Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE

A Bayesian latent class mixture model with censoring for correlation analysis in antimicrobial resistance across populations

Min Zhang[1] | Chong Wang*[1,2] | Annette O'Connor[2,3]

[1]Department of Statistics, Iowa State University, Iowa, United States of America
[2]Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Iowa, United States of America
[3] Department of Large Animal Clinical Sciences, Michigan State University, Michigan, United States of America

Correspondence
*Chong Wang. Email: chwang@iastate.edu

Present Address
3413 Snedecor Hall, 2438 Osborn Dr., Ames, Iowa, U.S., 50011.

Summary

Joint emergence of antimicrobial resistance across populations has been a global threat to public health as it involves multiple complications regarding the antibiotic consumption of agriculture, antimicrobial resistance transferred from food producing animals to human, and eventually human health. In this research, we developed method to quantify the correlation in antimicrobial resistance across populations from a new perspective besides the resistance proportion: the conventionally unnoticed mean shift of the susceptible bacteria. With the proposed Bayesian latent class mixture model with censoring and multivariate normal hierarchy, we address several challenges associated with analyzing the minimum inhibitory concentration data. Application of this approach to the surveillance data from National Antimicrobial Resistance Monitoring System leads to a detection of positive correlation in the central tendency of azithromycin resistance of the susceptible populations from *Salmonella* Typhimurium across food animal and human populations. Our proposed approach is accurate and robust as validated by simulation studies. Further implementation of this Bayesian model could serve as a useful tool to indicate the co-existence of antimicrobial resistance, and potentially a need of clinical intervention.

KEYWORDS:
Correlation, Antimicrobial resistance, Minimum inhibitory concentration, Bayesian latent class mixtures, Multivariate normal hierarchy, NARMS.

1 | INTRODUCTION

1.1 | Background

Antimicrobial resistance (AMR) has been a major threat to global public health for decades[1]. Surveillance programs form a critical part of the effort to identify and control the emergence of AMR. Knowledge of emerging resistance enables actions to mitigate the spread of emergence. For example, it was through surveillance systems that the emergence of ceftiofur resistance in poultry products and humans being associated with the introduction of the product into the poultry market and the impact of removal was detected. As a consequence, governments devote substantial funds to the maintenance of AMR surveillance programs.

In 1996, the National Antimicrobial Resistance Monitoring System (NARMS) was established to help assess the consequences to human health arising from the use of antimicrobial drugs in food animal production in the United States[2]. As a collaborative

work between the Centers for Disease Control and Prevention (CDC), the U.S. Food and Drug Administration (FDA), and the United States Department of Agriculture (USDA), this national surveillance system tracks changes in the antimicrobial susceptibility of certain enteric bacteria found in ill people, retail meats, and food animals in the United States. This task was achieved by testing for the minimum inhibitory concentration (MIC), which is the lowest concentration of a particular antibiotic that will inhibit the bacteria growth. MIC is currently measured from serial dilution experiments or obtained using the whole-genome sequencing based machine learning method[3].

Currently, data about AMR observed in NARMS are reported in an integrated report. The predominant approach of quantifying antimicrobial resistance level is to use the percentage of isolates resistant to a particular antibiotic, which was defined as the number of isolates resistant to the antibiotic divided by the total number of isolates being tested[4]. However, comparative analysis of resistance across either serotypes, antibiotics, or populations are not included in the report, although such comparative analysis could provide additional insights into resistance emergence. For example, are patterns in resistance in one serotype of *Salmonella* correlated with those in other *Salmonella* serotypes? Or, are patterns in resistance observed in one *Salmonella* serotype correlated between human and food animal populations?

Existing research based on NARMS and other surveillance programs pointed out that increased antimicrobial resistance in major food-borne and enteric bacteria (e.g. *Salmonella*, *Campylobacter* has been found in both human and food animal populations worldwide[5,6,7]. While many classes of antimicrobial agents used in food animals play important roles in growth enhancement and disease prevention, they tend to raise the possibility of development of cross-resistance in human bacterial pathogens[8]. The correlation in AMR level between human and food animal populations is an essential foundation for further studies on the transmission of antibiotic resistance across the populations. Positive correlations, which are of interest for antibiotic resistance surveillance data, refers to matched increases and/or decreases in a characteristic when compared across two or more populations. For MIC data, that characteristic could be either the resistant proportion or, a more interesting and comprehensive perspective, the mean MIC. The ability to evaluate multiple aspects of surveillance data certainly maximizes the value of society's investment in such programs. In this paper, it is our focus to develop a methodology that gives accurate and robust estimation of the correlation in AMR across populations.

1.2 | Literature review

In NARMS surveillance data, isolates were classified into susceptible, intermediate, and resistant components based on their MIC relative to the breakpoints adopted from the Clinical and Laboratory Standards Institute (CLSI)[9]. With the understanding that an intermediate susceptibility to some drugs would remove this drug as a clinical option, isolates with intermediate susceptibilities could be considered as susceptible[10].

So far, the cross-population correlation in AMR has been analyzed with the popular characteristic of MIC data: the resistant proportion of an organism to an antibiotic. For example, Wegener[11] calculated the prevalence of ceftiofur resistance for retail chicken and human *Salmonella* Heidelberg isolates by using the moving average of quarterly proportion of resistance. A close AMR pattern between food supply and human was concluded in this work solely by visualization of their similar trends. A more recent example of correlation analysis in AMR level can be found in Iwamoto *et al.*[12], where Spearman rank correlation was used to examine the relationship between the annual proportion of ceftriaxone resistance among Salmonella isolates from human, retail meats, and food animals. However, percent resistant method relies on dichotomization of isolates (i.e. susceptible or resistant), hence losing information of the MIC distribution[13], and preventing monitoring of correlation in mean MIC. Additionally, breakpoints can vary across years, making direct comparison of proportions inappropriate.

Gradual movement or shift of mean MIC in susceptible population, a phenomenon referred to as MIC creep/decline, has a different focus than changes in the resistant proportion or MIC geometric mean over a period of time[14]. In addition, correlation in mean MIC in susceptible population assists the identification of emerging joint resistance patterns. However, the challenge associated with the mean MIC estimation arises due to the censorship of the observations. For example, when observing an MIC of "=8" for an organism tested by some antibiotic, it actually indicates that the true MIC is > 4, ≤ 8, and ultimately unknown. Estimation of the mean MIC based on unadjusted observations tends to be overestimated due to the upward rounding of data[15]. Another challenge related to the mean MIC estimation is the underlying population heterogeneity of non-resistant and resistant isolates. As indicated by the frequency plot of the observed MIC, it is natural to represent the true values with a bimodal distribution to reflect the two overlapping components. Craig[16] resolved the censorship issue by integrating the uncertainty of the true $\log_2$MIC values in their underlying intervals and suggested modeling with a Gaussian mixture distribution. Under a different context, the isolates may be classified into wild and non-wild components; and sometimes the non-wild component

contains more than one component, which could be satisfied thanks to the flexibility of Gaussian mixture distribution. Subsequent research on estimating the full continuous scale MIC density with the semi-parametric [17] approach was conducted under Bayesian framework. Further analysis on the mean MIC creep/decline was studied by Zhang *et al.* [18] with a linear model in the non-resistant component by a fully parametric Bayesian method. Jaspers *et al.* [19] analyzed the joint distribution of MIC data on multiple antibiotics with Bayesian estimation of multivariate Gaussian mixtures, from which inference about the correlation between drug resistances within one year could be drawn. However, the multivariate means of MIC of each component were assumed as fixed from year to year, which again displayed ignorance to the potential changes in the mean MIC for the components where isolates have formed resistance.

