An Algorithm to Estimate the Importance of Bacterial Acquisition Routes in Hospital Settings

M. C. J. Bootsma† [1], M. J. M. Bonten[2,3,4], S. Nijssen[2], A.C. Fluit[3], O. Diekmann[1]

Affiliations
[1]*Department of Mathematics, Utrecht University, Utrecht, The Netherlands*
[2]*Department of Internal Medicine & Dermatology. University Medical Center Utrecht. Utrecht, The Netherlands*
[3]*Eijkman Winkler Institute for Microbiology, Infectious Diseases and Inflammation. University Medical Center Utrecht. Utrecht, The Netherlands*
[4]*Julius Center for Health research & Primary Care; University Medical Center Utrecht. Utrecht, The Netherlands*
† **Corresponding author**:
E-mail: bootsma@math.uu.nl,

Abstract

An algorithm is presented to calculate likelihoods of acquisition routes solely using individual patient data concerning period of stay and microbiological surveillance (without genotyping). The algorithm also produces estimates for the prevalence and the number of acquisitions by each route. The algorithm is applied to colonization data of third-generation-cephalosporin-resistant Enterobacteriaceae (CRE) in two Intensive Care units (ICUs), using genotyping and epidemiological linkage as reference standard. Surveillance cultures were obtained on admission and twice weekly thereafter. All CREs were genotyped. Based upon the reference standard, daily prevalence of CRE in ICU1 and ICU2 was 26.1% (standard deviation 15.4%) and 15.1% (standard deviation 13.4%), respectively, with five out of 23 (21.7%) and six out of 21 (28.6%) cases of acquired colonization being of exogenous origin, respectively. The algorithm predicted the likelihood of predominance of the endogenous over the exogenous route to be 99.7% and >99.9% for ICU1 and ICU2 respectively. The estimated number of acquisitions is 29.8 and 27.2 and the estimated prevalence is 27.6% and 17.6% for ICU1 and ICU2, respectively. Using longitudinal colonization data only, the algorithm can be used to determine the relative importance of acquisition routes and to quantify effects of interventions, taking patient-dependency into account.
Running title: Modeling acquisition routes
Keywords: maximum likelihood, transmission parameters, algorithm, small populations, Enterobacteriaceae

Abbreviations: MLE, Maximum Likelihood Estimate; CRE, cephalosporin resistant Enterobacteriaceae; ICU, Intensive Care Unit; MRSA, methicillin-resistant Staphylococcus aureus; VRE, vancomycin-resistant Enterococci; SD, standard deviation; CI, confidence interval; MCMC, Monte Carlo Markov Chain
Word count abstract: 198
Word count for the text, exclusive of references, tables, and figure legends: 6781 (including formulas).

Within health care settings, antibiotic resistance increasingly hampers successful treatment of infections, especially in intensive care units (ICUs) (1). For some pathogens (e.g., vancomycin-resistant *Staphylococcus aureus*, pan-resistant *Pseudomonas aeruginosa* and Acinetobacter species) the post-antibiotic era is approaching. With a limited armamentarium of antibiotics remaining available for treatment, infection prevention becomes more and more important. The epidemiology of antibiotic resistance in hospital settings, however, is complex and quantitative understanding of the dynamics is essential for designing efficient infection control strategies. As only a fraction of colonized patients will develop infections (2), the true volume of antibiotic resistance is best represented by asymptomatic carriage (i.e., colonization). Changes in the prevalence of colonization with antibiotic-resistant microorganisms within hospital settings may occur through different processes: admission and discharge of colonized and non-colonized patients; mutations, changing susceptible bacteria into resistant ones, followed by selection due to antibiotic pressure; and cross transmission, usually via temporarily contaminated hands of health care workers (3). A key characteristic of cross transmission is dependence among patients. The risk of acquisition (also called 'colonization pressure') is influenced by the colonization status of other patients (4). This has been demonstrated for methicillin-resistant Staphylococcus aureus (MRSA) (5), vancomycin-resistant Enterococci (VRE) (6) and Enterobacteriaceae (7).

Because of the typically small patient populations in ICUs (usually <20) and the rapid patient turnover, large fluctuations in proportions of colonized patients occur naturally (3). Also, the dependence created by cross transmission leads to overdispersion and autocorrelation in the number of colonized patients per day (8). So, the distribution of the number of patients colonized at a given day will be skewed and the variance to mean ratio of the number of patients colonized per day will exceed 1. Processes in which patients interact are usually called 'non-linear'. In contrast, mutations, selec-

tion of resistant flora and admission of colonized patients occur independently of the colonization status of other patients and these processes are called linear. For these processes there is still autocorrelation in the number of colonized patients per day (as patients stay in the unit for some time), but, when data cover a long time period, the number of patients colonized each day will be binomially distributed.

As the quantification of infection routes is relevant for the design of infection control strategies, as well as for the interpretation of the observed effects of interventions (8, 9), our aim here is to determine from the available data the relative importance of the various routes leading to detectable colonization. The Markov model proposed by Pelupessy et al. (10), uses longitudinal data concerning the number of patients colonized with a certain pathogen as input for maximum likelihood estimation of acquisition parameters. The extension introduced here uses data on individual patients, with the advantage that we

- can explicitly distinguish rates of admission of colonized patients from endogenous selection rates
- can use the actual changes in bed occupancy (as in (13)), i.e., there is no need to assume that all beds are occupied and that the length of stay is exponentially distributed
- can take the moments of obtaining cultures and the results of these cultures as the bookkeeping cornerstone of the model, while a stochastic model estimates the status of patients in-between culture sampling moments.
- allow for incorporation of other patient characteristics, e.g., antibiotic use.

