HHS Public Access

Author manuscript

J Occup Environ Hyg. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

J Occup Environ Hyg. 2022 March ; 19(3): 145–156. doi:10.1080/15459624.2022.2025998.

Application of Markov models to predict changes in nasal carriage of *Staphylococcus aureus* among industrial hog operations workers

Melissa G. Edmondson,
Division of Science Integration, National Institute for Occupational Safety and Health, Cincinnati, Ohio

Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Christopher D. Heaney,
Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Meghan F. Davis,
Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Department of Molecular and Comparative Pathobiology, Johns Hopkins School of Medicine, Baltimore, Maryland

Division of Infectious Diseases, Johns Hopkins School of Medicine, Baltimore, Maryland

Gurumurthy Ramachandran
Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland

Abstract

Industrial hog operation (IHO) workers can be occupationally exposed to *Staphylococcus aureus* and may carry the bacteria in their nares. Workers may persistently carry *S. aureus* or transition between different states of nasal carriage over time: no nasal carriage, nasal carriage of a human-associated strain, and nasal carriage of a livestock-associated strain. We aimed to develop a mathematical model to predict the proportion of IHO workers in each nasal carriage state over time, accounting for IHO worker mask use. We also examined data sufficiency requirements to inform development of models that produce reliable predictions. We used nasal carriage data from a cohort of 101 IHO workers in North Carolina, sampled every two weeks for 4 months, to develop a three state Markov model that describes the transition dynamics of IHO worker nasal carriage status over the study period and at steady state. We also stratified models by mask use to examine its impact on worker transition dynamics. If conditions remain the same, our

models predicted that 49.1% of workers will have no nasal carriage of *S. aureus*, 28.2% will carry livestock-associated *S. aureus*, and 22.7% will carry human-associated *S. aureus* at steady state. In stratified models, at steady state, workers who reported only occasional mask (<80% of the time) use had a higher predicted proportion of individuals with livestock-associated *S. aureus* nasal carriage (39.2%) compared to workers who consistently (≥80% of the time) wore a mask (15.5%). These results support epidemiological evidence that mask use may reduce IHO worker nasal carriage of livestock-associated *S. aureus*. We also evaluated how much longitudinal data is sufficient to create a Markov model that accurately predicts future nasal carriage states among the cohort. We did so by creating multiple models that withheld portions of the collected data and compared the model predictions to observed data. Our data sufficiency analysis indicated that models created with a small subset of the dataset (approximately 1/3 of observed data) perform similarly to models created using all observed data points. Mathematical models, such as Markov models, may have important utility to predict worker health status over time, even when limited longitudinal sampling data are available.

Keywords

mathematical modeling; stochastic modeling; transition probability matrix; zoonotic disease

INTRODUCTION

Staphylococcus aureus is an opportunistic pathogen that causes several adverse health outcomes, including skin and soft tissue infections in humans. *S. aureus* infection remains a significant cause of morbidity and mortality in the United States, causing nearly 120,000 bloodstream infections and 20,000 deaths in the United States in 2017 (Kourtis, Hatfield et al. 2019). Focused infection prevention and control efforts have resulted in reduced infection rates within the healthcare industry, but infection rates occurring outside of healthcare settings continue to persist. In addition to circulating community strains, an emergence of people becoming infected with livestock-associated strains of *S. aureus* has been observed in the last two decades, particularly among workers who have contact with animals (Fitzgerald 2012, Cuny, Wieler et al. 2015, Smith 2015).

Research indicates that swine production workers are at increased risk for nasal carriage of livestock-associated strains of *S. aureus* (Sahibzada, Hernandez-Jover et al. 2018). People who work with swine, including industrial hog operation workers, are often persistent or intermittent carriers of livestock associated *S. aureus* (Nadimpalli, Rinsky et al. 2015). Workers carrying livestock associated strains of *S. aureus* are significantly more likely to report skin and soft tissue infections compared to those who do not carry livestock associated strains (Nadimpalli, Stewart et al. 2016). There is also some evidence to support the hypothesis that livestock-associated *S. aureus* may be as or more pathogenic than common community-associated strains (Randad, Dillen et al. 2019).

Understanding the risk factors of *S. aureus* nasal carriage may inform interventions that reduce disease burden among exposed workers. Several epidemiological studies have investigated factors that influence persistence of livestock-associated *S. aureus* in the nares

of workers (Garcia-Graells, Antoine et al. 2012, Rinsky, Nadimpalli et al. 2013, Nadimpalli, Stewart et al. 2018). These studies, while informative, are costly, time-consuming and resource intensive, requiring a large number of study participants, many repeated measures, and laboratory analysis. Researchers and practitioners often do not have the resources necessary to collect samples from individuals longitudinally. Furthermore, it is unknown how often and how long it is necessary to sample a worker population to predict the steady state distribution of worker nasal carriage states under certain work conditions. If the amount of sampling data needed to make accurate predictions were known, researchers could design more cost-effective sampling strategies.

In this paper, we propose a Markov model as an approach to predict changes in worker nasal carriage of *S. aureus* over time. A Markov model comprises two or more states and corresponding transition probabilities that indicate the probability of moving from one state to another during a defined time period. The probability of moving into a different state, or staying in the same state, is dependent only on the current state. The simplest form of a Markov model is a Markov chain, where all states are observable. Markov chain models have previously been proposed as a generalized approach to describe changes in health status over time and as a tool for medical decision making (Sonnenberg and Beck 1993). Examples in the literature include breast cancer (Fujii, Mason et al. 2019, Huang, Li et al. 2020), infertility (Srinivasa Rao and Diamond 2020), hepatocellular carcinoma (Ishida, Wong et al. 2008), and HIV (Simpson, Strassburger et al. 2009). Batina and colleagues have previously used Markov Models to describe how residents of community nursing homes transition between states of methicillin-resistant *S. aureus* nasal carriage (Batina, Crnich et al. 2016). In this paper, we constructed a Markov chain model to describe how a cohort of industrial hog operation workers transition between states of nasal carriage over a period of 4 months. Since the use of face masks has been found to influence worker carriage status in this cohort previously (Nadimpalli, Stewart et al. 2018), we stratified our models by face mask use to demonstrate the effectiveness of exposure controls to modify long-term carriage rates. Finally, we selectively used different amounts of data to calculate transition probabilities between different states and to determine the minimum set of repeated, sequential measures needed to reasonably predict long term nasal carriage rates.

