Bayesian modeling of the impact of antibiotic resistance on the efficiency MRSA decolonization

Fanni Ojala[1◕], Mohamad R. Abdul Sater[2◕], Loren G. Miller[3], James A. McKinnell[3,4], Susan S. Huang[5], Yonatan H. Grad[2‡*], Pekka Marttinen[1‡*]

1 Department of Computer Science, Aalto University, Espoo, Finland
2 Department of Immunology and Infectious Diseases, Harvard TH Chan School of Public Health, Boston, MA, USA
3 Lundquist Institute, Torrance, CA
4 Expert Stewardship, Newport Beach, CA
5 Division of Infectious Diseases, University of California Irvine School of Medicine, Irvine, CA, USA

◕These authors contributed equally to this work.
‡These authors also contributed equally to this work.
* ygrad@hsph.harvard.edu (YHG); pekka.marttinen@aalto.fi (PM)

Abstract

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a major cause of morbidity and mortality and one of the most serious infectious disease threats. Colonization by MRSA increases the risk of infection and transmission, underscoring the importance of decolonization efforts. However, success of these decolonization protocols varies, raising the possibility that some MRSA strains may be more persistent than others. Here, we studied how the persistence of MRSA colonization correlates with genomic presence of antibiotic resistance genes. Our analysis using a Bayesian mixed effects survival model found that high-level resistance to mupirocin was strongly associated with failure of the decolonization protocol. However, we did not see a similar effect with resistance to chlorhexidine or other antibiotics. Including strain-specific random effects improved the predictive performance, indicating that some strain characteristics other than resistance also contributed to persistence. Study subject-specific random effects did not improve the model. Our results highlight the need to consider the properties of the colonizing MRSA strain when deciding which treatments to include in the decolonization protocol.

Author summary

Methicillin-resistant *Staphylococcus aureus* (MRSA) is responsible for a high burden of morbidity and mortality. MRSA colonization incurs risk of MRSA infection and transmission, highlighting the need for highly effective decolonization protocols. However, decolonization protocols have had mixed success. The extent to which this mixed success is attributable to MRSA strain variations and their resistance to antibiotics, including those like mupirocin that are commonly used for decolonization, versus study subject factors, has been unclear. Here, we characterized the effect of antibiotic resistance genes on the efficiency of an MRSA decolonization protocol. We found that mupirocin resistance and strain-specific effects were associated with reduced effectiveness of an MRSA decolonization protocol, but that resistance to other

antibiotics, including purported chlorhexidine resistance genes, and subject-specific effects had no discernible impact on protocol success. Our results highlight the need to consider the properties of the colonizing MRSA strain when deciding which treatments to include in the decolonization protocol.

Introduction

Staphylococcus aureus colonizes approximately 30% of the population [1] and is a leading cause of healthcare and community associated infections [2]. Healthcare-associated infections with MRSA are associated with higher mortality rates as well as increased cost and hospitalization duration compared to patients infected with methicillin-susceptible *S. aureus*. MRSA carriers have a higher predisposition for infection with a 35% risk of MRSA infection within one year following colonization [3–6]. The anterior nares are the main reservoir of *S. aureus*, and the skin, particularly the axilla and groin, and pharynx are also often colonized. The risk of infection is correlated with the extent of colonization, as determined by the number of body sites found to be colonized [7]. MRSA infections are most often caused by the colonizing strain [8]. Infection prevention and control strategies include reducing spread by preventing colonization of new individuals as well as decolonization of MRSA carriers.

Decolonization reduces carriage rates and subsequent infection by 30% [9–12]. However, the effectiveness of decolonization protocols varies. The extent to which this variation is due to the protocol, the features of the MRSA strains colonizing the study subjects, characteristics of the colonized individuals, and the interaction among these factors has been unclear. Moreover, most studies have lacked appropriate controls and/or have had limited sample sizes [13]; additionally, most studies of decolonization protocols have had limited if any analysis of the colonizing MRSA strains [14].

The CLEAR (Changing Lives by Eradicating Antibiotic Resistance) Trial is a randomized controlled clinical trial of MRSA carriers comparing hygiene education to education plus decolonization after hospital discharge. The intervention arm underwent repeated decolonization involving a 5-day decolonization regimen applied every other week for six months. The decolonization regimen involved chlorhexidine antiseptic for daily showering and twice-daily mouthwash plus twice-daily mupirocin (a topical antibiotic) treatment of bilateral nares. The education (control) arm received hygiene education alone. Body site samples were collected five times: at enrollment, at one, three- and six-months post-enrollment during the intervention phase, and at nine months, which was three months after the end of intervention. Swabs were obtained from multiple body sites including nares, throat, skin (axilla and groin), as well as accessible wounds, if present. While the trial demonstrated the benefit of the decolonization protocol with chlorhexidine and mupirocin, persistent colonization was noted in a subset of both trial arms [15]. The factors that contributed to persistent colonization were not addressed in the primary manuscript.

