Antibiotic usage in Flusurvey

Gwen, Seb (John? Ceire?)

Contents

1 Abstract

Background: Antibiotic resistance (ABR) is a growing public health problem caused by selection due to antibiotic use. Understanding who takes antibiotics, and why they do so, can help with the design and targeting of interventions for reducing prescription rates and hence unnecessary antibiotic use.

Methods: We used an internet-based open community cohort, Flusurvey, to assess levels of antibiotic use. Participants were asked whether they had taken antibiotics during each period in which they reported any symptoms. We developed regression models to explore population covariates and antibiotic use. Modelling and inference use the fully Bayesian approach via Markov Chain Monte Carlo (MCMC) simulation techniques.

Results: We analysed, 27,673 episodes of illness from 3,650 patients and found a rate of antibiotic consumption of 4% over the years 2012-2017. 16% of those taking antibiotics did not report visiting a medical service during the episode. Multivariate regression showed...

Conclusions: We found that antibiotics are used in 4% of illness episodes reported in a community cohort, and that a relatively high rate of antibiotic use was not associated with a visit to a medical centre (and hence potentially no prescription). Covariate importance... Intervention implication...

2 Introduction

Antibiotic resistance (ABR) is a growing public health problem (ref WHO). ABR is selected by the use of antibiotics, and hence the reduction in use is a public health priority. Understanding who uses antibiotics, and if there are any population covariates that make someone more likely to take antibiotics when unwell, can help to target interventions.

80% of antibiotics in the UK are prescribed in the community (ref). However, it is unclear what this population receiving antibiotics looks like and for what symptoms they are prescribed antibiotics (Ref). We used an existing internet-based open community cohort, Flusurvey, to determine if any covariates could be linked to likelihood of taking antibiotics.

- Previous work - what covariates are known to give high antibiotic use? (control for?)

- What could we do - previous intervention design in community? target GPs?

- Why Bayesian? (should we use the above information to inform our priors?)

3 Methods

3.1 Data

Included in this study were episodes of illness from any resident of the UK recruited into Flusurvey between 2012 and 2017. Flursurvey was approved by the London School of Hygiene and Tropical Medicine Ethics Committe (Application number 5530). Details of the survey structure and recruitment can be found in [1].

Briefly, participants recruited into Flusurvey were asked a set of background questions at the start of the influenza season. There was then a weekly email asking participants to complete a symptoms survey. If participants reported any symptoms in their weekly symptoms survey, they were asked "Did you take medication for these symptoms (tick all that apply)?". One possible medication was "Antibiotics". This record, of taking antibiotics or not, by episode is analysed here.

An episode is defined as:... ?? how much detail?

Episodes were excluded if there was no information on whether they had visited a medical service in this episode ("GP", "hospital", "A&E", "Other" or "waiting for an appointment") which removed 380 episodes (¡ 2% of total possible episodes). Further episodes were removed as there was no information on their age (201), whether they had received an influenza vaccine this year (35), whether they had a fever (30) or a health score (13).

3.2 Covariate identification

We considered all risk factors from Flusurvey that were potentially associated with differences in antibiotic usage rates. These were then plotted in a univariate analysis against antibiotic usage rates (see Supplementary). Only those where there were sufficient data on both the covariate and antibiotic usage rates were included (e.g. region was excluded).

The final covariates were: gender, age, influenza like illness (ILI) with fever, influenza vaccine received this year, frequent contact with children, underlying health issue (e.g. diabetes) and whether a participant visited a medical service during this episode. All covariates were binary except for age which was regularized by subtracting the mean and dividing by the standard deviation.

3.3 Analysis

We used the Flusurvey data, aggregated into patient episodes. For each of the n episodes, we had a set of observations $((y_i, w_i), i = 1, , n)$, where y_i was a binary response such that $y_i = 1$ if a participant in Flusurvey reported taking an antibiotic during this episode, and $y_i = 0$ if not. The $w_i = w_{i1}, ...w_{in}$ are the covariate values for each participant for this episode, mentioned above.

Our logistic regression model then estimated the binomial probability of receiving an antibiotic or not (y_i). We assumed Normal priors for all covariate coefficients (β_j) parameters with mean 0 and variance 10.

$$y_i \sim Binomial(1, \theta) \tag{1}$$

$$\theta = a + \sum_{0<j<n} \beta_j w_{ij} + \sum_{\substack{0<j<n \\ 0<k<n}} \beta_k w_{ij} w_{ik} \tag{2}$$

$$\beta_j \sim Normal(0, 10) \tag{3}$$

3.4 Models

We analysed the following logistic models for θ, where i is episode number:

$$M1 : \theta = a + \beta_1 w_{i,age} + \beta_2 w_{i,gender} \tag{4}$$

$$M2 : \theta = a + \beta_1 w_{i,age} + \beta_2 w_{i,gender} + \beta_3 w_{i,age} w_{i,gender} \tag{5}$$