METHODS

Data sources

The data used for this model were obtained from a four-month prospective longitudinal cohort study of *S. aureus* nasal carriage among industrial hog operation workers and household contacts from October 2013 through June 2014. Analysis of this data have been previously published and data collection methods are described in detail (Nadimpalli, Stewart et al. 2016, Nadimpalli, Stewart et al. 2018). In short, 101 industrial hog operation workers in North Carolina were recruited via snowball sampling by community organizers from the Rural Empowerment Association for Community Help. Study participants were mostly male (55%), Hispanic (88%), and 39 ± 11 years old (mean ± standard deviation). For more demographic information, see (Nadimpalli, Stewart et al. 2018) and Supplementary Table S1. Enrolled participants were screened for *S. aureus* nasal carriage at baseline

and every two weeks for four months via self-collected nasal swabs. Once samples were processed and plated, presumptive *S. aureus* colonies were strain typed and assessed for the staphylococcal protein A (spa) gene. At baseline, participants were administered a questionnaire assessing demographic information, habitual work activities, and household characteristics. During each follow-up visit, participants were administered questionnaires assessing their time-varying occupational activities, such as mask use, in each of the two one-week periods since the last visit. Figure 1 summarizes the study design. In their previously published analysis of this cohort, Nadimpalli and colleagues found that face mask use was associated with reduced odds of nasal carriage of livestock-associated *S. aureus* (Nadimpalli, Stewart et al. 2018). Accordingly, we also modeled the effect of mask use on the transition dynamics between *S. aureus* nasal carriage states. All data used in the analysis were de-identified and covered under the Johns Hopkins Bloomberg School of Public Health Institutional Review Board (IRB) number IRB00004608.

Model overview

We developed a discrete time Markov chain model to describe how industrial hog operation workers transition between *S. aureus* carriage states throughout the study period and to predict steady state carriage rates among this study population. We also assume that rates of nasal carriage of *S. aureus* among this cohort of IHO workers will eventually reach steady state levels and we use Markov modeling to predict which state(s) will dominate over time. Workers were categorized into a finite number of mutually exclusive health states and we estimated the probability of transitioning between these states during equal time intervals. We defined three discrete health states: no nasal carriage of *S. aureus* (NC), nasal carriage of human-associated *S. aureus* (HA), and nasal carriage of livestock-associated *S. aureus* (LA). A sample is considered positive for *S. aureus* if any culturable *S. aureus* is present in the nasal swab. To differentiate between livestock-associated and human-associated strains of *S. aureus*, we used the absence of the *scn* gene as a putative marker of livestock-association (Rinsky, Nadimpalli et al. 2013). The LA state is defined by nasal swab positive for *S. aureus* where the cultured strain did not carry the *scn* gene. The HA state includes all other workers with nasal swabs positive for *S. aureus* where the cultured strain did carry the *scn* gene. The NC state is that in which worker nasal swabs were culture negative for *S. aureus*. Although workers can carry both livestock-associated and human-associated strains at once, too few participants were co-carriers to accurately estimate the probability of transitioning to and from this state, so the small number of participants who carried both livestock and human-associated strains are included in the LA state. The three states in our model are represented in a state-transition diagram in Figure 2.

We represented the initial state of our Markov chain with an initial state vector $[N_0\ N_1\ N_2]$. The number of workers in each state during the baseline study visit are denoted N_i, where i is the nasal carriage state (NC=0, HA=1, and LA=2). We modeled time in equal intervals of two weeks to correspond with study visits. A worker can transition directly from any of the three states into any other state or stay in the same state during a given two-week Markov cycle. We assumed that only one single state transition could happen for each worker during each cycle. Accordingly, there are nine possible transitions that can occur in each Markov cycle, including remaining in the same state. The probabilities of each transition occurring

are expressed mathematically in the transition probability matrix $\begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}$ and in table form (see Table 1). The probability of moving between two states in one Markov cycle is represented by P_{ij}, where the worker transitions from the ith state at time t to the jth state at time t+1. By definition of a Markov process, we assume that a worker's probability of moving from one state to another state is dependent only on the current state of the worker (i.e., P_{ij} is constant for a given i and j). We used a straightforward process to determine transition probabilities (see supplementary Table S2). Since the health state of each worker was observed at the beginning and end of each two-week Markov cycle, we identified every single state transition that occurred among the study population during the study period. We then calculated the number of transitions from state i to state j divided by the total number of transitions that started at state i. This ratio is the transition probability from state i to state j.

Determining stationary probabilities to predict future distribution of *S. aureus* nasal carriage

The estimated distribution of worker carriage states after *t* Markov cycles is the product of the initial state vector (baseline distribution of nasal carriage status among the cohort) and the transition probability matrix raised to the power t.

$$\begin{bmatrix} N_0 & N_1 & N_2 \end{bmatrix} \times \begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}^{\mathrm{t}} = \begin{bmatrix} X_0 & X_1 & X_2 \end{bmatrix}$$

We assume that if there are no changes in workplace conditions, nasal carriage of *S. aureus* among this cohort will theoretically reach steady state over time. Given that our model has no absorbing states (that is, every state in the model is temporary and workers in any of the three states are expected to eventually leave that state in the long run), workers will continue to move between different nasal carriage states indefinitely and the model converges to a steady state distribution of workers in the three states. Consequently, the total proportion of workers in each nasal carriage state will remain approximately constant over time once the system reaches steady state. For steady state vector, $\begin{bmatrix} X_0 & X_1 & X_2 \end{bmatrix}$ and transition probability matrix, $\begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}$:

$$\begin{bmatrix} X_0 & X_1 & X_2 \end{bmatrix} \begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}^{\infty} = \begin{bmatrix} X_0 & X_1 & X_2 \end{bmatrix}$$

This mathematical property of Markov models allows us to predict what carriage state is expected to dominate this population at steady state if conditions remain the same.

To estimate the proportion of workers in each nasal carriage state throughout the study period, we multiplied the baseline vector (representing the distribution of workers in each nasal carriage state) with the transition probability matrix developed from the data collected during study visits raised to the t power, where t is the study visit number. To calculate the proportion of workers in each state at steady state, we multiplied the baseline vector by the transition probability matrix many times, using a brute force method, until the resulting vector ceased to change.

Stratified models based on mask use

We examined the influence of workers' self-reported mask use on the transition probabilities by developing stratified transition probability matrices for consistent mask use and for occasional mask use. At each study visit, individuals report average mask use within the two-week Markov cycle. We evaluated every single time-step (two week) transition that occurred over the study period. Each transition was matched with the worker's self-reported mask-use during that two-week interval. We stratified the observed transitions into two groups based on the average reported face mask use of the participant during the two-week Markov cycle in which the transition occurred. We defined "consistent use" as self-reported mask use of 80% or more during the prior two weeks and "occasional use" as self-reported mask use less than 80% in the prior two weeks. These definitions were aligned with definitions used in previous analysis of this data (Nadimpalli, Stewart et al. 2018). We also developed models for the subset of transitions in which individuals reported no mask use (0%) and complete mask use (100%) in the prior two weeks. Since workers reported their mask use at each study visit, a single worker could contribute transition data to more than one model. For example, a worker could report 100% mask use in one two-week period and report 0% mask use in the next two-week period. The first transition would be included in the consistent use and complete mask use models and the second transition would be included in the occasional use and no mask use models. We created transition probability matrices based on reported mask use in the prior two weeks that each transition occurred and compared steady state vectors for each model. 95% confidence intervals around the point estimates were computed by bootstrapping with 500 resampling runs and assuming an approximately normal underlying distribution.