The goal for the current investigation was to model the association between antibiotic resistance genes and the persistence of MRSA strains during a decolonization protocol. To study this, we used whole genome sequencing of 3901 isolates from 880 study subjects from CLEAR who had completed all follow-up visits. We used a Bayesian statistical framework (BaeMBac software) to define persistent strains [16] and formulated a Bayesian mixed effects survival model [17], where the survival outcome was the clearance of MRSA during a given study interval, and the lack of clearance represented persistent colonization. Resistances to different antibiotics as predicted by *Mykrobe predictor* [18] were the covariates in the fixed effects, and the study subject- and strain-specific random effects were included to quantify the impact of other subject and strain related factors on clearance. Our approach was fully Bayesian, which allowed

characterization of uncertainty of all quantities of interest and incorporation of prior knowledge [19].

Materials and methods

Data

See [9] for the details of the study protocol. In brief, subjects were selected for the study based on an MRSA positive culture within the hospitalization prior to enrollment (0-month). Isolates were collected from the study subjects at 1-month, 3-month, 6-month and 9-month visits from the start of the study. Per protocol, the decolonization regimen was stopped six months after discharge, and therefore data from the 9-month interval were excluded from the subsequent analysis. A resistance profile based on the presence or absence of resistance genes was estimated for each isolate using *Mykrobe predictor* [18], except for chlorhexidine for which BLAST [20] was used to detect the presence of *qac* genes as a marker for resistance.The mupirocin resistance is detected by *Mykrobe* using *mupA* and *mupB* genes and represents high-level resistance. Other information included the observation time, sequence type, hospital, study subject ID and body site of the swab. Subjects in the decolonization arm that did not fully adhere to the treatment were excluded. Contaminated isolates and isolates marked as MSSA were removed.

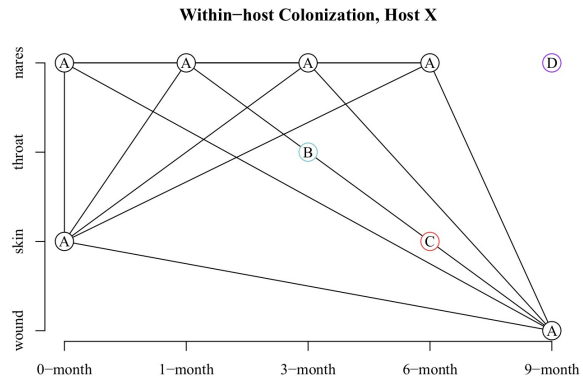
The MRSA positive isolates from a single study subject were divided into strains using the software BaeMBac [16], where a 'strain' is defined as a population of genetically closely related isolates. The software uses a Bayesian model based on the single nucleotide polymorphism (SNP) distance and time between consecutive visits to estimate the probability that a pair isolates collected from a study subject represent the same strain. The SNP distance of 45, estimated by BaeMBac using 10 percent of the education arm data, was used as a threshold (see S1 Fig) to divide the isolates from the same subject into strains. The MRSA isolates were primarily from ST5 (N=1337) and ST8 (N=1968) [16]; isolates from the remaining STs (N=533) were excluded because of the small number of samples. Most subjects were colonized with only one strain over the course of the study, but some were colonized with multiple strains (Table 1).

Table 1. Distribution of the number of strains colonizing a study subject.

#Strains	#Subjects	
	Decolonization	Education
1	160	287
2	24	78
3	1	27
4	1	4

Preprocessing

Our goal was to study the clearance of MRSA using survival models. For this purpose, we defined observations $(y_i, \mathbf{x}_i, t_i)$ in our survival data as follows. One observation i corresponded to one study interval (between consecutive follow-up visits) from one subject such that the subject was colonized by MRSA in the beginning of the interval. If the study subject cleared MRSA carriage during a given interval (e.g., from 1-month visit to 3-month visit), then $y_i = 1$, otherwise $y_i = 0$. The vector $\mathbf{x}_i$ specified the characteristics of the strain in the beginning of the interval, and it included the vector of indicators for resistance to different antibiotics. The time t_i was simply the length of



(a)

Strain-specific		
Interval		Observation
v_0	v_1	(y, x, t)
0-month	1-month	$(0, \mathbf{x}_A, 1)$
1-month	3-month	$(0, \mathbf{x}_A, 2)$
3-month	6-month	$(0, \mathbf{x}_A, 3)$
3-month	6-month	$(1, \mathbf{x}_B, 3)$

(b)

Site-specific			
Site	Interval		Observation
	v_0	v_1	(y, x, t)
Nares	0-month	1-month	$(0, \mathbf{x}_A, 1)$
Nares	1-month	3-month	$(0, \mathbf{x}_A, 2)$
Nares	3-month	6-month	$(0, \mathbf{x}_A, 3)$
Throat	3-month	6-month	$(1, \mathbf{x}_B, 3)$
Skin	0-month	1-month	$(1, \mathbf{x}_A, 1)$

(c)

Fig 1. Strains in a multiply-colonized study subject. a) An example of a study subject colonized by four separate strains, A, B, C, and D, over the study period. b) Observations in the *strain-specific* survival data formulation for the subject. The subject contributed three intervals to the survival data, since the 9-month visit was excluded. Strains B and C were cleared immediately after acquisition, whereas strain A was persistent throughout the study. c) Observations in the *site-specific* survival data formulation (see text for details).

the interval in months. We assessed clearance at 1-month, 3-month and 6-month visits. We denoted the starting visit by v_0 of the interval of interest (for example the 1-month visit) and v_1 the end visit (for example the 3-month visit).