Evidence was discovered in the cross-population correlation in the prevalence of resistance [12,20]. Unlike the increase in the central tendency of the MIC for the susceptible population, an increase in the prevalence of resistance would most likely indicate the dissemination of less susceptible clone(s) [14]. Another important perspective is to look into the cross-population correlation in the movement of the MIC distribution, but to our knowledge it has not been studied yet. In this research, it is our interest to fill in the gap by estimating correlation in the mean MIC through a Bayesian framework with a multivariate normal hierarchy linking different populations.

1.3 | Contributions

In this paper, a Bayesian latent class mixture model with censoring and multivariate normal hierarchy was proposed to determine the correlation in the unnoticed movement of the mean MIC of the susceptible isolates across two different populations. The proposed model was applied to the datasets from National Antimicrobial Resistance Monitoring System (NARMS) obtained from CDC (human data) and USDA (food producing animal data). In order to get a good estimation of the correlation in mean MIC across population, we considered monthly means of $\log_2$MIC in the non-resistant component and added a bivariate normal mixture model in the hierarchical structure to evaluate the correlation in $\log_2$MIC between populations. In the example of *Salmonella enterica* Typhimurium tested by azithromycin in human and food animal datasets, a significantly positive correlation in the central shift of non-resistant component was detected, suggesting entry of AMR to human through food animals. Simulations on the proposed method were carried out for strong, moderate, and mild correlations, to show the precision and robustness of the new approach. Applications of our model to other organisms or antibiotics across populations could serve as evidences of emerging or declining joint resistance, adding value to the AMR surveillance programs.

2 | METHODS

Our methodology estimates the correlation of the mean MIC for susceptible components across populations. In AMR surveillance programs, we observe MIC as the measurement of the antibiotic susceptibility of the isolates towards drugs. Two-fold serial dilution data, like the MIC observations, are commonly analyzed with base 2 logarithm transformation. At the upper limit of the serial dilution where the growth of the bacteria could not be inhibited even with the highest drug concentration, the MIC is right censored. The resistant components are usually highly right censored, meaning there is not enough information to extrapolate from. Also, the conventionally unnoticed MIC creep/decline takes place in the non-resistant components. As a consequence, it is necessary to focus the correlation analysis for the non-resistant components of bacteria.

In this section, a Bayesian latent class mixture model is introduced, and a straightforward analysis which we call the "naïve method" throughout this paper is brought up as a comparison.

2.1 | Model notations and assumptions

To fill the gap of estimating the cross-population correlation in the mean MIC of the susceptible component, we proposed a hierarchical Bayesian latent class mixture model and managed to address the challenges from the censored nature and the underlying distribution of the MIC data. To account for the censored nature, each observed MIC value was assumed to represent an interval where the true MIC value lies in, as depicted in Figure 1.

The notation used is as follows:

- $y^*_{s,i,j}$: the observed value of $\log_2$MIC for isolate j in month i from population s.
- $y_{s,i,j}$: the latent value of $\log_2$MIC for isolate j in month i from population s.

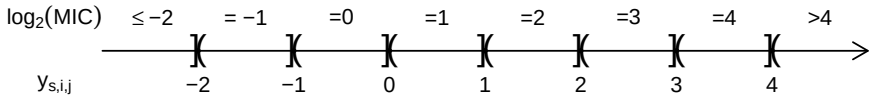

FIGURE 1 Relationship between the observed and latent $\log_2$MIC.

TABLE 1 Conversion table between the observed and latent $\log_2$MIC..

Observed $log_2MIC: y^*_{s,i,j}$	Censor type	Latent $log_2MIC: y_{s,i,j} \in (l_{s,i,j}, u_{s,i,j})$
$\leq a$	Left censored	$y_{s,i,j} \in (-\infty, a]$
$= a$	Interval censored	$y_{s,i,j} \in (a-1, a]$
$> a$	Right censored	$y_{s,i,j} \in (a, \infty)$

- $l_{s,i,j}$, $u_{s,i,j}$: the lower bound and upper bound of the latent true value $y_{s,i,j}$, and $y_{s,i,j} \in (l_{s,i,j}, u_{s,i,j})$. Conversion between $y^*_{s,i,j}$ with its sign and the interval is defined in Table 1.
- $c_{s,i,j}$: the latent indicator of the bacterial component from which the isolate j in month i from population s was drawn. $c = 0, 1$ represents non-resistant and resistant component, respectively.

The subscript $s = 1, 2$ represents the two populations whose correlation in the mean MIC are of interest to us; $i = 1, 2, ..., I$ is the time index, where I is the total number of months; $j = 1, 2, ..., n_{s,i}$ is the isolate index, where $n_{s,i}$ is the total number of observations for population s in month i. The support of the indices remain the same throughout the model description unless otherwise specified. In our example of Section 3, the populations across which the correlation was estimated were human and food animals.

2.2 | Model description

To study the correlation in the mean $\log_2$MIC in the susceptible component across populations, the Bayesian latent class mixture model with censoring and multivariate normal hierarchy is introduced in this section and it includes two levels. In the first level, the latent true values of $\log_2$MIC from each population within each month are modeled by a Gaussian mixture distribution with Bernoulli distributed weights. This approach is motivated by the bimodal distribution of the frequency plots of the $\log_2$MIC observations. In the hierarchical structure, a bivariate normal distribution (BVN) was imposed on the monthly means of $\log_2$MIC for the paired populations. The correlation parameter in the BVN distribution answers the question of whether the correlation in the mean MIC of the susceptible isolates exists across populations, hence is the most interesting parameter of the whole model.

The construction procedure of the proposed Bayesian hierarchical model that details the above structure is articulated below. For $s = 1, 2$; $i = 1, 2, ..., I$; and $j = 1, 2, ..., n_{s,i}$:

$$c_{s,i,j} | p_{s,i} \overset{ind}{\sim} Ber(p_{s,i}), \tag{1}$$

$$y_{s,i,j} | c_{s,i,j}, \beta_{0,s,i}, \beta_{1,s,i}, \sigma^2_{0,s}, \sigma^2_{1,s} \overset{ind}{\sim} \begin{cases} N(\beta_{0,s,i}, \sigma^2_{0,s}), c_{s,i,j} = 0 \\ N(\beta_{1,s,i}, \sigma^2_{1,s}), c_{s,i,j} = 1 \end{cases}. \tag{2}$$

The data model indicates that for population s, the true $\log_2$MIC of isolate j in month i is drawn from the resistant component with probability $p_{s,i}$, and drawn from the non-resistant component with probability $1 - p_{s,i}$. Given that isolate j belongs to the

non-resistant component, the true value of its $\log_2$MIC follows a normal distribution with mean $\beta_{0,s,i}$ and variance $\sigma^2_{0,s}$. Similarly, if that isolate j belongs to the resistant component, its $\log_2$MIC follows a normal distribution shifting to the right compared with the non-resistant normal curve. The variance parameters of the Gaussian mixture model vary across component and population. But observations from the same population and component are expected to have the same spread. In the data model, the isolates are classified by a "soft" probabilistic threshold without reliance on the pre-determined breakpoints.

