# USE OF STRAIN TYPING DATA TO ESTIMATE BACTERIAL TRANSMISSION RATES IN HEALTHCARE SETTINGS

Brian R. Jackson, MD, MS; Alun Thomas, PhD; Karen C. Carroll, MD; Frederick R. Adler, PhD; Matthew H. Samore, MD

## ABSTRACT

**OBJECTIVE:** To create an affordable and accurate method for continuously monitoring bacterial transmission rates in healthcare settings.

**DESIGN:** We present a discrete simulation model that relies on the relationship between in-hospital transmission rates and strain diversity. We also present a proof of concept application of this model to a prospective molecular epidemiology data set to estimate transmission rates for *Pseudomonas aeruginosa* and *Staphylococcus aureus*.

**SETTING:** Inpatient units of an academic referral center.

**PATIENTS:** All inpatients with nosocomial infections.

**INTERVENTION:** Mathematical model to estimate transmission rates.

**RESULTS:** Maximum likelihood estimates for transmission rates of these two species on different hospital units ranged from 0 to 0.36 transmission event per colonized patient per day.

**CONCLUSIONS:** This approach is feasible, although estimates of transmission rates based solely on strain typed clinical cultures may be too imprecise for routine use in infection control. A modest level of surveillance sampling substantially improves the estimation accuracy (*Infect Control Hosp Epidemiol* 2005;26:638-645).

Infection control interventions aim to reduce the rate of transmission of infectious agents between patients. Unfortunately, success in achieving this goal is difficult to measure directly, as conventional metrics such as the nosocomial infection rate are affected by several factors in addition to the transmission rate. Strain typing of bacterial isolates recovered from clinically directed or surveillance cultures provides additional information.[1,2] Typing is most commonly performed as part of an outbreak investigation, but it has also proved useful in identifying transmission in the absence of clinically apparent epidemics.[3] In both circumstances, strain typing data are typically interpreted qualitatively. The question usually posed is does a particular set of colonized or infected patients represent a cluster related to cross-infection or a common source?

Another approach to quantifying transmission rates involves use of serial surveillance cultures.[4,5] By detecting colonized but uninfected individuals and ascertaining the time interval of acquisition, this approach captures much more information about transmission. However, this strategy is impractical for routine use in most hospital settings, due primarily to the cost of obtaining and processing large numbers of cultures.

It would be desirable to have a method for estimating transmission rates based on more affordable data sets and that takes full advantage of the information available from strain typing. Rapid developments in automated molecular techniques make it practical to perform strain typing on all clinical bacterial isolates grown in a hospital laboratory. Our goal was to infer transmission rates based on these types of data, fitting an explicit transmission model for purposes of parameter estimation.

This article describes an application of this method to a study in which isolates from nosocomial infections at an academic medical center were routinely strain typed during a period of 8 months. With reliance on the simplifying assumptions of the model, Monte Carlo simulation was used to produce a test statistic across a range of possible parameter values. The Monte Carlo–generated test statistic was then compared with the actual test statistic to derive a maximum likelihood estimate of the transmission parameter. The benefits and drawbacks of this technique are described, along with suggestions for future work.

## METHODS

### *Patient Population and Molecular Typing Data*

A prospective study was conducted from January to August 2000 at a 400-bed, tertiary-care medical center. The study population consisted of all patients for whom a gram-negative rod, *Staphylococcus aureus,* or both were recovered

*Dr Jackson is from the Department of Medical Informatics and Dr Carroll is from the Department of Pathology, University of Utah; and ARUP Research Institute, Salt Lake City, Utah. Dr Thomas is from the Department of Medical Informatics and Center for High Performance Computing; Dr Adler is from the Departments of Mathematics and Biology; and Dr Samore is from Veteran's Administration Health Care and the Departments of Medical Informatics and Internal Medicine, University of Utah, Salt Lake City, Utah.*

*Address reprint requests to Brian R. Jackson, MD, MS, ARUP Research Institute, 500 Chipeta Way, Salt Lake City, UT 84108.*

