Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: www.tandfonline.com/journals/tstf20

A hierarchical Bayesian latent class mixture model with censorship for detection of linear changes and correlation analysis across populations in antimicrobial resistance

Jingheng Cai & Wei Zeng

To cite this article: Jingheng Cai & Wei Zeng (28 Feb 2026): A hierarchical Bayesian latent class mixture model with censorship for detection of linear changes and correlation analysis across populations in antimicrobial resistance, Statistical Theory and Related Fields, DOI: 10.1080/24754269.2026.2634423

To link to this article: https://doi.org/10.1080/24754269.2026.2634423

STATISTICAL THEORY AND RELATED FIELDS
https://doi.org/10.1080/24754269.2026.2634423

Taylor & Francis
Taylor & Francis Group

A hierarchical Bayesian latent class mixture model with censorship for detection of linear changes and correlation analysis across populations in antimicrobial resistance

Jingheng Cai and Wei Zeng

Department of Statistics, Sun Yat-sen University, Guangzhou, Guangdong Province, People's Republic of China

ABSTRACT
Antimicrobial resistance (AMR) has already been identified as an urgent issue affecting global health. Many research interests lie in monitoring the change in AMR in both human and animal populations. Moreover, it is important to study the correlation in AMR between the two populations. In this study, we develop a hierarchical latent class mixture model for the detection of linear changes and correlation analysis across populations with antimicrobial resistance. We propose Bayesian methods to estimate the unknown parameters in the proposed model. The simulation study is conducted to evaluate the empirical performance of the proposed method. Finally, we employ the proposed model and methodology to analyze the datasets obtained from the National Antimicrobial Resistance Monitoring System (NARMS).

ARTICLE HISTORY
Received 8 December 2024
Revised 25 March 2025
Accepted 12 February 2026

KEYWORDS
Antimicrobial resistance; Bayesian methods; censored data; correlated analysis; mixture distributions

1. Introduction

Antimicrobial resistance (AMR) occurs when bacteria, viruses, fungi and parasites no longer respond to antimicrobial medicines. As a result of drug resistance, antibiotics and other antimicrobial medicines become ineffective and infections become difficult or impossible to treat, increasing the risk of disease spread, severe illness, disability and death. In the past decades, AMR has been a significant concern in public health and received great research attention (Cohen, 2000; Spellberg et al., 2008). In 2015, the World Health Organization (WHO) has already identified AMR as an urgent problem affecting global health and development and discussed countermeasures (Mendelson & Matsoso, 2015). Drug-resistant strains have spread all over the world, especially in developing countries and emerging economies, where the medical impact is much more serious. The number of people in those low-income countries infected with drug-resistant strains shows an increasing trend year by year, and the time and medical costs required for treatment have also increased significantly. Moreover, at least 700,000 people die from AMR bacterial infections worldwide each year. If the world does not take action to control AMR, the number of deaths from AMR infections

---

CONTACT Jingheng Cai ✉ caijheng@mail.sysu.edu.cn Department of Statistics, Sun Yat-sen University, Guangzhou 510275, Guangdong Province, People's Republic of China

will exceed 10 million annually by 2050, with one person dying every 3 seconds. Africa and Asia will have 4.1 million and 4.7 million deaths from AMR infections each year, respectively, and this will result in a loss of over 100 trillion USD to the global healthcare system (Review on Antimicrobial Resistance, 2016). The World Bank predicts that by 2050, the high incidence rate of AMR will lead to an overall decline of 3.8% in global GDP, and will lead to 28.3 million people falling into extreme poverty, of which 26.2 million will be in low-income countries, which will make the goal of eradicating extreme poverty of the United Nations even more remote (Jonas et al., 2017). Therefore, it is essential to monitor the changes in AMR in the general population.

In Europe, a summary of AMR in zoonotic and indicator bacteria from humans and animals is reported by the European Centre for Disease Prevention and Control (ECDC) and the European Food Safety Authority (EFSA) every year. In the United States, the national surveillance system named the National Antimicrobial Resistance Monitoring System (NARMS), a national collaborative network of the U.S. Food and Drug Administration (FDA), the Centers for Disease Control and Prevention (CDC), the United States Department of Agriculture (USDA), as well as the public health laboratories in all 50 states and local health departments in three major cities, was developed to monitor changes in the antimicrobial susceptibility of certain foodborne bacteria found in animals, retail meats, and human illnesses.

The typical AMR data are the values of minimum inhibitory concentration (MIC) which is the minimum concentration of a specific antibiotic that will inhibit the growth of the bacteria. However, the observed MIC values are censored because only a discrete number of dilution levels are investigated and only concentrations above the true MIC will prevent growth of the isolate. For example, a bacterial isolate that is subjected to an antimicrobial at concentrations 1, 2, 4 μg/mL, shows inhibition of growth at 8 μg/mL. In such cases, if a 1 μg/mL dose of a drug does not inhibit growth but a 2 μg/mL dose does, then the observed MIC value is 1 whilst the true MIC value lies between 1 and 2. Therefore, an observed MIC value of 1 means that the true value of MIC is between 1 and 2 (interval censored), and so as 2, and 4. Moreover, given that the maximum record value is set to 8 μg/mL, an observed MIC value of 8 implies that the true value of MIC is larger than 8 (right censored). Cummings et al. (2016) pointed out that ignoring censorship would lead to biased estimates and be likely to overestimate the bacterial resistance to an antibiotic.