$$M3 : \theta = a + \beta_1 w_{i,age} + \beta_2 w_{i,gender} + \beta_3 w_{i,ili.fever} + \beta_4 w_{i,vaccine.this.year} \tag{6}$$

$$+ \beta_5 w_{i,freq.contact.children} + \beta_6 w_{i,underlying.risk} + \beta_7 w_{i,visit.medical.service} \tag{7}$$

$$M4 : \theta = a[visit.medical.service] + \beta_1 w_{i,age} + \beta_2 w_{i,gender} + \beta_3 w_{i,ili.fever} \tag{8}$$

$$+ \beta_4 w_{i,vaccine.this.year} + \beta_5 w_{i,freq.contact.children} + \beta_6 w_{i,underlying.risk} \tag{9}$$

$$+ \beta_7 w_{i,visit.medical.service} \tag{10}$$

$$a[visit.medical.service] \sim Normal(0, \sigma_v) \tag{11}$$

$$\sigma_v \sim Cauchy(0, 10) \tag{12}$$

$$\beta_j \sim Normal(0, 10) \tag{13}$$

The first and second model explore the relationship between age and gender in explaining antibiotic usage. There is a known relationship between high antibiotic usage in women (ref), and in the elderly (ref). The third model included all covariates with sufficient data.

Due to the need for a prescription to obtain antibiotics in the UK, there was a high association between a participant visit to a medical service and antibiotic usage. Hence, in the fourth model, the intercept was segregated by the binary covariate of whether a participant had made a medical visit in this episode.

We implemented the models in R 3.3.3 [ref R] using the "rethinking" package (ref) and Stan (ref). The theta parameter was converted to a probability using the logit function. The models were fit to the data using the "map2stan" function which uses Monte Carlo Markov Chain (MCMC) sampling to generate posterior distributions for all covariate coefficients. Most models were linear, but the interaction of age and gender was explored in Model 2. For the four models, 4,000 MCMC iterations are carried out with a burn-in sample of 1,000. Convergence was monitored by plotting trace and autocorrelation plots of the samples.

3.5 Model comparison

We compared the set of plausible models using the Watanabe-Akaike information criterion (WAIC) [ref]. The WAIC is an estimate of out-of-sample deviance. Small values of WAIC indicate a good fit. We also calculated pWAIC: the estimated effective number of parameters to give an idea of

how flexible each model is in fitting the sample, and the standard error of the WAIC estimate. To compare models, we calculate the Akaike weight, which is an estimate of the probability that the model will make the best predictions on new data, conditional on the set of models considered [ref]. The model with the greatest weight is likely to do the best at prediction.

4 Results

A total of 27,673 episodes of illness were included in this analysis for 3,650 patients. Antibiotics were taken in 11,51 (4%) of episodes. The maximum number of episodes per participant was 22, whilst the mean and standard deviation were 5.5 and 4.8.

16% of those who report taking an antibiotic for this episode, report that they did not visit a medical service. Of the 8% of episodes where participants did report visiting a medical service, 42% took an antibiotic during this episode.

Children (<18yo) and the elderly (>65) had higher rates of antibiotic usage than others (Figure 1), across all included Flusurvey seasons.
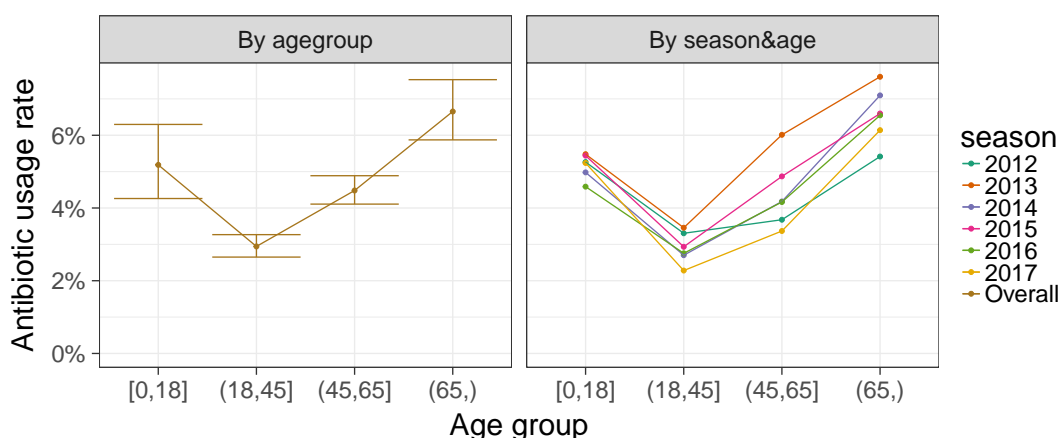

Figure 1: Antibiotic usage rate by age and season. Note that here antibiotic usage rate is per episode of illness.

4.1 Model comparison

When comparing the goodness of fit of the models, we find that Model 1 and 2 had no weight, whilst 3 & 4 had equal weighting (Table 1), suggesting that they are equally good.

