medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Modeling MRSA decolonization: Interactions between body sites and the impact of site-specific clearance

Onur Poyraz
Computer Science
Aalto University
School of Science
onur.poyraz@aalto.fi

Mohamad R. A. Sater
Immunology and Infectious Diseases
Harvard TH Chan
School of Public Health
msater@hsph.harvard.edu

Loren G. Miller & James A. McKinnell
Infectious Diseases
Lundquist Institute
Harbor-UCLA Medical Center
{lmiller,jmckinnell}@lundquist.org

Susan S. Huang
Infectious Diseases
University of California Irvine
School of Medicine
sshuang@hs.uci.edu

Yonatan H. Grad*
Immunology and Infectious Diseases
Harvard TH Chan
School of Public Health
ygrad@hsph.harvard.edu

Pekka Marttinen*
Computer Science
Aalto University
School of Science
pekka.marttinen@aalto.fi

Abstract

Methicillin-resistant *Staphylococcus aureus* (MRSA) colonizes multiple body sites, and carriage is an important risk factor for MRSA infection. Successful decolonization reduces disease incidence; however, the optimal decolonization protocol is unclear. Decolonization protocols include diverse sets of treatments that target multiple body sites, and standard methods can not estimate the impact of site-specific components from such data. Here, we formulated a Bayesian coupled hidden Markov Model (CHMM) for MRSA colonization dynamics that estimates interactions between body sites and quantifies the contribution of each component on successful decolonization. Importantly, this now makes it possible 1) to predict the efficiency of decolonization for any combination of site-specific treatments, including combinations not included in the data, and 2) to assess the impact of enhancing site-specific clearance on the success of decolonization. We applied the model to longitudinal data from a randomized controlled trial (RCT) of MRSA decolonization, including chlorhexidine (CHG) body and mouthwash and nasal mupirocin. Our findings confirmed that nares are central to MRSA persistence and act as a predominant source of colonization for other body sites. We estimated the CHG mouthwash as the least effective component, such that the combined effect of clearance on other sites accounted for as much as 97% (90% credible interval 90%-99%) of the full decolonization. Hence, the omission of CHG mouthwash still yields a highly efficient decolonization regimen while facilitating adherence. Finally, we used our model to estimate the effectiveness of hypothetical treatment improvements *in silico* and showed that enhancing MRSA clearance at nares will achieve the largest gains. This study demonstrates the use of advanced probabilistic modeling to go beyond what is typically achieved by RCTs, enabling evidence-based decision-making to streamline clinical protocols.

Keywords— MRSA, Decolonization Protocol, Machine Learning, Coupled Hidden Markov Model, Markov Chain Monte Carlo

1 Introduction

Methicillin-resistant *Staphylococcus aureus* is a common antimicrobial-resistant pathogen in community and healthcare settings[1, 2], causing an estimated 320,000 infections in hospitalized patients and over 10,000 deaths in the United States in 2017[3]. Progress in reducing invasive MRSA infections has slowed, underscoring the importance of continued innovation and effort to prevent disease[4]. As MRSA carriage is a major risk factor for invasive disease, efforts

*These authors contributed equally.

Preprint. Under review.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

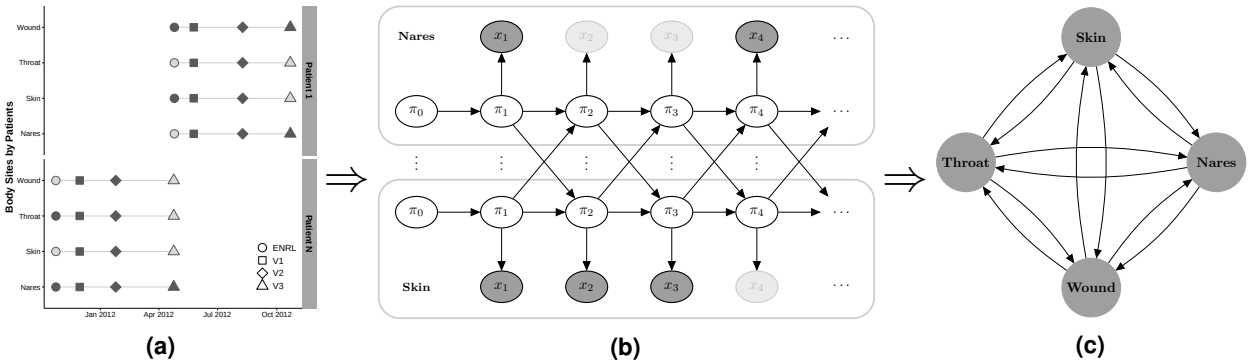


Figure 1 | Overview of the modeling strategy. (a) Visit records, (b) illustration of the coupled Hidden Markov Model (cHMM), and (c) estimated interactions between body sites. (a) An example with visit records for two participants. Visits approximately took place in 1 (V1), 3 (V2), and 6 (V3) months after enrollment (ENRL) in the trial. Filled markers correspond to the collected swabs, and faded markers represent samples missing due to trial exits or skipped visits[9]. **(b)** Here, for clarity, the CHMM is illustrated with two (nares and skin) of the four body sites (nares, skin, throat, and wound). In practice, the model includes all four sites. π_ts represent the unobserved true states (whether the site was colonized or not), and x_t represent observed states (whether MRSA was detected from the swab or not) at time t. The faded observation nodes correspond to missing observations, which the CHMM can naturally accommodate. The sequential model can be used to predict the duration of carriage. **(c)** The CHMM allows us to estimate and visualize interaction dynamics graphically, where edges represent the strength and direction of the interaction.

at prevention center on the promotion of decolonization and body hygiene as well as environmental cleaning[5]. The most common *S. aureus* carriage site is the anterior nares, but it can also colonize the perineum and groin, the axilla, the pharynx, as well as other body sites[6, 7]. While the anterior nares have been identified as a key reservoir for transmission and a risk factor for invasive disease[8], the extent of interaction among colonization sites and the importance of addressing each site in decolonization protocols remain unclear. Decolonization protocols primarily target the nares and can include other sites as well, but it would be ideal to simplify, where possible, without significantly impacting effectiveness. Achieving this goal requires a detailed understanding of the dynamic relationship among colonization sites and the extent to which colonization at one site affects colonization at other sites.

The CLEAR (Changing Lives by Eradicating Antibiotic Resistance) Trial demonstrated that the use of a post-discharge decolonization protocol in MRSA carriers reduces infection and hospitalization rates[9]. In the trial, 2,121 study participants were randomized into two groups to test the impact of the decolonization protocol: the *education* group (n=1,063) received an educational binder on hygiene, cleanliness, and MRSA transmission; the *decolonization* group (n=1,058) received this information and as well underwent decolonization for five days twice monthly for six months, with the protocol consisting of nasal mupirocin and chlorhexidine body and mouth wash. During these six months, swabs were collected from the participants at discharge from the hospitalization and three follow-up visits, which approximately took place at months 1, 3, and 6 after the discharge. Samples were taken from the nares, skin (axilla/groin), throat, and, if present, any wound. Participants had different numbers of observations because of trial exits or skipped visits. Overall, 20,506 samples were received in the first 6 months of follow-up. Additionally, the dataset includes reported adherence to the protocol, enabling the assessment of both real-world uptake and consideration of the ideal scenario of full compliance.

To model the interactions of MRSA colonization between body sites, we used an extension of a hidden Markov Model[10] (HMM, a well-known machine learning algorithm[11, 12]), a coupled HMM (CHMM)[13, 14, 15, 16, 17, 18, 19], where the probability of colonization at a given body site depends on the colonization of all body sites in the previous time step (Figure 1). Here, we developed two novel formulations: *Additive-CHMM* and *Or-CHMM*. In the former formulation, the probability of colonization at a given site is an additive function of colonization at the other sites. In the latter formulation, other body sites are considered jointly, and the probability of colonization in the body site of interest increases if *any* of the other sites is colonized. According to a predictive model selection criterion (leave-one-out cross-validation, LOO-CV)[20], the Additive-CHMM had superior accuracy compared to Or-CHMM or collection of site-specific standard HMMs (Supplementary Materials), and consequently, results of Additive-CHMM are presented and discussed in the main text. We provided an easy-to-use R-package that implements these models with an efficient Metropolis-within-Gibbs Markov Chain Monte Carlo (MCMC) algorithm, which yields Bayesian credible intervals (CI) for all model parameters (Methods).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

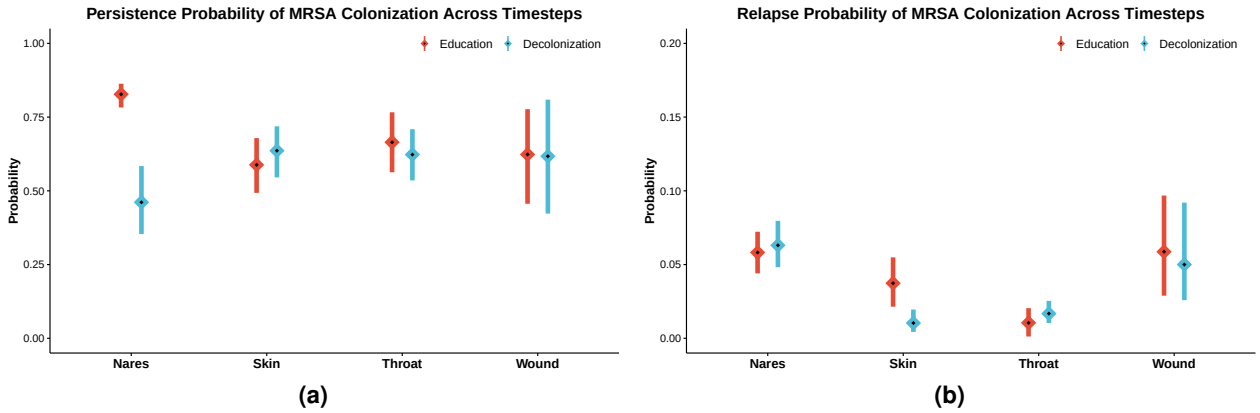


Figure 2 | Estimated (a) persistence and (b) relapse probability of MRSA colonization by body site. **(a)** Persistence probabilities are defined as the probability that a site will be colonized in the next time step when it is colonized in the current time step. **(b)** Relapse probabilities defined as the probability of a site getting colonized when it was not colonized in the previous time step. Persistence and relapse probabilities are calculated assuming other sites are not colonized. The calculation of the posterior distributions is explained in the Methods. Means and 90% credible intervals (CI) are represented in the figure by diamonds and lines, respectively.

