RESEARCH ARTICLE

A Bayesian approach to modeling antimicrobial multidrug resistance

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

1 Department of Statistics, Iowa State University, Ames, Iowa, United States of America, **2** Department of Veterinary Diagnostic and Production Animal Medicine, Iowa State University, Ames, Iowa, United States of America, **3** Department of Large Animal Clinical Sciences, Michigan State University, East Lansing, Michigan, United States of America

* chwang@iastate.edu

Abstract

Multidrug resistance (MDR) has been a significant threat to public health and effective treatment of bacterial infections. Current identification of MDR is primarily based upon the large proportions of isolates resistant to multiple antibiotics simultaneously, and therefore is a belated evaluation. For bacteria with MDR, we expect to see strong correlations in both the quantitative minimum inhibitory concentration (MIC) and the binary susceptibility as classified by the pre-determined breakpoints. Being able to detect correlations from these two perspectives allows us to find multidrug resistant bacteria proactively. In this paper, we provide a Bayesian framework that estimates the resistance level jointly for antibiotics belonging to different classes with a Gaussian mixture model, where the correlation in the latent MIC can be inferred from the Gaussian parameters and the correlation in binary susceptibility can be inferred from the mixing weights. By augmenting the laboratory measurement with the latent MIC variable to account for the censored data, and by adopting the latent class variable to represent the MIC components, our model was shown to be accurate and robust compared with the current assessment of correlations. Applying the model to *Salmonella* heidelberg samples isolated from human participants in National Antimicrobial Resistance Monitoring System (NARMS) provides us with signs of joint resistance to Amoxicillin-clavulanic acid & Cephalothin and joint resistance to Ampicillin & Cephalothin. Large correlations estimated from our model could serve as a timely tool for early detection of MDR, and hence a signal for clinical intervention.

Introduction

Background

The phrase, multidrug-resistant isolate, is typically used to refer to a bacterial isolate that is resistant to at least one antibiotic in three or more drug classes. The use of antibiotics for bacterial diseases of human and animals contributes to the selection of pathogenic bacteria resistance to multiple drugs [1]. In fact, multidrug resistance has become a significant threat to public health and the effective treatment of bacterial infections. Research has shown that the

in-hospital costs attributable to multidrug resistance isolates are alarmingly high, justifying the application of strict infection control measures in medical institutions with increased rate of multidrug-resistant infections [2]. It was also concluded by Giamarellos-Bourboulis et al. (2006) [3] that patients with infections by multidrug-resistant isolates were associated with decreased survival compared with infection by susceptible isolates.

In order to promote and protect public health by providing information about emerging bacterial resistance, resistance spreading pattern, and the impact of interventions, the National Antimicrobial Resistance Monitoring System (NARMS) was established in 1996. NARMS is a collaborative effort of three federal agencies: the Centers for Disease Control and Prevention (CDC), the U.S. Food and Drug Administration (FDA), and the United States Department of Agriculture (USDA), as well as state and local health departments in all 50 states [4]. NARMS tests isolates to determine their antibiotic susceptibility by finding the minimum inhibitory concentration (MIC, expressed in μg/mL), which is defined as the lowest concentration of a particular antibiotic that inhibits the visible growth of the bacteria. To date, dilution experiments [5] and multiple computational approaches (e.g., whole-genome sequencing based methods [6–9]) can be carried out for MIC determination. In NARMS, broth microdilution, as shown in Fig 1, on the Sensititre System from Trek Diagnostics was the method used to determine MICs for isolates [10]. Relative to the MIC breakpoints adopted from the Clinical and Laboratory Standards Institute (CLSI), isolates are classified as susceptible, intermediate, and resistant.

With societies investment in large-scale surveillance systems like NARMS, it is critical that we obtain as much value from the collected data as possible. Because multidrug resistance is particularly important, it would be desirable if the trends could be evaluated from multiple aspects when antibiotics appear to be jointly acquiring resistance. For example, knowledge that both drug A and drug B were showing increasing MIC levels in the same population of bacteria could trigger an investigation into why and possible early mitigation. Currently, this simultaneous increase in MIC of drug A and drug B would be detected only if both drugs increased resistance sufficiently to cause a change in the percentage of resistant bacteria. Such an approach to detection is very limited in scope. What is needed is a method that provides a more sensitive signal to public health officials of emerging trends that can trigger closer scrutiny earlier. For example, if both drug A and drug B have increasing MICs below the breakpoint threshold (or above the threshold) then current methods could not detect this change. Similarly, if resistance is increasing in two drugs, but only one at a level that changes organisms from being sensitive to resistant, current analysis methods could not detect this important joint emergence of resistance either. Therefore, there is a need to develop methods of detecting correlations in changes in MICs as well as changes in resistance category, which allows detection of joint resistance emergence in a more timely manner using already collected data.

Literature review

In practice, MIC data are often dichotomized to resistant and non-resistant and isolates with intermediate susceptibilities are often considered as susceptible [11]. A predominant approach to assessing the antimicrobial resistance (AMR) has focused on the proportion of resistant component, which can be found in research about AMR temporal trends [12, 13], cross-population correlation in AMR [14, 15], and multidrug resistance (MDR), etc.

For research in MDR, a survey on *Mannheimia* haemolytica isolated from respiratory diseased bovine shows an increasing pattern of antimicrobial MDR from 42% to 63% between 2009 and 2011 simply using descriptive statistics [16]. They also conducted mixed-effect logistic regression to analyze coresistance, i.e. the probability that resistance to a given agent was

Fig 1. Schematic of the microtiter plate for broth microdilution experiment to determine the minimal inhibitory concentration (MIC) [5]. A: MIC is recorded as = the lowest concentration that inhibits visible bacterial growth; B: MIC is recorded as > the highest concentration when growth occurs in all dilutions; C: MIC is recorded as ≤ the lowest concentration when no growth occurs in any concentrations.

associated with resistance to at least one other antimicrobial. More examples of using descriptive statistics to demonstrate MDR are found in the NARMS final report [10], as well as MDR studies in *Salmonella* isolates from humans in France [17] and Taiwan [18], etc. To examine and quantify the temporal trends of multidrug resistant gram-negative bacilli, resistant proportion data from a nine-year nosocomial surveillance study was analyzed using the chi-square test for trend, from which a rapid increasing pattern of resistance to three or more antimicrobials was discovered [19].