*Supported by National Library of Medicine Training Grant no. LM07124.*

**TABLE 1**
HOSPITAL WARDS, NUMBERS OF CLINICAL CULTURES, AND NUMBERS OF UNIQUE RIBOTYPES DURING THE 8-MONTH STUDY PERIOD

| Ward | *Pseudomonas aeruginosa* Clinical Cultures (Ribotypes) | *Staphylococcus aureus* Clinical Cultures (Ribotypes) | Total Clinical Cultures (Including Other Species) | Average Census | Average Length of Stay (d) |
|---|---|---|---|---|---|
| Burn ICU | 23 (13) | 24 (16) | 64 | 9.7 | 7.6 |
| Surgical ICU | 14 (10) | 16 (8) | 55 | 10.5 | 4.0 |
| Medical/surgical A | 7 (6) | 9 (9) | 46 | 31.9 | 3.9 |
| Neurology ICU | 1 (1) | 15 (10) | 36 | 9.2 | 3.2 |
| Physical medicine/rehabilitation | 4 (3) | 5 (4) | 33 | 17.7 | 12.4 |
| Medical/surgical B | 6 (5) | 8 (7) | 32 | 26.4 | 3.4 |
| Medical/surgical C | 3 (3) | 12 (8) | 27 | 17.3 | 4.4 |
| Newborn ICU | 1 (1) | 10 (7) | 25 | 31.3 | 26.0 |
| Medical ICU | 5 (5) | 9 (7) | 24 | 8.7 | 3.6 |
| Neurology | 3 (3) | 6 (3) | 23 | 22.1 | 3.3 |

ICU = intensive care unit.

from a clinical culture performed at least 2 days after hospital admission. In the case of multiple cultures from a single patient, only the first isolate of the infecting species was used for this modeling study. These strains were characterized by ribotyping using the RiboPrinter Microbial Characterization System (DuPont Qualicon, Wilmington, DE) with *Pvu*II as the restriction enzyme for *Pseudomonas* and *Eco*RI as the restriction enzyme for all other bacteria. Isolates were classified into ribogroups on the basis of banding pattern concordance, as determined by the computer software accompanying the RiboPrinter System. Overall, 391 isolates were ribotyped from 19 different hospital wards. These fell into 229 unique ribogroups. Approximately three-quarters of the isolates were from just three species: *Staphylococcus aureus* (120 isolates), *Escherichia coli* (85 isolates), and *Pseudomonas aeruginosa* (73 isolates).

Strain typing data were segregated into groups by species and hospital unit to apply the transmission model. Because the information content of the data is directly related to the number of data points, only data from the 10 wards with the largest numbers of clinical cultures were analyzed further. The data were further narrowed to *S. aureus* and *P aeruginosa*. These species were chosen because of the relatively large numbers of data points for them and because, in contrast to *E. coli*, these are species with which patients may be less likely to simultaneously harbor multiple strains,[6,7] which satisfies one of the key assumptions of the transmission model. The number of *S. aureus* clinical cultures during the study period per each of the 10 units studied ranged from 5 to 24; the number of *P aeruginosa* clinical cultures ranged from 1 to 23 (Table 1).

All patient-derived data were de-identified prior to use in this modeling study. The study design was reviewed by the University of Utah Institutional Review Board and determined to be exempt from federal regulation governing human research.

### *Model Description*

***Overview.*** A stochastic, discrete-time transmission model was constructed to simulate a closed hospital environment in which patients are admitted and discharged and share bacteria with each other (Appendix). Within the model, individual patients are traced so that the frequency distribution of specific strains can be followed. The parameters in the model include the number of unique strains in the community, the admission rate, the number of beds on the unit, the probability of colonization at admission, and the transmission rate (ie, the average number of transmission events per colonized patient per day). In addition to these parameters, the model incorporates the observed number and timing of clinical cultures. In this model, "colonized" refers to all patients harboring a strain of bacteria, and thus includes infected patients as well.

It is assumed that a large number of distinguishable strains of this species are present in the surrounding community. On admission, each patient who is colonized harbors a randomly selected strain. Bacterial strains are transmitted between patients on a hospital ward at a rate proportionate to the number of colonized patients. All patients are considered susceptible to acquisition of a new strain. Previously non-colonized patients who are the recipients of a transmission event become colonized with the strain of the transmitter. For previously colonized patients, a transmission event results in strain replacement with a new strain. It is assumed that colonized patients do not revert to the non-colonized state during their hospital stay. In this way, each patient harbors exactly zero strains or one strain of the species at any given point in time. Discharge is assumed to occur concurrently with admission (ie, discharge results in replacement of an individual with another patient from the community).