Another issue in modelling the distribution of MIC data is that bacteria isolates typically consist of a mixture of two components, indicating that the underlying distribution of the true MIC values consists of two overlapping components, which are called susceptible and resistant components (Zhang et al., 2020, 2021). Therefore, the analysis needs to account for unobserved population heterogeneity. Mixture models have been widely applied to solve this problem. Craig (2000) employed a Gaussian mixture model to handle the population heterogeneity of MIC. Jaspers, Lambert et al. (2016) proposed Bayesian mixture models with parametric and non-parametric components to estimate the distribution of MIC. Jaspers, Komárek et al. (2017) developed Bayesian multivariate mixture models with covariate-dependent mixing weights to monitor antimicrobial resistance.

The aforementioned researches only focus on analyzing the MIC data from the human population. However, AMR has increased in both human and animal populations (see, for example, Dutil et al., 2010; Hur et al., 2012; Voss-Rech et al., 2017, among others). It is also found that antibiotic-resistant bacteria can be passed from foodborne animals to humans through the food chain. For example, Van et al. (2012) pointed out that foodborne animals spread large numbers of antibiotic-resistant bacteria and bacteria that have not yet developed resistance but have resistance genes to humans through the food chain. Davis et al. (2015) revealed that retail meat is a potential vehicle for transmitting virulent, antibiotic-resistant K. pneumoniae from food animals to humans. Many classes of antimicrobials used in food-producing animals increase the likelihood of cross-resistance in human bacterial pathogens (Angulo et al., 2004); consequently, the ability to quantify the correlation in AMR levels between human and food-animal populations is an essential foundation for further studies on the transmission of antibiotic resistance across the populations (Zhang et al., 2021). Therefore, the analysis of cross-population correlation in AMR has been a hotspot issue in the field. Iwamoto et al. (2017) used Spearman rank correlation to assess the relationship between the annual proportion of ceftriaxone resistance among Salmonella isolates from humans, retail meats, and food animals. Zhang et al. (2021) proposed a Bayesian latent class mixture model for correlation analysis in antimicrobial resistance across populations. Yet the models proposed in the above literature assumed that the mean of MIC data does not change over time. Such assumptions are restrictive for many applications. Zhang et al. (2020) developed a latent class mixture model with a linear regression in the second level model to describe the time changes of MIC data; however, the model only concentrated on the susceptible isolates from the human population.

In this paper, we propose a hierarchical latent class mixture model with censorship for the detection of linear changes and correlation analysis across populations in antimicrobial resistance. At the first level, we propose mixture Gaussian distributions of MIC values for both human and animal populations. The correlation between the human and animal populations is taken into account in the second level model. As the mean values of MIC may not be constant over time (Zhang et al., 2020), linear regression with time as a covariate in the second level model is employed to deal with the change of MIC with time. Given the nature of mixture models and censoring, the observed likelihood function does not have a close form; consequently, it is hard to obtain the estimates of the unknown parameters via the maximum likelihood-based method. To solve this difficulty, we propose the sampling-based Bayesian method to obtain the estimates through the Markov chain Monte Carlo (MCMC) methods. To the best of our knowledge, no existing work has been developed to address all of the aforementioned issues jointly.

The remainder of this paper is organized as follows. Section 2 introduces the hierarchical Bayesian latent class mixture model. Section 3 proposes the Bayesian estimation procedure. The details of MCMC algorithms are also presented. Section 4 conducts a simulation study to assess the empirical performance of the proposed model and methodology. Section 5 applies the proposed model to analyze the data set from NARMS. Section 6 concludes the paper with a discussion. Appendix provides the full conditional distributions involved in the MCMC methods.

2. Model description

Let $y_{s,ti}$ be the observed value of $\log_2$ MIC for isolate i at time point t of population s, where $s = 1$ and 2 represent the human and animal populations, respectively; $t = 1, \ldots, T$ is the time index with T being the total number of time points; $i = 1, \ldots, n_{s,t}$ is the isolate index with $n_{s,t}$ being the number of observations for population s at time point t. Given that the observed value $y_{s,ti}$ is censored in practice, we introduce $y^*_{s,ti}$ as the true (latent) value of $\log_2$ MIC for isolate i time point t of population s. Since the observed MIC values are recorded as $\{1, 2, 4, 8, 16, \ldots\}$, the observed value of $\log_2$ MIC, that is $y_{s,ti}$, only takes values from

$\{0, 1, 2, 3, 4, \ldots\}$. In such cases, the relationship between observed (censored) variable $y_{s,ti}$ and true (latent) variable $y^*_{s,ti}$ is as follows:

$$y_{s,ti} = \begin{cases} \lfloor y^*_{s,ti} \rfloor, & \text{if } y_{s,ti} \text{ is interval censored,} \\ u, & \text{if } y_{s,ti} \text{ is right censored at } u, \end{cases} \tag{1}$$

where $\lfloor \cdot \rfloor$ is the floor function. Then a hierarchical Bayesian latent class mixture model (HBLCMM) is described as follows:

$$z_{s,ti} \sim \text{Bernoulli}(p_{st}), \tag{2}$$

$$y^*_{s,ti} | z_{s,ti} \sim \begin{cases} N(\beta_{st,0}, \sigma^2_{0s}), & z_{s,ti} = 0, \\ N(\beta_{st,1}, \sigma^2_{1s}), & z_{s,ti} = 1, \end{cases} \tag{3}$$

where $t = 1, \ldots, T$, $i = 1, \ldots, n_{s,t}$, and $s = 1$, 2 denoting the human and animal populations, respectively. The variable $z_{s,ti}$ is the hidden indicator of the bacterial component with $z_{s,ti} = 0$ and 1 representing susceptible and resistant components for human ($s = 1$) and animal ($s = 2$) populations, respectively. The parameters $\beta_{st,c}$'s are the expectations of $\log_2$ MIC for susceptible ($c = 0$) and resistant ($c = 1$) components of human ($s = 1$) and animal ($s = 2$) populations at time point t, respectively. Finally, σ^2_{cs}'s are the corresponding unknown variances.