Another approach to revealing MDR was to calculate the correlation between resistance to one drug and resistance to one or more other drugs within a bacterial population. The correlation can be calculated using data of binary classification of isolates' susceptibility (resistant v.s. non-resistant) or the quantitative MICs of several antibiotics [20, 21]. In an analysis of MDR in the European Union (EU), the multivariate binary resistance patterns were modeled with a generalized linear model whose within-subject correlation matrix defines the dependence among multiple antimicrobials [22]. The model parameters were estimated using a generalized estimating equations (GEE) approach, which led to the same correlation estimation as the Spearman rank correlation. All of the above-mentioned research articles on MDR were based upon MIC dichotomization defined by CLSI and the observed MIC values.

Methods based on categorization of quantitative data can cause information loss [23]. The CLSI breakpoints for several antibiotics have gone through changes during the NARMS program, leaving it inappropriate to compare of resistant proportions over time. Additionally, direct calculation with the observed MIC values could also be problematic, as they are all subject to the censoring issue. The observed "MIC = 16 μg/mL" of organism A in Fig 1 is interval censored, which actually indicates that the true MIC is >8 and ≤ 16 μg/mL" but ultimately unknown. Similarly, organism B (MIC > 32 μg/mL) and C (MIC ≤ 0.5 μg/mL) are right and left censored at the highest and lowest concentration of the serial dilution experiment, respectively. Therefore, what we actually observe from the MIC measurement is the interval where the true MIC lies in. Correlation in MIC between antibiotics could possibly be overestimated or underestimated if not adjusting for the censored nature of the data.

As part of detecting trends in increasing resistance in several antibiotics concurrently, it would be helpful to have an estimation on the correlation in the resistance level among different drugs, which includes two interesting perspectives:

i. Correlation in the continuous latent MICs, meaning that high MIC values to one drug indicate high MIC values to the other(s).

ii. Correlation in the binary classification of susceptibility, meaning that being resistant to one drug indicates being inclined to be resistant to the other(s).

Correlation in the continuous latent MICs can be obtained by modeling their density. When modeling the MIC density, Craig [24] proposed to integrate the uncertainty of the true $\log_2$MIC values within their underlying intervals to resolve the censorship issue. The author also suggested a Gaussian mixture distribution to reflect the resistant and non-resistant population components without reliance on specific breakpoints. These resistant and non-resistant components of the bacterial populations are indicated by the bimodal distributed frequency plot of the observed MIC. Following this idea, subsequent studies on AMR temporal trends [25] and cross-population correlation in AMR [26] have improved the previous ones that did not address the censored nature or the underlying mixture distribution. Jaspers et al. (2018) [27] developed a Bayesian method to model the joint MIC density of two or more antimicrobials, from which the correlation of type i can be inferred. However, their assumption that the mean MIC of the resistant and non-resistant components was fixed over years could be unrealistic, and inference about correlation of type ii cannot be made from their model.

To our knowledge, research that estimates both types of correlation in antibiotic resistance while accounting for the challenges of analyzing the MIC data has not been conducted yet. In this study, we filled in this gap by modeling for the joint distribution of MIC between antibiotics under a Bayesian framework with a Gaussian mixture model with covariance-dependent mixing weights. Most importantly, the method was able to provide inferences about two types of correlation in antibiotic resistance. Examples of *Salmonella* heidelberg isolates collected by CDC NARMS were analyzed for two pairs of antibiotics, pair 1: amoxicillin-clavulanic acid (AMC) & cephalothin (CEP); and pair 2: ampicillin (AMP) & cephalothin (CEP). These antibiotics are both β-lactam antibiotic with ampicillin and amoxicillin-clavulanic acid being in the penicillin class and cephalothin being a cephalosporin, so we would expect both type i and type ii correlation to exist. Our model was assessed by simulation studies to be accurate and robust in estimating the correlations in resistance. The practical contribution of our method is that a large correlation in the continuous latent MIC and/or the binary susceptibility could be used as a way for early detection of MDR phenomenon.

Methods

Model notations and assumptions

Our proposed methodology estimates the joint distribution of the antibiotic resistance between several drugs at the isolate level. The MIC data were log-transformed with base 2, which is a commonly seen technique for two-fold serial dilution data. Latent variables of the true $\log_2$MIC and the component indicator of the isolate were introduced to account for the data censorship and the unobserved components. The joint resistance was assumed to follow a multivariate Gaussian mixture model whose mixing weights contains the conditional dependence between the susceptibilities of the antibiotics. The notations used for model description are as follows:

- $y^*_{d,i}$: the observed but censored $\log_2$MIC for isolate i tested by drug d.

- $y_{d,i}$: the real but latent $\log_2$MIC for isolate i tested by drug d.

- $l_{d,i}$, $u_{d,i}$: the lower and upper bound of the real latent value $y_{d,i}$, and $y_{d,i} \in (l_{d,i}, u_{d,i}]$. The lower and upper bounds for an interval censored $\log_2$MIC are $y^*_{d,i} - 1$ and $y^*_{d,i}$; for right censored data, $y^*_{d,i}$ and $+\infty$; for left censored data, $-\infty$ and $y^*_{d,i}$.

- $c_{d,i}$: the latent indicator of the bacterial component from which the isolate i tested by drug d was drawn. $c_{d,i} = 0, 1$ indicates the susceptible and resistant component, respectively.

Ranges of the subscripts are $i = 1, 2, \ldots, I$, where I is the total number of isolates that were tested by multiple antibiotics simultaneously, and $d = 1, 2, \ldots, D$, where D is the number of antibiotics belonging to different classes that were considered in this research. The supports of the indices remain the same throughout the paper unless otherwise specified.