The covariates $\mathbf{x}_i$ included the presence of genetic markers for resistance of the colonizing strain at v_0 to the following antibiotics: ciprofloxacin, clindamycin, erythromycin, gentamicin, mupirocin, rifampicin, tetracycline, trimethoprim and chlorhexidine. Penicillin and methicillin resistance were excluded, as all isolates were expected to be resistant. Vancomycin and fusidic acid were also excluded, because there was no resistance to these antibiotics. Statistics of the survival data are given in Tables 2 and 3.

Table 2. Summary of the survival data.

	Decolonization	Education
No. intervals (i.e., observations)	270	911
No. cleared intervals (y = 1)	173	401
No. subjects for the intervals	186	396
No. strains in the intervals	215	540

The data were formulated for survival analysis in two ways, as illustrated in Fig 1. First, in a strain-specific analysis, clearance was defined such that at v_1 the subject was

Table 3. Summary of the resistance profiles in the survival data.

Resistant to	Decolonization	Education
Ciprofloxacin	227 (84.1%)	797 (87.5%)
Clindamycin	103 (38.1%)	394 (43.2%)
Erythromycin	223 (82.6%)	816 (89.6%)
Gentamicin	22 (8.1%)	72 (7.9%)
Mupirocin	30 (11.1%)	74 (8.1%)
Rifampicin	5 (1.9%)	31 (3.4%)
Tetracycline	8 (3.0%)	33 (3.6%)
Trimethoprim	8 (3.0%)	29 (3.2%)
Chlorhexidine*	33 (12.2%)	114 (12.5%)

Number of observations (i.e., intervals) in the strain-specific survival data corresponding to different antibiotic resistance types. (*resistance to chlorhexidine is not well characterized, here we used qac genes as an indicator)

not colonized by the strain at any site (see Fig 1b for illustration). Furthermore, in the strain-specific analysis there may have been multiple colonizing isolates at v_0 belonging to the same strain, and the covariate corresponding to a certain resistance was defined as present if at least one of these isolates was resistant; in practice, the isolates of the same strain were so closely related that their predicted full resistance profiles were identical in 86% of cases. Second, in a site-specific analysis, clearance of a strain was defined as the absence of the strain at v_1 on a body site of interest (either no strain was observed at v_1 or a strain different from the one at v_0 was observed). If no swab was collected on v_1, the observation was excluded from the survival analysis, except if the MRSA-positive v_0 swab was taken from wound, when it was considered cleared by v_1 (wound healed), see Fig 1c. The preprocessing pipeline is summarized in S2 Appendix.

Bayesian Survival Model

In survival analysis, the goal is to characterize time-to-event data in terms of the hazard of event or survival time until an event, affected by some covariates of interest. In our study, the fixed covariates were the presence or absence of resistance to each antibiotic, and the model included also subject- and strain-specific random effects. The parameters of interest were the fixed effect coefficients β, which denoted the magnitude of increase or decrease in risk or survival time for the covariates. Hence, formally, we were modeling the 'hazard' of clearance of an MRSA strain in a study interval from v_0 to v_1 when the strain was known to be resistant to some antibiotics at v_0. Consequently, an estimated hazard ratio $exp(\beta)$ of 1.5, for example, indicated that a unit increase in the corresponding covariate resulted in a 1.5-fold risk of the clearance.

Our observations were either interval- or right censored: observations corresponding to study intervals where the clearance occurred ($y = 1$) between v_0 and v_1 were modeled as interval censored, as the exact event time of the clearance within the interval was not known. Right censoring was used for observations corresponding to study intervals in which the clearance event did not occur ($y = 0$) by the end of the interval v_1.

The data consisted of

$$D_i = (y_i, \mathbf{x_i}, t_i), \quad i = 1, \ldots, N, \tag{1}$$

where i is an index for a visit interval such that $i = 1, \ldots, N$ comprised all visit intervals from all participants in the data set where the subject was colonized at the beginning of the interval, according to either the strain- or site-specific formulation, as

described under the section *Preprocessing*. The response variable y_i indicated whether the clearance happened within the interval, t_i was the length of the interval, and vector $\mathbf{x_i}$ held the resistance profile at the beginning of the interval. The data in the decolonization and education arms were analysed separately. We used an exponential survival model with the proportional hazards parameterization. We assumed that the clearance rate was constant during a given interval. In addition, conditionally on the fixed covariates, the subject, the strain and the fact that the study subject was colonized in the beginning of the interval, the clearance probability in the interval was independent of clearances on other intervals (this assumption follows from the 'memorylessness' property of the exponential distribution). Hence, the hazard was given by

$$h_i = h_0 \exp(\eta_i), \quad i = 1, \ldots, N, \tag{2}$$

where η_i was the linear predictor, defined as

$$\eta_i = \beta^T \mathbf{x}_i + \gamma_{s(i)} + \rho_{h(i)}, \quad i = 1, \ldots, N. \tag{3}$$