To account for the dependence of the human and animal populations, the second-level model is proposed as follows. Let $\boldsymbol{\alpha}_t = (\alpha_{1t}, \alpha_{2t})^\top$ where α_{st} is the logit transformation of p_{st}, $\boldsymbol{\beta}_{t0} = (\beta_{1t,0}, \beta_{2t,0})^\top$ and $\boldsymbol{\beta}_{t1} = (\beta_{1t,1}, \beta_{2t,1})^\top$ be the vectors including the mean values of $\log_2$ MIC for susceptible and resistant components of human and animal populations at time point t, respectively. For $t = 1, \ldots, T$,

$$\alpha_{st} = \log\left(\frac{p_{st}}{1 - p_{st}}\right), \ s = 1, 2, \quad \text{and} \quad \boldsymbol{\alpha}_t = \begin{pmatrix} \alpha_{1t} \\ \alpha_{2t} \end{pmatrix} \sim N(\boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha), \tag{4}$$

$$\boldsymbol{\beta}_{t0} = \begin{pmatrix} \beta_{1t,0} \\ \beta_{2t,0} \end{pmatrix} \sim N(\boldsymbol{\mu}_{t0}, \boldsymbol{\Sigma}_0), \quad \boldsymbol{\beta}_{t1} = \begin{pmatrix} \beta_{1t,1} \\ \beta_{2t,1} \end{pmatrix} \sim N(\boldsymbol{\mu}_{t1}, \boldsymbol{\Sigma}_1), \tag{5}$$

where $\boldsymbol{\mu}_\alpha$ is the unknown vector, $\boldsymbol{\Sigma}_\alpha$ and

$$\boldsymbol{\Sigma}_c = \begin{pmatrix} \Sigma_{c,11} & \Sigma_{c,12} \\ \Sigma_{c,21} & \Sigma_{c,22} \end{pmatrix}, \quad c = 0, 1,$$

are the unknown covariance matrices. The correlations of the human and animal populations are $\rho_c = \Sigma_{c,12}/\sqrt{\Sigma_{c,11}\Sigma_{c,22}}$, $c = 0$ and 1, for susceptible and resistant components, respectively. Moreover, following Zhang et al. (2020), we incorporate a linear trend into the model to detect the linear pattern of the mean $\log_2$ MIC with the changing of time:

$$\boldsymbol{\mu}_{tc} = \begin{pmatrix} \mu_{tc,1} \\ \mu_{tc,2} \end{pmatrix} = \begin{pmatrix} \lambda_{c1,0} \\ \lambda_{c2,0} \end{pmatrix} + \begin{pmatrix} \lambda_{c1,1} \\ \lambda_{c2,1} \end{pmatrix} \times x_t, \quad c = 0, 1, \tag{6}$$

where x_t is the time-index of time point t and λ's are the unknown coefficients. Letting $\boldsymbol{\lambda}_{cs} = (\lambda_{cs,0}, \lambda_{cs,1})^\top$, $c = 0, 1$, $s = 1, 2$, and $\mathbf{x}^*_t = (1, x_t)^\top$, the above formula can be rewritten as

$$\mu_{tc,s} = \boldsymbol{\lambda}^\top_{cs}\mathbf{x}^*_t, \quad c = 0, 1, \ s = 1, 2. \tag{7}$$

3. Bayesian analysis

3.1. Prior distributions

Let $\boldsymbol{\theta} = \{\sigma_{01}^2, \sigma_{02}^2, \sigma_{11}^2, \sigma_{12}^2, \boldsymbol{\lambda}_{01}, \boldsymbol{\lambda}_{02}, \boldsymbol{\lambda}_{11}, \boldsymbol{\lambda}_{12}, \boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha\}$ be the vector containing all unknown parameters. Following the common practice in the Bayesian analyzes, we assign the independent non-informative prior distributions for each component of $\boldsymbol{\theta}$, as follows:

$$\sigma_{cs}^2 \sim \text{Inv-Gamma}(1, 0.05), \quad c = 0, 1,\ s = 1, 2, \tag{8}$$

$$p(\boldsymbol{\lambda}_{cs}) \propto \text{constant}, \quad c = 0, 1,\ s = 1, 2, \tag{9}$$

$$p(\boldsymbol{\Sigma}_c) \propto |\boldsymbol{\Sigma}_c|^{-1.5}, \quad c = 0, 1, \tag{10}$$

$$p(\boldsymbol{\mu}_\alpha) \propto \text{constant}, \tag{11}$$

$$p(\boldsymbol{\Sigma}_\alpha) \propto |\boldsymbol{\Sigma}_\alpha|^{-1.5}. \tag{12}$$