Model description

Denote the latent $\log_2$MIC vector $(y_{1,i}, y_{2,i}, \ldots, y_{D,i})^T$ of isolate i tested by D antibiotics simultaneously as $\vec{y}_i$, the observed outcome as $\vec{y}_i^*$, and the indicator vector $(c_{1,i}, c_{2,i}, \ldots, c_{D,i})^T$ as $\vec{c}_i$. The joint antibiotic resistance was assessed at the isolate level as follows. For $i = 1, 2, \ldots, I$, and $d = 1, 2, \ldots, D$,

$$\vec{y}_i \overset{ind}{\sim} \text{MVN}_D(\vec{\mu}(\vec{c}_i), \Sigma), \tag{1}$$

where the d-th element $\mu_d(c_{d,i})$ of the mean vector $\vec{\mu}(\vec{c}_i)$ is

$$\mu_d(c_{d,i}) = \begin{cases} \mu_{0,d}, & \text{if } c_{d,i} = 0 \\ \mu_{1,d}, & \text{if } c_{d,i} = 1 \end{cases}. \tag{2}$$

The model of latent data assumes that the real $\log_2$MIC vector for isolate i tested by D antibiotics follows a D-dimensional multivariate normal distribution with a mean vector $\vec{\mu}$ and covariance matrix Σ. The d-th element of the mean vector μ_d is dependent on the component $c_{d,i}$ from which isolate i was drawn when treated with drug d. When considering drug d marginally, it is assumed that the classification of isolate i follows a Bernoulli distribution with probability p_d:

$$c_{d,i} \overset{ind}{\sim} \text{Bernoulli}(p_d). \tag{3}$$

Therefore, the mean $\log_2$MIC of isolate i against drug d takes value of $\mu_{1,d}$ with probability p_d, and takes value $\mu_{0,d}$ with probability $1 - p_d$. According to the meanings of being susceptible and resistant, the relationship of $\mu_{0,d} < \mu_{1,d}$ should always satisfy for all d. Some existing literature argues the phenomenon of antimicrobial multidrug resistance using the proportion of the isolates that are resistant to several antibiotic classes, where the resistant component is defined using the CLSI standards. The resistance classification $c_{d,i}$ in our method is different than the commonly used cut-off method in the sense that the assignment of component here is probabilistic rather than a hard dichotomization.

The covariance matrix Σ contains the variance of the latent $\log_2$MIC for each drug, regardless of the components. Such a simplification was assumed as it is impossible to accurately estimate variances for both components when there exist large proportions of left and/or right censored data, therefore variance estimation is not the primary goal of our work. The covariance matrix can be decomposed into a correlation matrix (with 1's on its diagonal and correlation parameters on its off-diagonal, denoted as R) sandwiched by the two scale matrices (with the standard deviations on its diagonal and 0's on its off-diagonal, denoted as S). This separation strategy was adopted for the sake of selection of prior distribution for Σ. When considering two drugs ($D = 2$),

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \tag{4}$$

$$=: SRS.$$

Distribution (3) implies independence in the component indicators among isolates for a given antibiotic d, which is reasonable. However, independence in the component indicators between antibiotics is not reasonable. When MDR occurs for a certain serotype, we expect not only significantly positive correlation in the latent $\log_2$MIC, but also conditional dependence in the binary susceptibility classification between different antibiotics. The former can be assessed by parameter ρ in the correlation matrix R, while the latter can be assessed by introducing the covariance parameter between antibiotic susceptibilities.

To illustrate the approach to modeling the conditional dependence between antibiotic susceptibilities [28], two drugs are given as an example ($D = 2$). Therefore, $\vec{c}_i \in \Gamma$, where $\Gamma = \{\{1, 1\}, \{1, 0\}, \{0, 1\}, \{0, 0\}\}$. Denote the covariance between the two antibiotic susceptibilities as δ. It can be shown that

$$
\begin{aligned}
P(c_{1,i} = 1, c_{2,i} = 1) &= p_1 p_2 + \delta \\
P(c_{1,i} = 1, c_{2,i} = 0) &= p_1 (1 - p_2) - \delta \\
P(c_{1,i} = 0, c_{2,i} = 1) &= (1 - p_1) p_2 - \delta \\
P(c_{1,i} = 0, c_{2,i} = 0) &= (1 - p_1)(1 - p_2) + \delta
\end{aligned}
\tag{5}
$$

Eq (5), the probabilities of the four components defined by the susceptibilities of two drugs, can be summarised to the following contingency table (Table 1). The feasible range of δ is

$$
-\min(p_1 p_2, (1 - p_1)(1 - p_2)) \leq \delta \leq \min(p_1, p_2) - p_1 p_2.
\tag{6}
$$

The binary classifications are independent between the two antibiotics when $\delta = 0$. Knowing the covariance δ is not very straightforward due to its uncommon range, therefore it was transformed to correlation, denoted as ϕ, in the binary susceptible classification:

$$
\phi = \frac{\delta}{\sqrt{p_1(1 - p_1)}\sqrt{p_2(1 - p_2)}}.
\tag{7}
$$

ϕ takes value between -1 and 1. A large ϕ implies that if a certain serotype is resistant to one antibiotic, then it is highly likely to be resistant to the other.

The parameter space involved in this Gaussian latent class mixture model is $\Theta = (\vec{\mu}, \Sigma, \vec{p}, \delta)$. Let $f(\cdot)$ be a generic representation for probability density function (pdf) or probability mass function (pmf). Also, denote all latent $\log_2$MIC values $(\vec{y}_1, \vec{y}_2, ..., \vec{y}_I)^T$ as $\mathbf{Y}$, the corresponding observed values as $\mathbf{Y}^*$, and the latent component indicators as $\mathbf{C}$. The mean vector of the bivariate normal distribution is $\vec{\mu} = (\mu_1, \mu_2)^T$, where $\mu_1 \in \{\mu_{0,d=1}, \mu_{1,d=1}\}$ and $\mu_2 \in \{\mu_{0,d=2}, \mu_{1,d=2}\}$. The marginal probabilities of being resistant form a vector $\vec{p} = (p_1, p_2)^T$.

Table 1. Contingency table of the components defined by the susceptibilities of two antibiotics.