Evaluating data sufficiency and model accuracy with varying amounts of data

To test the validity of our model, and to understand how much data are needed to create a reliable model (in other words, how much data is sufficient to make a reasonable prediction), we created multiple transition probability matrices using varying amounts of data and then compared each model's output to observed data. In other words, we created eight separate models, with the first model holding back the majority of the study data and subsequent models increasing the amount of study data used to define the transition probabilities. Specifically, the transition probability matrix for Model 1 was developed using transition data from baseline to visit 1. Model 2 was built using transition data from baseline to visit 1 *and* transition data from visit 1 to visit 2. Model 3 uses transition data from baseline to visit 1, from visit 1 to visit 2, and from visit 2 to visit 3, and so on. Each model is used to predict the distribution of nasal carriage status among the cohort after each Markov cycle. To calculate the expected distribution of workers in each nasal carriage state at the final study

visit (Visit 8), we multiplied the initial distribution of workers in each state, $[N_0\ N_1\ N_2]$, by the transition probability matrix developed for each Model raised to the 8th power, $\begin{bmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{bmatrix}^8$. 95% confidence intervals around the point estimates were computed by bootstrapping with 500 resampling runs and assuming an approximately normal underlying distribution. We calculate the correlation coefficient between observed data and model output for each of the 8 models. We also calculate the mean squared error between the observed and predicted values for the last study visit (follow up visit 8) as another measure of model performance.

RESULTS

Complete (unstratified) and stratified transition probability matrices

In the initial study visit, 58 (57.4%), 25 (24.8%), and 18 (17.8%) of workers were in the NC, HA, and LA nasal carriage states, respectively. Thus, the distribution of workers in each state is represented by the vector [0.574 0.248 0.178]. We observed 676 nasal carriage state transitions for 101 industrial hog operation workers over 8 Markov cycles. There were 319 transitions for which the individual reported wearing their mask during the two-week period at least 80% of the time and 312 transitions where mask use in the prior two weeks was reported as less than 80%. The remaining 45 observations were missing data on mask use during the prior two weeks and were not included in models stratified by mask use. Transition probability matrices for the complete worker nasal carriage model (that is, the model built using all observed transitions among all workers, regardless of mask use) and the models stratified by mask use are shown in Table 2. The transition probability matrix for the complete model used data from all 676 transitions observed and the stratified models used all transitions where data for mask use were available (631 transitions). The transition probability matrices indicate that under current conditions, most workers remain in the same nasal carriage state over a single two-week Markov cycle, and fewer workers will change nasal carriage states in a two-week time period. For the complete model, workers in the NC, HA, and LA states are expected to remain in the same carriage state 80%, 66%, and 69% of the time, respectively (Table 2a). In this model, transition dynamics between human associated and livestock associated *S. aureus* strains are similar, with the most individuals with *S. aureus* nasal carriage of either strain remaining in the same state over one Markov cycle (66% and 69%, respectively), relatively few transitions from one *S. aureus*-positive state to another (13% for both), and the remaining individuals transitioning to no nasal carriage. In the models stratified by mask use, most workers will also stay in the same state in each time step, except for individuals in the LA state who used masks consistently in the prior two weeks. When masks were used consistently, workers with nasal carriage of livestock associated *S. aureus* have only a 48% probability of remaining a carrier at the next time step, compared to 66.7% when mask use is inconsistent (Table 2b). Results of a sub analysis which only included transitions that occurred after complete (100%) mask use or no (0%) mask use show a similar pattern but may be less precise due to fewer transitions being included (Table 2c).

Impact of mask use on distribution of worker nasal carriage states

Steady state distributions for each of the three models are shown in Figure 3. The complete model, showing the worker population at their current level of mask usage, predicts that 28.2% (95% CI, 17.1–39.3%) of workers in this cohort are expected to be carriers of livestock associated *S. aureus* at steady state. In the consistent mask use model and the occasional mask use model, the number of workers with nasal carriage of livestock associated *S. aureus* at steady state is 15.5% (95% CI, 7.4–23.6%) and 39.2% (95% CI, 27.3–51.1%), respectively. More workers are expected to have no nasal carriage of *S. aureus* when masks are used consistently (57.6%; 95% CI, 45.4–69.8%) compared to when masks are use occasionally or not at all (42.6%; 95% CI, 30.5–54.7%). The consistent mask use model shows the highest proportion of individuals with human associated *S. aureus* (26.9%; 95% CI, 15.5–38.3%) compared to the complete model and the occasional mask use models (22.7% (95% CI, 11.7–33.7%) and 18.2% (95% CI, 8.8–27.6%) respectively). The occasional mask use model shows the highest proportion of individuals with livestock associated *S. aureus* (39.2%; 95% CI, 27.3–51.1%) compared to the complete and consistent mask use models (28.2% (95% CI, 17.2–39.3%) and 15.5% (95% CI, 7.4–23.6%) respectively).

Model accuracy and data sufficiency threshold

With a study population of 101 workers, 9 study visits, and 8 transitions per worker, there were 808 total transitions in 8 Markov cycles. Due to missing nasal swab or mask use data for some of the participants, 137 transitions were excluded, and 676 transitions were observed and included in our models. Table 3 summarizes the 8 models used in our data sufficiency analysis, including the number of transitions included in the model, the transition probability matrix for each model, the correlation coefficient when comparing the model to observed data over the 8 follow up visits, and the final study visit prediction. Graphs comparing the model's predictions versus the observed data for Models 1 and 8 are shown as an example (Figure 4). N.B.: Model 1 perfectly predicts study visit 1 solely because the model was created using only transitions from baseline to visit 1, and consequently matches week 1 exactly. Mean squared errors comparing the model predictions for the proportion of workers in each nasal carriage state after 8 Markov cycles with what was observed at follow-up visit 8 were calculated for Models 1 through 8 and are shown in Figure 5.

DISCUSSION

This study modeled changes in nasal carriage of *S. aureus* over time among industrial hog operation workers and investigated how mask use may play a role in strain specific carriage. Additionally, we investigated how much longitudinal data is necessary for developing reasonable models. Our models predicted that consistent mask use will result in reduced nasal carriage of *S. aureus* in general, and, more specifically, reduced carriage of livestock associated strains at 4 months and at a theoretical steady state. This result is consistent with prior epidemiological findings (Nadimpalli, Stewart et al. 2018) and has important health implications. Emerging research suggests increased pathogenicity of livestock associated *S. aureus* when compared to community associated strains (Randad, Dillen et al. 2019). IHO

workers who carry human associated and/or livestock associated *S. aureus* in their nares are at increased risk of skin and soft tissue infections (Wertheim, Melles et al. 2005).