At this point, the data model depicts the distribution of $\log_2$MIC values separately for each month and each population. In fact, different sampling conditions (e.g. institution, technician, time, etc.) could cause heterogeneity in the measurement of the antibiotic susceptibility of the isolates. Therefore, it could be helpful to borrow information and integrate uncertainty in the mean $\log_2$MIC values across months via the hierarchical structure. In order to find evidence of association across populations, the model parameters from (1) and (2) are joined together through multivariate normal distributions:

$$\text{logit}(p_{s,i}) = \log\left(\frac{p_{s,i}}{1-p_{s,i}}\right) = \alpha_{s,i}, \tag{3}$$

$$\boldsymbol{\alpha_i} \overset{iid}{\sim} MVN(\boldsymbol{\theta}, \boldsymbol{\Omega}), \text{ where } \boldsymbol{\theta} = \begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}, \boldsymbol{\Omega} = \begin{pmatrix}\tau_1^2 & \eta\tau_1\tau_2\\ \eta\tau_1\tau_2 & \tau_2^2\end{pmatrix}; \tag{4}$$

$$\boldsymbol{\beta_{0,i}} \overset{iid}{\sim} MVN(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}),\ \boldsymbol{\mu_0} = \begin{pmatrix}\mu_{0,1}\\ \mu_{0,2}\end{pmatrix}, \boldsymbol{\Sigma_0} = \begin{pmatrix}\gamma_{0,1}^2 & \rho_0\gamma_{0,1}\gamma_{0,2}\\ \rho_0\gamma_{0,1}\gamma_{0,2} & \gamma_{0,2}^2\end{pmatrix}; \tag{5}$$

$$\boldsymbol{\beta_{1,i}} \overset{iid}{\sim} MVN(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}),\ \boldsymbol{\mu_1} = \begin{pmatrix}\mu_{1,1}\\ \mu_{1,2}\end{pmatrix}, \boldsymbol{\Sigma_1} = \begin{pmatrix}\gamma_{1,1}^2 & \rho_1\gamma_{1,1}\gamma_{1,2}\\ \rho_1\gamma_{1,1}\gamma_{1,2} & \gamma_{1,2}^2\end{pmatrix}. \tag{6}$$

Expressions (3) to (6) are the model's hierarchical level, where the two populations are linked through a vector at a common time period and are modeled by bivariate normal distributions. In expression (4), $\boldsymbol{\alpha_i}$ refers to the vector of the resistant proportion of the two populations in month i after logit transformation: $(\alpha_{s=1,i}, \alpha_{s=2,i})^\intercal$; boldface Greek letters are used to represent vectors. The vector of transformed proportions of resistance follows a multivariate normal distribution with mean vector $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Omega}$; capitalized boldface Greek letters are used to represent matrices. Since the amount of populations we study once at a time is two, the multivariate normal distribution is simply bivariate. In the case of S populations where $S > 2$, the hierarchical structure could be extended to S-dimensional normal distributions.

It is important to remember that the objective of this paper is to find evidence of cross-population correlation in the monthly mean $\log_2$MIC, not considering the resistant isolates due to their heavy censorship. Expression (5) is about the non-resistant component: $\boldsymbol{\beta_{0,i}} = (\beta_{0,s=1,i}, \beta_{0,s=2,i})^\intercal$, the vector of the mean $\log_2$MIC of two populations in month i is assumed to follow a BVN centered at $(\mu_{0,1}, \mu_{0,2})^\intercal$; the parameters of the standard deviation $\gamma_{0,1}$ and $\gamma_{0,2}$ reflect the spread of the monthly mean $\log_2$MIC on the two dimensions of population. The correlation parameter ρ_0 reflects the degree to which the $\log_2$MIC means of the non-resistant isolates are linearly related across the populations, hence is the key to our research; $\rho_0 \in [-1, 1]$. An estimation of ρ_0 with small absolute value close to 0 indicates no or rather weak correlation in AMR across populations, while a large estimation close to 1 implies a strong positive correlation in AMR. When the latter scenario happens, our result could serve as an evidence of co-existence of emerging or declining AMR across populations. Similar with (5), expression (6) is an analog to the resistant component, but is of less interest to our study.

Consequently, the parameter space, denoted as Θ, is $\left(\boldsymbol{\sigma_0}, \boldsymbol{\sigma_1}, \boldsymbol{\theta}, \boldsymbol{\Omega}, \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}, \boldsymbol{\mu_1}, \boldsymbol{\Sigma_1}\right)$. To express the joint likelihood of the observations, we collapse all the observed and latent $\log_2$MIC in vectors $\boldsymbol{y}^*$ and $\boldsymbol{y}$, respectively. Let $f(\cdot|\cdot)$ be generic expression of the conditional density. Then the joint likelihood of the observed $\log_2$MIC is written out as the joint likelihood of the latent $\log_2$MIC integrated over the intervals where the discrete observations lay in.

$$f(\boldsymbol{y}^*|\Theta) = \int_{l_{2,I,n_{2,I}}}^{u_{2,I,n_{2,I}}} \cdots \int_{l_{1,1,1}}^{u_{1,1,1}} f(\boldsymbol{y}|\Theta)\, dy_{1,1,1} \cdots dy_{2,I,n_{2,I}}. \tag{7}$$

As indicated by expressions 4 to 6, with the parameter space Θ given, the data parameters and $y_{s,i,j}$ produced in month i are independent with those in month i', where $i \neq i'$. Hence, we have the joint likelihood of the latent $\log_2$MIC written as

$$f(\boldsymbol{y}|\Theta) = \prod_{i=1}^{I} f(\boldsymbol{y_i}|\Theta), \tag{8}$$

where $\boldsymbol{y} = (\boldsymbol{y_1}, \cdots, \boldsymbol{y_I})^\intercal$, and

$$f(\boldsymbol{y_i}|\Theta) = \int_{(\boldsymbol{\beta_{0,i}}, \boldsymbol{\beta_{1,i}}, \boldsymbol{\alpha_i})} \cdots \int f(\boldsymbol{y_i}|\boldsymbol{\alpha_i}, \boldsymbol{\beta_{0,i}}, \boldsymbol{\beta_{1,i}}, \boldsymbol{\sigma_0^2}, \boldsymbol{\sigma_1^2}) f(\boldsymbol{\alpha_i}|\boldsymbol{\theta}, \boldsymbol{\Omega}) f(\boldsymbol{\beta_{0,i}}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}) f(\boldsymbol{\beta_{1,i}}|\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})\, d\beta_{0,s_1,i} \ldots d\alpha_{s_2,i} \tag{9}$$

The $f(\boldsymbol{y_i}|\boldsymbol{\alpha_i}, \boldsymbol{\beta_{0,i}}, \boldsymbol{\beta_{1,i}}, \boldsymbol{\sigma_0^2}, \boldsymbol{\sigma_1^2})$ in Equation (9) corresponds to data model level in expressions (1) - (2), and can be expressed as the product of the likelihoods of all $y_{s,i,j}$ for $s = 1, 2$; $j = 1, ..., n_{s,i}$. The $f(\boldsymbol{\alpha_i}|\boldsymbol{\theta}, \boldsymbol{\Omega})$, $f(\boldsymbol{\beta_{0,i}}|\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0})$, and $f(\boldsymbol{\beta_{1,i}}|\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$ in Equation (9) correspond to the hierarchical level of the model. According to the Bayes rule that $f(\Theta|\boldsymbol{y}^*) \propto f(\boldsymbol{y}^*|\Theta) \times f(\Theta)$, inference of the parameters needs to be drawn from the posterior distribution. The choice of prior distributions $f(\Theta)$ is explained in the following subsection.