3.2. Posterior inference

Let $\mathbf{Y} = \{y_{s,ti} : s = 1, 2, t = 1, \ldots, T, i = 1, \ldots, n_{s,t}\}$ be the observations, and the Bayesian inference of $\boldsymbol{\theta}$ is based on the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{Y})$. However, given that the observations in $\mathbf{Y}$ are censored and the nature of the latent mixture model, the posterior distribution involves high dimensional integration and is not manageable. We employ data augmentation (Tanner & Wong, 1987) to attack this problem. More specifically, letting $\mathbf{Y}^* = \{y_{s,ti}^* : s = 1, 2, t = 1, \ldots, T, i = 1, \ldots, n_{s,t}\}$, $\mathbf{Z} = \{z_{s,ti} : s = 1, 2, t = 1, \ldots, T, i = 1, \ldots, n_{s,t}\}$, $\boldsymbol{\alpha}^* = \{\boldsymbol{\alpha}_t : t = 1, \ldots, T\}$, $\boldsymbol{\beta}^{*0} = \{\boldsymbol{\beta}_{t0} : t = 1, \ldots, T\}$, and $\boldsymbol{\beta}^{*1} = \{\boldsymbol{\beta}_{t1} : t = 1, \ldots, T\}$, observed data $\mathbf{Y}$ are augmented with the latent quantities, $(\mathbf{Y}^*, \mathbf{Z}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^{*0}, \boldsymbol{\beta}^{*1})$, in the posterior analysis. The joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{Y}^*, \mathbf{Z}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^{*0}, \boldsymbol{\beta}^{*1} \mid \mathbf{Y})$ does not involve high dimensional integrals. Then, a sufficiently large number of random samples are drawn from the joint posterior distribution $p(\boldsymbol{\theta}, \mathbf{Y}^*, \mathbf{Z}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^{*0}, \boldsymbol{\beta}^{*1} \mid \mathbf{Y})$. The Bayesian estimates and the standard error estimates can be obtained through the sample means and the sample covariance matrix. However, it is not straightforward to draw the samples from the joint posterior distribution. Markov chain Monte Carlo (MCMC) methods, that is, Gibbs sampler (Geman & Geman, 1984) and Metropolis-Hastings (MH) algorithm (Hastings, 1970; Metropolis et al., 1953), are needed. The full conditional distributions involved in the Gibbs sampler are presented in Appendix.

4. Simulation study

In this section, a simulation is conducted to examine the empirical performance of the proposed method. We took $n_1 = 180$, $n_2 = 150$, and $T = 48$ to represent 48 seasons (12 years) with $(x_1, \ldots, x_{48}) = (0, 0, 0, 0, 1, 1, 1, 1, 2, 2, 2, 2, \ldots, 11, 11, 11, 11)$ being the year-index. The latent values of $y_{s,ti}^*$'s were generated from the model described in Equations (2)–(6), and the observed (censored) value of $y_{s,ti}$ is obtained as follows:

$$y_{s,ti} = \begin{cases} \lfloor y_{s,ti}^* \rfloor, & \text{if } y_{s,ti}^* < 8 \text{ (interval censored)}, \\ 8, & \text{if } y_{s,ti}^* \geq 8 \text{ (right censored)}. \end{cases}$$

Table 1. Summaries of Bayesian estimates in the simulation study.

	$c = 0$				$c = 1$			
	$s = 1$		$s = 2$		$s = 1$		$s = 2$	
Par	BIAS	RMS	BIAS	RMS	BIAS	RMS	BIAS	RMS
σ_{cs}^2	0.085	0.115	0.051	0.065	0.003	0.016	0.018	0.050
$\lambda_{cs,0}$	−0.150	0.361	−0.018	0.286	−0.011	0.261	−0.014	0.253
$\lambda_{cs,1}$	0.007	0.048	−0.004	0.044	0.001	0.040	0.005	0.040

	$c = 0$		$c = 1$	
Par	BIAS	RMS	BIAS	RMS
$\Sigma_{c,11}$	0.339	0.465	0.167	0.309
$\Sigma_{c,12}$	0.066	0.221	0.091	0.233
$\Sigma_{c,22}$	0.123	0.262	0.191	0.294

The true values of the known parameters were chosen as follows: $\sigma_{01}^2 = 0.5, \sigma_{02}^2 = 0.3, \sigma_{11}^2 = 0.2, \sigma_{12}^2 = 0.4, \boldsymbol{\lambda}_{01} = (2, 0.5)^\top, \boldsymbol{\lambda}_{02} = (1, 0.3)^\top, \boldsymbol{\lambda}_{11} = (0, 0.3)^\top, \boldsymbol{\lambda}_{12} = (0, 0.2)^\top$, and

$$\boldsymbol{\mu}_\alpha = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \quad \boldsymbol{\Sigma}_\alpha = \begin{pmatrix} 1.0 & 0.3 \\ 0.3 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1 \end{pmatrix}, \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1.0 & 0.8 \\ 0.8 & 1 \end{pmatrix}.$$

We have tried a few test runs with different starting values and used the coda package (Plummer et al., 2006) in R to decide the number of burn-in iterations required for achieving convergence. The test runs showed that the MCMC algorithm converged within 2000 iterations. Hence, Bayesian results were obtained from 2000 MCMC samples after 2000 burn-in iterations. The simulation was conducted based on 100 replications. The summaries of the Bayesian estimates, including the averages of the bias (BIAS), and the root mean square error (RMS), of the important parameters are presented in Table 1. It can be seen that the proposed method can provide reasonable estimates of the unknown parameters.

5. Data application

5.1. Data description

Human data from the Centers for Disease Control and Prevention (CDC)

Centers for Disease Control and Prevention (CDC) released Antibiotic Resistance Threats in 2013, highlighting the threats posed by antibiotic-resistant germs on human health. CDC provides an interactive tool that contains antibiotic resistance data from bacteria isolated from humans as part of the NARMS (Centers for Disease Control & Prevention, 2022). Since human data from CDC are only available by year, we confirmed its monthly information from the National Center for Biotechnology Information (NCBI) database (National Library of Medicine, 2022). In CDC NARMS datasets, the largest number of the species is Salmonella enterica, which accounted for 60.8% by the end of 2021. We chose Salmonella enterica enteritidis as our research objective, because the serotype enteritidis accounts for 16.2% of the 51574 Salmonella enterica isolates from 1996 to 2021.

Food-producing animals data from the U.S. Food & Drug Administration (FDA)

The animal component of NARMS was initiated in 1997 and isolates were recovered from federally inspected slaughter and processing plants throughout the United States. The datasets can be downloaded from NARMS (U.S. Food & Drug Administration, 2022). The

Table 2. Bayesian estimates (Est) and the standard error estimates (SE) in the real analysis.