2 Results

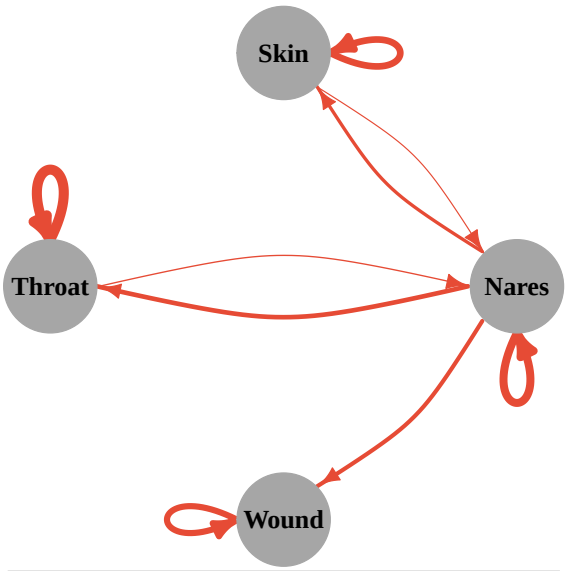
Nasal carriage of MRSA is the most persistent and most affected by the decolonization protocol. For each body site, we estimated the persistence of colonization by calculating the probability that the site would be colonized in the next time step when it was colonized in the previous time step, assuming other sites were not colonized. We found that in the education group, the estimated persistence was highest in the nares (Figure 2a). In the decolonization group, application of the protocol decreased persistence at each site, with the largest decrease in the nares[21] (Figure 2a). We also estimated the relapse probability, i.e., the probability of a site getting colonized when it was not colonized in the previous time step, assuming other sites were not colonized. However, this did not seem to be significantly affected by the decolonization protocol (Figure 2b). The number of samples from wounds was relatively small in the analysis, which caused the estimated parameters for the wound to have considerably wider CIs than estimates for the other sites. Repeating the analysis only on patients who reported adherence to the protocol led to almost identical results since most (70%) of the data were collected from adherent patients (Supplementary Materials).

The decolonization protocol reduces the transmission of MRSA between body sites. We quantified the strength of dependencies of MRSA colonization between body sites using the increase in the log-odds of colonization in the next time step caused by colonized sites in the previous time step (Figure 3). First, as expected, the most important determinant of colonization at any site was whether the site itself was colonized in the previous time step. Second, the nares was the only site linked to all the other sites in the education group, highlighting it as the likely source from which MRSA spreads to the other sites. Third, the decolonization protocol decreased the impact of the nares on the throat and removed the impact of the nares on wounds (Figure 3). Except for a minor increase in the impact of the nares on the skin, other dependencies between body sites were not affected by the treatment. These findings indicate two mechanisms responsible for the efficiency of the decolonization protocol: 1) the protocol decreased persistence in the nares, which is a central hub for colonization (Figure 2a), and 2) the protocol weakened links between body sites, reducing self-inoculation (Figure 3).

The CHMM accurately predicts the decrease in MRSA carriage over the study period. The key metric of success for decolonization is the extent to which it reduces MRSA carriage. To examine the CHMM's predictive accuracy, we simulated patient trajectories using parameters from the estimated models and compared the decrease in colonization in our model-based simulations to the observed decrease from the CLEAR trial data. The predicted reduction of the carriage was from 66% to 48% (90% CI is 45%-51%) in the education group and down to 26% (23%-28%) in the decolonization group (Figure 4). These predictions accurately match the reductions observed in the trial. However, a discrepancy between predictions and observations was seen in intermediate visits. Especially in the decolonization group, the observed carriage rate decreases more rapidly than predicted in the beginning and then more slowly. This outcome may reflect heterogeneity in treatment response among the study population: some patients responded very fast to the treatment, others more slowly.

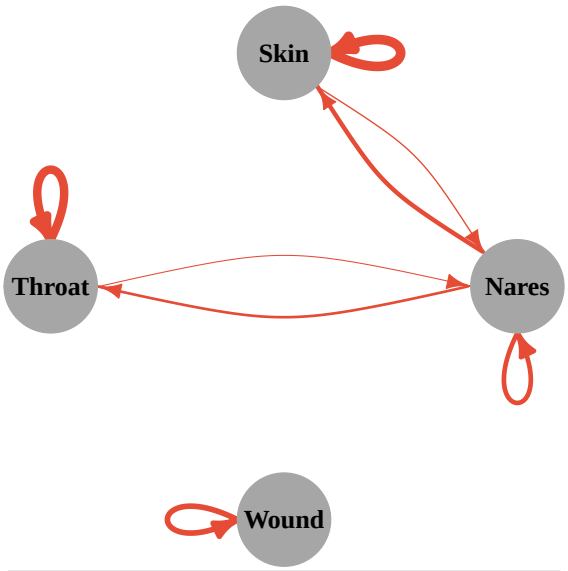
Treatments on nares, skin, and wound account for 97% of the effect of full decolonization. The importance of the nares has previously been noted in data from an intensive care unit[22]. However, the added value of decolonization

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



t \ t + 1		Nares	Skin	Throat	Wound
Nares	**mean** 90% CI	**4.4** [3.9, 4.8]	**1.8** [1.2, 2.6]	**2.5** [1.4, 4.5]	**2.3** [1.0, 3.4]
Skin	**mean** 90% CI	**0.6** [0.3, 0.9]	**3.6** [2.9, 4.5]	**0.6** [-0.1, 1.5]	**0.8** [-0.4, 2.0]
Throat	**mean** 90% CI	**0.7** [0.4, 1.1]	**0.2** [0.0, 0.4]	**5.5** [4.4, 7.5]	**0.0** [-0.8, 0.8]
Wound	**mean** 90% CI	**0.1** [-0.1, 0.3]	**0.0** [-0.1, 0.2]	**0.0** [-0.3, 0.1]	**3.4** [2.3, 4.4]

(a)



t \ t + 1		Nares	Skin	Throat	Wound
Nares	**mean** 90% CI	**2.6** [1.9, 3.2]	**2.4** [1.3, 3.5]	**1.5** [0.8, 2.1]	**-0.2** [-0.9, 0.2]
Skin	**mean** 90% CI	**0.8** [0.5, 1.0]	**5.2** [4.2, 6.2]	**0.8** [0.0, 1.3]	**-0.2** [-1.3, 0.4]
Throat	**mean** 90% CI	**0.6** [0.3, 0.9]	**0.3** [-0.1, 0.8]	**4.6** [3.9, 5.4]	**0.3** [-0.7, 2.0]
Wound	**mean** 90% CI	**0.1** [-0.1, 0.2]	**0.0** [-0.3, 0.3]	**0.0** [-0.3, 0.2]	**3.1** [1.3, 5.0]

(b)

Figure 3 | Estimated interactions among body sites in the (a) education and (b) decolonization groups. The red edges in the graphs show the increase in the log-odds of a site being colonized in the following time step if the linked site is colonized. Note there were no decreases in the log-odds. The edge thickness represents the expected change in the log-odds, and edges are excluded from the graph if the possibility for no effect is included in the 90% credible intervals (CI). Calculation of the log-odds is explained in the Methods. Tables show the mean and 90% CIs for all relations. We note that the number of wound samples was relatively small, which caused some wound-related parameters to have considerably wider CIs.

efforts that focus on body sites other than the nares is not well understood. To quantify the impact of decolonization efforts at each individual site, we predicted the efficiency of hypothetical simplified treatments, consisting of only a subset of the decolonization protocol (Figure 4). Nasal mupirocin decreased the estimated carriage from 66% to 31% (29%-34%), compared to 26% (23%-28%) of the full decolonization (Figure 4). Other single treatments were far less effective, reducing the carriage to 47% (44%-50%) with chlorhexidine mouthwash (CHG Oral) or 44% (41%-47%) with chlorhexidine bodywash (CHG Skin), a minor improvement compared to the 47% (44%-50%) of education alone. The best combination of two treatment components includes mupirocin and CHG Skin, i.e., leaving out CHG Oral, decreased carriage to 27% (25%-30%), corresponding 97% (90%-99%) of the decrease caused by the full decolonization protocol. In contrast, the combination of CHG Skin + CHG Oral and mupirocin + CHG Oral decreased the carriage only to 43%(40%-46%) and 30%(27%-33%), respectively (Suplementary Materials).

Enhancing clearance at the nares is predicted to achieve the largest gain in overall decolonization success. We used our model to predict the impact of future improvements to components of the decolonization protocol. We assumed optimal treatment components, defined as treatments that resulted in immediate clearance of a specific site, could be applied on top of the existing protocol. The results were consistent with the previous finding of the nares' central role in colonization and its clearance. In particular, the estimated final carriage with an optimized clearance of MRSA nasal carriage was 14% (13%-16%), while optimizing the throat or skin/wound treatments decreased the carriage to 21% (19%-23%) or 20% (18%-22%), respectively (Figure 4). Optimizing treatments both on nares + throat has the potential to reduce carriage further down to 10% (9%-11%) compared to 11% (10%-12%) of the nares + skin/wound, and 16% (14%-18%) of the throat + skin/wound (Suplementary Materials).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

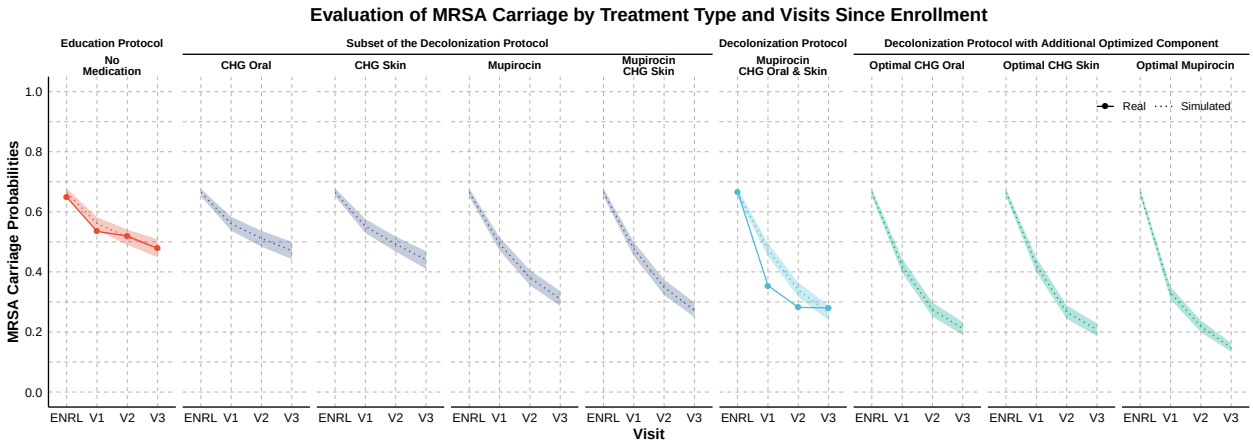


Figure 4 | The observed decrease in MRSA carriage over time compared to the decrease predicted by the model for actual and hypothetical treatments. In the trial, study subjects were in the education group or the decolonization group, which consisted of applying mupirocin to the nares, chlorhexidine mouthwash (CHG Oral) to the throat, and chlorhexidine body washes (CHG Skin) to the skin and, if present, wound. The predictions for hypothetical treatments assumed either that only a subset of medications was applied or that the decolonization protocol was used with an additional optimized intervention that resulted in immediate clearance of the corresponding target site. Dotted lines and shaded regions represent the mean and 90% credible intervals (CI) of simulations, and the solid lines represent values observed in the data.