2.3 | Prior distribution

Conjugate priors were assigned to the hyper mean parameters. We chose diffuse Gaussian priors for $\theta_1, \theta_2 \sim N(0, 10000)$ to reflect our lack of knowledge in the resistant proportion. But we chose moderately informative priors for $\mu_{0,s}, \mu_{1,s} \sim N(0, 100)$ since we know that the magnitude of $\log_2$MIC could hardly be smaller than -10 or greater than 10. The restriction $\mu_{0,s} < \mu_{1,s}$ $(s = 1, 2)$ was applied, because the non-resistant mean should always be smaller than the resistant mean. A default uniform prior was assigned to the data standard deviation $f(\boldsymbol{\sigma_c}) \propto 1$ $(c = 0, 1)$ in their positive spaces.

The selection of the prior for the covariance matrices is tricky. We adopted the separation strategy[21] over the popular inverse Wishart, since the latter distribution has a tendency to bias the posterior correlation downward[22]. Each of the covariance matrices was decomposed into a correlation matrix sandwiched by the scale matrices. The correlation matrix has 1's on its diagonal and correlation parameters on its off-diagonal positions; the scale matrix has the standard deviation on the diagonal. For example, the covariance matrix of the monthly means of $\log_2$MIC in the non-resistant component is written as:

$$\boldsymbol{\Sigma_0} = \begin{pmatrix} \gamma_{0,1}^2 & \rho_0\gamma_{0,1}\gamma_{0,2} \\ \rho_0\gamma_{0,1}\gamma_{0,2} & \gamma_{0,2}^2 \end{pmatrix} = \begin{pmatrix} \gamma_{0,1} & 0 \\ 0 & \gamma_{0,2} \end{pmatrix} \begin{pmatrix} 1 & \rho_0 \\ \rho_0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_{0,1} & 0 \\ 0 & \gamma_{0,2} \end{pmatrix} := \Gamma_0 \mathrm{R}_0 \Gamma_0. \tag{10}$$

Following the recommendation from the Stan development team[23], we used a prior for the correlation matrix called LKJ distribution[24], where $f(\mathrm{R}_0) \propto |\mathrm{R}_0|^{\nu-1}$; $\nu > 0$. By choosing LKJ $(\nu = 1)$ for the correlation matrices decomposed from Ω, Σ_0, and Σ_1, the densities of the priors are uniform over correlation matrices of their corresponding dimension d (in our case, $d = 2$), reflecting our *a priori* lack of knowledge for the correlations. Weakly informative half-Cauchy priors were assigned to the scale parameters $\tau_s, \gamma_{0,s}, \gamma_{1,s} \sim \text{Cauchy}^+(0, 2)$; $s = 1, 2$. Here, all prior distributions were assumed independent.

2.4 | Naïve calculation of correlation

As a comparison with the proposed Bayesian method, we introduce the "naïve method", a simple and straightforward frequentist approach. The naïve analysis for mean $\log_2$MIC ignores the nature of censoring of MIC data, and calculates the arithmetic average of $\log_2$MIC for the non-resistant isolates within each month. For example, if an observed MIC value was = 8 and was categorized as non-resistant according to the CLSI standards, then we treated $\log_2\text{MIC} = \log_2(8) = 3$ as the true $\log_2$MIC value and therefore used it for the naïve mean calculation for that month. For each population s, we could obtain a vector of monthly averages in the non-resistant component over I months. The Pearson correlation coefficient between the vectors of means was used to describe the strength of correlation in the mean $\log_2$MIC of the susceptible isolates between the two populations. Hence, the name "naïve method" comes from the fact that this frequentist calculation does not take into account the censorship or the underlying distribution of the MIC data.

In the following application and simulation sections, we will implement both the Bayesian and the frequentist methods to see how they perform for estimating the cross-population correlation in the mean MIC of susceptible components.

3 | REAL DATA ANALYSIS

3.1 | Data description and manipulation

3.1.1 | Centers for Disease Control and Prevention (CDC) human data

The human population of NARMS was launched in 1996 within the framework of CDC's Emerging Infections Program and the Food-borne Diseases Active Surveillance Network (FoodNet)[4]. *Salmonella* isolates, as the largest genus type among the four bacteria in NARMS (others are *Campylobacter*, *Shigella*, and *Escherichia coli* O157), were reported with year of collection, serotype, MIC value tested against multiple antibiotics, the test conclusion (resistant or not), etc. We limited our analysis to an important serotype: *S.* Typhimurium. It has 5398 isolates, accounting for 14.1% of the 38311 *Salmonella* isolates, ranking second to enteritidis (16.3%), and was tested for MIC for 28 antibiotics collected since 1996 till now.

3.1.2 | U.S. Department of Agriculture (USDA) food producing animal data

The food animal component of NARMS includes data from 1997 to 2015 with monitoring of Salmonella and later expanded to Campylobacter (1998), E. coli (2000), and Enterococcus (2003)[25], in which Salmonella also forms the largest proportion of the data. Isolates are recovered from samples obtained at federally inspected slaughter and processing plants[4]. The information related to *Salmonella* Typhimurium included the MIC for 23 antibiotics, the month and year of collection, the host the isolate was obtained from, etc, but no test conclusion. There were 1161 *S.* Typhimurium isolates, accounting for 3.0% of the 38867 *Salmonella* isolates, ranking third after *Salmonella* Kentucky (9.3%) and *Salmonella* Enteritidis (4.1%).

3.1.3 | Data manipulation

We selected *S.* Typhimurium treated by azithromycin (AZI) as an example for illustration. AZI, a clinically important macrolide antibiotic, is used to treat a wide variety of bacterial infections, and is often for the treatment of nontyphoidal *Salmonella* (NTS) when treatment is indicated[26]. According to the NARMS integrated report of 2015[27], AZI use for NTS is increasing, likely caused by the concerns about resistance to fluoroquinolones (e.g., ciprofloxacin)[28]. In livestock, macrolides are a commonly used antibiotic for treatment and control of disease, especially in cattle and swine, where they are highly effective for common diseases such as respiratory disease.

The pairs of the mean $\log_2$MIC during common time periods between human's and food animals' susceptible components allow us to calculate their cross-population correlation. Using monthly means of $\log_2$MIC, instead of yearly means, provides us with more pairs of data, hence it is beneficial for correlation calculation. For this reason, we acquired from CDC the month of isolate collection which is not publicly available, and could not be shared with data due to the CDC privacy constraints.

