Prevalence of antimicrobial resistant *Escherichia coli* in sympatric wild rodents varies by season and host.

Nicola J Williams[1], Christopher Sherlock[2], Trevor R Jones[3], Helen E Clough[1], Sandra E Telfer[4], Michael Begon[3], Nigel French[5], Charles A Hart[6], Malcolm Bennett[1].

[1]National Centre for Zoonosis Research, Institute of Infection and Global Health, University of Liverpool, UK. [2] Department of Mathematics and Statistics, Lancaster University, UK. [3] Institute of Integrative Biology, University of Liverpool, UK. [4] Institute of Biological and Environmental Sciences, University of Aberdeen, UK. [5] Institute of Veterinary, Animal and Biomedical Sciences, Massey University, New Zealand. [6] School of Host Immunity and Infection, University of Liverpool, UK.

A Statistical methodology and detailed results

A.1 Missingness assumption

On any given day when data were collected most of the 501 subjects were not caught and sampled. In order to perform a statistical analysis on the data the nature of any relationship between whether or not a subject was captured and whether or not it would test positive for antibiotic resistant *E. coli* conditional on presence of *E. coli* must be decided upon. An assumption that the missingness and response processes are independent seems reasonable in this scenario, although the statistical analysis which follows is valid even with the much less restrictive "missing at random" assumption (e.g. Diggle et al., 2002).

A.2 Notation and covariates

For each antibiotic, y_{ij} denotes j^{th} observation of the i^{th} subject, with $y_{ij} = 1$ indicating presence of bacteria resistant to that antibiotic. The vector of covariates for this observation is denoted $\mathbf{x}_{ij}$. Writing d for the number of days since 31^{st} December 2000, the vector of covariate parameters $\boldsymbol{\beta}$ comprises of

β_1	fixed intercept corresponding to an adult female bank vole
β_2	additional effect if subject is a wood mouse
β_3	additional effect if subject is male
β_4, β_5	additional effect if subject is a juvenile or sub-adult respectively
β_6, β_7	coefficient of linear time trend for bank voles and wood mice respectively
β_8, β_9	coefficients respectively of $\sin(2\pi d/365)$ and $\cos(2\pi d/365)$ (bank voles only)
β_{10}, β_{11}	coefficients respectively of $\sin(2\pi d/365)$ and $\cos(2\pi d/365)$ (wood mice only)

A.3 Model

For each of the four antibiotics, ampicillin, chloramphenicol, tetracycline and trimethoprim, the following Bernoulli model was fitted.

$$y_{ij} \sim \text{Bern}(p_{ij}) \quad , \ \text{logit}(p_{ij}) = \mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i \tag{1}$$

The subject specific random effect $\gamma_i \sim N(0, \sigma^2_\gamma)$ allows for the fact that even once covariates are taken into account some individuals will be more prone to a certain antibiotic resistant strain than other individuals. A further potential consideration is that individual infections might last longer than the sampling interval. In this case a subject that tested positive for antibiotic resistant *E. coli* at one time might be more likely to test positive the next time than would otherwise be the case given species, age, date and "robustness". Exploratory analysis via variograms of residuals from a fitted GLM showed that this was not a significant consideration.

A.4 Priors

A standard prior choice for the vector of covariate effects is

$$\boldsymbol{\beta} \sim N(\mathbf{m}_\beta, \mathbf{S}_\beta)$$

In our analysis flat (improper) priors were chosen: ($\mathbf{m}_\beta = \mathbf{0}$ and $\mathbf{S}_\beta^{-1} = \mathbf{0}$). This indicates no prior knowledge about the covariate effects and still leads to a proper posterior.

An improper prior for σ^2 would lead to an improper posterior and so a conjugate proper prior was chosen:

$$\left(\frac{\sigma^2}{n_\sigma S_\sigma^2}\right)^{-1} \sim \chi^2_{n_\sigma}$$

We chose $n_\sigma = 3$ since this provides the most diffuse prior with with a finite mean. The prior scaling S_σ was set to 1; this is the same order of magnitude as the contribution from significant fixed effects in a simple logistic regression fit to the data.

To test sensistivity of posterior parameter estimates to their priors, runs were also performed with both vaguer and tighter priors for σ^2 and with tighter priors for the covariate parameters β_i. Substantive inferences remained the same for each antibiotic whichever combination of priors was chosen.

A.5 Algorithm

The MCMC scheme is based upon the iteratively weighted least squares (IWLS) algorithm of Gamerman (1997). Updates are Metropolis-within-Gibbs and the m^{th} iteration proceeds as follows.

1. Use an IWLS update to sample $\boldsymbol{\beta}^{(m)}$ given $\boldsymbol{\gamma}^{(m-1)}$ and $\mathbf{y}$.
2. Cycle through the subjects; for the i^{th} subject use an IWLS update to sample $\gamma_i^{(m)}$ given $\boldsymbol{\beta}^{(m)}$, $\sigma^{2\ (m-1)}$ and $\mathbf{y}$.
3. Sample $\sigma^{2\ (m)}$ its exact (scaled inverse chi-square) distribution given $\boldsymbol{\gamma}^{(m)}$.