In the occasional mask use model, it is predicted that livestock associated strains of *S. aureus* will dominate at steady state, compared to the consistent mask use model where the human associated strain dominates. Interestingly, in the consistent mask use model, nasal carriage of human-associated *S. aureus* is predicted to be increased compared to the complete model. One possible explanation for this is that when exposure to livestock associated *S. aureus* is low due to mask use, the primary source of exposure to *S. aureus* may be to human-associated strains in non-occupational settings. When masks are not worn consistently, and occupational exposure to *S. aureus* is high, livestock-associated strains might outcompete human-associated strains, resulting in the apparent decrease in human-associated *S. aureus* when reported mask use is high.

Our analysis indicates that the Markov models perform reasonably well ($R^2 > 0.9$) in predicting future nasal carriage with as little as 4 weeks' worth of longitudinal data. Mean squared errors for Models 2 through 8 indicate that these models perform reasonably well (MSE < 0.005) at predicting nasal carriage states after 2 Markov cycles. In this dataset, reasonably accurate predictions were made with data from only three study visits. While more research is needed to establish how much data is needed to create an accurate transition probability matrix for changes in worker health over time, this research supports the fact that even limited amount of longitudinal data could have useful predictive power in research and occupational exposure monitoring. This might inform the way that future longitudinal studies are designed. Mathematical models, in combination with limited sampling data, may be used to understand workplace nasal carriage over time with reasonable accuracy, provided that workplace conditions and worker behaviors remain constant.

Our models also support the conclusions of prior studies that indicate that occupational mask use among this cohort may have a protective effect for nasal colonization of *S. aureus,* particularly livestock associated strains. Since Markov models are significantly less expensive than sample collection, models should be considered as an additional, complementary approach to estimation of the protective effect of workplace controls.

One purpose of our study is to illustrate the use of Markov modeling in industrial hygiene applications. To our knowledge, this is the first application of a Markov chain to model industrial hog operation worker nasal carriage of *S. aureus*. A Markov model is useful in this scenario because workers transition between a finite number of health (nasal carriage) states many times throughout the study period. Markov chains allow for future carriages states to be easily calculated, given several assumptions. However, these simplifying assumptions create limitations in our study. We assume, for example, that the probability of nasal carriage in the next time step is dependent only on the current state, and that mask usage is the predominate variable that explains variability in *S. aureus* acquisition. However, many other factors, including work tasks, hours worked, and host factors, can also contribute to transition rates. In the previous epidemiological study of this cohort, researchers found that recent pressure washing of barns and ever vs. never administering shots was associated with

scn-negative *S. aureus* nasal carriage. A limitation of this research is that our models does not further stratify based on these factors, because doing so would result in an increasingly small n. Increasing the complexity of models by including additional variables may, in some cases, increase model accuracy, but also may decrease the generalizability of the model. On the other hand, since study participants represent a heterogeneous population in terms of work environment and job duties (Nadimpalli, Rinsky et al. 2015), this may support the generalizability of our findings to the general population of IHO workers in North Carolina. We make the simplifying assumption that nasal carriage transitions happen in discrete two-week time intervals, and that each worker makes only one transition per Markov cycle. In reality, it is possible for multiple transitions to occur over a two-week period (Nadimpalli, Rinsky et al. 2015). Additionally, we assume that if conditions remain the same, nasal carriage of *S. aureus* among this cohort will theoretically reach steady state. Factors such as seasonality (Nehme, Létourneau et al. 2008) or, more likely, periodicity within the grow out cycle of swine production may play a role in nasal carriage of *S. aureus* and may prevent true steady state conditions from being achieved. Larger studies of longer duration are needed to identify longer term trends that could not be assessed in the current analysis. Another limitation is our poor understanding of self-reported mask use. Study participants reported the frequency of mask use but did not provide details on the type of mask, how often masks were replaced, how they were stored or laundered, and whether they were taken home. Differentiating N-95 respirators from dust masks and other face coverings could further explain the seemingly protective effect of mask use and would inform policy recommendations for reducing worker exposure.

Due to resource limitations and worker privacy concerns, quality repeated measures data on worker nasal carriage may be impractical to obtain. Therefore, there are many advantages to using models to predict what future carriage distributions look like using a limited number of data points. This study estimated the occupational risk of *S. aureus* among industrial hog operations workers using a new approach that required less data but produced reasonably accurate predictions in the short term.

CONCLUSION

We developed a Markov model to predict nasal carriage of *S. aureus* in a cohort of workers over 4 months. We also made steady state predictions about the distribution of livestock-associated and human-associated *S. aureus* among this occupational group. Our study suggests that under current conditions, about half of workers (50.9%) will have nasal carriage of *S. aureus* at steady state, with 55.4 % of those workers having livestock associated *S. aureus* nasal carriage. In the model representing a consistent mask use scenario among the work force, the number of workers with nasal carriage is slightly decreased, and the ratio of livestock-associated *S. aureus* carriage to human associated *S. aureus* carriage is decreased. Under the occasional mask use model, the opposite is true. This supports previous epidemiological findings that mask use may reduce worker exposure to *S. aureus* in industrial hog operation environments.

Robust longitudinal data for occupational cohorts are typically scarce. If movement of workers between various infection states can be accurately predicted, researchers may be

able to determine the long-term impact of interventions using fewer data points. Given that models that were developed using one third of the data points produced similar results to models that used all data, researchers may be able to design more cost-effective sampling strategies in the future. Markov models have practical applications for occupational settings and should be considered as a tool for predicting changes in worker health states and evaluating the effectiveness of workplace controls.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors would like to thank the workers and community members who participated in the study from which this data were obtained. The authors also thank the community-based organization members who made essential contributions to this research and without whom, this study would not be possible.

