neelon, B., Preisser, J. S. & Maciejewski, M. L. (2017), 'A marginalized two-part model for longitudinal semicontinuous data', *Statistical Methods in Medical Research* **26**(4), 1949–1968. PMID: 26156962.
URL: *https://doi.org/10.1177/0962280215592908*

Stan Development Team (2018), *Stan Modeling Language User's Guide and Reference Manual.* Version 2.18.0.
URL: *http://mc-stan.org*

Stroustrup, B. (2013), *The C++ Programming Language*, 4th edn, Addison-Wesley Professional.

Turnheim, K. (1998), 'Drug dosage in the elderly', *Drugs & Aging* **13**(5), 357–379.
URL: *https://doi.org/10.2165/00002512-199813050-00003*

Vehtari, A., Gelman, A. & Gabry, J. (2017*a*), 'Pareto smoothed importance sampling'.

Vehtari, A., Gelman, A. & Gabry, J. (2017*b*), 'Practical bayesian model evaluation using leave-one-out cross-validation and waic', *Statistics and Computing* **27**(5), 1413–1432.
URL: *https://doi.org/10.1007/s11222-016-9696-4*

Walker, S. H. & Duncan, D. B. (1967), 'Estimation of the probability of an event as a function of several independent variables', *Biometrika* **54**(1-2), 167–179.
URL: *https://doi.org/10.1093/biomet/54.1-2.167*

Watanabe, S. (2013), 'A widely applicable bayesian information criterion', *Journal of Machine Learning Research* **14**(1), 867–897.
URL: *http://dl.acm.org/citation.cfm?id=2567709.2502609*

WHO (2018). WHO report on surveillance of antibiotic consumption: 2016-18 early implementation. Geneva: World Health Organisation; 2018. Licence: CC BY-NC-SA 3.0 IGO.

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York.
URL: *https://ggplot2.tidyverse.org*

Wickham, H., Franois, R., Henry, L. & Mller, K. (2019), *dplyr: A Grammar of Data Manipulation.* R package version 0.8.1.
URL: *https://CRAN.R-project.org/package=dplyr*

Wilke, C. O. (2019), *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'.* R package version 0.9.4.
URL: *https://CRAN.R-project.org/package=cowplot*

World Bank (2017). Drug-Resistant Infections: A Threat to Our Economic Future. Washington, DC: World Bank. License: Creative Commons Attribution CC BY 3.0 IGO.

Appendices

A Data Preparation

Figure 28 shows that around 36% of the data was missing, primarily in the resistance proportion. All data was dropped. For the covariates, the little data missing is insignificant. For the resistance proportion, the data was missing in the sense that it was never measured, **not** because there was some mechanism that caused values to not be reported. This is because the clinical isolates were not relevant to the drug.
We grouped the route into "Invasive" and "Non-Invasive" groups, due to limited observations for rectal transmission. Additionally, we removed a row which had a DDD of 0 due to irrelevance.

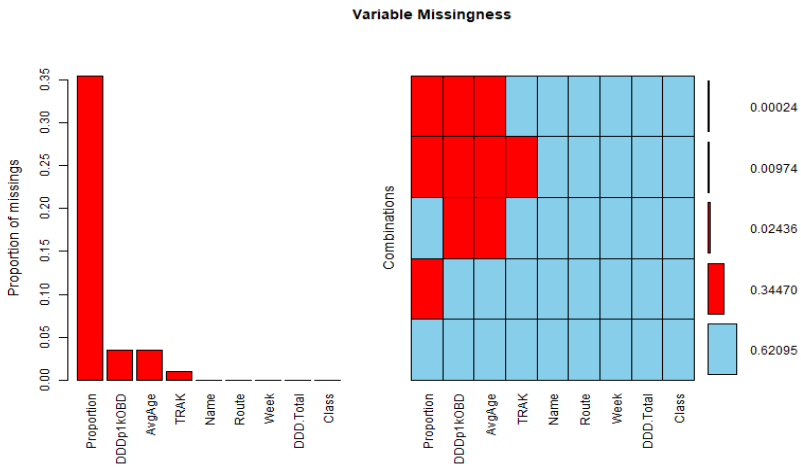

Figure 28: Missingness

Figure 28 was created using the `VIM` package by Kowarik & Templ (2016).