In Eq 3, $\gamma_s, s = 1, \ldots, S$, and $\rho_h, h = 1, \ldots, H$, were the the strain- and subject-specific random effects and S and H were the numbers of strains and study subjects, respectively. Functions $h(i)$ and $s(i)$ specified the subject (i.e. 'host') and the strain corresponding to interval i, respectively. The priors for the random effects were defined as

$$\gamma_s \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\gamma^2), \quad \text{and} \quad \rho_h \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\rho^2), \quad \text{for all } s, h. \tag{4}$$

The survival function for interval i was thus

$$S_i(t_i) = \exp(-t_i h_i), \tag{5}$$

and it represented the probability that the study subject was still colonized by the same strain (i.e., "survived from clearance") at the end of the interval. By letting θ denote jointly all model parameters, the log-likelihood function was defined as

$$l(\theta) = \sum_{i=1}^{N} \{ I(y_i = 0) \log(S_i(t_i)) + I(y_i = 1) \log(1 - S_i(t_i)) \}. \tag{6}$$

The priors for the coefficient β for the fixed effects and the intercept term β_0 were defined to be relatively non-informative, i.e., to have a variance that exceeded the range of effects expected in the data, as follows:

$$\beta \sim N_D(0, 2.5I) \quad \text{and} \quad \beta_0 \sim N(0, 20), \tag{7}$$

where D was the dimension of the fixed effects and I the identity matrix. The priors on the hyperparameters for the random effects (σ_γ and σ_ρ) were determined from the decomposition of the covariance matrix of the random effects into a correlation matrix Ω, a simplex π, and a scale parameter τ. Details of this decomposition can be found in *rstanarm* documentation and the Stan user guide [21, 22]. We set the hyperparameters as follows:

$$\Omega \sim \text{LKJ}(\zeta = 1), \quad \pi \sim \text{Dirichlet}(\phi = 1), \quad \text{and} \quad \tau \sim \text{Gamma}(1, 1). \tag{8}$$

We estimated the posterior distributions for the parameters of interest by drawing samples from the posterior with an MCMC sampler, implemented using *rstanarm*'s function *stan_surv*. The R package *rstanarm* is an extension of the Stan programming language developed specifically as a platform for statistical analysis and Bayesian inference [21]. We ran the No-U-Turn sampler (a variant of Hamiltonian Monte

Carlo) [23] for 7500 iterations for the strain-specific models and for 10000 iterations for the site-specific models over four chains. We increased the target average acceptance probability in the presence of divergent transitions as suggested by *rstanarm* documentation [24]. We assessed the convergence with $\hat{R}$ values and by visual inspection of traceplots. The full model included all antibiotics and both strain and study subject random effects. We compared the full model with different random effect configurations using the 10-fold cross-validation (CV). The coefficients β were estimated separately for the decolonization and education arms. The coefficient β represents the logarithm of the hazard ratio, which here means that the 'instantaneous rate' of clearance happening for a resistant strain is exp() times the rate for non-resistant strains.

Results

Exploratory data analysis: Before estimating the hazard ratios using the Bayesian approach, we conducted an exploratory analysis by calculating the clearance probability at v_1 given resistance at v_0 directly from the observation counts (intervals in the survival data) for resistant and non-resistant strains. This approach only considered one type of resistance at a time and neglected the different interval lengths. We saw that only 32% of the mupirocin-resistant observations were cleared at v_1 in the decolonization arm while 67% of the non-resistant cases were cleared (Fig 2). The difference was significant with non-overlapping 95% confidence intervals. In the education arm, there was no difference in the clearance probability between resistant and non-resistant strains. Further, the clearance probability of the non-resistant strains is considerably larger in the decolonization arm than in the education arm, reflecting the overall efficiency of the protocol [9].

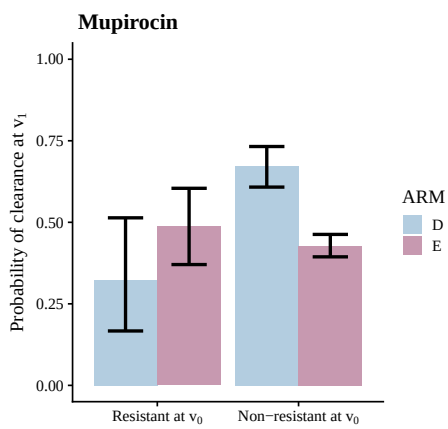


Fig 2. Clearance probabilities calculated from the counts of observations. Clearance probabilities given mupirocin resistance, computed directly from the counts of intervals colonized with resistant or non-resistant observations. On the y-axis, we have the clearance probability at the end of an interval, i.e., at v_1, and the x-axis shows the resistance status at v_0. The probability of clearance was calculated by dividing the numbers of persistent and cleared cases with the numbers of resistant or non-resistant observations in the data. The probability of clearance was lower for mupirocin-resistant strains than for non-resistant strains in the decolonization arm (D; blue). In the education arm (E; lavender), the probability of clearance (i.e., spontaneous loss of carriage) was the same regardless of the resistance status.