Funding for this study was provided by National Institute for Occupational Safety and Health (NIOSH) grant K01OH010193; Johns Hopkins NIOSH Education and Research Center grant T42OH008428; a directed research award from the Johns Hopkins Center for a Livable Future; award 018HEA2013 from the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases Discovery Program at the Johns Hopkins University, School of Medicine, Department of Medicine, Division of Infectious Diseases; and National Science Foundation (NSF) grant 1316318 as part of the joint NSF-National Institutes of Health (NIH)-U.S. Department of Agriculture Ecology and Evolution of Infectious Diseases program. C.D.H. and G.R. were supported by NIOSH grant K01OH010193, E.W. “Al” Thrasher Award 10287, NIEHS grant R01ES026973, and NSF grant 1316318. M.F.D. was supported by the NIH Office of the Director (K01OD019918). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

REFERENCES

Batina NG, Crnich CJ, Anderson DF and Dopfer D. (2016). “Models to predict prevalence and transition dynamics of methicillin-resistant Staphylococcus aureus in community nursing homes.” Am J Infect Control 44(5): 507–514. [PubMed: 26777286]

Cuny C, Wieler LH and Witte W. (2015). “Livestock-Associated MRSA: The Impact on Humans.” Antibiotics (Basel) 4(4): 521–543. [PubMed: 27025639]

Fitzgerald JR (2012). “Livestock-associated Staphylococcus aureus: origin, evolution and public health threat.” Trends Microbiol 20(4): 192–198. [PubMed: 22386364]

Fujii T, Mason J, Chen A, Kuhn P, Woodward WA, Tripathy D, Newton PK and Ueno NT (2019). “Prediction of Bone Metastasis in Inflammatory Breast Cancer Using a Markov Chain Model.” Oncologist 24(10): 1322–1330. [PubMed: 30952823]

Garcia-Graells C, Antoine J, Larsen J, Catry B, Skov R. and Denis O. (2012). “Livestock veterinarians at high risk of acquiring methicillin-resistant Staphylococcus aureus ST398.” Epidemiol Infect 140(3): 383–389. [PubMed: 22082716]

Huang Y, Li Q, Torres-Rueda S. and Li J. (2020). “The Structure and Parameterization of the Breast Cancer Transition Model Among Chinese Women.” Value Health Reg Issues 21: 29–38. [PubMed: 31634794]

Ishida H, Wong JB, Hino K, Kurokawa F, Nishina S, Sakaida I, Okita K, Tamesa T, Oka M, Torimura T, Sata M, Takahashi S, Chayama K. and Inoue Y. (2008). “Validating a Markov model of treatment for hepatitis C virus-related hepatocellular carcinoma.” Methods Inf Med 47(6): 529–540. [PubMed: 19020689]

Kourtis AP, Hatfield K, Baggs J, Mu Y, See I, Epson E, Nadle J, Kainer MA, Dumyati G, Petit S, Ray SM, Ham D, Capers C, Ewing H, Coffin N, McDonald LC, Jernigan J. and Cardo D. (2019). “Vital

Signs: Epidemiology and Recent Trends in Methicillin-Resistant and in Methicillin-Susceptible Staphylococcus aureus Bloodstream Infections - United States." MMWR Morb Mortal Wkly Rep 68(9): 214–219. [PubMed: 30845118]

Nadimpalli M, Rinsky JL, Wing S, Hall D, Stewart J, Larsen J, Nachman KE, Love DC, Pierce E, Pisanic N, Strelitz J, Harduar-Morano L. and Heaney CD (2015). "Persistence of livestock-associated antibiotic-resistant Staphylococcus aureus among industrial hog operation workers in North Carolina over 14 days." Occup Environ Med 72(2): 90–99. [PubMed: 25200855]

Nadimpalli M, Stewart JR, Pierce E, Pisanic N, Love DC, Hall D, Larsen J, Carroll KC, Tekle T, Perl TM and Heaney CD (2016). "Livestock-Associated, Antibiotic-Resistant Staphylococcus aureus Nasal Carriage and Recent Skin and Soft Tissue Infection among Industrial Hog Operation Workers." PLoS One 11(11): e0165713.

Nadimpalli ML, Stewart JR, Pierce E, Pisanic N, Love DC, Hall D, Larsen J, Carroll KC, Tekle T, Perl TM and Heaney CD (2018). "Face Mask Use and Persistence of Livestock-associated Staphylococcus aureus Nasal Carriage among Industrial Hog Operation Workers and Household Contacts, USA." Environ Health Perspect 126(12): 127005.

Nehme B, Létourneau V, Forster RJ, Veillette M. and Duchaine C. (2008). "Culture-independent approach of the bacterial bioaerosol diversity in the standard swine confinement buildings, and assessment of the seasonal effect." Environmental Microbiology 10(3): 665–675. [PubMed: 18237302]

Randad PR, Dillen CA, Ortines RV, Mohr D, Aziz M, Price LB, Kaya H, Larsen J, Carroll KC, Smith TC, Miller LS and Heaney CD (2019). "Comparison of livestock-associated and community-associated Staphylococcus aureus pathogenicity in a mouse model of skin and soft tissue infection." Scientific Reports 9(1): 6774. [PubMed: 31043631]

Rinsky JL, Nadimpalli M, Wing S, Hall D, Baron D, Price LB, Larsen J, Stegger M, Stewart J. and Heaney CD (2013). "Livestock-associated methicillin and multidrug resistant Staphylococcus aureus is present among industrial, not antibiotic-free livestock operation workers in North Carolina." PLoS One 8(7): e67641.

Sahibzada S, Hernandez-Jover M, Jordan D, Thomson PC and Heller J. (2018). "Emergence of highly prevalent CA-MRSA ST93 as an occupational risk in people working on a pig farm in Australia." PLoS One 13(5): e0195510.

Simpson KN, Strassburger A, Jones WJ, Dietz B. and Rajagopalan R. (2009). "Comparison of Markov model and discrete-event simulation techniques for HIV." Pharmacoeconomics 27(2): 159–165. [PubMed: 19254048]

Smith TC (2015). "Livestock-associated Staphylococcus aureus: the United States experience." PLoS Pathog 11(2): e1004564.

Sonnenberg FA and Beck JR (1993). "Markov models in medical decision making: a practical guide." Med Decis Making 13(4): 322–338. [PubMed: 8246705]

Srinivasa Rao ASR and Diamond MP (2020). "Deep Learning of Markov Model-Based Machines for Determination of Better Treatment Option Decisions for Infertile Women." Reprod Sci 27(2): 763–770. [PubMed: 31939200]

Wertheim HF, Melles DC, Vos MC, van Leeuwen W, van Belkum A, Verbrugh HA and Nouwen JL (2005). "The role of nasal carriage in Staphylococcus aureus infections." Lancet Infect Dis 5(12): 751–762. [PubMed: 16310147]

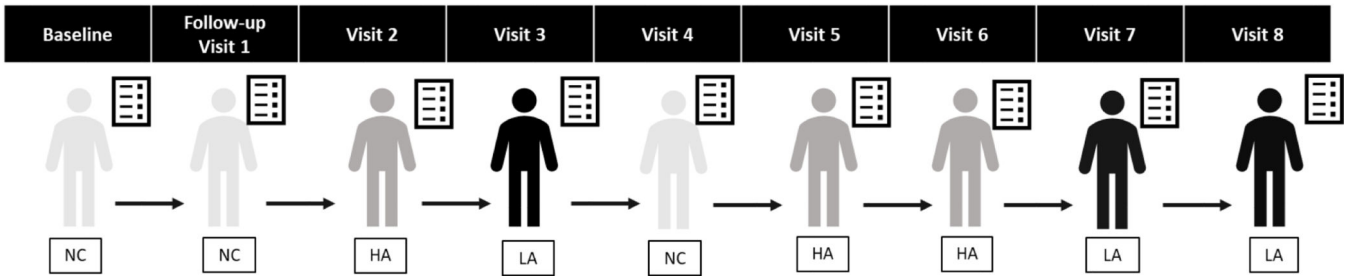


Figure 1.