By the time of this work, NARMS data after year 2017 are preliminary and the isolate collection and/or testing are still in progress, thus not included in our analysis. Since the AZI test for Salmonella started officially in 2011[27], the relevant trial samples occurred before 2011 were removed; that is a removal of 25 human isolates and 20 food animal isolates collected in 2008. For the cases where the month information or the MIC results of AZI testing are missing, the isolates were excluded from our analysis. After such elimination, 1333 human isolates from 2011 to 2017 and 572 food animal isolates from 2011 to 2015 remained, leading to 60 pairs of monthly mean of $\log_2$MIC for the non-resistant component spanning from January 2011 to December 2015. On average, there were 16 human observations and 10 food animal observations per month during this five-year period. The scatter plot of the monthly arithmetic mean $\log_2$MIC of non-resistant isolates is displayed in Figure 2 for human and food animal populations from January 2011 to December 2015, where a moderate cross-population correlation in mean $\log_2$MIC could be seen.

In the CDC human data, the test conclusion for *Salmonella* AZI used the cutoff of $>16\,\mu g/ml$ between 2011 and 2015, and $\geq 32\,\mu g/ml$ for the other years. In the USDA food animal data, Salmonella isolates tested by AZI were right censored at $>16\,\mu g/ml$ consistently through all years, but the test conclusions are not included in the dataset. We adopted $>16\,\mu g/ml$ as the cutoff rule for our example data during 2011 to 2015, even though it is stated in the NARMS integrated report[27] that Salmonella isolates from humans, retail meats, and food animals used the CLSI investigational breakpoint of $\geq 32\,\mu g/ml$ in order to determine susceptibilities to AZI. This is an example of inconsistent breakpoint and dilution spectrum. Based on the data manipulation described above and the breakpoint of $>16\,\mu g/ml$, there was only one azithromycin-resistant isolate in each of the two populations during this period. This is a very low level of resistance, leaving the analysis based on resistant proportion unreliable. For situations like this, assessing the correlations in the mean MIC of the non-resistant isolates (major component) could add useful information for the decisions makers.

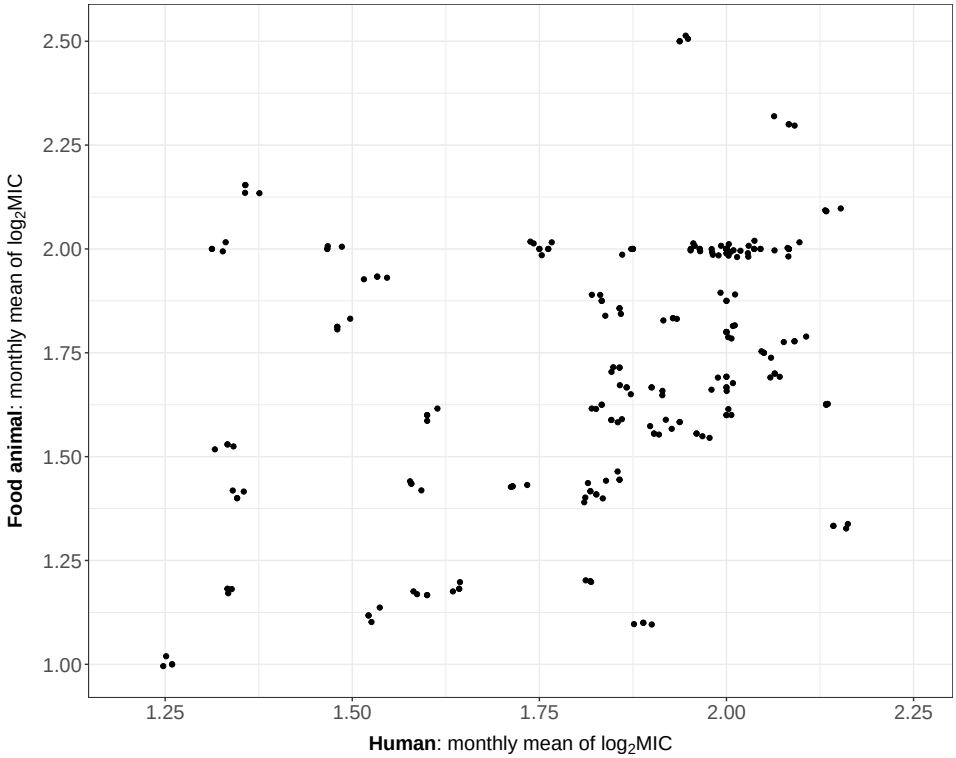

FIGURE 2 Scatter plot with jittering of the monthly arithmetic mean $\log_2$MIC of non-resistant *Salmonella* Typhimurium isolates tested by azithromycin in human and food animal populations between January 2011 and December 2015.

3.2 | Implementation

To draw inference from the proposed model, a Bayesian analysis of Markov Chain Monte Carlo (MCMC) was conducted in R environment (version 3.3.5) with package *rstan*[29]. To sample effectively from the posterior $f(\Theta|\boldsymbol{y}^*)$ that has no closed form, No-U-Turn Sampler (NUTS)[30], an extension of the powerful Hamiltonian Monte Carlo (HMC), was implemented; it is also the default and preferred sampling algorithm in Stan. The Stan script is provided in Supplementary material 8. All relevant R scripts, including cleaning of the NARMS data, model construction, model implementation, and results visualization, can be found on Github [1].

The choice of the initial values of the MCMC was based upon the selected example dataset of *S.* Typhimurium tested by AZI from NARMS. For population s, the initial values of the variances $\sigma^2_{c,s}$ in equation (2) were calculated by the variance of all observed $\log_2$MIC values in component c, regardless of the month the isolate was obtained. The initial value of the monthly resistant proportion after logit transformation for population s ($\alpha_{s,i}$) was calculated by dividing the number of resistant isolates with the total number of isolates in that month followed by logit transformation. When all isolates appeared to be under or above the breakpoint in some months, a very small proportion was added to 0 or deducted from 1, to make sure $\alpha_{s,i} = \log(\frac{p_{s,i}}{1-p_{s,i}})$ are finite values. And the starting point of the monthly mean of $\log_2$MIC for population s in component c ($\beta_{c,s,i}$) was calculated as the arithmetic average of the observed $\log_2$MIC in the corresponding subset of data. If for some particular population in some months, one component has no observation, then they were imputed by the average of the other monthly means in the same component of that population.

The initial values of the above $\alpha_{s,i}$, $\beta_{c=0,s,i}$, and $\beta_{c=1,s,i}$ can be arranged as $I \times 2$ data frames with the row representing month and the column representing population. Their means over the time index (i.e. column means of the data frames) were used as the initial values of the bivariate mean vectors $\boldsymbol{\theta}$, $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ in equations (4), (5) and (6). Similarly, their standard deviations over month (i.e. column standard deviations of the data frames) were assigned as the initial values of the scale parts of matrices $\boldsymbol{\Omega}$, $\boldsymbol{\Sigma_0}$ and $\boldsymbol{\Sigma_1}$. And the Pearson correlations across populations were plugged into the correlation matrices as starting values.