		$c_{2,i}$	
		0	1
$c_{1,i}$	0	$(1 - p_1)(1 - p_2) + \delta$	$p_2(1 - p_1) - \delta$
	1	$p_1(1 - p_2) - \delta$	$p_1 p_2 + \delta$

Then under such a context of two antibiotics, the data likelihood can be written as

$$
\begin{aligned}
f(\mathbf{Y}^*|\Theta) &= \prod_{i=1}^{I} f(\vec{y}_i^*|\Theta) \\
&= \prod_{i=1}^{I} \int_{l_{2,i}}^{u_{2,i}} \int_{l_{1,i}}^{u_{1,i}} \sum_{\vec{c}_i \in \Gamma} f(\vec{y}_i, \vec{c}_i|\Theta) dy_{1,i} dy_{2,i} \\
&= \prod_{i=1}^{I} \int_{l_{2,i}}^{u_{2,i}} \int_{l_{1,i}}^{u_{1,i}} \sum_{\vec{c}_i \in \Gamma} f(\vec{y}_i|\vec{\mu}, \Sigma, \vec{c}_i) \cdot f(\vec{c}_i|\vec{p}, \delta) dy_{1,i} dy_{2,i}.
\end{aligned}
\tag{8}
$$

In (8), the first equality is due to independence among the submitted isolates. The second equality is to sum the latent component indicators over their possible values, and to integrate the continuous $\log_2$MIC over the intervals of which the discrete measurements lay in. The two parts in $\sum_{\vec{c}_i \in \Gamma} f(\vec{y}_i|\vec{\mu}, \Sigma, \vec{c}_i) \cdot f(\vec{c}_i|\vec{p}, \delta)$ are defined by the bivariate normal distribution (1) and the contingency Table 1, which forms a Gaussian mixture distribution with mixing weights containing the covariance between the antibiotic susceptibility. Consequently, a Bayesian model is obtained:

$$
\begin{aligned}
f(\Theta, \mathbf{Y}, \mathbf{C}|\mathbf{Y}^*) &= f(\vec{\mu}, \Sigma, \vec{p}, \delta, \mathbf{Y}, \mathbf{C}|\mathbf{Y}^*) \propto f(\mathbf{Y}^*, \mathbf{Y}, \mathbf{C}, \vec{\mu}, \Sigma, \vec{p}, \delta) \\
&= f(\mathbf{Y}^*|\mathbf{Y}, \mathbf{C}, \vec{\mu}, \Sigma, \vec{p}, \delta) \cdot f(\mathbf{Y}, \mathbf{C}|\vec{\mu}, \Sigma, \vec{p}, \delta) \cdot f(\vec{\mu}, \Sigma, \vec{p}, \delta) \\
&= f(\mathbf{Y}^*|\mathbf{Y}, \mathbf{C}, \vec{\mu}, \Sigma, \vec{p}, \delta) \cdot f(\mathbf{Y}, \mathbf{C}|\vec{\mu}, \Sigma, \vec{p}, \delta) \cdot f(\vec{\mu}) \cdot f(\Sigma) \cdot f(\vec{p}) \cdot f(\delta)
\end{aligned}
\tag{9}
$$

In (9), the pmf $f(\mathbf{Y}^*|\mathbf{Y}, \mathbf{C}, \vec{\mu}, \Sigma, \vec{p}, \delta)$ associates $\mathbf{Y}^*$, the observed $\log_2$MIC, with its underlying latent values $\mathbf{Y}$, and corresponds to the lowest hierarchy of the model. The term $f(\mathbf{Y}^*|\mathbf{Y}, \mathbf{C}, \vec{\mu}, \Sigma, \vec{p}, \delta) \cdot f(\mathbf{Y}, \mathbf{C}|\vec{\mu}, \Sigma, \vec{p}, \delta)$ is expressed in Eq (8). Selection of the prior distributions for the model parameters is described in the following Subsection of Prior specification.

Prior specification

A joint prior distribution of the unknown model parameters $\Theta = (\vec{\mu}, \Sigma, \vec{p}, \delta)$ is required by the full Bayesian analysis. If assuming independence, it can be expressed as the product of each individual prior distribution as shown in the last equality in (9). Diffuse Gaussian priors were selected for $\vec{\mu}$ due to the *a priori* lack of knowledge about the means: $\mu_{0,1}, \mu_{1,1}, \mu_{0,2}, \mu_{1,2} \sim N(0, 10000)$. The constraint that $\mu_{0,d} < \mu_{1,d}$ was applied to reflect the relationship between the susceptible and resistant component. Similarly, a non-informative prior distribution Beta(0.5, 0.5) was determined for p_1 and p_2. A default uniform prior was assigned to the covariance parameter δ on its feasible range: $f(\delta) \propto 1$ for $-\min(p_1 p_2, (1 - p_1)(1 - p_2)) \le \delta \le \min(p_1, p_2) - p_1 p_2$.

As for the prior distribution of the covariance matrix Σ, since the widely-used inverse Wishart distribution tends to bias downward the posterior correlation [29], the separation strategy [30] was adopted instead. According to the recommendation of the Stan development team [31], an LKJ prior was determined for the correlation matrix [32]: $f(R) \propto |R|^{\nu-1}$, $\nu > 0$. By choosing LKJ ($\nu = 1$) for the correlation matrix R, the density of its prior is uniform over the correlation matrices of dimension $D = 2$. We assigned weakly informative half-Cauchy priors to the scale parameters: $\sigma_1, \sigma_2 \sim \text{Cauchy}^+(0, 2)$.

Alternative approaches to assess correlation

It is often of interest to compare a new approach with alternative methods. As a comparison to the Bayesian approach to estimating the correlation in latent $\log_2$MIC and in susceptibility, alternative methods [20, 21] were brought up and used for comparison. Two alternative approaches for the correlation related to MIC (type i correlation referred to above) are Pearson correlation and Spearman rank correlation. We used Spearman rank correlation for the observed $\log_2$MIC because the rank approach acknowledges the ordinal nature of the MIC. For the binary correlation (i.e. type ii), the Spearman and Pearson methods give the same result—that is they calculate the correlation in the binary classification expressed as $c_{d,i}$ in the model—as the estimate using the GEE approach [22].

Real data analyses

Data description

The dataset used to apply the Bayesian approach to modeling the MDR is the human population of NARMS, which was launched in 1996 within the framework of CDC's Emerging Infections Program and the Foodborne Diseases Active Surveillance Network (FoodNet). Participating public health laboratories submit their every 20th non-typhoidal *Salmonella*, *Campylobacter*, *Shigella*, and *Escherichia coli* O157 isolates to CDC for antibiotic susceptibility testing [33]. The data were published with year of collection, genus, serotype, MIC value tested against multiple antibiotics, the test conclusion (resistant or not), etc. Currently, CDC NARMS routinely tests for susceptibility to 18 antibiotic agents that belong to 12 classes of drugs.