Model comparison: We used the 10-fold cross-validation to compare the

prediction accuracy of the different random effect combinations (no random effects, study subject, strain, study subject and strain). We quantified the results using the expected log-predictive density (`elpd`) [25], which is a metric for prediction accuracy. In both education and decolonization arms, including strain random effects improved the model considerably (Table 4). In contrast, including the subject-specific random effects did not improve the model, but instead slightly decreased the `elpd` value in the education arm. Because this decrease was minor and not significant, we decided to use the complete model to characterize both the strain and study subject random effects in the following section. The estimates for the fixed effects representing the impact of antibiotic resistance types on clearance were approximately the same regardless of whether the study subject random effects were included in the model or not (compare Fig 3 and S3 Fig).

Table 4. Model comparison.

Model	Decolonization		Education	
	elpd (mean)	elpd (SE)	elpd (mean)	elpd (SE)
Fixed	-208.0	9.6	-689.1	12.9
Fixed + Subject + Strain	-174.0	7.1	-599.4	14.5
Fixed + Subject	-174.1	7.5	-671.6	14.9
Fixed + Strain	-174.1	7.0	-598.7	14.4

A model selection criterion, *expected log-predictive density*, measuring the prediction accuracy of the model, estimated with 10-fold cross-validation (mean and standard error) for different model formulations. Larger values of `elpd` indicate a better model.

Impact of antibiotic resistance on persistence: In the decolonization arm, the mupirocin resistance coefficient was -2.6 (95% CI is -4.0 to -1.3), indicating that the clearance rate of resistant strains was approximately 0.07 times the clearance rate of the non-resistant strains (Fig 3). Mupirocin resistance thus was correlated with greater MRSA persistence in the decolonization arm. However, this effect was not observed in the education arm (i.e., spontaneous loss was similar regardless of resistance). In contrast, chlorhexidine-resistant strains were not more persistent in the decolonization arm than the non-resistant strains, despite the use of chlorhexidine mouth and body-wash as part of the decolonization protocol. Furthermore, ciprofloxacin (–0.71, 95% CI: [-1.31, -0.12]) and erythromycin (–0.97, 95% CI: [-1.67, -0.29]) resistances were weakly associated with increased persistence in the education arm. Resistance to other antibiotics was not significantly associated with clearance, but the number of samples corresponding to some resistance types was limited (see Table 3), leading to wide credible intervals.

Study subject and strain random effects: There was more variation in the strain random effects than in the study subject random effects in both the decolonization and education arms (Fig 4), which means that antibiotic resistance alone does not fully explain the variability in persistence. Furthermore, the variation in the strain random effects was larger in the education arm than in the decolonization arm. The study subject random effects were small in both arms. However, we note that many subjects were colonized by one strain only (see Table 1) and for those cases the effects of the strain and subject are statistically indistinguishable. Sequence type did not correlate with strain random effects (S4 Fig).

Site-specific analysis: In the decolonization arm, mupirocin resistance was again strongly associated with a reduced rate of clearance (i.e., increased persistence) in the nares (–2.26, 95% CI: [-3.8, -0.87]) but did not significantly correlate with clearance at other body sites (Fig 5, S5 Fig). Ciprofloxacin resistance and gentamicin resistance were weakly associated with increased persistence (–0.99, 95% CI: [-1.67, -0.32] and –1.12, [-2.06, -0.21]) in the nares in the education arm. In addition, we saw possible

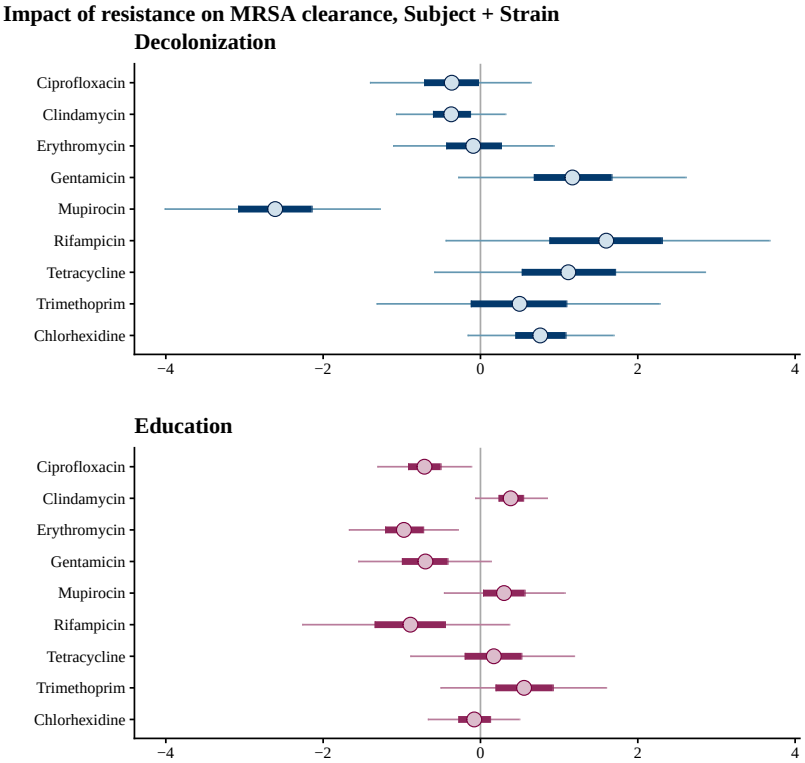


Fig 3. Credible intervals for the effects of antibiotic resistance types in the decolonization and education arms. 95% posterior credible intervals for the β parameters, representing the impact of each antibiotic resistance type on clearance. The model has study subject and strain random effects included and resistance types as fixed effects. A lower coefficient indicates a decreased rate (hazard) of clearance.