	$c=0$				$c=1$			
	$s=1$		$s=2$		$s=1$		$s=2$	
Par	Est	SE	Est	SE	Est	SE	Est	SE
σ^2_{cs}	0.188	0.046	0.136	0.032	0.140	0.051	0.110	0.026
$\lambda_{cs,0}$	2.283	0.771	2.465	0.687	0.779	0.913	1.663	0.923
$\lambda_{cs,1}$	0.059	0.399	0.178	0.306	0.619	0.456	0.290	0.412

	$c=0$		$c=1$	
Par	Est	SE	Est	SE
$\Sigma_{c,11}$	4.180	2.575	5.920	3.050
$\Sigma_{c,12}$	1.017	1.377	1.576	1.775
$\Sigma_{c,22}$	2.367	1.705	4.245	2.256

strain of Salmonella enterica enteritidis was selected, which accounts for 9.7% of the 32284 Salmonella enterica isolates from 2006 to 2018.

5.2. Real data analysis

Chloramphenicol (CHL) is a common antibiotic that is widely used in developing countries because of its low cost and significant antimicrobial capacity. Therefore, Salmonella enteritidis isolates tested by CHL are selected as our research objectives. To calculate the cross-populations correlation of the quarterly average of $\log_2$ MIC for the human and food-producing animal, both human and food-producing animal datasets were required to be compared in the same periods. Since the quarterly data from the human dataset can only be confirmed from 2014 to 2021 by the NCBI database, and information about isolates from the food-producing animals is not publicly available after 2018, all data collected before 2014 and after 2018 were removed and only isolates collected from 2014 to 2018 were retained. For the cases that range from 2014 to 2018, we also eliminated the data lacking quarterly information as well as MIC results tested by CHL. Finally, 327 human isolates and 2001 food-producing animal isolates were obtained for the following data analysis. The Bayesian estimates (Est) and standard error estimates (SE) obtained from 2000 MCMC samples after 2000 burn-in iterations are presented in Table 2.

The summaries of the results are as follows. (i) It can be calculated that $\rho_0 = \frac{1.017}{\sqrt{4.180\times 2.367}} = 0.323$ and $\rho_1 = \frac{1.576}{\sqrt{5.920\times 4.245}} = 0.314$. In addition, the posterior probability of $\rho_1 > \rho_0$ is about 0.5, which indicates that the correlations of the human and animal populations are close between susceptible and resistant components. (ii) The posterior probabilities of $\lambda_{11,1} > \lambda_{01,1}$ and $\lambda_{12,1} > \lambda_{02,1}$ are respectively 0.850 and 0.593, indicating that the linear trend pattern in the resistant component is stronger than that in the susceptible component of both human and animal populations.

We further used the complete Deviation Information Criteria (DIC; Celeux et al., 2006) to compare our proposed model with linear regression, M_l, with its counterpart, M_c, which assumes that the mean vectors $\boldsymbol{\mu}_{t0}$ and $\boldsymbol{\mu}_{t1}$ are not related to the time-index x_t. More specifically, all $\lambda_{cs,1}$'s are fixed at zero in M_c. The complete DIC values of M_l and M_c are 41128 and 57773, respectively, suggesting that M_l is the better model and also affirming the linear pattern of the mean MIC with the changing of time.

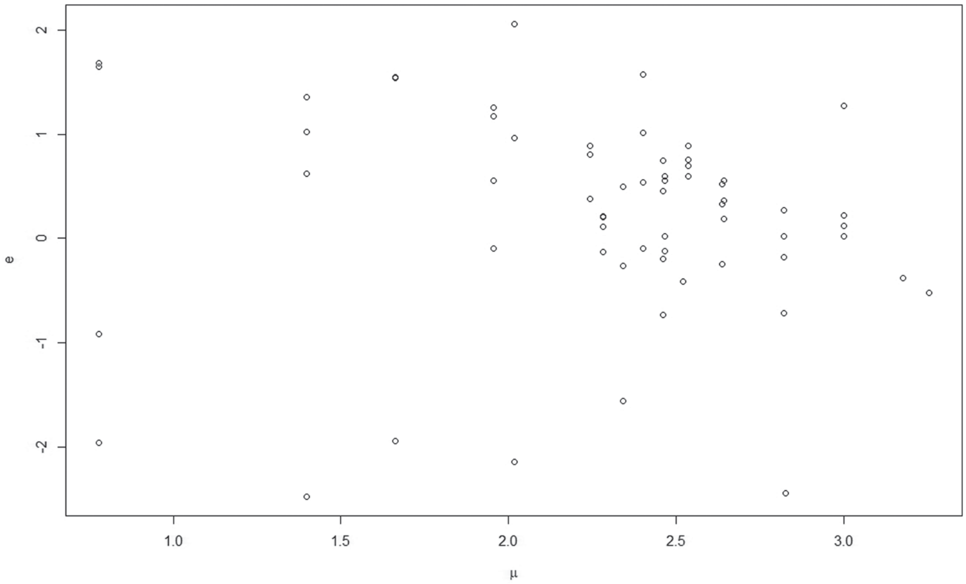

Figure 1. Residual plot.