3 Discussion

We presented a model for the dynamics and clearance of MRSA colonization. Our model estimated interactions between multiple body sites, disentangled the impact of individual treatment components on the overall efficiency of decolonization protocol, and accurately predicted the decrease in carriage percentage over the study period. We found that not all treatment components contribute equally to successful decolonization. Among the CLEAR Trial components, mupirocin was estimated as the most effective treatment component, and our results indicated that omission of chlorhexidine mouthwash still yields a highly effective decolonization regimen. More broadly, this analysis indicates that the impact of individual interacting components of complex clinical protocols and RCT data can be rigorously deconvoluted and assessed *in silico* by machine learning tools like those used here, thereby enabling the design of interventions that are more efficient, easier to adhere to, and thus more likely to be successful.

As with all modeling, our analysis made several simplifying assumptions. First, we only considered the presence or absence of MRSA, ignoring the type and the burden of MRSA[23]. However, heterogeneity in treatment responses among patients and strain type of MRSA might affect the colonization duration. Second, the analysis focused on individual MRSA clearance, but the decolonization protocol may have other benefits, i.e., reduction in transmission between patients[24, 25]. Third, our model assumed that patients had their follow-up visits at the same time intervals, while in practice, there was some variation. Fourth, we assumed that whether an observation was missing did not depend on the colonization status of the body site, which is likely not always correct; for example, if a previously colonized wound is healed and not colonized anymore, no additional samples would be taken from the wound. Finally, our analysis is based on a Bayesian approach and required a specification of prior probabilities on model parameters to express what is believed about their values before seeing the data. We used relatively uninformative priors, and a comparison included in the Suplementary Materials shows that the model fit was not sensitive to the exact values of the priors. Despite these limitations, the model successfully estimated interactions between body sites, quantified the impact of individual treatment components, and predicted the decrease in MRSA carriage over the study period. Additionally, it allowed us to estimate the sensitivity and specificity of the detection of MRSA colonization, as those are immediately available as the emission parameters of the CHMM (Suplementary Materials). This analytical approach thus demonstrates how data from randomized controlled trials can inform on biological processes as well as guide improvements in clinical protocols and decision-making.

4 Methods

Isolate Collection MRSA isolates were collected as part of the CLEAR (Changing Lives by Eradicating Antibiotic Resistance) Trial. The trial was designed to compare the impact of a repeated decolonization protocol plus education on general hygiene and environmental cleaning to education alone on MRSA infection and hospitalization [3]. Study

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

subjects in the trial were recruited from hospitalized patients based on an MRSA positive culture or surveillance swabs. After recruitment, swabs were obtained from different body parts of subjects (nares, skin, throat, and wound) around the time of hospital discharge (ENRL) and at 1,3,6 and 9 months (V1-V4, respectively) following the initial visit. The decolonization lasted only for 6 months, and consequently, we only modeled visits until V3 in this study. We note that some enrolled study subjects, despite a history of MRSA colonization or infection, did not have swabs positive for MRSA at the first time point (ENRL). More details about sample collection and data preprocessing can be found in the [3].

Hidden Markov Model Markov models are a well-known tool for time series analysis. Hidden Markov Models (HMM) are a special case where Markov process is observed indirectly through noisy observations. Discrete-time discrete-state Hidden Markov model with a latent sequence $\boldsymbol{\pi}$, $\pi_t \in \{1, \ldots, \mathrm{K}\}$ and observations $\mathbf{x}$, $x_t \in \{1, \ldots, \mathrm{L}\}$ for $t \in \{1, \ldots, \tau\}$, is defined as

$$p\left(\pi_{1:\tau}, x_{1:\tau}\right) = \underbrace{p\left(\pi_1\right)}_{\text{initial state}} \left[\prod_{t=2}^{\tau} \underbrace{p\left(\pi_t \mid \pi_{t-1}\right)}_{\text{transition}}\right] \left[\prod_{t=1}^{\tau} \underbrace{p\left(x_t \mid \pi_t\right)}_{\text{emission}}\right] \tag{1}$$

where K is number of latent states, τ is the number of time steps, and L the number of possible observed states. Parameters of the model are the initial state probability π_0, transition probability T, and emission probability E, defined as

$$\pi_0(i) = p(\pi_1 = i), \tag{2}$$
$$T(i,j) = p(\pi_t = j \mid \pi_{t-1} = i), \tag{3}$$
$$E(i,j) = p(x_t = j \mid \pi_t = i). \tag{4}$$

Note that the rows of T and E are probability distributions by definition. We will jointly denote the model parameters as $\theta = \{\pi_0, T, E\}$. If θ is known, then we can estimate the latent states π conditional on parameters θ and observations $\mathbf{x}$ using the forward-filtering backward-sampling algorithm.

Forward-Filtering Backward-Sampling Filtering is the estimation of the current hidden state by using all observations so far, $p\left(\pi_t \mid x_{1:t}\right)$ [12]. This is known as *forward-filtering* and it can be represented as follows:

$$p\left(\pi_t \mid \mathbf{x}_{1:t}\right) \propto \sum_{\pi_{t-1}} p\left(\pi_t, \pi_{t-1}, x_t \mid \mathbf{x}_{1:t-1}\right) \tag{5}$$
$$= p\left(x_t \mid \pi_t\right) \sum_{\pi_{t-1}} p\left(\pi_t \mid \pi_{t-1}\right) p\left(\pi_{t-1} \mid \mathbf{x}_{1:t-1}\right) \tag{6}$$

We define $\alpha\left(\pi_t\right) = p\left(\pi_t \mid \mathbf{x}_{1:t}\right)$ and substitute that into Equation 6, which yields the following recursive equation known as α-*recursion* or *forward recursion*;

$$\alpha\left(\pi_t\right) \propto \underbrace{p\left(x_t \mid \pi_t\right)}_{\text{corrector}} \underbrace{\textstyle\sum_{\pi_{t-1}} p\left(\pi_t \mid \pi_{t-1}\right) \alpha\left(\pi_{t-1}\right)}_{\text{predictor}} \tag{7}$$

The recursion starts with $\alpha\left(\pi_1\right) = p\left(\pi_1 \mid x_1\right) = p\left(x_1 \mid \pi_1\right) p\left(\pi_1\right) / p\left(x_1\right)$. This shows that the filtered distribution $\alpha(.)$ is propagated through to the next time-step, where it acts like a **'prior'**. In other words, at each time step the calculated posterior becomes the new prior for the next time-step [26].

To estimate the posterior distribution of latent states, we sample from the joint distribution of the latent sequence, $p(\boldsymbol{\pi}_{1:\tau} \mid \mathbf{x}_{1:\tau})$:

$$p(\boldsymbol{\pi}_{1:\tau} \mid \mathbf{x}_{1:\tau}) = p(\pi_\tau \mid \mathbf{x}_{1:\tau}) \prod_{t=1}^{\tau-1} p(\pi_t \mid \pi_{t+1}, \mathbf{x}_{1:t}) \tag{8}$$
$$\propto p(\pi_\tau \mid \mathbf{x}_{1:\tau}) \prod_{t=1}^{\tau-1} p(\pi_{t+1} \mid \pi_t) p(\pi_t \mid \mathbf{x}_{1:t}) \tag{9}$$
$$= \alpha\left(\pi_\tau\right) \prod_{t=1}^{\tau-1} p(\pi_{t+1} \mid \pi_t) \alpha\left(\pi_t\right) \tag{10}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

According to Equation 8, to be able to compute the posterior distribution of latent states, we need **'time-reversed'** transitions $p(\pi_t \mid \pi_{t+1}, x_{1:t})$, which are obtained using α-*recursion*s from the forward filtering. The simulation starts from the end of the sequence and proceeds backwards recursively. Starting point is

$$\hat{\pi}_\tau \sim \alpha(\pi_\tau) = p\left(\pi_\tau \mid \mathbf{x}_{1:\tau}\right) \tag{11}$$

For each time step t, the unnormalized sampling probability can be calculated using the previously sampled latent state $\pi_{t+1} = \hat{\pi}_{t+1}$, and the $\hat{\pi}_t$ can be sampled using following:

$$\hat{\pi}_t \sim p\left(\pi_{t+1} = \hat{\pi}_{t+1} \mid \pi_t\right) \alpha(\pi_t) \tag{12}$$

This procedure is known as forward-filtering backward-sampling[26].

Learning Model Parameters Conditionally on the latent states and observations the parameters $\theta = \{\pi_0, T, E\}$ of the model can be estimated. The posterior distribution $p(\pi_0, T, E \mid \mathbf{x}_{1:\tau}^{1:N})$ is given by

$$p(\pi_0, T, E \mid \mathbf{x}_{1:\tau}^{1:N}, \boldsymbol{\pi}_{1:\tau}^{1:N}) \propto p(\pi_0, T, E) p(\mathbf{x}_{1:\tau}^{1:N}, \boldsymbol{\pi}_{1:\tau}^{1:N} \mid \pi_0, T, E) \tag{13}$$

$$\propto \underbrace{p(\pi_0, T, E)}_{\text{prior}} \underbrace{\prod_{n=1}^{N} p(\mathbf{x}_{1:\tau}^{n}, \boldsymbol{\pi}_{1:\tau}^{n} \mid \pi_0, T, E)}_{\text{likelihood}}. \tag{14}$$

Factorizing the likelihood term in Equation 14 as in Equation 1, we can get the following result:

$$\prod_{n=1}^{N} p(\mathbf{x}_{1:\tau}^{n}, \boldsymbol{\pi}_{1:\tau}^{n} \mid \pi_0, T, E) = \prod_{n=1}^{N} \left\{ p\left(\pi_1^n \mid \pi_0\right) \prod_{t=2}^{\tau} p\left(\pi_t^n \mid \pi_{t-1}^n, T\right) \prod_{t=1}^{\tau} p\left(x_t^n \mid \pi_t^n, E\right) \right\} \tag{15}$$

$$= \prod_{k=1}^{K} \left\{ \pi_0(k)^{\pi_{0_k}^{\text{counts}}} \prod_{l=1}^{K} T(k,l)^{T_{k,l}^{\text{counts}}} \prod_{l=1}^{L} E(k,l)^{E_{k,l}^{\text{counts}}} \right\}. \tag{16}$$

Since, π_0, rows of the T, and rows of the E are probability distribution, [.]'s inside the Equation 16 corresponds to Multinomial distribution with parameters $\pi_0^{\text{counts}}, T_{k,:}^{\text{counts}}$ and $E_{k,:}^{\text{counts}}$, where π_0^{counts} corresponds to the initial state counts, T^{counts} corresponds to the state to state transition counts, and E^{counts} corresponds to the state to observation emission counts. Since the Dirichlet is conjugate prior for the Multinomial distribution, we set a Dirichlet prior for $\theta = \{\pi_0, T, E\}$ with hyperparameters $\pi_0^{\text{prior}}, T_{k,:}^{\text{prior}}$, and $E_{k,:}^{\text{prior}}$, respectively. Because of the conjugacy, the resulting posterior in Equation 14 is also a Dirichlet distribution:

$$p\Big(\pi_0 \mid \pi_0^{\text{counts}}, \pi_0^{\text{prior}}\Big) = Dirichlet\Big(\pi_0^{\text{counts}} + \pi_0^{\text{prior}}\Big), \tag{17}$$

$$p\left(T_{k,:} \mid T_{k,:}^{\text{counts}}, T_{k,:}^{\text{prior}}\right) = Dirichlet\left(T_{k,:}^{\text{counts}} + T_{k,:}^{\text{prior}}\right), \tag{18}$$

$$p\left(E_{k,:} \mid E_{k,:}^{\text{counts}}, E_{k,:}^{\text{prior}}\right) = Dirichlet\left(E_{k,:}^{\text{counts}} + E_{k,:}^{\text{prior}}\right). \tag{19}$$