[1] Github repository: https://github.com/MinZhang95/AMR-Correlation

TABLE 2 Correlations in non-resistant $\log_2$MIC across populations for Salmonella enterica typhimurium isolates estimated by frequentist and Bayesian methods.

Population 1	Population 2	Bayesian estimation	CI of Bayesian method	Frequentist estimation
human	animal	0.4600	(0.1017, 0.7583)	0.3688

For the MCMC, we conducted in three chains simultaneously with ten thousand iterations and 50% of the length was burn-in. We used the potential scale reduction factor of Gelman and Rubin[31] to assess convergence of the three chains. Additionally, we calculated the number of effective samples for each sampled parameter to ensure reasonable accuracy in the tails of the posterior distribution. The point estimates of the model parameters were determined by the means of the posterior draws after the burn-in sessions. The ends of the 95% credible intervals (CI) were obtained from the 2.5th and 97.5th percentiles of the posterior distributions. The results were summarized in the following Result Section.

4 | RESULT

The correlation in the mean $\log_2$MIC of the non-resistant *S.* Typhimurium isolates tested by azithromycin across the human and food animal populations was estimated through the application of the proposed Bayesian latent class mixture model with censoring and multivariate normal hierarchy on the NARMS datasets from January 2011 to December 2015. The Bayesian estimation of the correlation parameter, its 95% credible interval, and the frequentist estimation through the naïve method are listed in Table 2.

The point estimation of 0.46 together with its 95% credible interval (0.1017, 0.7583) from the Bayesian approach indicate that there exists a significant positive correlation in the mean $\log_2$MIC between the susceptible isolates in the human population and those in the food animal population. It implies that the conventionally unnoticed MIC creep occurred in these two populations moved towards similar directions. Interestingly, we found that the Bayesian point estimation is greater than the frequentist estimate by 0.0912. It is conjectured that the estimation from the proposed Bayesian approach is more accurate compared with the frequentist method, because the former one takes into account the censorship issue and makes the full use of the data information. To validate this conjecture, a simulation study was conducted and the performance of the two methods were compared.

5 | SIMULATION

In this section, a simulation study was conducted to assess the performance of the proposed hierarchical Bayesian latent class mixture model and the frequentist naïve method, by comparing their estimation results with the underlying data generators.

In the following description of data simulation, we denote the known model parameters ($\boldsymbol{\sigma_c}$, $\boldsymbol{\theta}$, $\boldsymbol{\mu_c}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Sigma_c}$; $c = 0, 1$) with a "hat" on top of the Greek letters. These parameters were given the pre-determined values that were estimated from the application of the Bayesian approach on the human-food animal example in Section 3, so that the simulated datasets are close to what we might observe in the real world. In particular, we are most interested in the correlation parameter $\hat{\rho}_0$, which is the correlation part of $\boldsymbol{\Sigma_0}$, was given the true value of 0.46. In each simulated dataset, there are 60 months, 16 human isolates and 10 animal isolates per month; this is to ensure a similar size of observations with the real dataset.

For $s = 1, 2$; $i = 1, 2, ..., I$; $j = 1, 2, ..., n_{s,i}$; where $I = 60$, $n_{s=1,i} = 16$ and $n_{s=2,i} = 10$:

(i) Generate $\boldsymbol{\alpha_i} = \left(\alpha_{s=1,i}, \alpha_{s=2,i}\right)^{\intercal} \overset{iid}{\sim} MVN_2(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Omega}})$.

(ii) Convert $\boldsymbol{\alpha_i}$ to $\boldsymbol{p_i}$, the vector of monthly proportion of resistant isolates, through expit transformation (inverse of logit): $\boldsymbol{p_i} = \left(p_{s=1,i}, p_{s=2,i}\right)^{\intercal} = \frac{1}{1+exp^{-\alpha_i}}$.

(iii) Generate the vector of monthly mean $\log_2$MIC in the non-resistant and resistant population, respectively:

TABLE 3 Simulation results based on different true values of correlations, estimated by frequentist and Bayesian methods.

True $\hat{\rho}_0$ †	Method	Mean estimation bias	Root of mean squared error
0.80	Bayes	−0.0219	0.0927
	Naïve	−0.1972	0.2134
0.46	Bayes	−0.0025	0.1461
	Naïve	−0.1002	0.1510
0.30	Bayes	+0.0115	0.1621
	Naïve	−0.0561	0.1360

†The parameter $\hat{\rho}_0$ is the correlation in the non-resistant $\log_2$MIC across populations.

$$\boldsymbol{\beta_{0,i}} = \left(\beta_{c=0,s=1,i}, \beta_{c=0,s=2,i}\right)^{\intercal} \overset{iid}{\sim} MVN_2(\hat{\boldsymbol{\mu}}_{\mathbf{0}}, \hat{\boldsymbol{\Sigma}}_{\mathbf{0}});$$
$$\boldsymbol{\beta_{1,i}} = \left(\beta_{c=1,s=1,i}, \beta_{c=1,s=2,i}\right)^{\intercal} \overset{iid}{\sim} MVN_2(\hat{\boldsymbol{\mu}}_{\mathbf{1}}, \hat{\boldsymbol{\Sigma}}_{\mathbf{1}}).$$

(iv) Generate the latent variable of class indicator for each population s and month i: $c_{s,i,j} \overset{ind}{\sim} Ber(p_{s,i})$.

(v) Generate the latent value of $\log_2$MIC for each population s and month i: $y_{s,i,j} \overset{ind}{\sim} \begin{cases} N(\beta_{c=0,s,i}, \sigma^2_{0,s}), c_{s,i,j} = 0 \\ N(\beta_{c=1,s,i}, \sigma^2_{1,s}), c_{s,i,j} = 1 \end{cases}$.

The $\log_2$MIC obtained from step (v) are continuous quantities drawn from a Gaussian mixture model and need to be censored by following the conversion rule described in Table 1. Step (vi) describes the censoring operation for the example of *S.* Typhimurium tested by azithromycin.

(vi) Convert the latent value of $\log_2$MIC $y_{s,i,j}$ to the censored value $y^*_{s,i,j}$: $y^*_{s,i,j} = \begin{cases} -2, y_{s,i,j} \leq -2 \\ \lceil y_{s,i,j} \rceil, -2 < y_{s,i,j} \leq 4 \\ 4, y_{s,i,j} > 4 \end{cases}$; $\lceil \cdot \rceil$ represents the ceiling of a number.

According to the dilution spectrum we observed from the application data, the most susceptible isolates were from human samples with MIC = 0.5 µg/ml. It indicates that if $y_{s,i,j} \leq \log_2(0.5) - 1 = -2$, it will be left censored as $y^*_{s,i,j} \leq -2$; that is $l_{s,i,j} = -\infty$ and $u_{s,i,j} = -2$. The most resistant isolates came from human and food animal samples with MIC greater than 16 µg/ml. Similarly, it means that if $y_{s,i,j} > \log_2(16) = 4$, it will be right censored as $y^*_{s,i,j} > 4$; that is $l_{s,i,j} = 4$ and $u_{s,i,j} = +\infty$. A latent value of $\log_2$MIC in between -2 and 4 will be interval censored with its upper bound being the nearest integer to the right and lower bound being the nearest integer to the left.