Cultures occur at specified points in time and allow the strain colonizing the patient to be observed. A culture

TABLE 2
RESULTS OF THE SENSITIVITY ANALYSIS

| Parameter | Range Tested | Maximum Likelihood Estimate* | Mean Simpson's Index of Diversity | Mean Colonization Prevalence |
|---|---|---|---|---|
| Probability of colonization at admission | 0.05–0.5 | 0.15–0.6 | 0.71–0.72 | 5.06–8.21 |
| Census, no. of patients | 5–15 | 0.20–0.55 | 0.71–0.71 | 7.28–7.29 |
| Average length of stay, d | 4–12 | 0.15–0.97 | 0.59–0.86 | 5.64–8.1 |
| Window size on statistic, d | 10–285 | 0.16–0.28 | | |
| No. of community strains | 10–100,000 | 0.16–0.26 | 0.68–0.71 | 7.27–7.29 |

*Transmissions per colonized patient per day.

is modeled by randomly selecting a colonized patient (as the data set includes only positive cultures) and recording the strain harbored by that patient at that point in time. The model assumes that each patient is cultured at most once during the hospital stay.

***Simulations.*** The behavior of the model was explored by simulation. Strain diversity and colonization prevalence were assessed at the end of each simulation. Strain diversity was measured using Simpson's index of diversity,[8] which is the probability of finding two different strains on choosing twice, with replacement, from a set of colonizing isolates. Colonization prevalence was the proportion of patients in the simulation who were colonized at that time.

A base-case model was developed, using parameter values derived from the following sources. The per-bed admission rate, which equaled the discharge rate, was set to the reciprocal of the observed average length of stay for each unit. The number of patients was set to the average census of each unit, rounded to the nearest integer. The number and timing of clinical cultures was taken directly from observed hospital infection data. The probability of colonization at admission was estimated from the literature for a particular species. In our study, the baseline estimate of the probability of colonization at admission was set to 0.2 for *S. aureus*,[9] and 0.3 for *P. aeruginosa*.[10] The number of unique strains in the community was set to an arbitrary large number (1,000 in the simulation experiments reported here) so that the probability of two patients being independently admitted with the same strain was much lower than the probability of two patients sharing a strain due to a transmission event. In principle, this parameter could be set to a number reflective of the discriminatory power of the strain typing method being used.

The model was programmed in Java (Sun Microsystems, Santa Clara, CA) and all simulations were run on a personal computer with a Pentium-III processor (Intel, Santa Clara, CA). As part of validating the program, a special case of the model was solved analytically using Markov chain theory.[11] The theoretic steady-state values predicted by the Markov chain were compared with results generated via Monte Carlo simulation.

### *Fitting the Model to the Data*

***Statistical Approach.*** Monte Carlo integration was used to determine maximum likelihood estimates for the target parameter (the transmission rate), as the complexity of the model made direct calculation of likelihoods impractical. A range of transmission rates was entered into the model, setting other parameters to values used in the base-case model. For each rate, a large number of simulations were performed, and the results were summarized by means of a test statistic. If this matched the test statistic observed in the original data set, the simulation was accepted; otherwise it was rejected.[12-15] The likelihood profile was then obtained by plotting the acceptance rate against the transmission rate.

***Test Statistic.*** A test statistic was devised to reflect the relative proportion of observed isolates with recent same-strain matches. More specifically, given a species and hospital ward, it equals the number of clinical isolates that match the strain of at least one previous isolate from a different patient on that ward within a fixed time window, divided by the total number of isolates for which at least one previous isolate was observed within that time window. The use of this test statistic assumed that cultures were not performed in response to a suspected cluster. Examples of valid data sets for this type of test statistic would include strain typing of all clinical isolates, random surveillance cultures, or both.

The rationale for this statistic is that in the absence of transmission between patients, bacteria within the hospital would be expected to show the same level of strain diversity as bacteria in the community. As the transmission rate increases, diversity would be expected to decrease, thus increasing the probability that two randomly selected isolates would be the same strain. As with the model itself, the test statistic presented here should be thought of as a proof of concept, rather than as an exhaustively studied optimal solution.

For the experiments reported here, the time window on the test statistic was fixed at 14 days; this choice was somewhat arbitrary and represents an empirically guided compromise. On one hand, the smaller the number of days between same-strain infections, the stronger the evidence for a high transmission rate. For this reason, decreasing the size of the window up to a certain point decreases the width of the confidence interval around the maximum likelihood estimate. On the other hand, as the window size is narrowed beyond a certain point, the decrease in the num-

ber of same-species infections begins to increase the width of the confidence interval. For most of the ward–species pairs studied here, the width of the confidence interval was minimized at approximately 14 days, and so this was fixed as the window size for all subsequent experiments.

***Monte Carlo Integration.*** Each simulation included a 100-day run-in period to reach steady-state probabilities for strain sharing among patients. Steady-state strain diversity and colonization prevalence are achieved by approximately day 25, and so the use of 100 days was conservative. This was followed by an 8-month period during which sampling occurred. For each simulation, it was recorded whether the number of same-strain infections within 14 days of each other matched the corresponding number from the ribotyping data set. The proportion of simulations with matching results was then plotted against the transmission rate, yielding the likelihood profile. The peak of this curve represents the maximum likelihood estimate for the transmission rate, and the curvature is related to the precision of the estimate. Ninety-five percent confidence intervals were approximated from support intervals according to the method of Hudson[16] as the range of transmission rates for which the natural log of the likelihood was within 2.0 of the maximum log-likelihood.