We limited our analyses to an important *Salmonella* serotype: *S.* heidelberg, which has 1196 isolates, accounting for 3.12% of the total 38311 *Salmonella* enterica isolates. The pairs of antibiotics we focused on are AMC (Amoxicillin-clavulanic acid, β–lactam combination agent class) & CEP (Cephalothin, Cephalosporin class) and AMP (Ampicillin, Penicillin class) & CEP. Since the susceptibility of heidelberg to CEP was not tested after 2003, there were 498 isolates recorded for each pair. The scatter plots of the observed $\log_2$MIC of the two pairs are shown in Fig 2. The CLSI cut-off value after $\log_2$ transformation is 5 for all drugs studied here. Thus, there were 3.82% isolates jointly resistant to AMC & CEP, and 5.22% jointly resistant to AMP & CEP.

Implementation

A Bayesian analysis of Markov Chain Monte Carlo (MCMC) was implemented in R version 3.6.3 using the *rstan* package [34] in order to draw inference from the posterior distribution f $(\Theta, \mathbf{Y}, \mathbf{C}|\mathbf{Y}^*)$ which has no closed form. The sampling algorithm adopted was the No-U-Turn sampler variant of Hamiltonian Monte Carlo which is efficient, and hence the preferred method in Stan [35].

For each of the two example pairs, the initial values were determined based upon the $\log_2$-MIC observations to avoid random initialization. More specifically, the mean vector $\vec{\mu}$ was obtained by taking the arithmetic mean of the censored observations within each component for each drug. The standard deviations in the scale matrix S were each calculated directly from the observations of each drug. The Spearman rank correlation of the observations between the paired drugs was used as the initial value of type i correlation ρ. The proportions of the resistant isolates were regarded as the initial values of $\vec{p}$. The difference between the jointly resistant proportion and the product of the marginal resistant proportions is used as the initial value for type ii covariance δ. Here, the resistant isolates are classified by CLSI standards.
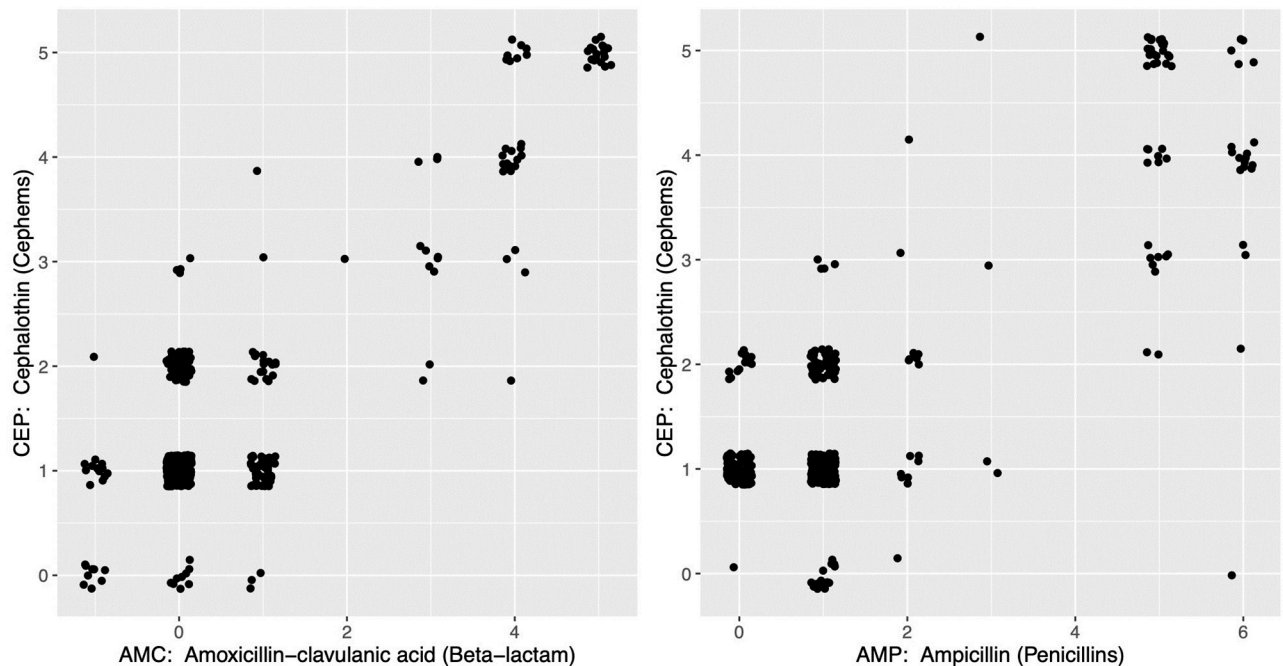
Fig 2. Scatter plots of the observed log$_2$MIC (MIC in unit μg/mL) with jittering of *Salmonella* heidelberg isolates tested by AMC & CEP (left) and AMP & CEP (right) in NARMS human population during 1996-2003.

The model implementation was based on three chains with 10,000 MCMC iterations, disregarding the first 50% realizations as the burn-in session. The MCMC convergence was examined with the Gelman–Rubin diagnostic [36]. The point estimates and the 95% credible intervals of the model parameters were obtained by taking the means and the 2.5th and 97.5th percentiles of the posterior draws after the burn-in portion. The application results were summarized and can be found in the following subsection. Model specification, implementation, and other relevant R scripts are available in a public GitHub repository (GitHub: https://github.com/MinZhang95/AMR-MDR).

Application results

The correlations in both the continuous latent log$_2$MIC and the binary susceptibility are of particular interest when modeling the joint distribution of resistance to several antibiotics. Their estimation results can be found in Table 2, where the point estimation, standard deviation, and credible interval for the two parameters of each example pair are presented.