Inference of HMM with MCMC Sampling There are several ways to estimate the parameters of HMM when both hidden states and model parameters θ are unknown, for example the expectation-maximization algorithm, variational Bayes, or MCMC sampling. In this paper, we will use MCMC. The goal of MCMC is to produce draws from the posterior $p(\boldsymbol{\pi}, \theta \mid \mathbf{x})$. The Gibbs sampler is an MCMC algorithm suitable for high dimensional problems, such as the HMM, and it samples parameters one-by-one using their distributions conditional on the other parameters. In the HMM, it will alternate between sampling the model parameters θ conditional on the latent sequence $\boldsymbol{\pi}$ from $p(\theta \mid \boldsymbol{\pi}, \mathbf{x})$, and the latent sequence $\boldsymbol{\pi}$ conditional on the model parameters θ from $p(\boldsymbol{\pi} \mid \theta, \mathbf{x})$. The algorithm iterates the following steps N times, which gives N draws from the posterior:

1. Sample the model parameters $\theta^* = \{\pi_0^*, T^*, E^*\}$ from Equation 17, Equation 18, and Equation 19, respectively. (Note that they are independent of each other given latent sequence.)
2. Sample the latent sequence of HMM using the updated model parameters θ using the Forward-Filtering Backward-Sampling algorithm.

Before calculating the posteriors, the first N_0 samples are discarded as a convergence (warm-up) period of the Markov chain. The plate diagram of the HMM is given in the Figure 5a, and the algorithm is described in Algorithm 1.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Algorithm 1 Hidden Markov Model

Require: Prior of π_0, T, E; Initialize $\boldsymbol{\pi}$ for each chain
Ensure: Posterior parameters

```
function GIBBS SAMPLER(x, π, π_0, T, E)
    for n in 1 : N do
        π*_0 ~ p(π_0 | π)
        T* ~ p(T | π)
        E* ~ p(E | π, x)
        π* ~ p(π | x, π*_0, T*, E*)
    end for
end function
```

Coupled Hidden Markov Model In the ordinary HMM, transition parameters do not change. However, in the CHMM, transitions in one chain are affected by other chains. In principle, it would be possible to define a single joint HMM where the latent state represents the latent states of all individual chains jointly and adapt the solution for the HMM described above. However, this is inefficient where there are many chains because the number of states would grow as $O(K^C)$, where K is the number of hidden states in a chain and C is the number of chains. On the high level, our key idea is that the transition matrix of each chain is modeled conditionally on the states of the other chains (and not as a single large joint transition matrix). A similar formulation is considered by [15]; however, the details are different. Note that a transition matrix T for a specific chain will not be time-independent anymore but changes at each time step depending on the states of the other chains. The graphical representation of the model (showing just two chains) is given in Figure 1, and the plate diagram of the CHMM is given in Figure 5b. There are $O(KC)$ parameters in this formulation, which is much more efficient for the increased number of chains.

Defining the transition matrix is a critical part of the algorithm, and it is essential to carry as much information about the other chains as possible. We model the dependencies between the chains with parameters $\boldsymbol{\beta}$. Assume there are C chains and let $\mathbb{C}$ denote the set of chains. Further, assume that each chain can be in one of K possible states, and let $\mathbb{K}$ denote the set of states. Finally, assume that there is a baseline state $\hat{k}$ such that if a chain is in that state, it does not affect other chains (in our application, this state corresponds to the absence of MRSA colonization in the respective body site). The transition matrix for the chain $\hat{c}$, at time t, denoted as $T_t^{[\hat{c}]}$, is defined by;

$$U_t^{[\hat{c}]} = \beta_0^{[\hat{c}]} + \sum_{c \in \{\mathbb{C} \setminus \hat{c}\}} \sum_{k \in \{\mathbb{K} \setminus \hat{k}\}} \beta_k^{[\hat{c} \leftarrow c]} \mathbb{I}\left[\pi_{t-1}^{[c]} = k\right], \text{ and} \tag{20}$$

$$T_t^{[\hat{c}]} = \sigma_{\text{row}}(U_t^{[\hat{c}]}). \tag{21}$$

The transition matrix $T_t^{[\hat{c}]}$ is obtained in Equation 21 by applying a row-wise softmax operator σ_{row} to the unnormalized transition matrix $U_t^{[\hat{c}]}$. Parameter $\beta_0^{[\hat{c}]}$ corresponds to an intercept matrix and it specifies the transition matrix of the target chain $\hat{c}$ when all other chains are in the baseline state $\hat{k}$. Parameter $\beta_k^{[\hat{c} \leftarrow c]}$ represents the impact of chain c on the target chain $\hat{c}$ and it is added to $\beta_0^{[\hat{c}]}$ whenever chain c was in state $k \neq \hat{k}$ in the previous time step. We will denote all the parameters for target chain $\hat{c}$ as $\boldsymbol{\beta}^{[\hat{c}]}$.

In Equation 20, the unnormalized transition probability U is an additive function of the latent states of the other chains; hence we call it the *Additive-CHMM*. We also implemented a simple formulation than the Additive-CHMM, which we call *Or-CHMM*. In the Or-CHMM, the target chain is not affected by the other chains individually; instead, there the target chain is affected whenever *any* of the other chains is in some state different from the baseline state, and the unnormalized transition matrix is given by:

$$U_t^{[\hat{c}]} = \beta_0^{[\hat{c}]} + \sum_{k \in \{\mathbb{K} \setminus \hat{k}\}} \beta_k^{[\hat{c}]} \mathbb{I}\left[\exists c \in \{\mathbb{C} \setminus \hat{c}\} : \pi_{t-1}^{[c]} = k\right] \tag{22}$$

In our application we have $K = 2$, corresponding to the presence and absence of colonization in a given body site. In this case, Equation 22 has a simple form: If all other chains are in state $\hat{k}$, the output is only $\beta_0^{[\hat{c}]}$, and if any of the other chains is in a state other $k \neq \hat{k}$, then the output is the sum of $\beta_0^{[\hat{c}]}$ and $\beta_k^{[\hat{c}]}$.

Design and Interpretation of the β parameters In the CHMM, the transition matrix T for each chain is modeled as a function of β parameters, where each β_k is a matrix of the same size as the transition matrix. Therefore, the β

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

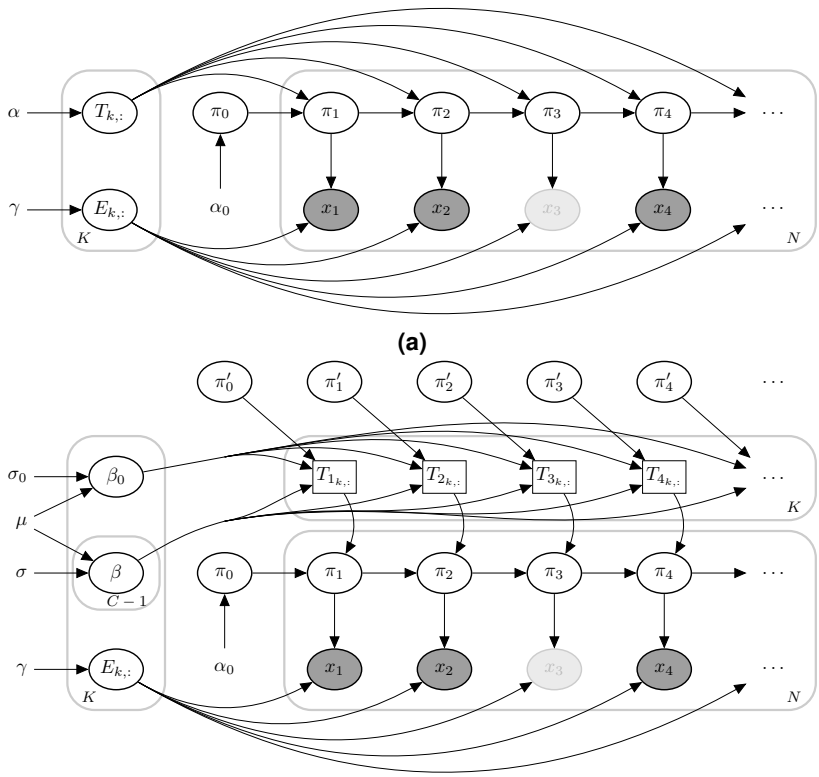


Figure 5 | Plate diagram of (a) Hidden Markov Model (HMM) and (b) Coupled Hidden Markov Model (CHMM). Here, for clarity, we illustrate only one chain. π_t represents the unobserved true states, and x_t represents the noisy observations at time t. The faded observation nodes correspond to missing values. Plain parameters are hyperparameters of the model. Edges from π_t to x_t represent *emission probabilities*. In the CHMM, the next state in one site depends on the previous states of the other sites. It is illustrated by π'. Therefore, *transition probabilities* change at each time step.

parameters are a list of matrices. Since the rows of a transition matrix T are assumed independent, we also model β parameters such that the rows of those matrices are independent. However, to avoid redundant parameters, the rows on each row of β are assumed to sum to zero. In practice, we sample the first $K-1$ parameter on a given row and set the Kth element to equal the negative of the sum of all the other parameters. This constraint ensures that the mapping from β to T is one-to-one.