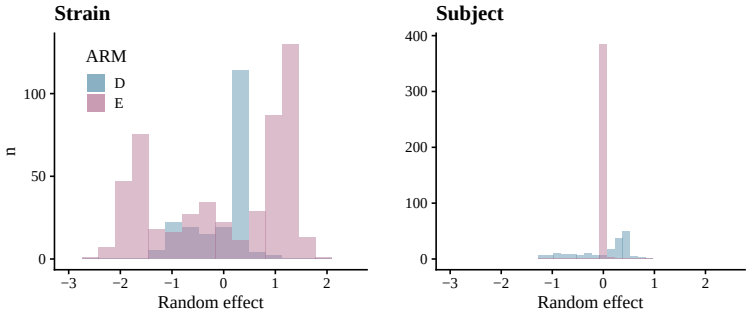


Fig 4. Estimated study subject and strain random effects. The figure shows histograms of the estimated strain and study subject-specific random effects. In both the decolonization (D) and education (E) arms, there was more variability in the strain random effects than in the study subject random effects.

weak associations between chlorhexidine resistance and decreased persistence in the throat in the decolonization arm (2.07, 95% CI: [0.13, 4.00]), and between tetracycline resistance and increased persistence in the wound in the education arm (–1.65, 95% CI: [-3.32, -0.11]) (S5 Fig). The variation in the strain random effects was again greater than in the subject random effects (S6 Fig). Furthermore, this effect was clearest in the

nares, from which most of the samples were obtained.

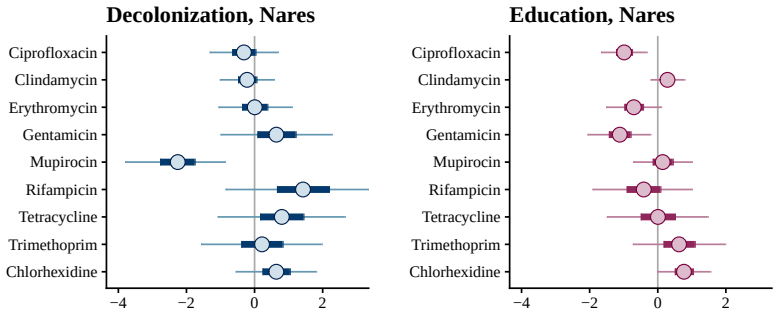


Fig 5. Results of the site-specific analysis The figure shows 95% credible intervals for the effect of each antibiotic resistance type on clearance in both study arms. The results for the nares are shown here, as it had the largest effect, and for the other sites in S5 Fig

Discussion

We applied Bayesian survival analysis on a dataset of sequenced MRSA samples collected from colonized patients after hospital discharge at given intervals during a follow-up period. Our results showed that mupirocin-resistant MRSA strains were more persistent than non-resistant strains in the decolonization arm, but not in the education only arm. When we looked at each body site separately, the effect of mupirocin was detected only in the nares, and not in the skin, throat, or wound. Since mupirocin is administered intranasally as part of the decolonization protocol and nares is the most prominent site of MRSA colonization [26], this result seems expected. However, despite chlorhexidine also being part of the decolonization protocol, chlorhexidine resistance did not seem to be associated with decolonization failure. This could be because chlorhexidine is applied to the throat (mouth wash), and skin and wound (baths), but not intranasally, but is most likely because genetic correlates of chlorhexidine resistance are quite poor. In the site-specific analysis, we saw a possible association between chlorhexidine and *decreased* persistence in the throat; however, the credible interval only slightly excluded zero, and hence this could be due to noise or otherwise reflect an unknown interaction between resistance types and persistence. Our study also confirmed the previously reported overall increased probability of clearance in the decolonization arm [9].

In the education arm, resistance to ciprofloxacin and erythromycin were found associated with increased persistence. A plausible biological mechanism is that some antibiotics had been used by the study subjects, giving resistant strains an advantage over non-resistant strains. A similar explanation could apply to the findings from the site-specific analysis, where we observed an increase in persistence in the education arm for gentamicin and ciprofloxacin-resistant strains in the nares and for tetracycline-resistant strains in wounds. Another possible explanation is that the order of causation is reversed: a strain that is persistent might be more likely to become antibiotic-resistant than a non-persistent strain, perhaps due to increased antibiotic exposure or shared biological mechanism. Further studies are needed to distinguish among these hypotheses.