Study Design. 101 industrial hog operation workers were enrolled in the original study. Study visits occurred two weeks apart. The figure above represents one individual as she or he transitions between different nasal carriage states over the course of the study. Transitions between these states over the two-week time intervals are represented by an arrow. NC = no nasal carriage of *S. aureus*; HA = nasal carriage of human-associated *S. aureus*; LA = nasal carriage of livestock-associated *S. aureus*.

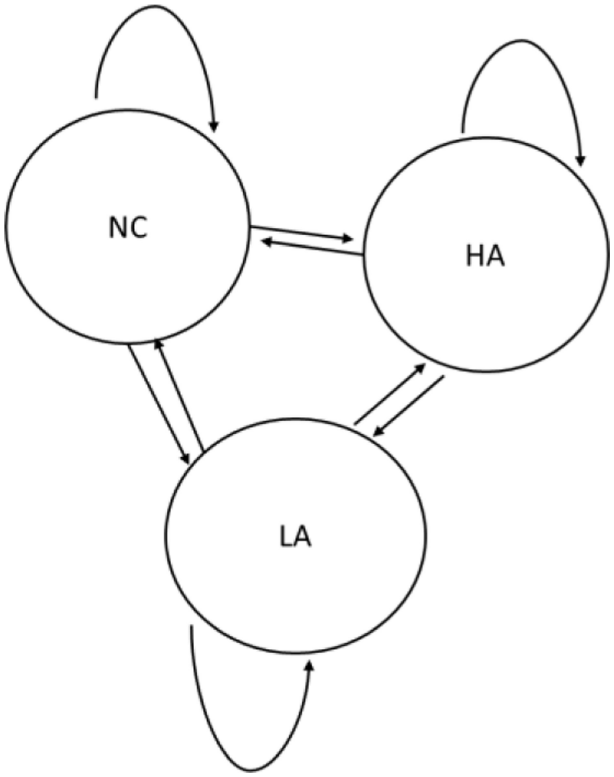


Figure 2.

State transition diagram. Arrows indicate possible transitions that can occur over one Markov cycle. Arrows leading from a circle back to itself indicates that a worker may stay in the same state over a Markov cycle. NC = no nasal carriage of S. aureus; HA = nasal carriage of human-associated *S. aureus*; LA = nasal carriage of livestock-associated *S. aureus*.

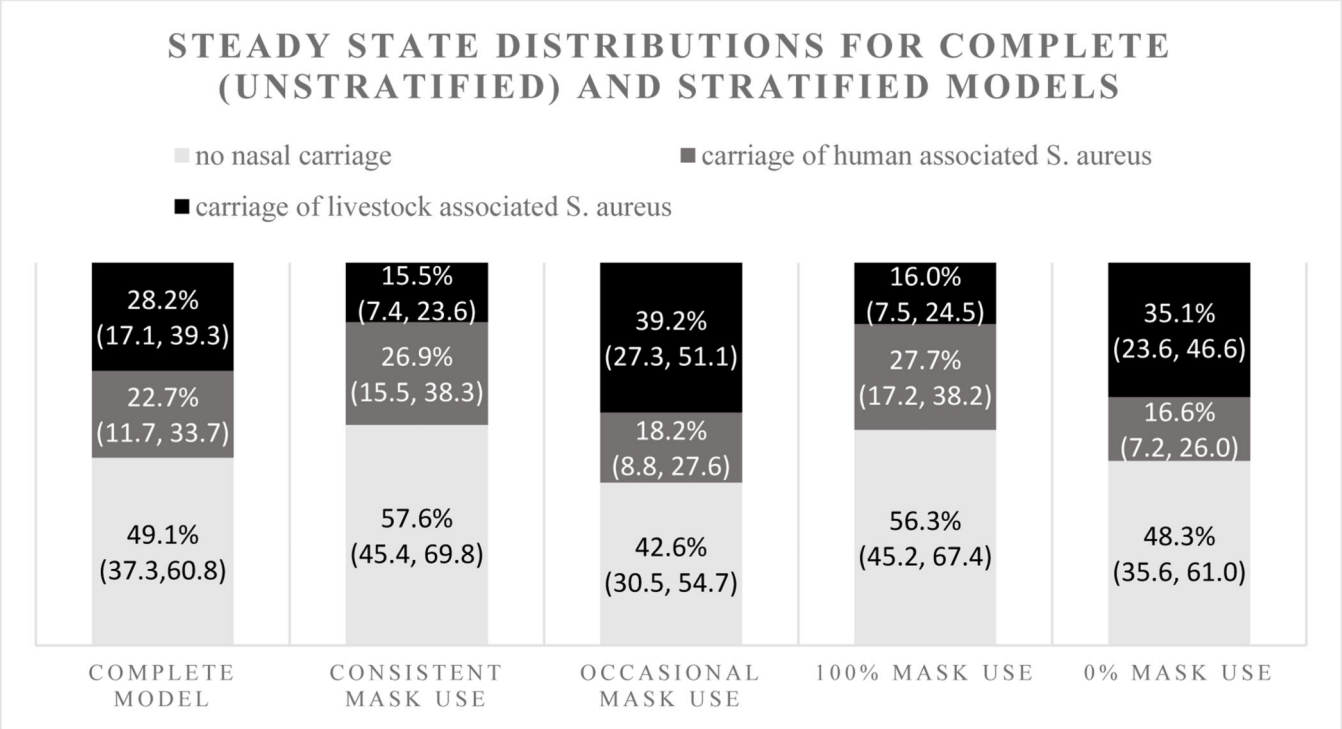
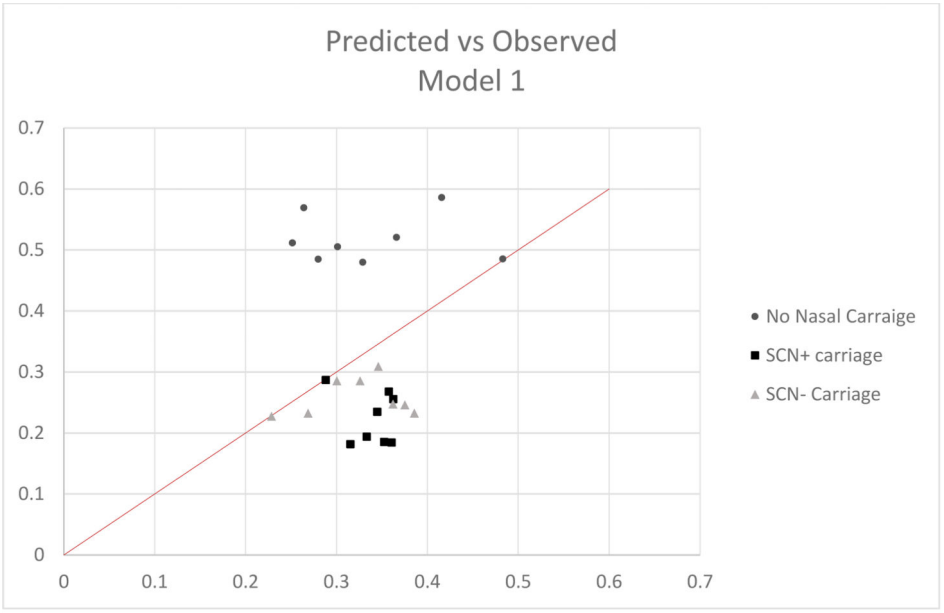