To give an interpretation to the parameters β, we start with Equation 21 and write the softmax for a single element of a transition matrix $T(i,j)$:

$$T_t(i,j) = \frac{\exp\left(U_t(i,j)\right)}{\exp\left(U_t(i,j)\right) + \sum_{l\in\{\mathbb{K}\setminus k\}} \exp\left(U_t(i,l)\right)}, \tag{23}$$

from which it follows after straightforward algebra:

$$\text{logit}(T_t(i,j)) = U_t(i,j) - \log\left(\sum_{l\in\{\mathbb{K}\setminus k\}} \exp\left(U_t(i,j)\right)\right). \tag{24}$$

In our application $K=2$ and sums of the rows of β_k are assumed equal to 0, so consequently also the rows in the U_t sum to zero, which gives us:

$$\text{logit}(T_t(i,j)) = U_t(i,j) - U_t(i,\neg j) \tag{25}$$

$$= 2U_t(i,j), \tag{26}$$

where $\neg j$ refers to the other element on the row that is not j. Therefore, in this 2-dimensional case, all β_0 values correspond to half of the log-odds of the respective transition probability, and similarly, parameters β_k representing the interactions between the chains correspond to half of the change in the log-odds because of the presence of colonization in the other chain.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Adaptive Metropolis-Hastings Within Gibbs Sampling So far, we have described the design specifications of the CHMM. Since transitions are changing at each time step, it is not feasible to calculate their sufficient statistics as in Equation 16. To resolve this, we sample the $\beta^{[\hat{c}]*}$ using a Metropolis-Hastings(MH) step within the Gibbs sampler because conditional on the latent states of all the chains, the transition parameters are fully determined by the $\beta^{[\hat{c}]}$ parameters. To get a draw from $p\big(\beta^{[\hat{c}]} \mid \boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}\big)$, where $\boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}$ states of the other chains, we use the following steps;

1. Make a proposal $\beta^{[\hat{c}]*} \sim q\big(\beta^{[\hat{c}]}\big)$
2. Transform $\beta^{[\hat{c}]*}$ samples to transition probabilities $T_{1:\tau}^{[\hat{c}]*}$ according to latent states of the other chains, π'
3. Accept the proposal $\beta^{[\hat{c}]*}$ with probability

$$\alpha\left(T_{1:\tau}^{[\hat{c}]*}\right) = \min\left\{1, \frac{p\left(\beta^{[\hat{c}]*} \mid \boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}\right) q\left(\beta^{[\hat{c}]} \mid \beta^{[\hat{c}]*}\right)}{p\left(\beta^{[\hat{c}]} \mid \boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}\right) q\left(\beta^{[\hat{c}]*} \mid \beta^{[\hat{c}]}\right)}\right\} \tag{27}$$

where

$$p\left(\beta^{[\hat{c}]} \mid \boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}\right) \propto p\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]} \mid \beta^{[\hat{c}]*}, \boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}\right) p\left(\beta^{[\hat{c}]*}\right) \tag{28}$$

The first quantity on the right-hand side of the Equation 28 can be directly calculated with the transition parameters $T_{1:\tau}^{[\hat{c}]*}$ for corresponding chain. We used the Gaussian proposal in which the previously sampled $\beta^{[\hat{c}]}$ is the mean such that $q\big(\beta^{[\hat{c}]*} \mid \beta^{[\hat{c}]}\big) = N\big(\beta^{[\hat{c}]*} \mid \beta^{[\hat{c}]}, c^2\Sigma\big)$. As a variance parameter, we used a fixed diagonal covariance matrix during the warm-up period and then used the covariance matrix estimated from the previous samples. We initially set $c \approx 2.38/\sqrt{d}$ as metropolis jumping scaling factor since it is theoretically the most efficient scaling factor[27], where d is the dimension of the sampling. Then we adaptively scaled the scaling factor c as described in the [28]. Since this proposal distribution is symmetric (i.e., Normal), $q\big(\beta^{[\hat{c}]} \mid \beta^{[\hat{c}]*}\big)$ and $q\big(\beta^{[\hat{c}]*} \mid \beta^{[\hat{c}]}\big)$ in Equation 27, cancel out each other.

To sample the latent sequences we need to modify the Forward-Filtering Backward-Sampling algorithm[18], since we want samples from $p\big(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]} \mid \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}, \mathbf{x}_{1:\tau}^{[\hat{c}]}\big)$ instead of $p\big(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]} \mid \mathbf{x}_{1:\tau}^{[\hat{c}]}\big)$. Because in our application the within chain dependencies are stronger than between chain dependencies, we will use the Forward-Filtering Backward-Sampling with a modified model where the outgoing edges from the current chain are neglected as the proposal distribution q, and adjust for the discrepancy between the proposal and the true conditional distribution by using the following MH acceptance probability[18]:

$$\alpha\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]*}\right) = \min\left\{1, \frac{p\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]*}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]} \mid \mathbf{x}_{1:\tau}^{[\mathbb{C}]}\right) q\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]} \mid \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}, \mathbf{x}_{1:\tau}^{[\hat{c}]}\right)}{p\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]}, \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]} \mid \mathbf{x}_{1:\tau}^{[\mathbb{C}]}\right) q\left(\boldsymbol{\pi}_{1:\tau}^{[\hat{c}]*} \mid \boldsymbol{\pi}_{1:\tau}^{[-\hat{c}]}, \mathbf{x}_{1:\tau}^{[\hat{c}]}\right)}\right\} \tag{29}$$

where $p(.)$ distribution on the right-hand side of the Equation 29 is the joint posterior probability of the latent sequences, which is straightforward to calculate conditionally on the current CHMM parameters, and $q(.)$ is the probability of samples for particular chain that can be calculated during the backward sampling process (for details, see [18]). Algorithm 2 summarizes the whole algorithm. To validate the implementation, we used the algorithm to estimate parameters in simulated data. The results in Supplementary Figure 1 show that the algorithm estimated the parameters correctly and yielded well-calibrated posterior distributions.

Implementation details We set the initial covariance matrix for the metropolis proposal as $0.01 \times I$, where I is the identity matrix, which corresponds to a step size giving the optimal acceptance rate of $\approx 23\%$ [29]. We set the prior of β_0 as $N(\beta_0 \mid 0, 5)$, which is almost uninformative so that the estimates are not affected strongly by the prior. For the rest of the β parameters, denoted by β_k, we used sparsity encouraging Horseshoe prior with mean and scale parameters are 0 and 1, respectively (see Suplementary Materials for details). We used a uniform prior on the initial state probabilities π_0, and very weak Dirichlet priors for the rows of emission probabilities E such that we set the value to 0.9 if the latent state emits the same observation and set the value to $0.1/(n-1)$, n is the number of states, if the latent state emits a different observation. In such a formulation, except for the initialization, the prior has a negligible effect because it is summed with observation counts during inference. We drew 15,000 MCMC samples, and we set the warm-up length as 7,500. Posterior probabilities are calculated using the remaining MCMC samples. HMM and CHMM implicitly assume that time intervals between observations are the same, which is not the case in our data. Therefore, during the model training, we assumed that there are missing observations at 2, 4, and 5 months after enrollment.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Algorithm 2 Coupled Hidden Markov Model

Require: Prior of π_0, β, E; Initialize $\boldsymbol{\pi}$ for each chain
Ensure: Posterior parameters

```
function METROPOLIS-HASTINGS WITHIN GIBBS SAMPLER(x, π, π_0, β, E)
    for n in 1 : N do
        for ĉ in ℂ do
            π_0^[ĉ]* ~ p(π_0^[ĉ] | π^[ĉ])
            E^[ĉ]* ~ p(E^[ĉ] | π^[ĉ], x^[ĉ])
            β^[ĉ]* ~ Metropolis-Hastings(β^[ĉ]* | β^[ĉ])
            T^[ĉ]* ← σ_row(∑ β^[ĉ]*)
        end for
        for ĉ in ℂ do
            π^[ĉ]* ~ Metropolis-Hastings(π^[ĉ] | π^[−ĉ], x^[ĉ], π_0^[ĉ]*, T^[ĉ]*, E^[ĉ]*)
        end for
    end for
end function
```

Data Availability

The CLEAR (Changing Lives by Eradicating Antibiotic Resistance) Trial demonstrated that the use of a post-discharge decolonization protocol in MRSA carriers reduces infection and hospitalization rates; ClinicalTrial.gov number `NCT01209234`, see reference [9] for informed consent and institutional review board approvals. The data set used in the present article is available on request from the corresponding author Susan Huang of the [9], a co-author of this study.

Code Availability

The R package is available at https://github.com/onurpoyraz/chmmMCMC.

References

[1] Gorwitz, R. J. *et al.* Changes in the prevalence of nasal colonization with staphylococcus aureus in the united states, 2001–2004. *The J. infectious diseases* **197**, 1226–1234, DOI: 10.1086/533494 (2008).

[2] Weiner-Lastinger, L. M. *et al.* Antimicrobial-resistant pathogens associated with adult healthcare-associated infections: summary of data reported to the national healthcare safety network, 2015–2017. *Infect. Control. & Hosp. Epidemiol.* **41**, 1–18, DOI: 10.1017/ice.2019.296 (2020).

[3] for Disease Control, C. & (US), P. *Antibiotic resistance threats in the United States, 2019* (Centres for Disease Control and Prevention, US Department of Health and Human Services, 2019).

[4] Kourtis, A. P. *et al.* Vital signs: epidemiology and recent trends in methicillin-resistant and in methicillin-susceptible staphylococcus aureus bloodstream infections—united states. *Morb. Mortal. Wkly. Rep.* **68**, 214, DOI: 10.15585/mmwr.mm6809e1 (2019).

[5] Johnson, P. D. *et al.* Efficacy of an alcohol/chlorhexidine hand hygiene program in a hospital with high rates of nosocomial methicillin-resistant staphylococcus aureus (mrsa) infection. *Med. J. Aust.* **183**, 509–514, DOI: 10.5694/j.1326-5377.2005.tb07151.x (2005).

[6] Kluytmans, J., Van Belkum, A. & Verbrugh, H. Nasal carriage of staphylococcus aureus: epidemiology, underlying mechanisms, and associated risks. *Clin. microbiology reviews* **10**, 505–520, DOI: 10.1128/CMR.10.3.505 (1997).

[7] McKinnell, J. A., Huang, S. S., Eells, S. J., Cui, E. & Miller, L. G. Quantifying the impact of extranasal testing of body sites for methicillin-resistant staphylococcus aureus colonization at the time of hospital or intensive care unit admission. *Infect. control hospital epidemiology* **34**, 161–170, DOI: 10.1086/669095 (2013).

[8] Perl, T. M. *et al.* Intranasal mupirocin to prevent postoperative staphylococcus aureus infections. *New Engl. J. Medicine* **346**, 1871–1877, DOI: 10.1056/NEJMoa003069 (2002).

[9] Huang, S. S. *et al.* Decolonization to reduce postdischarge infection risk among mrsa carriers. *New Engl. J. Medicine* **380**, 638–650, DOI: 10.1056/NEJMoa1716771 (2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[10] Rabiner, L. R. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE* **77**, 257–286, DOI: 10.1109/5.18626 (1989).

[11] Murphy, K. P. *Machine learning: a probabilistic perspective* (MIT press, 2012).

[12] Barber, D. & Cemgil, A. T. Graphical models for time-series. *IEEE Signal Process. Mag.* **27**, 18–28, DOI: 10.1109/MSP.2010.938028 (2010).

[13] Brand, M., Oliver, N. & Pentland, A. Coupled hidden markov models for complex action recognition. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, 994–999, DOI: 10.1109/CVPR.1997.609450 (IEEE, 1997).

[14] Brand, M. Coupled hidden markov models for modeling interacting processes (1997).

[15] Sherlock, C., Xifara, T., Telfer, S. & Begon, M. A coupled hidden markov model for disease interactions. *J. Royal Stat. Soc. Ser. C (Applied Stat.* **62**, 609–627, DOI: 10.1111/rssc.12015 (2013).

[16] Rezek, I., Sykacek, P. & Roberts, S. J. Learning interaction dynamics with coupled hidden markov models. *IEE Proceedings-Science, Meas. Technol.* **147**, 345–350 (2000).

[17] Zhong, S. & Ghosh, J. A new formulation of coupled hidden markov models. *Dept. Elect. Comput. Eng., Univ. Austin, Austin, TX, USA* (2001).

[18] Touloupou, P., Finkenstädt, B. & Spencer, S. E. Scalable bayesian inference for coupled hidden markov and semi-markov models. *J. Comput. Graph. Stat.* **29**, 238–249, DOI: 10.1080/10618600.2019.1654880 (2020).