For each species–ward pair, the model was initially run for 10,000 simulations per transmission rate, varying the transmission rate over a logarithmic scale. The resulting likelihood curves were all unimodal with unambiguous peaks. The locations of the peaks and the limits of the confidence intervals were then determined more precisely by rerunning the model at 50,000 simulations per transmission rate and varying the transmission rate linearly over regions around the peak, lower confidence limit, and upper confidence limit predicted by the initial runs. Because the log-likelihood curves were approximately linear in the regions around the confidence interval limits, the values for these limits were determined by linear interpolation.

### *Sensitivity Analysis*

For sensitivity analysis, the burn intensive care unit *P. aeruginosa* parameters were used as a baseline. Each of the input parameters, as well as the length of the run-in period, were varied across a biologically plausible range for the burn intensive care unit. The effects of these changes on the maximum likelihood estimates, on mean strain diversity, and on mean colonization prevalence are presented in Table 2. The maximum likelihood estimates were derived using the method described above; the Simpson's index of diversity and mean colonization prevalence were studied separately by running 10,000 simulations for each set of parameters.

To investigate the benefit of increasing the amount of information available to the model, the effect of performing weekly surveillance cultures on the confidence limits of the estimate of the transmission rate was assessed. Non-infected patients in the model were randomly selected at specified times and strains were recorded. A maximum of one culture was assumed per patient per hospital stay, regardless of whether it was a clinical culture or a surveillance culture. The surveillance culture simulations used the parameters and *P. aeruginosa* infection data from the burn intensive care unit, with the test statistic fixed at values that would give the same maximum likelihood estimates for the transmission rate as the unaugmented burn intensive care unit data. The model was then applied to the augmented burn intensive care unit data, including clinical plus simulated surveillance cultures, to determine the confidence intervals around the maximum likelihood estimates.

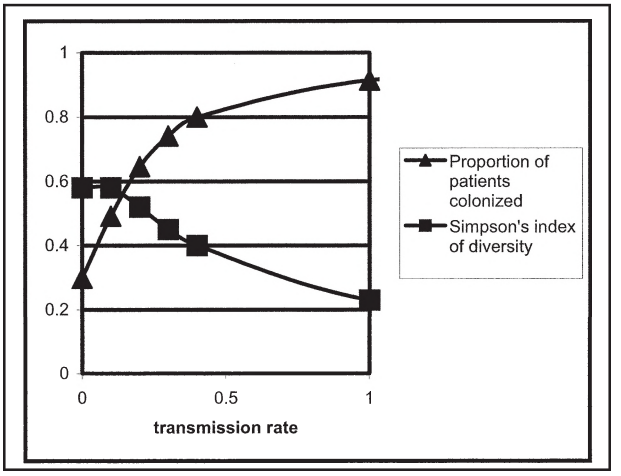


**FIGURE 1.** Predicted colonization prevalence and strain diversity versus transmission rate.

## RESULTS

### *Characteristics of the Model*

With the use of the burn intensive care unit *P. aeruginosa* parameters as a baseline, the model predictions for colonization prevalence and strain diversity were studied by calculating these values at the end of each of a large number of simulations. The mean colonization prevalence predicted by the model was 72.7%, with 2.5th and 97.5th percentile values of 40% and 100%, respectively. The mean Simpson's index of diversity was 0.46, with 2.5th and 97.5th percentile values of 0.00 and 0.78, respectively. The effects of varying the transmission rate on mean colonization prevalence and mean Simpson's index are shown in Figure 1.

### *Application to Hospital Data Set*

Across the hospital wards studied, the estimated transmission rate for *S. aureus* ranged from 0.0 to 0.36 transmission per colonized patient per day, and the estimated rate for *P. aeruginosa* ranged from 0.13 to 0.28 (Table 3). There was substantial overlap among the 95% confidence intervals. For several of the ward–species subsets, there were no same-strain clinical cultures within 14 days of each other, and so the maximum likelihood estimate for the transmission rate was 0. For wards