Figure 3.

Steady state distributions for the complete model (i.e. unstratified model using data from all observed transitions) and the models stratified by mask use. Percentages are shown with 95% confidence intervals in parenthesis. Consistent mask use is defined as self-reported mask use ≥ 80% of the time and occasional mask use is defined as self-reported mask use <80%.



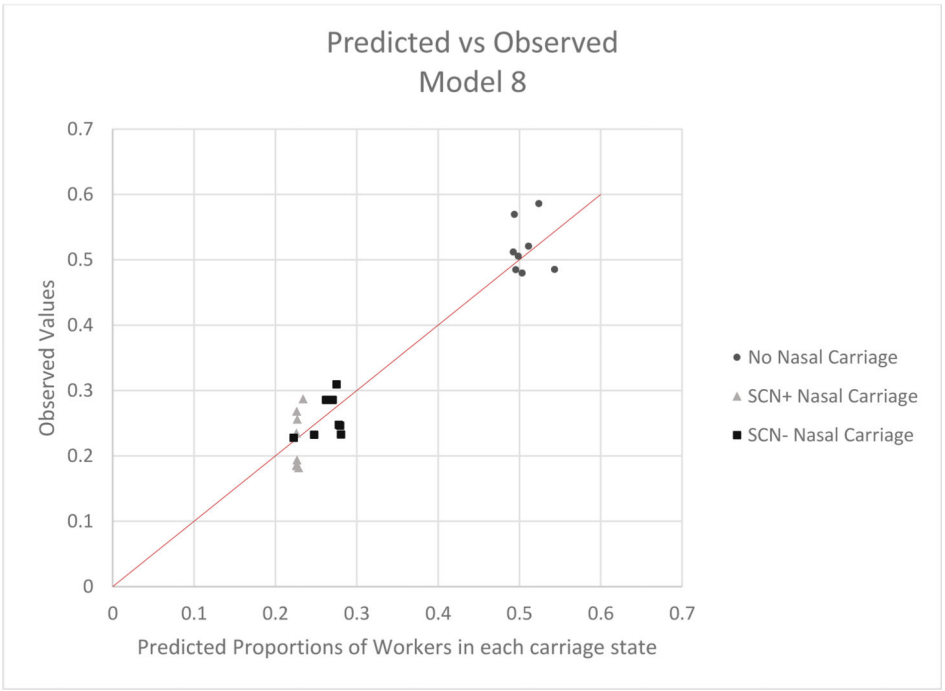


Figure 4.
Predicted versus observed values for Model 1 and Model 8.

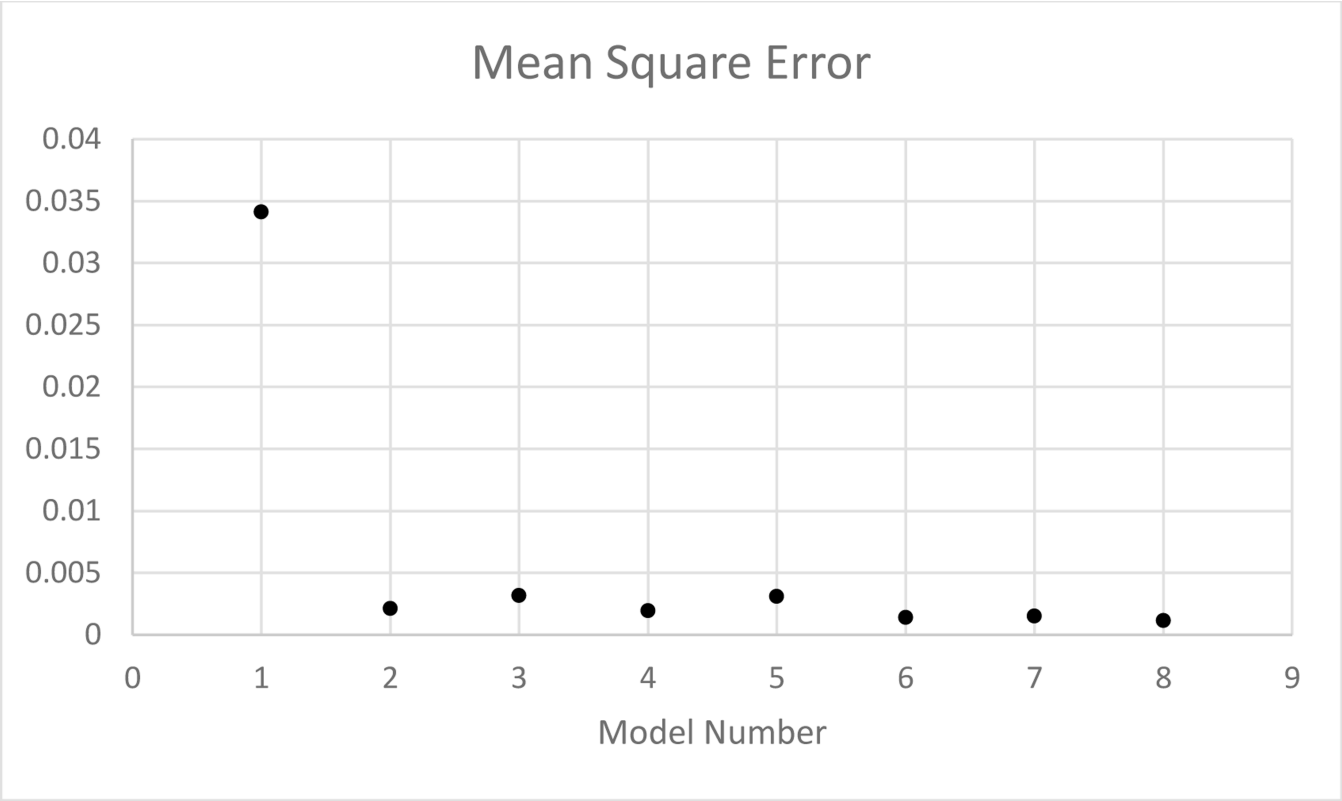


Figure 5.
Mean square error for Models 1–8 for predicting the proportion of individuals in each nasal carriage state at the final study visit. The mean squared error is computed by calculating the difference between each pair of predicted and observed values and averaging the squared differences.