Strain-specific random effects improved the ability of the model to predict persistence, which suggests that another genomic factor beyond the resistance

determinants may affect persistence. The sequence type (ST5 vs ST8) of the MRSA strain did not seem to be associated with persistence S4 Fig. Including the study subject-specific random effects in addition to the strain-specific effects did not improve the model, in contrast with reports that study subject related factors affect MRSA colonization (for example the nasal microbiota [27]). However, we note that some strain and subject-specific effects were overlapping (when there was only one strain from one subject), and consequently some subject effects might be explained as part of the strain effects.

In future studies, a larger dataset could confirm associations between resistance and persistence that did not reach statistical significance in this study. This could also help to identify in more detail the genetic determinants of persistence that were represented here by the strain-specific random effects. While the random effects for study subjects were not significant in our study, including explicit characteristics of the subjects might add power to find some features that affect persistence. In addition, the use of antibiotics other than those that were part of the protocol during the study period was not considered in this study, but including them as covariates might reveal further insights about the relationship between resistance and persistence.

Conclusion

We showed that genetic variants for mupirocin-resistance in MRSA were associated with a large drop in the efficiency of a decolonization protocol that includes intranasal mupirocin. Therefore, alternative decolonization protocols for patients with mupirocin-resistant MRSA colonization should be considered, such as nasal iodophor in place of mupirocin, although mupirocin is a superior treatment of the two in general [28]. However, we did not see a similar effect for chlorhexidine body and mouth wash, another part of the decolonization protocol, nor for any other antibiotic, which supports chlorhexidine as a reliable component of a skin decontamination protocol even when genetic correlates of chlorhexidine resistance are identified. In general, these findings point to the potential utility of improving the efficiency of decolonization protocols by characterizing an individual's colonizing strain to determine its resistance profile.

Supporting information

S1 Fig. BaeMBac output for same strain probability Contour plot detailing the same strain probability with SNP distance $d*$ on the x-axis and the time between consecutive visits (in generations) on the y-axis. The probability of 0.5 was used to decide the threshold distance of 45 SNPs that was used to classify a pair of MRSA isolates observed in consecutive visits as the same or different strain. The BaeMBas was run using 10 percent of randomly selected isolates from the education arm. The threshold was not sensitive to the amount of data, and the decolonization arm was not used as the BaeMBac software assumes a model of neutral evolution when calculating the same strain probability.

S2 Appendix. Preprocessing pipeline Preprocessing steps to format the data for the survival analysis. n_g represents the number of genomes (isolates), n_i the number of intervals and n_s the number of study subjects with sequencing data for the colonizing strains available. The number of observations is smaller after "Restructure data", because the survival data are considered by interval: recruitment to 1-month, 1-month to 3-month, 3-month to 6-month and 6-month to 9-month. For example, in the original data one body site could have an isolate at each of the visits, contributing in total five observations in the original data. However, we only have four intervals to consider in the case of survival data, because we are interested in the clearance status of an MRSA strain between or at the end interval of two consecutive isolates.

S3 Fig. Estimated fixed effects with only strain random effects in the model The 95% credible intervals for parameters β for all antibiotics with strain random effects included in the model, but excluding the subject-specific random effects.

S4 Fig. Estimated strain random effects for sequence types ST5 and ST8. The strain-specific survival model was used to estimate the strain random effects. Histograms show the distributions of the strain random effect posterior means in the decolonization and education arms.

S5 Fig. Throat, skin and wound posterior intervals 95% posterior intervals.

S6 Fig. Estimated study subject and strain random effects by site. Strain random effects have more variation in the posterior means than study subject random effects, most notably in the nares.

References

1. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. Staphylococcus aureus infections: Epidemiology, pathophysiology, clinical manifestations, and management. Clinical Microbiology Reviews. 2015;28(3):603–661. doi:10.1128/CMR.00134-14.

2. Sievert DM, Ricks P, Edwards JR, Mph AS, Patel J, Srinivasan A, et al. Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2009–2010. Source: Infection Control and Hospital Epidemiology. 2013;34(1):1–14. doi:10.1086/668770.

3. Huang SS, Rifas-Shiman SL, Warren DK, Fraser VJ, Climo MW, Wong ES, et al. Improving Methicillin-Resistant Staphylococcus aureus Surveillance and Reporting in Intensive Care Units. The Journal of Infectious Diseases. 2007;195(3):330–338. doi:10.1086/510622.

4. von Eiff C, Becker K, Machka K, Stammer H, Peters G. Nasal Carriage as a Source of Staphylococcus aureus Bacteremia . New England Journal of Medicine. 2001;344(1):11–16. doi:10.1056/NEJM200101043440102.

5. Huang SS, Hinrichsen VL, Datta R, Spurchise L, Miroshnik I, Nelson K, et al. Methicillin-resistant staphylococcus aureus infection and hospitalization in high-risk patients in the year following detection. PLoS ONE. 2011;6(9). doi:10.1371/JOURNAL.PONE.0024340.

6. McDanel JS, Murphy CR, Diekema DJ, Quan V, Kim DS, Peterson EM, et al. Chlorhexidine and mupirocin susceptibilities of methicillin-resistant Staphylococcus aureus from colonized nursing home residents. Antimicrobial Agents and Chemotherapy. 2013;57(1):552–558. doi:10.1128/AAC.01623-12.