TABLE 3
INFECTION NUMBERS AND MAXIMUM LIKELIHOOD ESTIMATES FOR TRANSMISSION RATES

| Hospital Ward | Species | All Infections Within 14 Days of Previous Infection | Infections Within 14 Days of Previous Same-Strain Infection | Crude Infection Rate* | Maximum Likelihood Estimate for Transmission Rate ($CI_{95}$)† |
|---|---|---|---|---|---|
| Burn ICU | *Pseudomonas aeruginosa* | 17 | 7 | 0.076 | 0.26 (0.09–1.19) |
| | *Staphylococcus aureus* | 18 | 3 | 0.079 | 0.07 (0.01–0.14) |
| Surgical ICU | *P. aeruginosa* | 9 | 1 | 0.021 | 0.13 (0.005–0.69) |
| | *S. aureus* | 11 | 4 | 0.024 | 0.36 (0.11–2.1) |
| Medical/surgical A | *S. aureus* | 3 | 0 | 0.0046 | 0.0 (0.0–2.58) |
| Neurology ICU | *S. aureus* | 9 | 1 | 0.021 | 0.13 (0.006–0.58) |
| Medical/surgical B | *S. aureus* | 3 | 0 | 0.0038 | 0.0 (0.0–4.21) |
| Medical/surgical C | *S. aureus* | 5 | 1 | 0.012 | 0.18 (0.01–2.19) |
| Newborn ICU | *S. aureus* | 4 | 0 | 0.035 | 0.0 (0.0–0.18) |

ICU = intensive care unit; $CI_{95}$ = 95% confidence interval.
*Infections per patient-day.
†Transmissions per colonized patient per day.

and species with only a few clinical cultures, the confidence intervals were so wide as to render the maximum likelihood estimate of the transmission rate essentially meaningless. Thus, the only results reported here are those from ward–species subsets with confidence interval widths of less than 5 transmissions per colonized patient per day.

The widths of the confidence intervals are directly related to the numbers of cultures for each species–ward subset. This is shown most clearly when the simulated surveillance cultures are added to the clinical cultures. In Figure 2, the confidence intervals around the transmission rate estimate are plotted against the total number of cultures performed. Given this transmission model and the observed data, it appears that adding a modest number of random surveillance cultures (equal to one or two times the number of clinical cultures) would markedly narrow the width of the confidence interval. On the other hand, adding further surveillance cultures beyond that number would be of limited benefit.

The sensitivity analysis (Table 2) showed maximum likelihood estimates to be most sensitive to the length of stay, probability of colonization at admission, and census. Of these parameters, length of stay and census can be directly observed in practice, significantly reducing their potential to introduce error into the final estimate. Maximum likelihood estimates were fairly stable regarding the number of community strains and the size of the time window.

## DISCUSSION

This study addressed the need for an efficient method to quantitatively monitor transmission rates in healthcare settings. We have described a new dynamic transmission model that incorporates strain typing data and allows both non-colonized and colonized patients to acquire new organisms from other colonized patients. Calculations on the model are performed by Monte Carlo integration. Several simplifying assumptions are imposed, but the fallacy of conventional statistical methods that assume infectious events in individuals are independent is avoided. The model is presented as a proof of concept; further work would be required to create an application to meet the needs of a hospital epidemiology service.

An advantage of this methodology is that it is feasibly applied to data from relatively small numbers of patients. We demonstrated its use by analyzing results of a prospective study in which all unique clinical isolates of gram-negative rods and *S. aureus* were ribotyped. The model-derived rates of transmission for different organisms across intensive care units varied substantially, with *S. aureus* in the surgical intensive care unit ranked the highest. However, the point estimates exhibited wide confidence limits. This imprecision is not unexpected, given the relatively small numbers of patients on each ward, the even smaller numbers of infections, and the resulting stochastic variability in the model. Other hospital transmission models have likewise shown a high degree of stochastic variation.[17] It is encouraging that the precision of these estimates can be markedly improved by selecting one or two patients per week on the ward for surveillance cultures.

Several other authors have estimated nosocomial transmission rates by fitting data to mathematical models.[18-21] Most of these previous studies relied on information gathered from extensive surveillance culturing.[18-20] Such culturing strategies have not become widespread outside of research settings, in part due to the costs involved. In contrast, Cooper and Lipsitch have proposed a hidden Markov model approach that can be used with clinical culture data only.[21] Their transmission rate estimates