B Models Compared

This section outlines all of the models compared in this report. brms syntax was used to keep things concise, and we shortened the formula further using individual letters. The corresponding terms are in the first table. All models used 3 chains. Runtimes varied between 10 and 30 minutes for all models except the overfit model - which took over 1 hour. The machine was a Dell Inspiron 15-7559, running Windows 10 Home 64-bit, an Intel(R) Core(TM) i5-6300 HQ CPU @ 2.30 GHz, with 4 cores. Note that chains were run on separate cores. Expressions included prior to | within brackets are group-level effects, while those outside are constant. `adapt_delta` is a parameter controlling the step size of the HMC algorithm. The maximum value is 1, and higher values make for slower computation, but help to avoid divergent transitions.

Letter	Stands For
A	Average age (standardised)
C	Class
CG	Grouped Class
D	log DDD per 1000 OBD
N	Drug name
P	Grouped Proportion
R	Route

Table 6: Drug Class Groupings

Name	brms formula	warmup	iterations	adapt_delta
mp1	`P ~ A + D*R + (1|CG)`	2500	6000	0.9
mc1	`P ~ A + D*R*N`	1000	4000	0.9
mc2	`P ~ (1 + A + D*R|N)`	1000	4500	0.98
mc31	`P ~ A + D*R + (1|C)`	1000	4500	0.91
mc32	`P ~ A + (1+D*R|C)`	2000	5000	0.99
mc33	`P ~ A + R + (1+D|C)`	2000	5000	0.95
mc34	`P ~ A + R + D + (1|C)`	2000	5000	0.9
mc41	`P ~ A + (1+D*R|N)`	1000	4500	0.91
mc42	`P ~ A + D*R + (1|N)`	1000	4500	0.95
mc43	`P ~ A + R + (1+D|N)`	1000	4500	0.98
mc44	`P ~ A + R + D + (1|N)`	1000	4500	0.9

Table 7: Drug Class Groupings

C R Session Information

```
R version 3.5.3 (2019-03-11)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 10 x64 (build 17134)

Matrix products: default

locale:
[1] LC_COLLATE=English_United States.1252  LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252 LC_NUMERIC=C
[5] LC_TIME=English_United States.1252

attached base packages:
[1] grid      stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] xtable_1.8-4      brms_2.9.3        Rcpp_1.0.1        dplyr_0.8.1       VIM_4.8.0
[6] colorspace_1.4-1  loo_2.1.0         bayesplot_1.7.0   data.table_1.12.2 cowplot_0.9.4
[11] ggplot2_3.1.1

loaded via a namespace (and not attached):
[1] class_7.3-15         rio_0.5.16           ggridges_0.5.1       rsconnect_0.8.13
[5] markdown_0.9         base64enc_0.1-3      rstudioapi_0.10      rstan_2.18.2
[9] audio_0.1-6          DT_0.6               mvtnorm_1.0-10       ranger_0.11.2
[13] bridgesampling_0.6-0 robustbase_0.93-5    shinythemes_1.1.2    packrat_0.5.0
[17] shiny_1.3.2          compiler_3.5.3       tictoc_1.0           backports_1.1.4
[21] assertthat_0.2.1     Matrix_1.2-15        lazyeval_0.2.2       cli_1.1.0
[25] later_0.8.0          beepr_1.3            htmltools_0.3.6      prettyunits_1.0.2
[29] tools_3.5.3          igraph_1.2.4.1       coda_0.19-2          gtable_0.3.0
[33] glue_1.3.1           reshape2_1.4.3       carData_3.0-2        cellranger_1.1.0
[37] nlme_3.1-137         crosstalk_1.0.0      lmtest_0.9-37        laeken_0.5.0
[41] stringr_1.4.0        ps_1.3.0             openxlsx_4.1.0       mime_0.6
[45] miniUI_0.1.1.1       gtools_3.8.1         DEoptimR_1.0-8       MASS_7.3-51.1
[49] zoo_1.8-5            scales_1.0.0         colourpicker_1.0     hms_0.4.2
[53] promises_1.0.1       Brobdingnag_1.2-6    parallel_3.5.3       inline_0.3.15
[57] shinystan_2.5.0      curl_3.3             gridExtra_2.3        StanHeaders_2.18.1
[61] stringi_1.4.3        dygraphs_1.1.1.6     e1071_1.7-1          boot_1.3-20
[65] pkgbuild_1.0.3       zip_2.0.2            rlang_0.3.4          pkgconfig_2.0.2
[69] matrixStats_0.54.0   lattice_0.20-38      purrr_0.3.2          rstantools_1.5.1
[73] htmlwidgets_1.3      labeling_0.3         tidyselect_0.2.5     processx_3.3.1
[77] plyr_1.8.4           magrittr_1.5         R6_2.4.0             pillar_1.4.0
[81] haven_2.1.0          foreign_0.8-71       withr_2.1.2          xts_0.11-2
[85] abind_1.4-5          sp_1.3-1             nnet_7.3-12          tibble_2.1.1
[89] crayon_1.3.4         car_3.0-3            readxl_1.3.1         callr_3.2.0
[93] forcats_0.4.0        threejs_0.3.1        vcd_1.4-4            digest_0.6.19
[97] tidyr_0.8.3          httpuv_1.5.1         stats4_3.5.3         munsell_0.5.0
[101] viridisLite_0.3.0    shinyjs_1.0
```