7. Sim BLH, McBryde E, Street AC, Marshall C. Multiple Site Surveillance Cultures as a Predictor of Methicillin-Resistant Staphylococcus aureus Infections . Infection Control & Hospital Epidemiology. 2013;34(8):818–824. doi:10.1086/671273.

8. Huang SS, Diekema DJ, Warren DK, Zuccotti G, Winokur PL, Tendolkar S, et al. Strain-relatedness of methicillin-resistant Staphylococcus aureus isolates recovered from patients with repeated infection. Clinical Infectious Diseases. 2008;46(8):1241–1247. doi:10.1086/529381.

9. Huang SS, Singh R, McKinnell JA, Park S, Gombosev A, Eells SJ, et al. Decolonization to Reduce Postdischarge Infection Risk among MRSA Carriers. New England Journal of Medicine. 2019;380(7):638–650.

10. Climo MW, Yokoe DS, Warren DK, Perl TM, Bolon M, Herwaldt LA, et al. Effect of Daily Chlorhexidine Bathing on Hospital-Acquired Infection. New England Journal of Medicine. 2013;368(6):533–542.

11. Milstone AM, Elward A, Song X, Zerr DM, Orscheln R, Speck K, et al. Daily chlorhexidine bathing to reduce bacteraemia in critically ill children: a multicentre, cluster-randomised, crossover trial. The Lancet. 2013;381(9872):1099–1106. doi:10.1016/S0140-6736(12)61687-0.

12. Huang SS, Septimus E, Kleinman K, Moody J, Hickok J, Avery TR, et al. Targeted versus Universal Decolonization to Prevent ICU Infection. New England Journal of Medicine. 2013;368(24):2255–2265.

13. Golubchik T, Batty EM, Miller RR, Farr H, Young BC, Larner-Svensson H, et al. Within-Host Evolution of Staphylococcus aureus during Asymptomatic Carriage. PLoS ONE. 2013;8(5). doi:10.1371/JOURNAL.PONE.0061319.

14. Muthukrishnan G, Lamers RP, Ellis A, Paramanandam V, Persaud AB, Tafur S, et al. Longitudinal genetic analyses of Staphylococcus aureus nasal carriage dynamics in a diverse population. BMC Infectious Diseases. 2013;13(1). doi:10.1186/1471-2334-13-221.

15. Miller LG, Singh R, Eells SJ, Gillen D, McKinnell JA, Park S, et al. Chlorhexidine and Mupirocin for Clearance of Methicillin Resistant Staphylococcus aureus Colonization After Hospital Discharge: A Secondary Analysis of the CLEAR Trial. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2022;doi:10.1093/CID/CIAC402.

16. Järvenpää M, Abdul Sater MR, Lagoudas GK, Blainey PC, Miller LG, McKinnell JA, et al. A Bayesian model of acquisition and clearance of bacterial colonization incorporating within-host variation. PLoS Computational Biology. 2019;15(4):1–25. doi:10.1371/journal.pcbi.1006534.

17. Klein JP, Houwelingen HCV, Ibrahim JG, Scheike TH. Handbook of Survival Analysis; 2014.

18. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, et al. Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. Nature Communications. 2015;6:1–14. doi:10.1038/ncomms10063.

19. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis Third edition (with errors `[U+FB01]`xed as of 13 February 2020). 2013;(February):677.

20. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403–410. doi:https://doi.org/10.1016/S0022-2836(05)80360-2.

21. Gabry J, Brilleman S, Buros J, Wood S, Team RCD, Bates D. Bayesian Applied Regression Modeling via Stan. Package 'rstanarm'. 2020;.

22. Brilleman SL, Elci EM, Ag B, Wolfe R. Bayesian Survival Analysis Using the rstanarm R Package. 2020;.

23. Hoffman MD, Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research. 2014;15:1593–1623.

24. Gabry J, Goodrich B. How to Use the rstanarm Package • rstanarm; 2020. Available from: `https://mc-stan.org/rstanarm/articles/rstanarm.html`.

25. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing. 2017;27(5):1413–1432. doi:10.1007/s11222-016-9696-4.

26. Poyraz O, Sater MRA, Miller LG, McKinnell JA, Huang SS, Grad YH, et al. Modeling MRSA decolonization: Interactions between body sites and the impact of site-specific clearance. Journal of the Royal Society Interface. 2022; p. 2020.11.30.20239798. doi:10.1101/2020.11.30.20239798.

27. Lee AS, De Lencastre H, Garau J, Kluytmans J, Malhotra-Kumar S, Peschel A, et al. Methicillin-resistant Staphylococcus aureus. Nature Reviews Disease Primers. 2018;4. doi:10.1038/NRDP.2018.33.

28. Huang SS, Septimus E, Kleinman K, Heim L, Moody J, Avery TR, et al. 4. 137 Hospital Cluster-Randomized Trial of Mupirocin-Chlorhexidine vs Iodophor-Chlorhexidine for Universal Decolonization in Intensive Care Units (ICUs) (Mupirocin Iodophor Swap Out Trial). Open Forum Infectious Diseases. 2021;8(Supplement$_1$) : $S3 - -S4.doi : 10.1093/ofid/ofab466.004$.