Since the credible interval for each correlation parameter of each case is above 0, we are able to conclude significant correlations. In other words, the correlations in the latent log$_2$MIC

Table 2. Estimations of the correlation in the latent log$_2$MIC (ρ) and correlation in susceptibility classification (ϕ) using *Salmonella* heidelberg isolates of NARMS human population tested by AMC & CEP and AMP & CEP during 1996-2003.

Pair	Parameter	Estimation	Standard deviation	Credible interval
AMC & CEP	ρ	0.6303	0.0450	(0.5352, 0.7115)
	ϕ	0.9677	0.0188	(0.9210, 0.9935)
AMP & CEP	ρ	0.2818	0.0934	(0.0884, 0.4628)
	ϕ	0.9174	0.0295	(0.8524, 0.9663)

and in the binary susceptibility are significantly positive for the paired antibiotics. The estimations in the correlation of binary susceptibility ϕ are especially high, with values greater than 0.9 for both cases. The point estimations of ρ and ϕ from the Spearman correlation are 0.3986 and 0.9042 for AMC & CEP, 0.1316 and 0.6381 for AMP & CEP, which has some discrepancy from the Bayesian estimations. To assess the performance of the hierarchical Bayesian model, with the Spearman correlation as a comparison, simulation studies were conducted.

Simulation

To assess the performance of the proposed Bayesian approach to estimating the correlation in antibiotic resistance, three simulation studies were conducted using the estimated results from the two application examples in the Real data analyses Section and a negative control as the underlying data generators. The negative control shared the same parameters as the application results of AMC & CEP except that it assumed null correlations between the two antibiotics. Greek letters with "hats" on top are adopted to denote the known model parameters $\hat{\Theta} = (\hat{\vec{\mu}}, \hat{\Sigma}, \hat{\vec{p}}, \hat{\delta})$.

To simulate $I = 498$ isolates (which is the size of the real data examples):

1. Generate I categorical realizations following the multinomial distribution and denote the counts within each category as $(n_{0,0}, n_{0,1}, n_{1,0}, n_{1,1})^T \sim \text{Multinomial}(I, \hat{P}(c_{1,i} = 0, c_{2,i} = 0), \hat{P}(c_{1,i} = 0, c_{2,i} = 1), \hat{P}(c1, i = 1, c_{2,i} = 0), \hat{P}(c_{1,i} = 1, c_{2,i} = 1))$. Each $\hat{P}(\cdot)$ is calculated by entering $\hat{\vec{p}}$ and $\hat{\delta}$ into the probabilities shown in the contingency Table 1.

2. Generate **Y**, the latent $\log_2 \text{MIC}$, which are composed of
$n_{0,0}$ data from $\text{MVN}_2\left(\begin{pmatrix} \hat{\mu}_{0,1} \\ \hat{\mu}_{0,2} \end{pmatrix}, \hat{\Sigma}\right)$, $n_{0,1}$ data from $\text{MVN}_2\left(\begin{pmatrix} \hat{\mu}_{0,1} \\ \hat{\mu}_{1,2} \end{pmatrix}, \hat{\Sigma}\right)$, $n_{1,0}$ data from $\text{MVN}_2\left(\begin{pmatrix} \hat{\mu}_{1,1} \\ \hat{\mu}_{0,2} \end{pmatrix}, \hat{\Sigma}\right)$, $n_{1,1}$ data from $\text{MVN}_2\left(\begin{pmatrix} \hat{\mu}_{1,1} \\ \hat{\mu}_{1,2} \end{pmatrix}, \hat{\Sigma}\right)$.

3. Convert the latent $\log_2 \text{MIC}$ to the censored $\mathbf{Y}^*$:
$$y_{d,i}^* = \begin{cases} \text{lowest concentration,} & y_{d,i} \leq \text{lowest concentration} \\ \lceil y_{d,i} \rceil, \text{lowest} & < y_{d,i} \leq \text{highest concentration} \\ \text{highest concentration,} & y_{d,i} > \text{highest concentration} \end{cases}; \lceil \cdot \rceil \text{ represents the ceiling of a}$$
number.
The lowest and highest concentrations refer to the two extremes of the dilution range, which is drug-dependent. Each of the three antibiotics studied here has its own spectrum of serial dilution.

4. Assign the binary conclusions (non-resistant v.s. resistant) according to the censored data using the CLSI breakpoints. The censored $\log_2 \text{MIC}$ data simulated from the above steps are considered as susceptible if below the breakpoint which is 5; otherwise, classified as resistant.

We obtained $n = 100$ simulated datasets by repeating the above procedure 100 times for each case. Both the proposed Bayesian approach and the Spearman methods were applied to the simulated datasets. Therefore, 100 estimation results were attained for each pair. Since we are interested in comparing the accuracy and robustness of the two methods, the mean absolute error (MAE) and the root of mean squared error (RMSE) were calculated for each set of

Table 3. Comparisons of the estimated correlations for latent $\log_2$MIC (ρ) and binary classification (ϕ) between the Bayesian and Spearman method through simulations.

Scenario	Parameter	Truth	Method	Estimation	SD	MAE	RMSE
Scenario 1	ρ	0.6303	Bayesian	0.6304	0.0409	0.0331	0.0407
			Spearman	0.7040	0.0388	0.0756	0.0832
	ϕ	0.9677	Bayesian	0.9461	0.0174	0.0228	0.0276
			Spearman	0.6302	0.0746	0.3375	0.3455
Scenario 2	ρ	0.2818	Bayesian	0.2649	0.0655	0.0522	0.0671
			Spearman	0.1933	0.0472	0.0892	0.1001
	ϕ	0.9174	Bayesian	0.8926	0.0218	0.0269	0.0329
			Spearman	0.7231	0.0493	0.1943	0.2004
Scenario 3	ρ	0.0000	Bayesian	-0.0074	0.0590	0.0478	0.0592
			Spearman	-0.0096	0.0481	0.0390	0.0489
	ϕ	0.0000	Bayesian	0.0174	0.0416	0.0344	0.0449
			Spearman	0.0055	0.0459	0.0370	0.0460

results, as displayed in Table 3. Scenarios 1-3 represent data generators of applications AMC & CEP, AMP & CEP, and the null case, respectively.