[19] Ghahjaverestan, N. M. *et al.* Coupled hidden markov model-based method for apnea bradycardia detection. *IEEE journal biomedical health informatics* **20**, 527–538, DOI: 10.1109/JBHI.2015.2405075 (2015).

[20] Gelman, A., Hwang, J. & Vehtari, A. Understanding predictive information criteria for bayesian models. *Stat. computing* **24**, 997–1016, DOI: 10.1007/s11222-013-9416-2 (2014).

[21] Von Eiff, C., Becker, K., Machka, K., Stammer, H. & Peters, G. Nasal carriage as a source of staphylococcus aureus bacteremia. *New Engl. J. Medicine* **344**, 11–16, DOI: 10.1056/NEJM200101043440102 (2001).

[22] Hall, M. D. *et al.* Improved characterisation of mrsa transmission using within-host bacterial sequence diversity. *Elife* **8**, e46402, DOI: 10.7554/eLife.46402 (2019).

[23] Datta, R. *et al.* High nasal burden of methicillin-resistant *Staphylococcus aureus* increases risk of invasive disease. *J. Clin. Microbiol.* **52**, 312–314, DOI: 10.1128/JCM.01606-13 (2014).

[24] Cepeda, J. A. *et al.* Isolation of patients in single rooms or cohorts to reduce spread of mrsa in intensive-care units: prospective two centre study. *The Lancet* **365**, 295–304, DOI: 10.1016/S0140-6736(05)17783-6 (2005).

[25] Huskins, W. C. *et al.* Intervention to reduce transmission of resistant bacteria in intensive care. *New Engl. J. Medicine* **364**, 1407–1418, DOI: 10.1056/NEJMoa1000373 (2011).

[26] Barber, D. *Bayesian reasoning and machine learning* (Cambridge University Press, 2012).

[27] Gelman, A. *et al. Bayesian data analysis* (CRC press, 2013).

[28] Sherlock, C., Fearnhead, P., Roberts, G. O. *et al.* The random walk metropolis: linking theory and practice through a case study. *Stat. Sci.* **25**, 172–190 (2010).

[29] Roberts, G. O., Rosenthal, J. S. *et al.* Optimal scaling for various metropolis-hastings algorithms. *Stat. science* **16**, 351–367, DOI: 10.1214/ss/1015346320 (2001).

[30] Watanabe, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010).

Author Contributions

YHG and PM perceived the study. OP, PM, and YHG designed the model. OP implemented the model and run the analyses. OP, SSH, YHG, and PM interpreted the results. MRAS, LGM, JAM, and SSH prepared the data. OP, PM, YHG wrote the manuscript with comments from all authors.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Materials

Model Comparison We compared three models: the site-specific Standard-HMMs as a baseline (i.e., modeling each site independently with a different HMM), the Or-CHMM, and the Additive-CHMM model using six different initial parameterizations. First, we considered prior $N(\beta_k \mid 0, 5)$, as in β_0, and prior $N(\beta_k \mid 0, 1)$, to see how sensitive the performance is to the strength of the prior. In addition to these Gaussian priors, we also considered the Laplace and Horseshoe distributions, which are sparse priors that encourage values close to zero. Note that β_0 always had the same uninformative Gaussian prior because the β_0 values represent the impact of the history of a chain on its future values, which is expected to be non-sparse. On the other hand, β_k values correspond to the effects of the other chains, which may be zero in practice.

We compared the models with three different information criteria (IC): 1) The Akaike Information Criteria (AIC), which is the summation of the negative log-likelihood of the model and the number of parameters, 2) the Watanabe-Akaike Information Criteria (WAIC), also known as the Widely applicable information criteria, which is the generalized version of the AIC onto singular statistical models [30] and 3) the leave-one-out cross-validation (LOO-CV), which measures model performance by estimating pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values[20]. WAIC and LOO-CV are more computationally intensive than AIC. Still, they are theoretically preferred over AIC for Bayesian models[20]. For all criteria, a lower score is better. Supplementary Table 1 presents all the scores used to evaluate model performance.

IC	AIC			WAIC			LOO-CV		
Prior \ Model	Standard HMMs*	Or CHMM	Additive CHMM	Standard HMMs*	Or CHMM	Additive CHMM	Standard HMMs*	Or CHMM	Additive CHMM
Gaussian $\sigma = 1$	10790	10693	10492	10787	10729	10657	10787	10730	10653
Gaussian $\sigma = 5$		10690	10486		10730	10659		10730	10655
Laplace $\sigma = 1$		10673	10488		10715	10651		10716	10646
Laplace $\sigma = 5$		10700	**10486**		10735	10659		10735	10653
Horseshoe $\sigma = 1^{\dagger}$		10676	10499		10715	**10646**		10716	**10643**
Horseshoe $\sigma = 5$		10704	10492		10738	10662		10739	10656

(a)

IC	AIC			WAIC			LOO-CV		
Prior \ Model	Standard HMMs*	Or CHMM	Additive CHMM	Standard HMMs*	Or CHMM	Additive CHMM	Standard HMMs*	Or CHMM	Additive CHMM
Gaussian $\sigma = 1$	8716	8637	8503	8714	8662	8599	8714	8662	8598
Gaussian $\sigma = 5$		8644	8517		8666	8610		8666	8609
Laplace $\sigma = 1$		8667	**8501**		8678	8604		8679	8603
Laplace $\sigma = 5$		8645	8502		8670	8607		8670	8606
Horseshoe $\sigma = 1^{\dagger}$		8670	8514		8679	**8593**		8679	**8592**
Horseshoe $\sigma = 5$		8656	8503		8671	8600		8671	8600

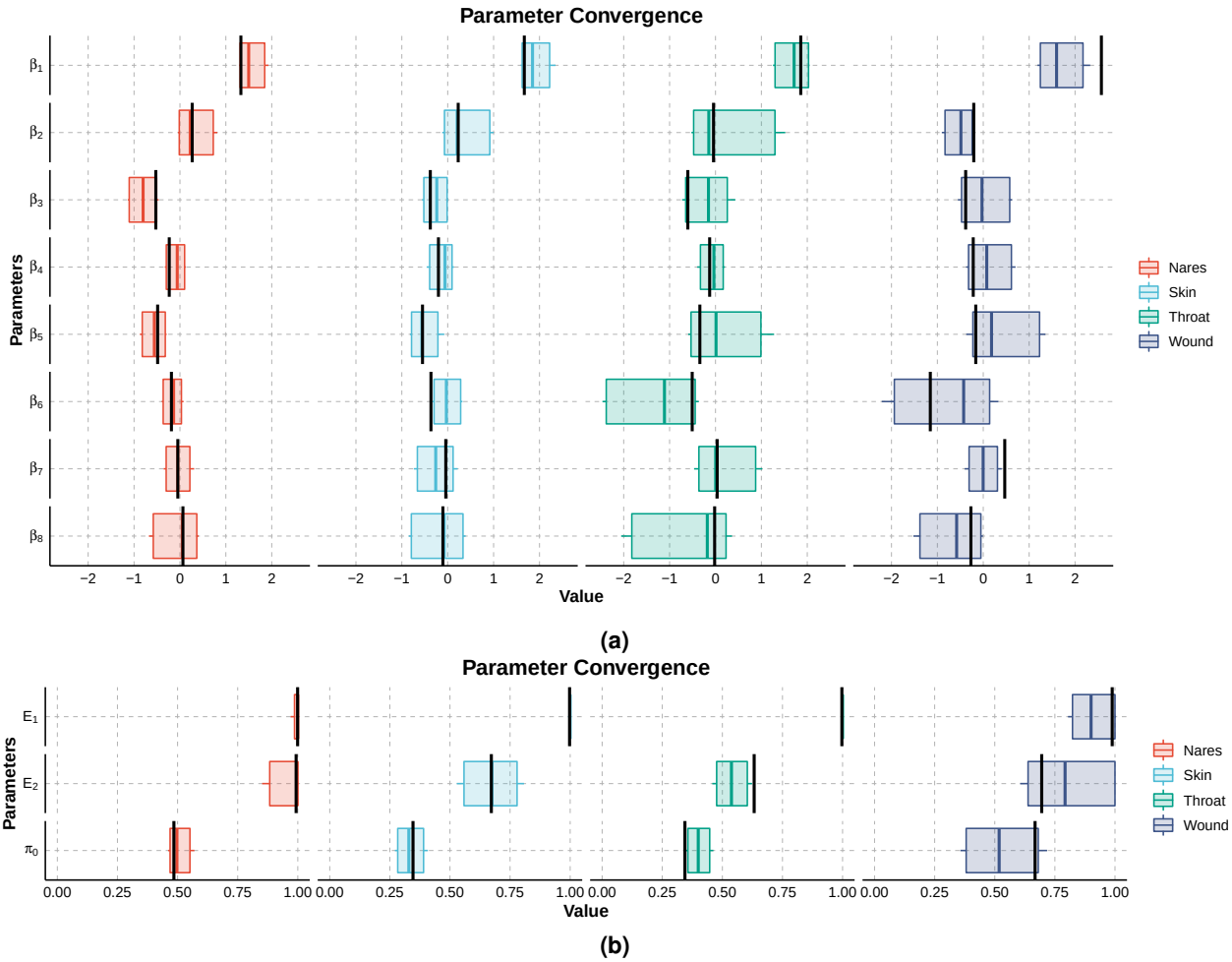
(b)

Supplementary Table 1 | Comparison of models using information criteria in (a) education and (b) decolonization groups. All priors are 0 mean, and σ's represent the scale parameters of the distributions. The best formulation is represented by bold text for each information criterion. For all criteria, a lower score is better.