D Code & Data Access

Code and data will soon be available on `https://github.com/S1889112/EdinburghMSc`, after personal information and copyrighted papers have been removed, as well as a general cleanup of file structure.

D.1 Stan Code Example

For those wishing to learn Stan, I have pasted the code underlying model `mc41` below from using the `stancode` function in `brms`. This may also help in understanding the role of the `disc` parameter in ordered logit. Note that Stan is written in $C++$ (Stroustrup, 2013), so variables and types are explicitly declared – knowing this may help with syntax understanding.

```
// generated with brms 2.9.3
functions {
    /* cumulative-logit log-PDF for a single response
    * Args:
    *   y: response category
    *   mu: linear predictor
    *   thres: ordinal thresholds
    *   disc: discrimination parameter
    * Returns:
    *   a scalar to be added to the log posterior
    */
    real cumulative_logit_lpmf(int y, real mu, vector thres, real disc) {
    int ncat = num_elements(thres) + 1;
    real p;
    if (y == 1) {
        p = inv_logit(disc * (thres[1] - mu));
    } else if (y == ncat) {
        p = 1 - inv_logit(disc * (thres[ncat - 1] - mu));
    } else {
        p = inv_logit(disc * (thres[y] - mu)) - inv_logit(disc * (thres[y - 1] - mu));
    }
    return log(p);
    }
}
data {
    int<lower=1> N;  // number of observations
    int<lower=2> ncat;  // number of categories
    int Y[N];  // response variable
    int<lower=1> K;  // number of population-level effects
    matrix[N, K] X;  // population-level design matrix
    int<lower=1> K_disc;  // number of population-level effects
    matrix[N, K_disc] X_disc;  // population-level design matrix
    // data for group-level effects of ID 1
    int<lower=1> N_1;  // number of grouping levels
    int<lower=1> M_1;  // number of coefficients per level
    int<lower=1> J_1[N];  // grouping indicator per observation
    // group-level predictor values
    vector[N] Z_1_1;
    int prior_only;  // should the likelihood be ignored?
}
transformed data {
```

```
    int Kc = K;
    matrix[N, Kc] Xc;  // centered version of X
    vector[Kc] means_X;  // column means of X before centering
    for (i in 1:K) {
        means_X[i] = mean(X[, i]);
        Xc[, i] = X[, i] - means_X[i];
    }
}
parameters {
    vector[Kc] b;  // population-level effects
    // temporary thresholds for centered predictors
    ordered[ncat - 1] temp_Intercept;
    vector[K_disc] b_disc;  // population-level effects
    vector<lower=0>[M_1] sd_1;  // group-level standard deviations
    // standardized group-level effects
    vector[N_1] z_1[M_1];
}
transformed parameters {
    // actual group-level effects
    vector[N_1] r_1_1 = (sd_1[1] * (z_1[1]));
}
model {
    // initialize linear predictor term
    vector[N] mu = Xc * b;

    // initialize linear predictor term
    vector[N] disc = X_disc * b_disc;
    for (n in 1:N) {
        // add more terms to the linear predictor
        mu[n] += r_1_1[J_1[n]] * Z_1_1[n];
        disc[n] = exp(disc[n]);
    }

    // priors including all constants
    target += normal_lpdf(temp_Intercept | 0,3);
    target += normal_lpdf(b_disc | 0,3);
    target += exponential_lpdf(sd_1 | 1);
    target += normal_lpdf(z_1[1] | 0, 1);
    // likelihood including all constants
    if (!prior_only) {
        for (n in 1:N) {
        target += cumulative_logit_lpmf(Y[n] | mu[n], temp_Intercept, disc[n]);
        }
    }
}
generated quantities {
    // compute actual thresholds
    vector[ncat - 1] b_Intercept = temp_Intercept + dot_product(means_X, b);
}
```