Discussion and conclusion

In this work, we proposed a Bayesian latent class mixture model as a tool to assist with assessment of the antibiotic multidrug resistance phenomenon. When MDR occurs, we expect strong correlation in the latent $\log_2$MIC between antibiotics, and also strong correlation in the binary susceptibility classification. This Bayesian approach is capable of addressing both characteristics while modeling the resistance level for multiple drugs. The approach offers the opportunity to obtain more from the MIC data than is currently being extracted. On the same dataset, using the proposed model, researchers and public health officials are able to (a) assess the correlation of MIC for antibiotics even below the resistant threshold; (b) assess the binary classification of organisms resistant or non-resistance in a manner independent of the predetermined threshold values; and (c) account for the censored nature of the MIC data.

One of the goals of the surveillance programs is to early detect issues with the hope that interventions can mitigate risk. The purpose of correlation estimation is not merely to test for significant correlations, but also to identify newer correlations that are emerging. The question could be "Are two antibiotics that were previously not correlated starting to show positive correlations?". And the answers inferred from our approach could trigger an investigation into the mechanism of that correlation. Similarly, if the correlation between two classes of antibiotic increases, this might also warrant an investigation. As such, we envision that most of the findings from our analyses will be hypotheses generating. Further, the presence of correlation should be combined with other information available in the dataset. In our examples, only 3.82% of isolates are jointly resistant to AMC & CEP, and 5.22% jointly resistant to AMP & CEP, so concerns about correlation might be tempered by the low resistant prevalence. Public health officials would need to balance the three pieces of information: type i correlation, type ii correlation and resistant prevalence when making decisions.

The currently used analytical methods, Spearman correlation in $\log_2$MIC and the binary conclusions, were compared against the proposed model in Table 3. What can be seen from the bias that arises is that although the Spearman calculation is straightforward, it does not consider the censored nature of the MIC data, or the mixture structure of susceptible and

resistant isolates. As shown in the simulation results in Table 3, when correlations exist, the advantages in the Bayesian model are obvious in terms of the mean absolute error (MAE) and the root of mean squared error (RMSE), meaning that the Bayesian method provides a more accurate and robust estimation of the correlations. When correlations do not exist, such advantages become less obvious and the two methods perform comparably well. It can be found that the alternative method tends to underestimate the correlation in the susceptibility status when the ground truth is not null, which could be misleading in real life. Though the Spearman correlation of 0.6302 in the binary conclusions for the AMC & CEP example is quite high, it is not as close to the truth as the Bayesian estimation. However, the proposed model does require a mixture of susceptible and resistant components in order to obtain an accurate estimation, and we found this to be true for many (but not all) of the organism-antibiotic combinations in the NARMS datasets.

The proposed Bayesian model was articulated for the two dimensional case where two antibiotics were considered simultaneously. This work can be generated to three or higher dimensions and applied to more antibiotics. Application of this model to three drugs and more could promote our understanding to the surveillance data and serve as an alert when MDR emerges.

Author Contributions

Conceptualization: Chong Wang, Annette O'Connor.

Formal analysis: Min Zhang, Chong Wang.

Funding acquisition: Chong Wang, Annette O'Connor.

Methodology: Min Zhang, Chong Wang, Annette O'Connor.

Resources: Annette O'Connor.

Software: Min Zhang.

Supervision: Chong Wang, Annette O'Connor.

Validation: Min Zhang, Annette O'Connor.

Visualization: Min Zhang.

Writing – original draft: Min Zhang, Chong Wang, Annette O'Connor.

Writing – review & editing: Min Zhang, Chong Wang, Annette O'Connor.

References

1. Nikaido H. Multidrug resistance in bacteria. Annual Review of Biochemistry. 2009 Jul; 78:119–46. https://doi.org/10.1146/annurev.biochem.78.082907.145923 PMID: 19231985

2. Tansarli GS, Karageorgopoulos DE, Kapaskelis A, Falagas ME. Impact of antimicrobial multidrug resistance on inpatient care cost: an evaluation of the evidence. Expert Review of Anti-infective Therapy. 2013 Mar; 11(3):321–31. https://doi.org/10.1586/eri.13.4 PMID: 23458771

3. Giamarellos-Bourboulis EJ, Papadimitriou E, Galanakis N, Antonopoulou A, Tsaganos T, Kanellako-poulou K, et al. Multidrug resistance to antimicrobials as a predominant factor influencing patient survival. International Journal of Antimicrobial Agents. 2006 Jun; 27(6):476–81. https://doi.org/10.1016/j.ijantimicag.2005.12.013 PMID: 16707252

4. Karp BE, Tate H, Plumblee JR, Dessai U, Whichard JM, Thacker EL, et al. National antimicrobial resistance monitoring system: two decades of advancing public health through integrated surveillance of antimicrobial resistance. Foodborne Pathogens and Disease. 2017 Oct; 14(10):545–57. https://doi.org/10.1089/fpd.2017.2283 PMID: 28792800

5. Wiegand I, Hilpert K, Hancock RE. Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nature Protocols. 2008 Feb; 3(2):163. https://doi.org/10.1038/nprot.2007.521 PMID: 18274517

6. Nguyen M, Long SW, McDermott PF, Olsen RJ, Olson R, Stevens RL, et al. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. Journal of Clinical Microbiology. 2019 Feb; 57(2). https://doi.org/10.1128/JCM.01260-18 PMID: 30333126

7. Nguyen M, Brettin T, Long W, Musser JM, Olsen RJ, Olson R, et al. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumoniae. Scientific reports. 2018 Jan; 8(1), 1–11.

8. Eyre DW, De Silva D, Cole K, Peters J, Cole MJ, Grad YH, et al. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. Journal of Antimicrobial Chemotherapy. 2017 Mar; 72(7), 1937–47. https://doi.org/10.1093/jac/dkx067 PMID: 28333355

9. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Walker H, et al. Penicillin-binding protein transpeptidase signatures for tracking and predicting β-lactam resistance levels in Streptococcus pneumoniae. MBio. 2016 Jun; 7(3), e00756–16. https://doi.org/10.1128/mBio.00756-16 PMID: 27302760

10. CDC. National Antimicrobial Resistance Monitoring System for Enteric Bacteria (NARMS): Human isolates final report, 2010. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC. 2012.