Each Markov chain was run for a total of 20000 iterations and for all runs convergence was achieved after a burn in period of less than 500 iterations. Plots in the main text and parameter estimates presented in Section A.7 use the last 19500 iterations of output.

A.6 Seasonal signal

Parameters were incorporated into the model so as to discern separate seasonal cycles for each species. For the Gamerman algorithm to be applicable parameters must be combined linearly and so annual cycles were represented as superposition of sine/cosine waves, for example for bank voles:

$$\beta_8 \sin(2\pi d/365) + \beta_9 \cos(2\pi d/365)$$

This is equivalent to the more intuitive

$$A\cos(2\pi d/365 - \theta)$$

with

$$A = \sqrt{\beta_8^2 + \beta_9^2} \quad \text{and} \quad \theta = \begin{cases} \cos^{-1}(\beta_9/A) & (\beta_8 \geq 0) \\ 2\pi - \cos^{-1}(\beta_9/A) & (\beta_8 < 0) \end{cases}$$

Here A as is the amplitude of the seasonal signal and $d := 365 \times \theta/2\pi$ is the day of the year at which the signal is at its peak. Posterior samples for d for bank voles (wood mice) were generated from the posterior samples for β_8 and β_9 (β_{10} and β_{11}) using the above equations.

A.7 Parameter estimates and credible intervals

Table A.7.1 shows a point estimate (the posterior median) with 95% credibility limits for each of the covariate effects for each of the antibiotics. In a Bayesian context an effect is usually deemed statistically significant if this credibility interval does not contain 0. We therefore conclude that for all antibiotics prevalence of resistance (conditional on presence of *E.coli*) is higher amongst wood mice than bank voles, and higher amongst adults than sub-adults (except for tetracycline, and perhaps trimethoprim). There is little evidence however for any difference between the sexes or between adults and juveniles. The prevalence of resistance in the wood mouse population is clearly increasing with time; there is also evidence of a weaker increasing trend amongst bank voles. Superimposed upon this trend is a seasonl cycle, which is stronger for wood mice, most probably peaking around early July, and weaker for bank voles, most probably peaking around late Summer or early Autumn.

	ampicillin	chloramphenicol	tetracycline	trimethoprim
β_1: offset	-3.37 (-4.16,-2.62)	-5.18 (-6.36,-4.11)	-4.11 (-5.01,-3.28)	-4.53 (-5.59,-3.61)
β_2: WM - BV	1.38 (0.53,2.22)	2.32 (1.16,3.52)	1.6 (0.66,2.53)	1.59 (0.58,2.66)
β_3: M - FM	0.08 (-0.28,0.44)	0.34 (-0.09,0.76)	0.09 (-0.28,0.45)	0.09 (-0.29,0.48)
β_4: Ju. - Ad.	-0.14 (-0.89,0.54)	0.05 (-0.88,0.85)	0.10 (-0.67,0.79)	-0.21 (-1.10,0.57)
β_5: SA - Ad.	-0.68 (-1.38,-0.06)	-1.09 (-2.3,-0.16)	0.12 (-0.50,0.69)	-0.51 (-1.26,0.16)
β_6: BV trend	0.88 (0.24,1.52)	1.53 (0.67,2.38)	1.55 (0.91,2.23)	1.74 (1.02,2.50)
A: BV seas.	0.36 (0.07,0.81)	0.95 (0.32,1.66)	0.60 (0.16,1.10)	0.67 (0.18,1.22)
d: BV seas.	262*	236 (192,276)	243 (187,295)	268 (217,323)
β_7: WM trend	0.53 (0.06,1.01)	0.36 (-0.19,0.92)	0.95 (0.47,1.43)	1.38 (0.9,1.88)
A: WM seas.	0.42 (0.15,0.70)	0.95 (0.58,1.31)	0.60 (0.31,0.88)	0.77 (0.45,1.09)
d: WM seas.	189 (137,234)	172 (145,197)	172 (136,205)	190 (162,214)

*Table A.7.1; posterior median, 2.5% and 97.5% quantiles for model parameters for each antibiotic. * Regarding the presence of ampicillin resistant E. coli in bank voles there was at least some posterior mass throughout the whole year. Posterior quantiles cannot be sensibly defined in this context and so the posterior mode is presented.*

Table A.7.2 shows correlations between the posterior means of the subject specific random effects. Monte Carlo permutation tests against a null hypothesis of no correlation showed each of these to be significant with a p-value of less than 0.0001. We therefore conclude that there are one or more unmeasured covariates or unobserved mechanisms which have a similar effect on resistance for each antibiotic.

	ampicillin	chloramphenicol	tetracycline
chloramphenicol	0.52		
tetracycline	0.71	0.61	
trimethoprim	0.70	0.58	0.74

Table A.7.2; correlations between posterior means of the subject specific random effects for the four antibiotics.

References

Gamerman D. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing.* 1997; 7: 57-68.

Diggle P. J., Heagerty P. Liang K., and Zeger S. L. Analysis of longitudinal data (2nd ed.). 2002, *Oxford University Press*, Oxford, UK.