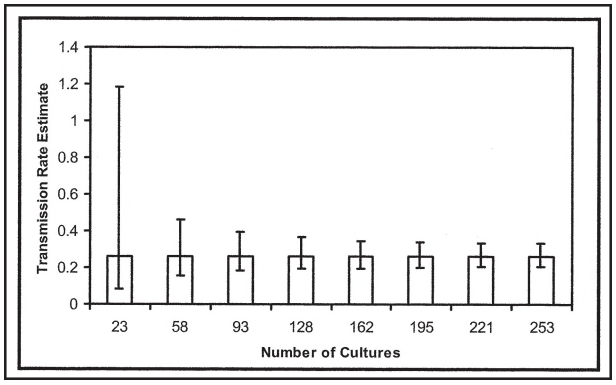


**FIGURE 2.** Transmission rate estimates with 95% confidence intervals versus number of cultures (burn intensive care unit *Pseudomonas aeruginosa* with simulated surveillance cultures).

were associated with tighter confidence intervals than those in this article, but they used 40 months of data for each estimate as compared with 8 months in this article. Thus, it is difficult to directly compare the results.

A key difference between the model proposed in this article and other models of nosocomial infection is that it is based on strain diversity, rather than on the numbers of clinically infected or colonized individuals. In the model reported here, the dates of detection of clinical infections are fixed to occur on the dates that organisms were recovered. Conditioning on the actual infection date assumes that factors that influence clinical infection are independent of the underlying transmission dynamics and that different strains are equally likely to cause clinical infection. This is an advantageous assumption to the extent that clinical infection rates are sensitive to many sources of variation besides transmission, such as underlying illness and medical device use. This variation is controlled for by this methodology.

Another key distinction between this model and several previously published transmission models is the use of discrete event simulation. In this approach, subjects are individually traced to allow monitoring of each colonized patient's strain type. Although deterministic models (including those based on differential equations) may be more mathematically satisfying as well as more familiar to readers, it is not clear how well they model the highly nondeterministic transmission patterns seen at the scale of an individual hospital unit.[17]

The transmission rate as presented here is related to, but not identical to, the basic transmission parameter $R_0$. $R_0$ is defined as the average number of secondary cases of infection, or sometimes colonization, resulting from a single primary case in a fully susceptible population.[22] The transmission rate as presented here can be thought of as $R_0$ divided by the average post-colonization length of stay. Examples of published $R_0$ estimates in intensive care settings include 0.14 for *P. aeruginosa*[20] and 0.69 for vancomycin-resistant enterococci.[19] When length of stay is adjusted for, the estimates produced by this model correspond to $R_0$ values ranging from 0 to 2.1.

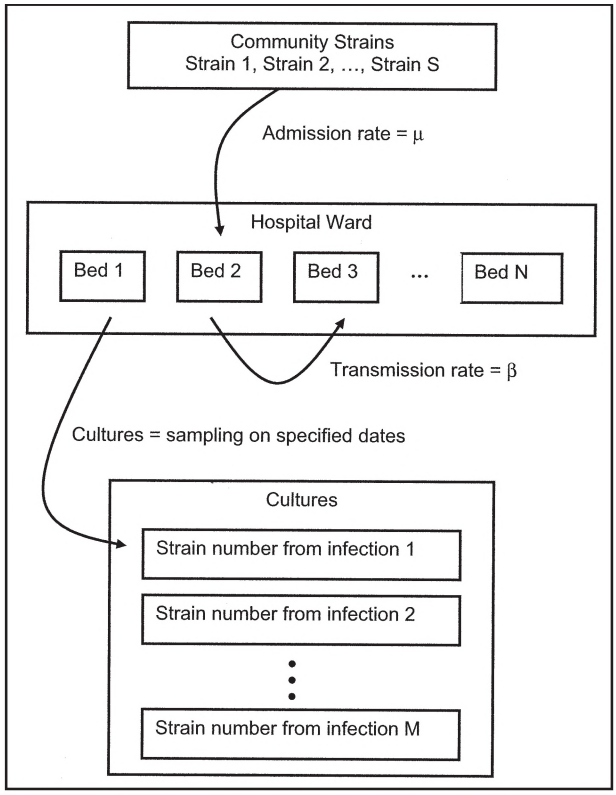


**FIGURE 3.** Diagram of the transmission model.

This model includes several simplifying assumptions. It assumes a constant admission rate regardless of infection status, so that lengths of stay follow an exponential distribution. It assumes constant transmissibility and infectivity across strains and across patients, and equal prevalence of each community strain. (An example of a violation of this last assumption would be patients returning to the hospital still colonized from the previous visit.) The model also assumes that colonized patients have exactly one strain, are as susceptible as non-colonized patients to acquiring a new strain, and are all equally likely to transmit to new patients (eg, without regard to infection status of the transmitter). Assessing the significance of this set of limitations would require comparing the performance of this model with that of a somewhat more complex strain carriage and transmission model. Finally, it assumes that all patients are admitted directly to the ward of interest from the community, with no transmission between patients on different wards. This last assumption in particular is violated to some degree on certain hospital units such as intensive care units. Assessing the significance of this violation will require prospective assessment of colonizing strain diversity on admission to different units. All of these aspects of the model could in principle be made more biologically realistic, but at the expense of adding additional parameters, the estimation of which would add additional sources of error.[23]

This model was not designed to be able to detect emergence of new strains, whether the result of selective pressure such as antibiotic use or importation of strains from other communities. In fact, the model assumes that the number and distribution of strains do not change over time. One practical impact of this might be the need for periodic reassessment of community strain diversity to recalibrate the model.