*Presented priors in the table are not applicable to Standard-HMMs

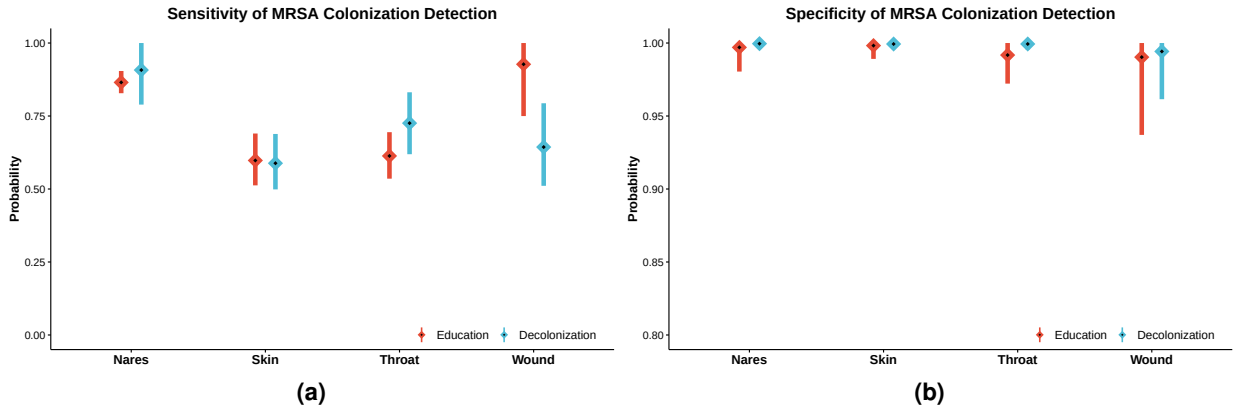
†Presented model in the main text

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

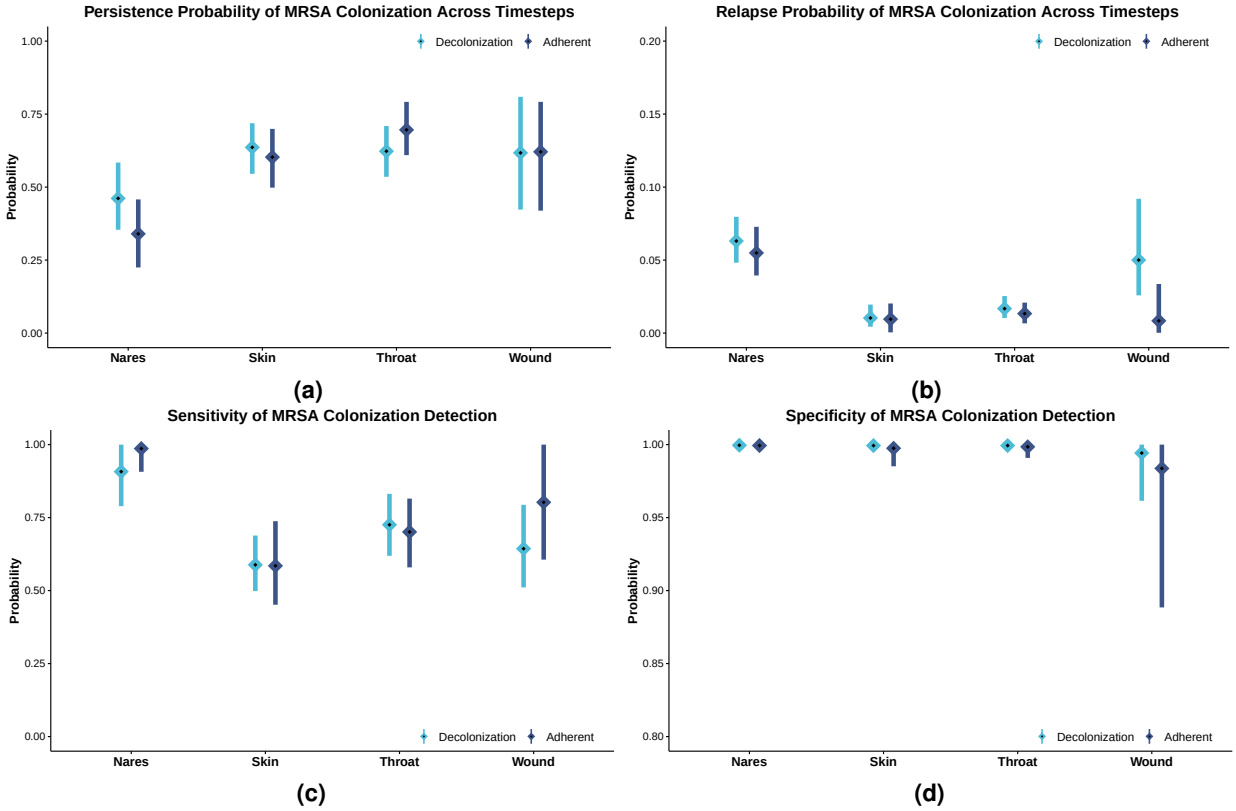


Supplementary Figure 1 | Parameter convergence in the simulated data: (a) β parameters correspond to change in the log-odds (see Design and Interpretation of the β parameters for details), and they determine the transition probabilities at given site, while **(b)** π_0 and E parameters are probability distributions, where π_0 shows the colonization probability for a given chain at hospital discharge and E_1 and E_2 correspond to specificity and sensitivity of the MRSA colonization detection, respectively. To create realistic test data, we used parameters estimated with the real data to simulate the test data. Further, the simulated data mimics the real data in terms of sample size and the pattern of observed samples: There are no samples in the second, fourth, and fifth months (because there were no corresponding visits in the real data), and some samples were not collected because of trial exits, skipped visits, or the patient did not have a wound. In the figures, black vertical lines represent the true parameter value, and bars correspond to 90% credible intervals (CI). We see that the algorithm estimates the parameters correctly, and the CIs cover the true value for all except five parameters, which corresponds to what one would expect from well-calibrated 90% CIs ($5/44 \approx 0.11$).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

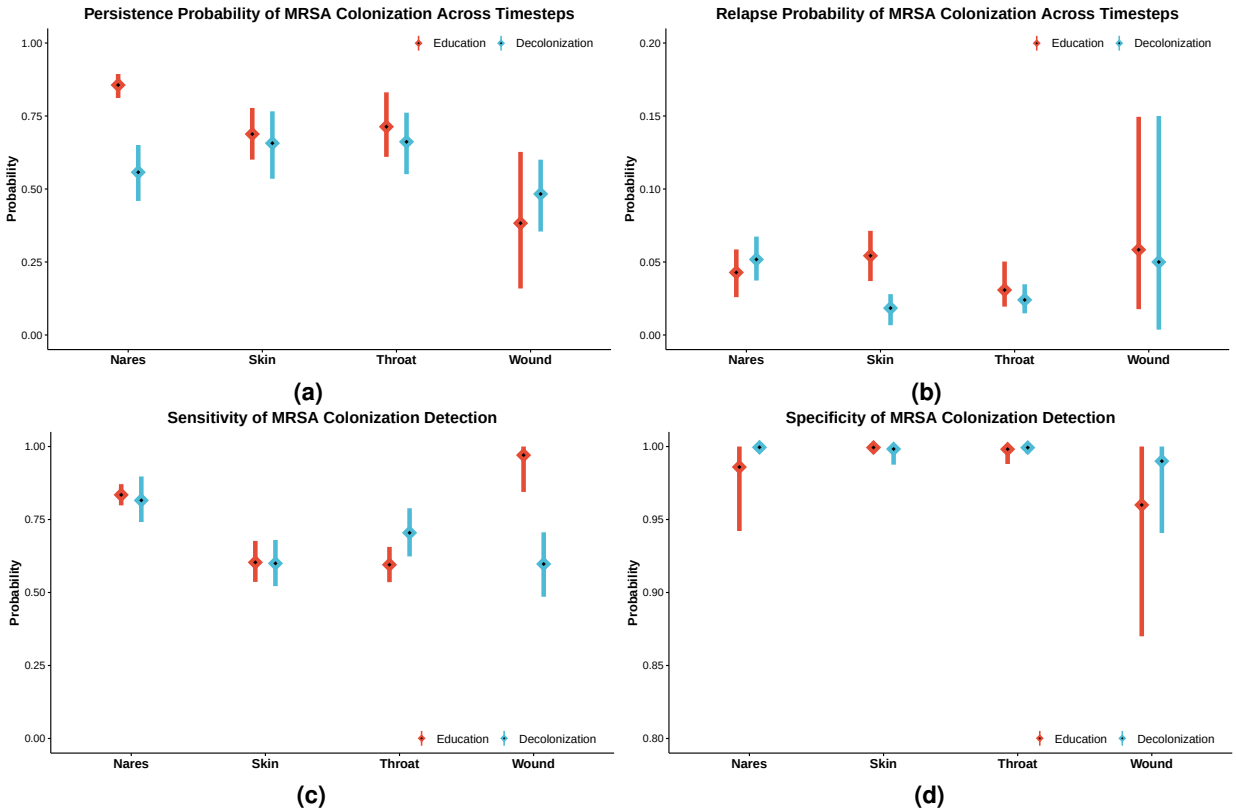


Supplementary Figure 2 | Additional results obtained by the Additive-CHMM: Estimated (a) sensitivity and (b) specificity of MRSA colonization detection by body site. **(a)** Sensitivities (the probability of correctly detected colonized sites) and **(b)** specificities (the probability of correctly detected clear sites) are estimated from the model's emission probabilities. Calculation of the posterior distributions is explained in the Methods. Means and 90% credible intervals (CI) are represented in the figure by squares and lines, respectively.



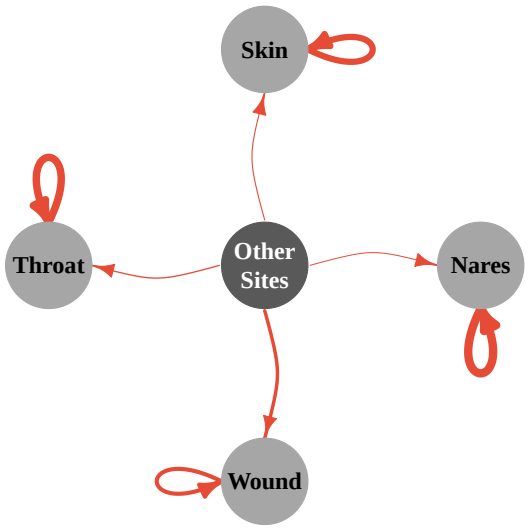
Supplementary Figure 3 | Effect of adherency in comparison with decolonization regarding Additive-CHMM: Estimated (a) persistence probability and (b) relapse probability of MRSA colonization, and (c) sensitivity and (d) specificity of MRSA colonization detection by body site. The **(a)** persistence probabilities (the probability of the site is remaining colonized) and **(b)** relapse probabilities (the probability of the site is getting colonized) are estimated from the model's transition parameters. **(c)** Sensitivities (the probability of correctly detected colonized sites) and **(d)** specificities (the probability of correctly detected clear sites) are estimated from the model's emission probabilities. Calculation of the posterior distributions is explained in the Methods. Means and 90% credible intervals (CI) are represented in the figure by squares and lines, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



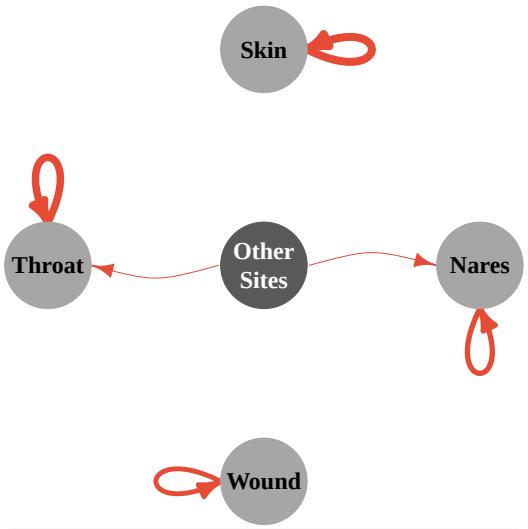
Supplementary Figure 4 | Results of Or-CHMM: Estimated (a) persistence probability and (b) relapse probability of MRSA colonization and (c) sensitivity and (d) specificity of MRSA colonization detection by body site. The **(a)** persistence probabilities (the probability of the site is remaining colonized) and **(b)** relapse probabilities (the probability of the site is getting colonized) are estimated from the model's transition parameters. **(c)** Sensitivities (the probability of correctly detected colonized sites) and **(d)** specificities (the probability of correctly detected clear sites) are estimated from the model's emission probabilities. Calculation of the posterior distributions is explained in the Methods. Means and 90% credible intervals (CI) are represented in the figure by squares and lines, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



t \ t + 1		Nares	Skin	Throat	Wound
Other Sites	**mean**	**1.6**	**1.7**	**4.4**	**1.7**
	90% CI	(0.8,3.1)	(0.7,3.8)	(0.8,11.1)	(-0.9,6.8)
Self Effect	**mean**	**4.5**	**3.0**	**7.4**	**5.4**
	90% CI	(3.2,6.4)	(1.4,5.6)	(3.5,14.7)	(1.5,15.2)