Finally, we assess the goodness-of-fit of the proposed model by considering the residual plot. Let $\lambda_{cs,0}^{(l)}$, $\lambda_{cs,1}^{(l)}$, and $\beta_{st,c}^{(l)}$ be the MCMC samples of $\lambda_{cs,0}$, $\lambda_{cs,1}$, and $\beta_{st,c}$'s, respectively. Moreover, let $\mu_{tc,s}^{(l)} = \lambda_{cs,0}^{(l)} + \lambda_{cs,1}^{(l)} x_t$ and $e_{tc,s}^{(l)} = \beta_{st,c}^{(l)} - \mu_{tc,s}^{(l)}$ be MCMC samples of the predicted values of $\beta_{st,c}$ and the corresponding residuals, respectively. Figure 1 presents the residual plot generated by the sample means of $\mu_{tc,s}^{(l)}$ and $e_{tc,s}^{(l)}$. The residual plot shows that the model fits the data well.

6. Discussion

This paper proposes a hierarchical latent class mixture model with censorship for the detection of linear changes and correlation analysis across populations in antimicrobial resistance. Bayesian estimates of the unknown parameters are obtained via the MCMC methods.

Despite the above contributions, this study has limitations. First, we propose a linear model to describe the linear pattern of the mean $\log_2$ MIC with the changing of time. However, the effects of time may not be linear. A promising attempt is to employ a non-parametric model to assess the functional effects of time on the mean $\log_2$ MIC. Second, we assume that the true distribution of $\log_2$ MIC is normal. This normality assumption might be violated in practice. The future research interests will lie in relaxing the normality assumption with more sophisticated techniques. Finally, we employed a residual plot to evaluate the goodness-of-fit of the proposed model. It is important to develop an appropriate goodness-of-fit statistic to assess the model fit.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This paper is supported by the grants from National Natural Science Foundation of China [Grant no. 12426503].

ORCID

Jingheng Cai http://orcid.org/0000-0001-8267-3361

References

Angulo, F. J., Baker, N. L., Olsen, S. J., Anderson, A., & Barrett, T. J. (2004). Antimicrobial use in agriculture: Controlling the transfer of antimicrobial resistance to humans. *Seminars in Pediatric Infectious Diseases*, *15*(2), 78–85. https://doi.org/10.1053/j.spid.2004.01.010

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*(4), 651–674.

Centers for Disease Control and Prevention (2022). National Antimicrobial Resistance Monitoring System (NARMS) Now: Human Data. Accessed 2022, August 10. https://www.cdc.gov/narmsnow

Cohen, M. L. (2000). Changing patterns of infectious disease. *Nature*, *406*(6797), 762–767. https://doi.org/10.1038/35021206

Craig, B. A. (2000). Modeling approach to diameter breakpoint determination. *Diagnostic Microbiology and Infectious Disease*, *36*(3), 193–202. https://doi.org/10.1016/S0732-8893(99)00130-3

Cummings, K. J., Perkins, G. A., Khatibzadeh, S. M., Warnick, L. D., V. A. Aprea, & Altier, C. (2016). Antimicrobial resistance trends among *Salmonella* isolates obtained from horses in the northeastern United States (2001–2013). *American Journal of Veterinary Research*, *77*(5), 505–513. https://doi.org/10.2460/ajvr.77.5.505

Davis, G. S., Waits, K., Nordstrom, L., Weaver, B., Aziz, M., Gauld, L., Grande, H., Bigler, R., Horwinski, J., Porter, S., Stegger, M., Johnson, J. R., Liu, C. M., & Price, L. B. (2015). Intermingled *Klebsiella pneumoniae* populations between retail meats and human urinary tract infections. *Clinical Infectious Diseases*, *61*(6), 892–899. https://doi.org/10.1093/cid/civ428

Dutil, L., Irwin, R., Finley, R., Ng, L. K., Avery, B., Boerlin, P., Bourgault, A. M., Cole, L., Daignault, D., Desruisseau, A., Demczuk, W., Hoang, L., Horsman, G. B., Ismail, J., Jamieson, F., Maki, A., Pacagnella, A., & Pillai, D. R. (2010). Ceftiofur resistance in *Salmonella enterica* serovar Heidelberg from chicken meat and humans, Canada. *Emerging Infectious Diseases*, *16*(1), 48–54. https://doi.org/10.3201/eid1601.090729

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109. https://doi.org/10.1093/biomet/57.1.97

Hur, J., Jawale, C., & Lee, J. H. (2012). Antimicrobial resistance of *Salmonella* isolated from food animals: A review. *Food Research International*, *45*(2), 819–830. https://doi.org/10.1016/j.foodres.2011.05.014

Iwamoto, M., Reynolds, J., Karp, B. E., Tate, H., Fedorka-Cray, P. J., Plumblee, J. R., Hoekstra, R. M., Whichard, J. M., & Mahon, B. E. (2017). Ceftriaxone-resistant nontyphoidal *Salmonella* from humans, retail meats, and food animals in the United States, 1996–2013. *Foodborne Pathogens and Disease*, *14*(2), 74–83. https://doi.org/10.1089/fpd.2016.2180

Jaspers, S., Komárek, A., & Aerts, M. (2017). Bayesian estimation of multivariate normal mixtures with covariate-dependent mixing weights, with an application in antimicrobial resistance monitoring. *Biometrical Journal*, *60*(1), 7–19. https://doi.org/10.1002/bimj.v60.1

Jaspers, S., Lambert, P., & Aerts, M. (2016). A Bayesian approach to the semiparametric estimation of a minimum inhibitory concentration distribution. *The Annals of Applied Statistics*, *10*(2), 906–924. https://doi.org/10.1214/16-AOAS918

Jonas, O. B., Irwin, A., Berthe, F. C. J., Le Gall, F. G., & Marquez, P. V. (2017). *A Drug-Resistant Infections: A Threat to Our Economic Future (Vol. 2 of 2): Final report*. HNP/Agriculture Global Antimicrobial Resistance Initiative. World Bank Group. http://documents.worldbank.org/curated/en/323311493396993758

Mendelson, M., & Matsoso, M. P. (2015). The World Health Organization global action plan for antimicrobial resistance. *South African Medical Journal*, *105*(5), 325–325. https://doi.org/10.7196/SAMJ.9644

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machine. *The Journal of Chemical Physics*, *21*(6), 1087–1091. https://doi.org/10.1063/1.1699114

National Library of Medicine (2022). National Center for Biotechnology Information. Accessed 2022, October 31. https://www.ncbi.nlm.nih.gov/

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, *6*, 7–11.