The model is only moderately sensitive to the number of distinguishable strains in the community, which in turn is a function of the discriminatory power of the genotyping method. Thus, methods with more or less discriminatory power than the RiboPrinter would be expected to produce somewhat narrower or wider confidence intervals, respectively. The fact that the model, or for that matter the Simpson's index, is not more sensitive to the number of community strains may be counterintuitive. This is perhaps best explained by the fact that the probability of two infections being of the same strain in the absence of transmission is proportional to the inverse of the number of community strains. These inverses are all small relative to the probabilities of detecting transmission-related same-strain infections, at least within the ranges of parameters tested. For a few species, one or two ribotypes may account for a large proportion of colonization in a community.[24] For such species, this model could give skewed results.

Given these sources of structural and parametric uncertainty, it is difficult at this point to assess the absolute accuracy of the transmission rate estimates produced by this model. For example, it currently would be premature to use these estimates to draw conclusions about the relative transmission rates of *S. aureus* versus *P. aeruginosa* in different hospital locations. The need for accuracy depends on the application under consideration. For quality improvement, the key issue is the ability to detect changes over time, and so precision is more important than absolute accuracy. For comparison between wards or between institutions, the key issues are precision and avoidance of differential biases. Further validation of the model and assessment of its accuracy will require a prospective clinical study with serial culturing as a gold standard to determine transmission rates.

These results suggest that bacterial strain typing can be combined with a mathematical model to estimate transmission rates in healthcare settings. They also suggest that analysis based on strain typed clinical cultures alone is probably insufficient to produce precise enough transmission rate estimates for use in routine infection control. One obvious way to extend the model, as shown in Figure 2, is to add a modest number of surveillance cultures. We are currently working on a more general statistical model using Markov chain Monte Carlo techniques.[25,26] This will allow incorporation of additional data, such as time from admission to infection and results of repeat surveillance cultures on selected patients, to improve precision and control for variable distributions of strains at admission.