Table 1.

General structure of transition probability matrices and initial state vectors for Markov models used in this study. The probability of moving between two states in one Markov cycle is represented by P_{ij}, where the worker transitions from the ith state at time t to the jth state at time t+1.

Transition Probability Matrix				
		State at time t+1		
		No nasal carriage of *S. aureus* (NC)	**Nasal carriage of human-associated *S. aureus* (HA)**	**Nasal carriage of livestock-associated *S. aureus* (LA)**
State at time t	NC	P_{00}	P_{01}	P_{02}
	HA	P_{10}	P_{11}	P_{12}
	LA	P_{20}	P_{21}	P_{22}
Initial State Vector				
State at time t = 0		N_0	N_1	N_2

Table 2.

Transition probability matrices developed using (a) all transitions observed during the study period, (b) transitions observed when workers self-reported mask use in prior two weeks, either consistent (≥ 80% reported mask use) or occasional (<80% reported mask use), and (c) the subset of transitions where workers reported complete mask use (100% reported mask use) or no mask use (0% reported mask use). For 45 observed transitions, data on mask use in the prior two weeks are missing. Gray shaded cells represent transitions to the same state (i.e. the line of identity).

(a) Transition Probability Matrix for All Observed Transitions (676 transitions)

	State at time t+1		
State at time t	**NC**	**HA**	**LA**
NC	0.802	0.084	0.114
HA	0.208	0.658	0.134
LA	0.178	0.130	0.692

(b) Transition Probability Matrix Stratified by Mask Use

Consistent Mask Use (319 transitions)	State at time t+1		
State at time t	**NC**	**HA**	**LA**
NC	0.783	0.128	0.089
HA	0.266	0.627	0.107
LA	0.344	0.172	0.484

Occasional Mask Use (312 transition)	State at time t+1		
State at time t	**NC**	**HA**	**LA**
NC	0.689	0.108	0.203
HA	0.242	0.516	0.242
LA	0.225	0.108	0.667

(c) Transition Probability Matrix for Models Stratified by Mask Use

100% Mask Use (310 transitions)	State at time t+1		
State at time t	**NC**	**HA**	**LA**
NC	0.776	0.132	0.092
HA	0.266	0.627	0.107
LA	0.328	0.180	0.492

0% Mask Use (74 transitions)	State at time t+1		
State at time t	**NC**	**HA**	**LA**
NC	0.7	0.075	0.225
HA	0.111	0.445	0.444
LA	0.36	0.16	0.48

Table 3.

Transition probability matrices and predicted final study visit prediction for Models 1 through 8 of the data sufficiency analysis. Each model was created using varying amounts of data, with the first model holding back the majority of the study data and subsequent models increasing the amount of study data used to define the transition probabilities. The transition probability matrix for Model 1 was developed using transition data from baseline to visit 1. Model 2 was built using transition data from baseline to visit 1 and transition data from visit 1 to visit 2. Model 3 uses transition data from baseline to visit 1, from visit 1 to visit 2, and from visit 2 to visit 3, and so on. Each model is used to predict the distribution of nasal carriage status among the cohort at the final study visit (follow up visit 8). The initial distribution of workers in each state, [0.574 0.248 0.178], is multiplied by the transition probability matrix raised to the 8^{th} power to produce the final study visit prediction. Transition probability matrices are given in the format displayed in Table 1. For example, for Model 1:

$$[0.574\ 0.248\ 0.178] \times \begin{bmatrix} 0.789 & 0.123 & 0.088 \\ 0.120 & 0.800 & 0.080 \\ 0.000 & 0.111 & 0.889 \end{bmatrix}^{8} = [0.252\ 0.363\ 0.385]$$

The correlation coefficient reported here refers to the correlation between the model's predictions for follow up visits 1 through 8 compared to what was observed among the cohort. The final study visit prediction refers to the models' output when predicting the distribution of nasal carriage statuses of the cohort at follow up visit 8, given as NC (95% CI); HA (95% CI); LA (95% CI). NC = no nasal carriage of S. aureus; HA = nasal carriage of human-associated *S. aureus*; LA = nasal carriage of livestock-associated *S. aureus*.

Model	Number of Transitions Used to form Matrix	Transition Probability Matrix	Correlation Coefficient (R^2)	Visit 8 Prediction (95% C.I.) NC HA LA
Model 1	100	0.789 0.123 0.088 0.120 0.800 0.080 0.000 0.111 0.889	0.0059	0.252 (0.147, 0.357) 0.363 (0.247, 0.479) 0.385 (0.263, 0.507)
Model 2	198	0.837 0.096 0.067 0.226 0.623 0.151 0.171 0.073 0.756	0.933	0.545 (0.430, 0.660) 0.191 (0.101, 0.281) 0.264 (0.157, 0.371)
Model 3	295	0.801 0.081 0.118 0.186 0.686 0.128 0.188 0.078 0.734	0.9075	0.486 (0.364, 0.608) 0.203 (0.106, 0.300) 0.311 (0.197, 0.425)
Model 4	385	0.781 0.083 0.136 0.159 0.727 0.114 0.220 0.088 0.692	0.9233	0.470 (0.343, 0.597) 0.237 (0.131, 0.343) 0.293 (0.184, 0.402)
Model 5	463	0.793 0.073 0.134 0.192 0.702 0.106 0.186 0.088 0.726	0.9058	0.478 (0.356, 0.600) 0.211 (0.110, 0.312) 0.311 (0.200, 0.422)
Model 6	547	0.790 0.086 0.124 0.190 0.698 0.112 0.186 0.114 0.700	0.926	0.474 (0.357, 0.591) 0.242 (0.147, 0.337) 0.284 (0.176, 0.392)
Model 7	613	0.797 0.083 0.120 0.209 0.657 0.134 0.183 0.124 0.693	0.9268	0.491 (0.362, 0.620) 0.222 (0.127, 0.317) 0.287 (0.170, 0.404)
Model 8	676	0.801 0.084 0.115 0.208 0.658 0.134 0.178 0.130 0.692	0.929	0.493 (0.376, 0.610) 0.227 (0.117, 0.337) 0.280 (0.169, 0.391)

Model	Number of Transitions Used to form Matrix	Transition Probability Matrix	Correlation Coefficient (R^2)	Visit 8 Prediction (95% C.I.) NC HA LA
Observed Data	--	--	1	0.508 0.246 0.246