Review on Antimicrobial Resistance (2016). *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*. Accessed 2016, May 19. https://wellcomecollection.org/works/thvwsuba

Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., & Edward, J. (2008). The epidemic of antibiotic-resistant infections: A call to action for the medical community from the infectious diseases society of america. *Clinical Infectious Diseases*, *46*(2), 155–164. https://doi.org/10.1086/524891

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–550. https://doi.org/10.1080/01621459.1987.10478458

U.S. Food & Drug Administration (2022). The national antimicrobial resistance monitoring system: Integrated reports/summaries. Accessed 2022, February 11. https://www.fda.gov/animal-veterinary/national-antimicrobial-resistance-monitoring-system/integrated-reportssummaries

Van, T. T. H., Nguyen, H. N. K., Smooker, P. M., & Coloe, P. J. (2012). The antibiotic resistance characteristics of non-typhoidal *Salmonella enterica* isolated from food-producing animals, retail meat and humans in South East Asia. *International Journal of Food Microbiology*, *154*(3), 90–106. https://doi.org/10.1016/j.ijfoodmicro.2011.12.032

Voss-Rech, D., Potter, L., Vaz, C. S. L., Pereira, D. I. B., Sangioni, L. A., Vargas, Á. C., & de Avila Botton, S. (2017). Antimicrobial resistance in nontyphoidal *Salmonella* isolated from human and poultry-related samples in Brazil: 20-year meta-analysis. *Foodborne Pathogens and Disease*, *14*(2), 116–124. https://doi.org/10.1089/fpd.2016.2228

Zhang, M., Wang, C., & O'Connor, A. M. (2020). A hierarchical Bayesian latent class mixture model with censorship for detection of linear temporal changes in antibiotic resistance. *PloS One*, *15*(1), e0220427. https://doi.org/10.1371/journal.pone.0220427

Zhang, M., Wang, C., & O'Connor, A. M. (2021). A Bayesian latent class mixture model with censoring for correlation analysis in antimicrobial resistance across populations. *BMC Medical Research Methodology*, *21*(1), 186. https://doi.org/10.1186/s12874-021-01384-w

Appendix. Full conditional distributions

A.1 Full conditional distribution of $z_{s,ti}$

For $s = 1, 2$, $t = 1, \dots, T$, and $i = 1, \dots, n_{s,t}$,

$$p(z_{s,ti} = 1 \mid \cdot) \propto \frac{\exp(\alpha_{st})}{1 + \exp(\alpha_{st})} \times \exp\left\{ -\frac{(y^*_{s,ti} - \beta_{st,1})^2}{2\sigma^2_{1s}} \right\}. \tag{A1}$$

A.2 Full conditional distribution of $y^*_{s,ti}$

For $s = 1, 2, t = 1, \ldots, T$, and $i = 1, \ldots, n_{s,t}$, $y^*_{s,ti}$ follows the following truncated normal distribution:

$$(y^*_{s,ti} \mid \cdot) \sim \begin{cases} N(\beta_{st,z_{s,ti}}, \sigma^2_{z_{s,ti},s})\mathrm{I}(y_{s,ti}, y_{s,ti} + 1), & \text{if } y_{s,ti} \text{ is interval censored,} \\ N(\beta_{st,z_{s,ti}}, \sigma^2_{z_{s,ti},s})\mathrm{I}(y_{s,ti}, +\infty), & \text{if } y_{s,ti} \text{ is right censored.} \end{cases} \tag{A2}$$

A.3 Full conditional distribution of β_{tc} and σ^2_{cs}

For $t = 1, \ldots, T, s = 1, 2$, and $c = 0, 1$,

$$(\beta_{st,c} \mid \cdot) \sim N(\beta^{\text{post}}_{st,c}, \sigma^{2\ \text{post}}_{st,c}), \tag{A3}$$

where

$$\begin{aligned} \sigma^{2\ \text{post}}_{1t,c} &= (\sigma^2_{1,c}/n^*_{1,tc} + \Sigma_{c,11} - \Sigma^2_{c,12}/\Sigma_{c,22})^{-1}, \\ \sigma^{2\ \text{post}}_{2t,c} &= (\sigma^2_{2,c}/n^*_{2,tc} + \Sigma_{c,22} - \Sigma^2_{c,21}/\Sigma_{c,11})^{-1}, \\ \beta^{\text{post}}_{1t,c} &= \left(\frac{\bar{y}^*_{1,tc}}{\sigma^2_{1,c}/n^*_{1,tc}} + \frac{\mu_{tc,1} + \Sigma_{c,12} \times (\beta_{2t,c} - \mu_{tc,2})/\Sigma_{c,22}}{\Sigma_{c,11} - \Sigma^2_{c,12}/\Sigma_{c,22}} \right) \times \sigma^{2\ \text{post}}_{1t,c}, \\ \beta^{\text{post}}_{2t,c} &= \left(\frac{\bar{y}^*_{2,tc}}{\sigma^2_{2,c}/n^*_{2,tc}} + \frac{\mu_{tc,2} + \Sigma_{c,21} \times (\beta_{1t,c} - \mu_{tc,1})/\Sigma_{c,11}}{\Sigma_{c,22} - \Sigma^2_{c,21}/\Sigma_{c,11}} \right) \times \sigma^{2\ \text{post}}_{2t,c}, \end{aligned}$$

with $n^*_{s,tc} = \sum_{i=1}^{n_{s,t}} \mathrm{I}(z_{s,ti} = c)$ and $\bar{y}^*_{s,tc} = \frac{1}{n^*_{s,tc}} \sum_{i=1}^{n_{s,t}} \mathrm{I}(z_{s,ti} = c) \times y^*_{s,ti}$.