11. Rabatsky-Ehr T, Whichard J, Rossiter S, Holland B, Stamey K, Headrick ML, et al. NARMS Working Group. Multidrug-resistant strains of Salmonella enterica Typhimurium, United States, 1997–1998. Emerging Infectious Diseases. 2004 May; 10(5):795. https://doi.org/10.3201/eid1005.030209 PMID: 15200811

12. Hanon JB, Jaspers S, Butaye P, Wattiau P, Méroc E, Aerts M, et al. A trend analysis of antimicrobial resistance in commensal Escherichia coli from several live-stock species in Belgium (2011–2014). Preventive Veterinary Medicine. 2015 Dec; 122(4):443–52. https://doi.org/10.1016/j.prevetmed.2015.09.001 PMID: 26423778

13. Tadesse DA, Zhao S, Tong E, Ayers S, Singh A, Bartholomew MJ, et al. Antimicrobial drug resistance in Escherichia coli from humans and food animals, United States, 1950–2002. Emerging Infectious Diseases. 2012 May; 18(5):741. https://doi.org/10.3201/eid1805.111153 PMID: 22515968

14. Iwamoto M, Reynolds J, Karp BE, Tate H, Fedorka-Cray PJ, Plumblee JR, et al. Ceftriaxone-resistant nontyphoidal Salmonella from humans, retail meats, and food animals in the United States, 1996–2013. Foodborne Pathogens and Disease. 2017 Feb; 14(2):74–83. https://doi.org/10.1089/fpd.2016.2180 PMID: 27860517

15. Wegener HC. Antibiotic resistance—linking human and animal health. Improving Food Safety Through a One Health Approach: Workshop Summary. 2012 Sep (p. 331). National Academies Press.

16. Lubbers BV, Hanzlicek GA. Antimicrobial multidrug resistance and coresistance patterns of Mannheimia haemolytica isolated from bovine respiratory disease cases—a three-year (2009–2011) retrospective analysis. Journal of Veterinary Diagnostic Investigation. 2013 May; 25(3):413–7. https://doi.org/10.1177/1040638713485227 PMID: 23572456

17. Weill FX, Guesnier F, Guibert V, Timinouni M, Demartin M, Polomack L, et al. Multidrug resistance in Salmonella enterica serotype Typhimurium from humans in France (1993 to 2003). Journal of Clinical Microbiology. 2006 Mar; 44(3):700–8. https://doi.org/10.1128/JCM.44.3.700-708.2006 PMID: 16517842

18. Lauderdale TL, Aarestrup FM, Chen PC, Lai JF, Wang HY, Shiau YR, et al. Multidrug resistance among different serotypes of clinical Salmonella isolates in Taiwan. Diagnostic Microbiology and Infectious Disease. 2006 Jun; 55(2):149–55. https://doi.org/10.1016/j.diagmicrobio.2006.01.002 PMID: 16529903

19. D'Agata EM. Rapidly rising prevalence of nosocomial multidrug-resistant, Gram-negative bacilli: a 9-year surveillance study. Infection Control & Hospital Epidemiology. 2004 Oct; 25(10):842–6. https://doi.org/10.1086/502306 PMID: 15518026

20. Chang HH, Cohen T, Grad YH, Hanage WP, O'Brien TF, Lipsitch M. Origin and proliferation of multiple-drug resistance in bacterial pathogens. Microbiology and Molecular Biology Reviews. 2015 Mar; 79(1):101–16. https://doi.org/10.1128/MMBR.00039-14 PMID: 25652543

21. Oggioni MR, Coelho JR, Furi L, Knight DR, Viti C, Orefici G, et al. Significant differences characterise the correlation coefficients between biocide and antibiotic susceptibility profiles in Staphylococcus aureus. Current Pharmaceutical Design. 2015 May; 21(16):2054. https://doi.org/10.2174/1381612821666150310103238 PMID: 25760337

22. Jaspers S, Ganyani T, Ensoy C, Faes C, Aerts M. Development and application of statistical methodology for analysis of the phenomenon of multi-drug resistance in the EU: demonstration of analytical approaches using antimicrobial resistance isolate-based data. EFSA Supporting Publications. 2016 Sep; 13(9):1084E.

23. 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 Jan; 15(1):44–54. https://doi.org/10.1089/fpd.2017.2301 PMID: 29039983

24. Craig BA. Modeling approach to diameter breakpoint determination. Diagnostic Microbiology and Infectious Disease. 2000 Mar; 36(3):193–202. https://doi.org/10.1016/S0732-8893(99)00130-3 PMID: 10729662

25. 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 Jan; 15(1):e0220427. https://doi.org/10.1371/journal.pone.0220427 PMID: 32004341

26. Zhang M, Wang C, O'Connor A. A Bayesian latent class mixture model with censoring for correlation analysis in antimicrobial resistance across populations. BMC Medical Research Methodology. 2021 Sep; 21(1):1–12. https://doi.org/10.1186/s12874-021-01384-w PMID: 33397292

27. 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 Jan; 60(1):7–19. https://doi.org/10.1002/bimj.201600253 PMID: 28898442

28. Dendukuri N, Joseph L. Bayesian approaches to modeling the conditional dependence between multiple diagnostic tests. Biometrics. 2001 Mar; 57(1):158–67. https://doi.org/10.1111/j.0006-341X.2001.00158.x PMID: 11252592

29. Alvarez I, Niemi J, Simpson M. Bayesian inference for a covariance matrix. Preprint arXiv:1408.4050. 2014 Aug.

30. Barnard J, McCulloch R, Meng XL. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. Statistica Sinica. 2000 Oct:1281–311.

31. Stan Development Team. Stan User's Guide version 2.22. Online https://mc-stan.org/docs/2_22/stan-users-guide-2_22.pdf. 2019.

32. Lewandowski D, Kurowicka D, Joe H. Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis. 2009 Oct; 100(9):1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

33. National Antimicrobial Resistance Monitoring System. National Antimicrobial Resistance Monitoring System (NARMS) methodology. Online https://www.fda.gov/media/101741/download. 2018.

34. Stan Development Team. RStan: the R interface to Stan. R package version 2.21.2 Online http://mc-stan.org/. 2020.

35. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014 Apr; 15(1):1593–623.

36. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical science. 1992; 7(4):457–72. https://doi.org/10.1214/ss/1177011136