(a)

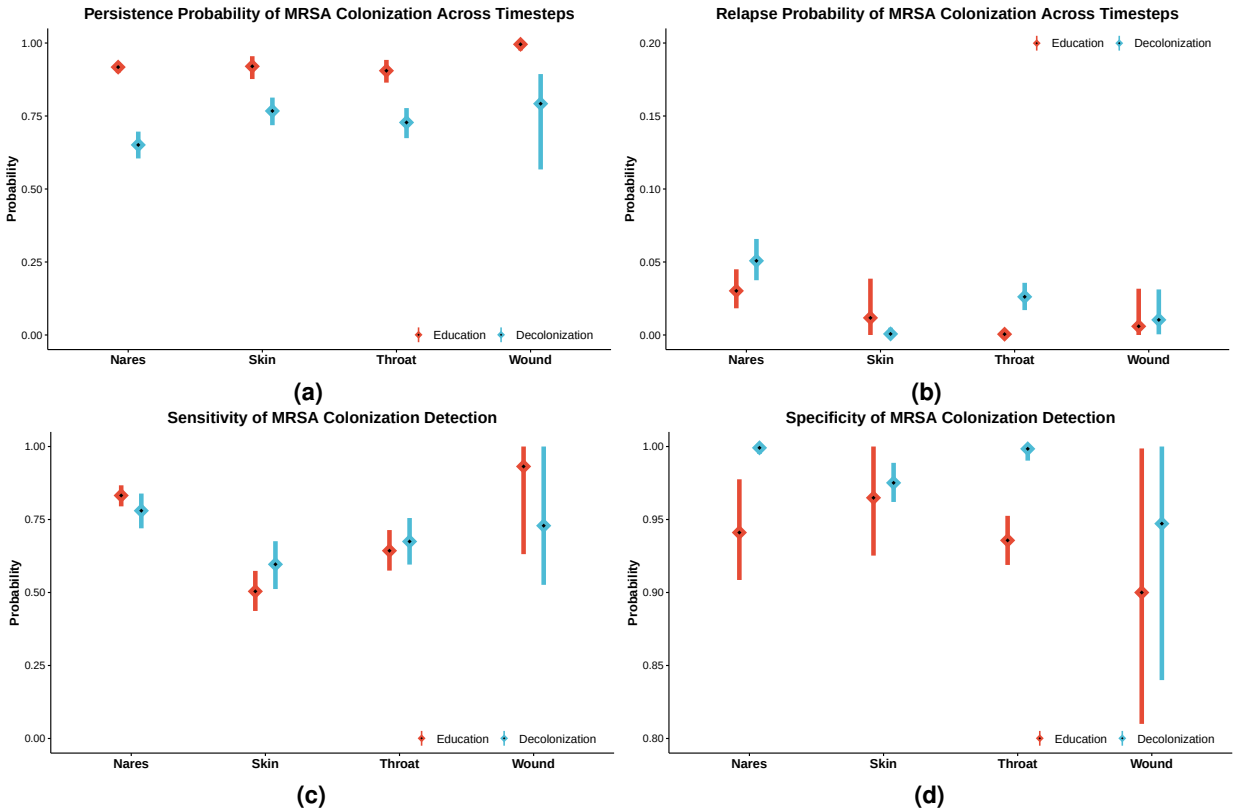


t \ t + 1		Nares	Skin	Throat	Wound
Other Sites	**mean**	**0.8**	**1.2**	**1.1**	**0.2**
	90% CI	(0.5,1.1)	(0.2,2.9)	(-0.1,3.1)	(-4.1,3.3)
Self Effect	**mean**	**2.3**	**3.1**	**4.3**	**7.0**
	90% CI	(1.9,2.8)	(1.6,5.6)	(2.7,6.8)	(2.3,17.2)

(b)

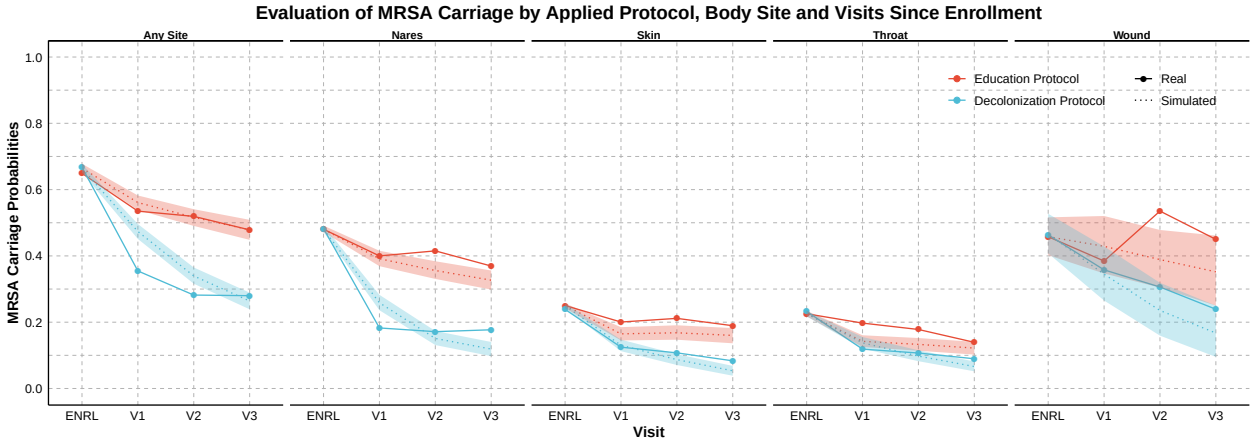
Supplementary Figure 5 | Estimated interactions between the body sites in the (a) education and (b) decolonization groups according to the Or-CHMM. Red edges in the graphs show the increase in the log-odds of a site being colonized in the following time step if the linked site is colonized. In contrast, blue edges (there were none) show the decrease in the log-odds. The thickness of the edges represents the mean of the change in the log-odds, and edges are excluded if the 90% credible intervals (CI) includes value 0. Tables show the mean and 90% CI values for all relations. *Other sites* is a variable that equals one if any of the sites other than the target of the edge is positive, i.e., *Other sites* refers to a different variable for each target site.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

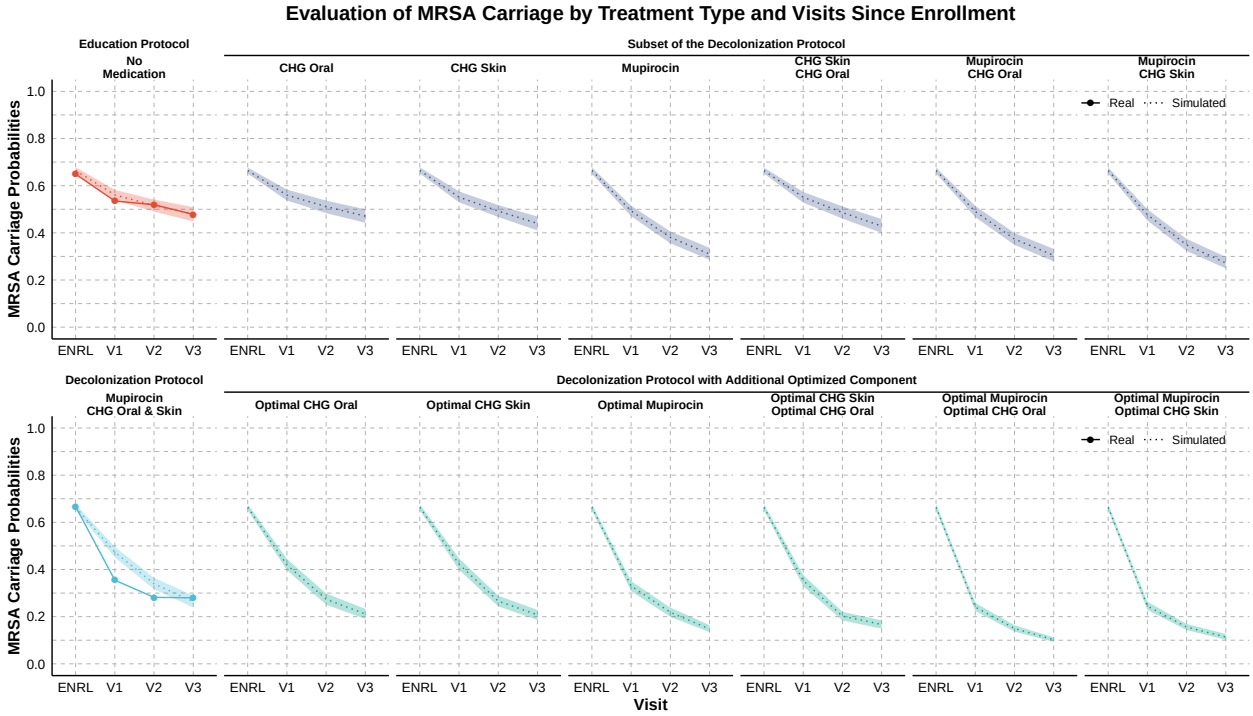


Supplementary Figure 6 | Results of Standard-HMMs: Estimated (a) persistence probability and (b) relapse probability of MRSA colonization, and (c) sensitivity and (d) specificity of MRSA colonization detection by body site. The **(a)** persistence probabilities (the probability of the site is remaining colonized) and **(b)** relapse probabilities (the probability of the site is getting colonized) are estimated from the model's transition parameters. **(c)** Sensitivities (the probability of correctly detected colonized sites) and **(d)** specificities (the probability of correctly detected clear sites) are estimated from the model's emission probabilities. Calculation of the posterior distributions is explained in the Methods. Means and 90% credible intervals (CI) are represented in the figure by squares and lines, respectively.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.30.20239798; this version posted March 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



(a)



(b)

Supplementary Figure 7 | The observed decrease in MRSA carriage over time (a) by body site and study arm, and (b) by actual and hypothetical treatments compared to the decrease predicted by the model. In the trial, study subjects were in the education group or the decolonization group, which consisted of applying mupirocin to the nares, chlorhexidine mouthwash (CHG Oral) to the throat, and chlorhexidine body washes (CHG Skin) to the skin and, if present, wound. **(a)** Shows the site-specific clearance rate for education and decolonization group along with model predictions. **(b)** We conducted our predictive analysis across all possible combinations of hypothetical treatments assumed either that only a subset of medications was applied or that the decolonization protocol was used with an additional optimized intervention that resulted in immediate clearance of the corresponding target site. We note that, for ease of comparison, **(b)** includes all the outcomes presented in the Figure 4. Dotted lines and shaded regions represent the mean and 90% credible intervals (CI) of simulations, and the solid lines represent values observed in the data.