For $c = 0, 1$ and $s = 1, 2$,

$$(\sigma^2_{cs} \mid \cdot) \sim \text{Inv-Gamma}\left[1 + \frac{1}{2} \sum_{t=1}^{T} n^*_{s,tc},\ 0.05 + \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{n_{s,t}} \mathrm{I}(z_{s,ti} = c) \times (y^*_{s,ti} - \beta_{st,c})^2 \right]. \tag{A4}$$

A.4 Full conditional distribution of $\boldsymbol{\Sigma}_c$

For $c = 0, 1$,

$$(\boldsymbol{\Sigma}_c \mid \cdot) \sim \text{Inv-Wishart}(T, \mathbf{W}_c), \tag{A5}$$

where $\mathbf{W}_c = \sum_{t=1}^{T} (\boldsymbol{\beta}_{tc} - \boldsymbol{\mu}_{tc})(\boldsymbol{\beta}_{tc} - \boldsymbol{\mu}_{tc})^\top$.

A.5 Full conditional distribution of α_t

For $t = 1, \ldots, T$,

$$p(\boldsymbol{\alpha}_t \mid \cdot) \propto \exp\left\{ -\frac{1}{2} (\boldsymbol{\alpha}_t - \boldsymbol{\mu}_\alpha)^\top \boldsymbol{\Sigma}_\alpha^{-1} (\boldsymbol{\alpha}_t - \boldsymbol{\mu}_\alpha) + \sum_{s=1}^{2} n^*_{s,t1} \alpha_{st} - n_{s,t} \log(1 + \exp(\alpha_{st})) \right\}. \tag{A6}$$

This is a non-standard distribution. MH algorithm (Hastings, 1970; Metropolis et al., 1953) is employed to generate the MCMC samples. The details are as follows: At the l-th iteration with a current value of $\boldsymbol{\alpha}_t^{(l)}$, a new candidate $\boldsymbol{\alpha}_t^*$ is generated from the proposal distribution $N(\boldsymbol{\alpha}_t^{(l)}, \boldsymbol{\Sigma}^*_\alpha)$ where

$$\boldsymbol{\Sigma}_\alpha^{*-1} = \begin{pmatrix} \dfrac{1}{4} \displaystyle\sum_{t=1}^{T} n_{1,t} & 0 \\ 0 & \dfrac{1}{4} \displaystyle\sum_{t=1}^{T} n_{2,t} \end{pmatrix} + \boldsymbol{\Sigma}_\alpha^{-1}.$$

Then the new candidate is accepted with the following probability

$$\min\left[1, \frac{p(\boldsymbol{\alpha}_t \mid \cdot)}{p(\boldsymbol{\alpha}_t^* \mid \cdot)}\right].$$

A.6 Full conditional distribution of μ_α and Σ_α

$$(\boldsymbol{\Sigma}_\alpha \mid \cdot) \sim \text{Inv-Wishart}(T-1, \mathbf{W}_\alpha^{-1}), \quad (\boldsymbol{\mu}_\alpha \mid \cdot) \sim N(\bar{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_\alpha / T), \tag{A7}$$

where

$$\bar{\boldsymbol{\alpha}} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{\alpha}_t, \quad \mathbf{W}_\alpha = \sum_{t=1}^{T} (\boldsymbol{\alpha}_t - \bar{\boldsymbol{\alpha}})(\boldsymbol{\alpha}_t - \bar{\boldsymbol{\alpha}})^\top.$$

A.7 Full conditional distribution of λ_{cs}

For $s = 1, 2$ and $c = 0, 1$,

$$(\boldsymbol{\lambda}_{cs} \mid \cdot) \sim N(\boldsymbol{\lambda}_{cs}^{\text{post}}, \boldsymbol{\Sigma}_{\lambda,cs}^{\text{post}}), \tag{A8}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_{\lambda,c1}^{\text{post}} &= (\Sigma_{c,11} - \Sigma_{c,12}^2 / \Sigma_{c,22}) \cdot \mathbf{W}_\lambda, \\
\boldsymbol{\Sigma}_{\lambda,c2}^{\text{post}} &= (\Sigma_{c,22} - \Sigma_{c,21}^2 / \Sigma_{c,11}) \cdot \mathbf{W}_\lambda, \\
\boldsymbol{\lambda}_{c1}^{\text{post}} &= \Sigma_{c,12} \Sigma_{c,22}^{-1} \boldsymbol{\lambda}_{c2} + \mathbf{W}_\lambda \sum_{t=1}^{T} \mathbf{x}_t^* \left(\beta_{1t,c} - \Sigma_{c,12} \Sigma_{c,22}^{-1} \beta_{2t,c}\right), \\
\boldsymbol{\lambda}_{c2}^{\text{post}} &= \Sigma_{c,21} \Sigma_{c,11}^{-1} \boldsymbol{\lambda}_{c1} + \mathbf{W}_\lambda \sum_{t=1}^{T} \mathbf{x}_t^* \left(\beta_{2t,c} - \Sigma_{c,21} \Sigma_{c,11}^{-1} \beta_{1t,c}\right),
\end{aligned}$$

with $\mathbf{W}_\lambda^{-1} = \sum_{t=1}^{T} \mathbf{x}_t^* \mathbf{x}_t^{*\top}$.