## REFERENCES


1. Chetchotisakd P, Phelps CL, Hartstein AI. Assessment of bacterial cross-transmission as a cause of infections in patients in intensive care units. *Clin Infect Dis* 1994;18:929-937.
2. Bischoff WE, Reynolds TM, Hall GO, Wenzel RP, Edmond MB. Molecular epidemiology of vancomycin-resistant *Enterococcus faecium* in a large urban hospital over a 5-year period. *J Clin Microbiol* 1999;37:3912-3916.
3. Almuneef MA, Baltimore RS, Farrel PA, Reagan-Cirincione P, Dembry LM. Molecular typing demonstrating transmission of gram-negative rods in a neonatal intensive care unit in the absence of a recognized epidemic. *Clin Infect Dis* 2001;32:220-227.
4. Samore MH, Bettin KM, DeGirolami PC, et al. Wide diversity of *Clostridium difficile* types at a tertiary referral hospital. *J Infect Dis* 1994;170:615-621.
5. Bauer TM, Ofner E, Just HM, Daschner FD. An epidemiologic study assessing the relative importance of airborne and direct contact transmission of microorganisms in a medical intensive care unit. *J Hosp Infect* 1990;15:301-309.
6. Thuong M, Arvaniti K, Ruimy R, et al. Epidemiology of *Pseudomonas aeruginosa* and risk factors for carriage acquisition in an intensive care unit. *J Hosp Infect* 2003;53:274-282.
7. Cespedes C, Said-Salim B, Miller M, et al. The clonality of *Staphylococcus aureus* nasal carriage. *J Infect Dis* 2005;191:444-452.
8. Peet RK. The measurement of species diversity. *Annual Review of Ecology and Systematics* 1974;5:285-307.
9. Jernigan JA, Pullen AL, Flowers L, Bell M, Jarvis WR. Prevalence of and risk factors for colonization with methicillin-resistant *Staphylococcus aureus* at the time of hospital admission. *Infect Control Hosp Epidemiol* 2003;24:409-414.
10. Bonten MJM, Bergmanns DCJJ, Speijer H, Stobberingh EE. Characteristics of polyclonal endemicity of *Pseudomonas aeruginosa* colonization in intensive care units. *Am J Respir Crit Care Med* 1999;160:1212-1219.
11. Cinlar E. *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall; 1975:152.
12. Marjoram P, Molitor J, Plagnol V, Tavare S. Markov chain Monte Carlo without likelihoods. *Proc Natl Acad Sci U S A* 2003;100:15324-15328.
13. Fu YX, Li WH. Estimating the age of a common ancestor of a sample of DNA sequences. *Mol Biol Evol* 1997;14:195-199.
14. Weiss G, von Haeseler A. Inference of population history using a likelihood approach. *Genetics* 1998;149:1539-1546.
15. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol Biol Evol* 1999;16:1791-1798.
16. Hudson DJ. Interval estimation from the likelihood function. *Journal of the Royal Statistical Society* 1971;33:256-262.
17. Cooper BS, Medley GF, Scott GM. Preliminary analysis of the transmission dynamics of nosocomial infections: stochastic and management effects. *J Hosp Infect* 1999;43:131-147.
18. Sebille V, Valleron A-J. A computer simulation model for the spread of nosocomial infections caused by multidrug-resistant pathogens. *Computers and Biomedical Research* 1997;30:307-322.
19. Austin DJ, Bonten MJM, Weinstein RA, Slaughter S, Anderson RM. Vancomycin-resistant enterococci in intensive-care hospital settings: transmission dynamics, persistence, and the impact on infection-control programs. *Proc Natl Acad Sci U S A* 1999;96:6908-6913.
20. Pelupessy I, Bonten MJM, Diekmann O. How to assess the relative importance of different colonization routes of pathogens within hospital settings. *Proc Natl Acad Sci U S A* 2002;99:5601-5605.
21. Cooper B, Lipsitch M. The analysis of hospital infection data using hidden Markov models. *Biostatistics* 2004;5:223-237.
22. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press; 1991.
23. Hilborn R, Mangel M. *The Ecological Detective: Confronting Models With Data*. Princeton, NJ: Princeton University Press; 1997:153-160.
24. Diekema DJ, Pfaller MA, Turnidge J, et al. Genetic relatedness of multidrug-resistant, methicillin (oxacillin)-resistant *Staphylococcus aureus* bloodstream isolates from SENTRY Antimicrobial Resistance Surveillance Centers worldwide, 1998. *Microb Drug Resist* 2000;6:213-221.
25. Auranen K, Arjas E, Leino T, Takala AK. Transmission of pneumococcal carriage in families: a latent Markov process model for binary longitudinal data. *Journal of the American Statistical Association* 2000;95:1044-1053.
26. O'Neill PD, Balding DJ, Becker NG, Eerola M, Mollison D. Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Applied Statistics* 2000;49(part 4):517-542.

# APPENDIX

## DESCRIPTION OF THE MODEL (FIG. 3)

### *Variables*

S = number of unique strains of the species of interest in the community. All are assumed to exist at an equal frequency.

N = number of hospital beds. All beds are assumed to be filled at all times, so that when a patient is discharged, a new one is simultaneously admitted.

$\mu$ = admission rate.

p = probability that a randomly chosen patient is colonized at the time of admission. Thus, the probability that a randomly chosen patient is not colonized at admission equals 1 - p, and the probability that a randomly chosen patient is colonized at admission with a specified strain equals p/S.

$\beta$ = transmission rate = average number of transmission events per colonized patient per day.

### *Model Events*

Admission/Discharge: An admission/discharge event consists of randomly selecting a bed, deleting the strain identifier associated with that bed (if any), and then assigning a new strain identifier to that bed with probability p. Admission/discharge events take place randomly throughout each simulated model day according to a Poisson process with parameter $\mu$.

Transmission: For each transmission event, one patient is randomly selected from among all admitted, currently colonized patients to be the transmitting patient. A different patient is then selected from among all currently admitted patients to be the receiving patient. The receiving patient's colonization status is then set to true (regardless of whether this patient had previously been colonized), and the strain number is set to the strain number of the transmitting patient. Transmission events take place randomly throughout each simulated model day according to a Poisson process with parameter $\beta$.

Culture: Culture dates are set according to the dates in the observed data set. Thus, if the observation data included cultures on each of the days 1, 2, and 10, then the model would simulate a culture on days 1, 2, and 10 of each simulation. Patients are assumed to have a maximum of one culture per hospital stay. Each simulated culture consists of randomly choosing one patient from among all currently colonized patients who had not previously had a culture and recording the strain number carried by that patient.