End of simulation.

By repeating the above procedure for 100 times, we obtained 100 simulated datasets, each of which was estimated with both the proposed Bayesian model and the naïve method. The two approaches were assessed and compared by the mean bias of ρ_0, and its root of mean squared error (RMSE). These two metrics are the indicators of a model's precision and robustness by measuring the average bias to the truth and the average deviation around the truth. Since it is important to assess the model performance under different strengths of correlation, we also conducted simulations for $\hat{\rho}_0 = 0.8$ (strong correlation) and $\hat{\rho}_0 = 0.3$ (weak correlation) with the other settings unchanged. The mean estimation bias and RMSE from the simulations can be found in Table 3.

6 | DISCUSSION

When estimating the cross-population correlation in the mean $\log_2$MIC of the susceptible isolates, the proposed Bayesian latent class mixture model shows advantages compared with the frequentist method. In the simulation result (Table 3), the absolute value of the mean estimation bias of the Bayesian method is fairly small compared with the scale of $\log_2$MIC data, and is much

smaller than those of the frequentist method. It indicates that the proposed model has high accuracy for estimating correlation, especially for strong and moderate correlation. By comparing the RMSE, we can find that the Bayesian method is rather robust for the strong correlation case. For moderate and weak correlations, the robustness of the two methods are comparable since their RMSE are of similar scale.

According to Annis and Craig [15], neglecting the censored nature of the MIC data in the naïve method leads to overestimation in the mean of $\log_2$MIC, while they did not investigate its impact in correlation. Intuitively, when the two populations are positively correlated in the monthly mean of the latent $\log_2$MIC, the inherent correlation could be dwarfed by the upward rounding procedure, as shown by the underestimation of the naïve method in the simulation study. This points to the importance of the censorship adjustment included in the Bayesian model.

Another superiority of the Bayesian latent class mixture model lies in the independence of the methods from the the breakpoint value. The naïve method subsets the dataset by relying on the NARMS-established breakpoints and considering only the non-resistant isolates under the cutoffs. It is not a rare case where the CLSI breakpoints were updated for some antimicrobial agents during the NARMS history. For example, CDC Human's Salmonella resistance to streptomycin was adjusted from $\geq 64\,\mu g/ml$ to $\geq 32\,\mu g/ml$ in 2014, hence the current breakpoint could not be applied to the previous years due to limited concentrations tested [32]. Apart from this, Mouton [33] argued that there is a major difference between clinical and microbiological breakpoints, where the former is an indicator for clinical success while the latter is for detecting resistant populations. Therefore, despite the simplicity of the naïve method, it has obvious defects compared with the proposed hierarchical Bayesian model.

It is within our expectation that the mean estimation biases of the Bayesian approach are negative for relatively large correlation cases. The prior distribution for the correlation matrix R_0 was chosen as LKJ (1) whose marginal density of every correlation is Beta $\left(\frac{d}{2}, \frac{d}{2}\right)$ distribution on (-1, 1), where $d \times d$ is the dimension of the correlation matrix [24]. It implies that this prior distribution may lead to some shrinkage of correlation estimation towards zero. Such a property does not form a concern for our research question as the accuracy of the Bayesian model is quite promising. When the number of populations extends to higher dimensions ($d > 10$), the LKJ prior concentrates mass near zero correlations and could be problematic [22].

In summary, the proposed Bayesian latent class mixture model addresses the challenges associated with the MIC data analysis, and provides accurate and robust estimation of correlation in mean $\log_2$MIC of susceptible isolates across populations. It performs especially well for large underlying correlation cases, which is crucial to public health as co-resistance of antibiotics across populations could signal needs of remedial actions.

7 | CONCLUSION AND FUTURE WORK

In this work, we proposed a Bayesian latent class mixture model with censoring and multivariate normal hierarchy for inference of the correlation in antimicrobial resistance across populations. Besides the cross-population correlation on the resistance proportion which has been the focus in the existing literature, we also targeted the correlation in the mean MIC of the susceptible populations. By applying the model to the NARMS data, we detected positively correlated mean MIC shift in the non-resistant component between human and food animal populations. This means that for the susceptible isolates, the monthly MIC means in human with that in food animal are changing more in the same direction than in different directions. This result indicates a possibility for the introduction of azithromycin resistance of *S.* Typhimurium in human through food animals.

In future work, we are going to study several extended questions about the correlated antimicrobial resistance. One aspect is to evaluate the correlation in the mean MIC across antibiotics, which could help provide evidence of co-resistance among drugs. Another aspect is to see whether the mean MIC for an antibiotic in two serotypes are positively correlated. If the answer is yes and the serotypes possess different resistance gene, this suggests the genes share a similar mechanism of inducing resistance. The ability to assess correlations from more perspectives would enable increased information to be extracted from the AMR surveillance programs, and create further value to the public health.

ACKNOWLEDGMENTS

We appreciate the Centers for Disease Control and Prevention (CDC) for providing us with the month information of the isolate collection in the National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS) data.

Author contributions

Min Zhang
Formal analysis, Methodology, Software, Validation, Visualization, Writing original draft, Writing review and editing.
Chong Wang
Conceptualization, Funding acquisition, Methodology, Supervision, Writing review and editing.
Annette O'Connor
Conceptualization, Funding acquisition, Project administration, Supervision, Writing original draft, Writing review and editing.

Financial disclosure

None reported.

Conflict of interest

The authors declare no potential conflict of interests.

8 | STAN CODE OF THE PROPOSED BAYESIAN MODEL

```
functions {
  // Define log probability density function for an interval censored normal
  real intervalCensoredNormal_lpmf(int y, real mu, real sigma, vector breaks) {
    real p;
    if (y == 0) { // from -inf to the first break
      p = normal_lcdf(breaks[1] | mu, sigma);
    } else if (y == rows(breaks)) { // from the last break to +inf
      p = normal_lccdf(breaks[rows(breaks)] | mu, sigma);
    } else { // between two breaks
      p = log_diff_exp(normal_lcdf(breaks[y+1] | mu, sigma), normal_lcdf(breaks[y] | mu, sigma));
    }
    return p;
  }
}

data {
  int<lower = 0> T; // number of month
  int<lower = 0> n; // number of obs
  int time[n]; // observed month col
  int pop[n]; // observed population col
  int y[n]; // observed location wrt the breaks
  int<lower = 0> k;
  vector[k] intervalBreaks; // a vector with k elements
}

parameters {
  matrix[T, 2] alpha;
  matrix[T, 2] beta0;
  matrix[T, 2] beta1;
  vector<lower = 0>[2] sigma0;
  vector<lower = 0>[2] sigma1;
```

```
  vector[2] theta;
  vector<lower = 0>[2] tau;
  corr_matrix[2] eta;

  ordered[2] mua;
  vector<lower = 0>[2] gamma0;
  corr_matrix[2] rho0;

  ordered[2] mub;
  vector<lower = 0>[2] gamma1;
  corr_matrix[2] rho1;
}

transformed parameters {
  vector[2] mu0 = [mua[1], mub[1]]';
  vector[2] mu1 = [mua[2], mub[2]]';

  matrix[T, 2] prop;
  for (i in 1:T) {
    for (j in 1:2) {
      prop[i, j] = inv_logit(alpha[i, j]);
    }
  }
}

model {
  mua ~ normal(0, 10);
  mub ~ normal(0, 10);
  theta ~ normal(0, 10);

  gamma0 ~ cauchy(0, 2);
  gamma1 ~ cauchy(0, 2);
  tau ~ cauchy(0, 2);

  rho0 ~ lkj_corr(0.001);
  rho1 ~ lkj_corr(0.001);
  eta ~ lkj_corr(0.001);

  for (r in 1:T) {
    beta0[r] ~ multi_normal(mu0, quad_form_diag(rho0, gamma0));
    beta1[r] ~ multi_normal(mu1, quad_form_diag(rho1, gamma1));
    alpha[r] ~ multi_normal(theta, quad_form_diag(eta, tau));
  }

  for (i in 1:n) {
    int t;
    int s;
    real p;

    t = time[i];
```

```
    s = pop[i];
    p = prop[t, s];

    target += log_mix(1-p,
                      intervalCensoredNormal_lpmf(y[i] | beta0[t, s], sigma0[s], intervalBreaks),
                      intervalCensoredNormal_lpmf(y[i] | beta1[t, s], sigma1[s], intervalBreaks));
  }
}
```