Model	WAIC	pWAIC	Difference in WAIC	Akaike weight	SE	Difference in SE
M4	5305.3	8.3	NA	0.52	144.16	NA
M3	5305.5	8.3	0.1	0.48	144.33	0.27
M2	9521.5	3.1	4216.2	0	207.59	152.26
M1	9523.4	4.5	4218.1	0	207.75	152.36

Table 1: Model comparison output

The parameter estimates for the coefficients are shown in Table 2 and Figure 2. Most are highly similar between the two models except for the intercept (which varies in model formulation).

The biggest parameter is for the coefficient of ILI with fever - suggesting that those with the most serious illness are the most likely to take an antibiotic.

The fit of the model to the data is shown in Figure 3. ?? why model 3/4 so bad?

Figure 2: Posterior parameter estimates for Model 3 (haven't generated for Model 4 yet - gives me weird extraction samples (only 4...) when use same method as Model 3, to do).
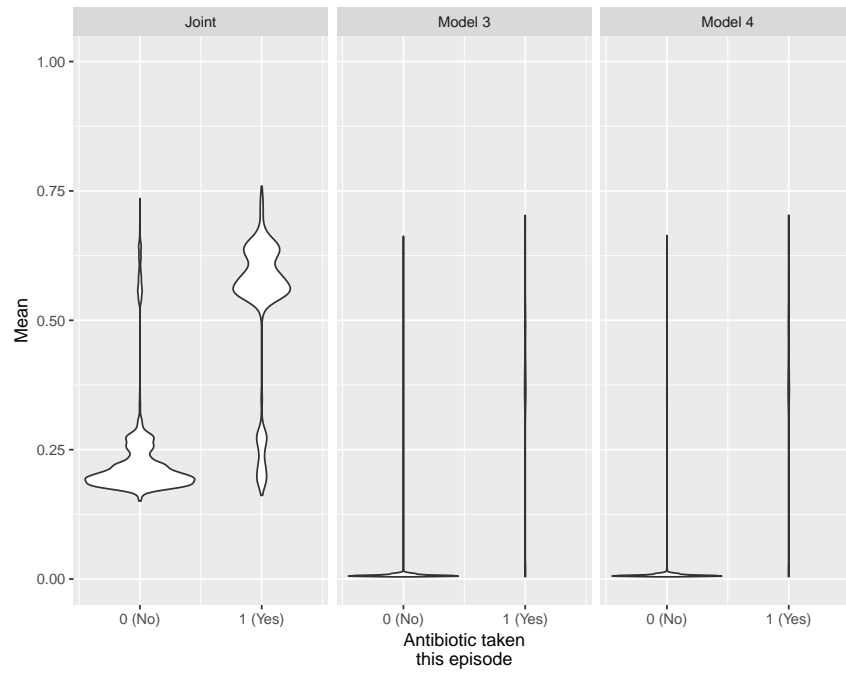
Figure 3: Model fits. Samples from the posterior distribution are taken and used to generate mean antibiotic exposure rates for each episode. The left plot uses the weights from the model fits to estimate the joint output. The middle and right plot show the distribution of means for each episode over 1,000 samples for Model 3 and 4.

Para.	Description	Model 3			Model 4		
		Mean	Lower	Upper	Mean	Lower	Upper
a	Intercept	0.47	0.43	0.5	0.88	0	1
β_1	Coefficient of age	0.52	0.49	0.55	0.52	0.49	0.55
β_2	Coefficient of gender	0.53	0.51	0.54	0.53	0.51	0.54
β_3	Coefficient of ILI fever	0.65	0.59	0.7	0.65	0.59	0.7
β_4	Coefficient of vaccine this year	0.48	0.44	0.51	0.48	0.45	0.51
β_5	Coefficient of frequent contact with children	0.57	0.53	0.6	0.57	0.53	0.61
β_6	Coefficient of underlying risk	0.39	0.36	0.43	0.39	0.35	0.42
β_7	Coefficient of visit to medical service	0.01	0.01	0.01	0.06	0	0.99
σ_v	Variance in intercept (a)	NA	NA	NA	1	0.51	1

Table 2: Estimates of parameters from Models 3 & 4. The model outputs are converted here from the fitted log-odds scale using the logistic function. Note that the intercept (a), is a mean value for Model 4.

5 Discussion

Random thoughts:

- Interest in "inappropriate" antibiotic use. Although Flusurvey asks about flu / antibiotic use it doesn't give us a handle on this as diagnosis is not clear.

- only look at Flusurvey "seasons": peak prescribing time?

- can we see whether socioeconomic or health reasons are more importance?

- can we change the agegroups = some more differentiation in the lower age band?

- Do for other countries and compare?

References

[1] Alma J Adler, Ken TD Eames, Sebastian Funk, and W John Edmunds. Incidence and risk factors for influenza-like-illness in the uk: online surveillance using flusurvey. *BMC infectious diseases*, 14(1):232, 2014.