References

1. World Health Organization . *Antimicrobial resistance: global report on surveillance*. World Health Organization . 2014.
2. Karp BE, Tate H, Plumblee JR, et al. National Antimicrobial Resistance Monitoring System: two decades of advancing public health through integrated surveillance of antimicrobial resistance. *Foodborne Pathogens and Disease* 2017; 14(10): 545–557.
3. Nguyen M, Long SW, McDermott PF, et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. *Journal of Clinical Microbiology* 2019; 57(2): e01260–18.
4. NARMS . The National Antimicrobial Resistance Monitoring System: Enteric Bacteria. 2018.
5. Gupta A, Nelson JM, Barrett TJ, et al. Antimicrobial resistance among campylobacter strains, United States, 1997–2001. *Emerging Infectious Diseases* 2004; 10(6): 1102–1109.
6. Hur J, Jawale C, Lee JH. Antimicrobial resistance of Salmonella isolated from food animals: A review. *Food Research International* 2012; 45(2): 819–830.
7. Williamson DA, Lane CR, Easton M, et al. Increasing antimicrobial resistance in nontyphoidal Salmonella isolates in Australia from 1979 to 2015. *Antimicrobial Agents and Chemotherapy* 2018; 62(2): e02012–17.
8. Angulo FJ, Baker NL, Olsen SJ, Anderson A, Barrett TJ. Antimicrobial use in agriculture: controlling the transfer of antimicrobial resistance to humans. *Seminars in Pediatric Infectious Diseases* 2004; 15(2): 78–85.
9. National Antimicrobial Resistance Monitoring System . *NARMS breakpoints*. 2018. https://www.fda.gov/media/108180/download/.
10. Rabatsky-Ehr T, Whichard J, Rossiter S, et al. Multidrug-resistant Strains of Salmonella enterica Typhimurium, United States, 1997–1998. *Emerging Infectious Diseases* 2004; 10(5): 795–801.
11. Wegener HC. Antibiotic resistance—linking human and animal health. In: Choffnes ER, Relman DA, Olsen LA, Hutton R, Mack A., eds. *Improving food safety through a one health approach: Workshop summary.*Washington, DC: National Academies Press. 2012 (pp. 331–349).
12. Iwamoto M, Reynolds J, Karp BE, et al. Ceftriaxone-resistant nontyphoidal Salmonella from humans, retail meats, and food animals in the United States, 1996–2013. *Foodborne Pathogens and Disease* 2017; 14(2): 74–83.
13. Mazloom R, Jaberi-Douraki M, Comer JR, Volkova V. Potential information loss due to categorization of Minimum Inhibitory Concentration frequency distributions. *Foodborne Pathogens and Disease* 2018; 15(1): 44–54.
14. Sader HS, Fey PD, Fish DN, et al. Evaluation of vancomycin and daptomycin potency trends (MIC creep) against methicillin-resistant Staphylococcus aureus isolates collected in nine US medical centers from 2002 to 2006. *Antimicrobial Agents and Chemotherapy* 2009; 53(10): 4127–4132.
15. Annis DH, Craig BA. Statistical properties and inference of the antimicrobial MIC test. *Statistics in Medicine* 2005; 24(23): 3631–3644.

16. Craig BA. Modeling approach to diameter breakpoint determination. *Diagnostic Microbiology and Infectious Disease* 2000; 36(3): 193–202.

17. Jaspers S, Lambert P, Aerts M, others . A Bayesian approach to the semiparametric estimation of a minimum inhibitory concentration distribution. *The Annals of Applied Statistics* 2016; 10(2): 906–924.

18. Zhang M, Wang C, O'Connor A. A hierarchical Bayesian latent class mixture model with censorship for detection of linear temporal changes in antibiotic resistance. *PLOS ONE* 2020; 15(1): e0220427.

19. Jaspers S, Komárek A, Aerts M. Bayesian estimation of multivariate normal mixtures with covariate-dependent mixing weights, with an application in antimicrobial resistance monitoring. *Biometrical Journal* 2018; 60(1): 7–19.

20. Vieira AR, Collignon P, Aarestrup FM, et al. Association between antimicrobial resistance in Escherichia coli isolates from food animals and blood stream isolates from humans in Europe: an ecological study. *Foodborne Pathogens and Disease* 2011; 8(12): 1295–1301.

21. Barnard J, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 2000: 1281–1311.

22. Alvarez I, Niemi J, Simpson M. Bayesian inference for a covariance matrix. *Preprint arXiv:1408.4050v2* 2016. In Proceedings of 26th Annual Conference on Applied Statistics in Agriculture. April 27-29, 2014.

23. Stan Development Team . Stan User's Guide version 2.22. 2019. http://mc-stan.org/.

24. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 2009; 100(9): 1989–2001.

25. USDA . *USDA NARMS report*. 2018.

26. DuPont HL. Bacterial diarrhea. *New England Journal of Medicine* 2009; 361(16): 1560–1569.

27. US Food and Drug Administration . The National Antimicrobial Resistance Monitoring System (NARMS) Integrated Report, 2015. 2017.

28. Sjölund-Karlsson M, Joyce K, Blickenstaff K, et al. Antimicrobial susceptibility to azithromycin among Salmonella enterica isolates from the United States. *Antimicrobial Agents and Chemotherapy* 2011; 55(9): 3985–3989.

29. Stan Development Team . RStan: the R interface to Stan. 2019. R package version 2.19.1. http://mc-stan.org/.

30. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.. *Journal of Machine Learning Research* 2014; 15(1): 1593–1623.

31. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; 7(4): 457–472.

32. CDC . *antibiotics-tested*. 2018.

33. Mouton JW. Breakpoints: current practice and future perspectives. *International Journal of Antimicrobial Agents* 2002; 19(4): 323–331.

How to cite this article: Zhang M., Wang C., and O'Connor A. (2020), A Bayesian latent class mixture model with censoring for correlation analysis in antimicrobial resistance across populations, *Statistics in Medicine*, *2017;00:1–6???*.