So the model formulation is data driven from the very beginning and incorporates all the information that is available. It yields maximum likelihood estimates (MLE's), as well as confidence regions, for acquisition parameters, thus enabling the identification of the dominant acquisition route. Moreover, the probability that a specific patient is colonized at a given time can be determined. This allows for calculation of relevant quantities, e.g., the prevalence and the expected number of acquisitions in the unit.

Here, we have performed a 'proof of principle' by making a prospective comparison of model predictions on the relative importance of endogenous and exogenous acquisition of third-generation cephalosporin-resistant Enterobacteriaceae (CRE) in two ICUs to the reference standard derived from extensive surveillance and genotyping data, that are available in this particular case. We have chosen CRE as marker pathogen because, according to the literature, both endogenous and exogenous acquisition can contribute to its epidemiology (11, 12). This is in contrast to the epidemiology of MRSA and VRE, where exogenous acquisition is known to be a much more dominant acquisition route.

The underlying acquisition model

An important building block for our algorithm is a mechanistic acquisition model, which governs the changes in colonization status. It incorporates the different infection routes and, preferably, requires the specification of only few parameters. The mechanistic model that we employ in the present study is characterized by the following assumptions:

- patients can be in two states: either a patient is colonized, i.e., he/she carries the pathogen of interest at a level that is, in principle, detectable or a patient is uncolonized, i.e., does not carry the infective agent at a detectable level.
- once a patient becomes colonized, he/she remains colonized during the rest of the stay.
- uncolonized patients can acquire colonization exogenously by transmission or can go through an endogenous process in which the (already present) pathogen grows to detectable levels.
- when we know the colonization status of all patients at a certain time, we know the probability per unit of time for uncolonized patients to acquire colonization. More precisely, we assume that the probability per susceptible patient to turn into a colonized patient per infinitesimal small unit of time Δt is $(\alpha + \beta I/n)\Delta t$, where α represents the endogenous term and $\beta I/n$ the cross transmission term, with I the number of colonized patients in the ICU and n the total number of patients in the ICU. Both α and β should be non-negative and are to be estimated from the data. (Note incidentally that the parameter for cross transmission β can also be expressed in terms of a reproduction number R_N which gives the average number of secondary infected cases if all individuals in the ICU are non-colonized (23). For small values of β, R_N is approximately equal to βD, with D the average length of stay in the unit.)
- as only the days of culturing/admission/discharge are known, and not the exact moments, we use a day as the smallest time unit in our model and pretend that admission and discharge always occur at one and the same hour (say, 12.00).
- we assume that uncolonized patients can acquire colonization only from those patients who were already colonized at the hour of admission and discharge and not in a two step procedure during one and the same day (i.e., patients who become colonized can not infect other patients during the same day). This leads to a per diem probability per susceptible patient to acquire colonization of $1 - e^{-(\alpha+\beta\frac{I}{n})}$.

The algorithm

As input data we need for each patient:

- day of admission
- day of discharge
- days at which a sample is taken (which is cultured)
- results of cultures (assumed, for the time being, to be 100 percent reliable)
- the colonization status at admission

The output are MLE's (and confidence intervals) of the acquisition parameters α and β, MLE's for the probabilities that patients are colonized on each day and related quantities as the total number of acquisitions and the fraction of the acquisitions that can be ascribed to each acquisition route.

For those patients for which the colonization status is unknown, we work with the probability of being colonized. From day to day these probabilities evolve according to the

mechanistic acquisition model. So once a culture result becomes available, we know how likely it was, given the parameters in the acquisition model. This "knowledge" then serves as the basis for the MLE. The rest is bookkeeping: we need to incorporate that patients are discharged and that new patients are admitted (note that it is difficult to assign a probability of being colonized to a newly admitted patient, but that it is straightforward if patients are cultured on admission).

A detailed technical description of the representation of the ICU state and the various operations that update this state on the basis of the mechanistic model and the data is given in the appendix. There we also explain how to use the algorithm to calculate relevant epidemiological quantities, e.g., the prevalence per day and the expected number of acquisitions that can be ascribed to each acquisition route.

Justification of the use of the algorithm

The standard method (15) to calculate confidence sets is only asymptotically correct (when the length of the study period approaches infinity) and requires that the true parameters are not on the boundary of the domain, hence that both of the colonization routes are of importance. To test how well the asymptotic theory performs for finite study periods and when the true acquisition parameters are on the boundary of the domain, we simulated an ICU with 10 beds, which are always occupied. We varied the relative importance of the acquisition routes, but kept the mean prevalence in the ICU constant at 20 percent, while 5 percent of the patients was colonized on admission. The length of stay in the ICU was exponentially distributed with a mean of 8 days. Observation of colonization was assumed to be perfect. Results are based on 100,000 simulations. For each simulation, we applied our algorithm to calculate 95% confidence sets for the acquisition parameters and we calculated the fraction of the simulations for which the true acquisition parameters where contained in the calculated confidence set.

For the clinical study, a goodness of fit χ^2-test (with two free parameters) based on the MLE was performed to test whether the model fitted the data accurately.

Setting of the clinical study

Colonization with CRE was studied in a medical (ICU-1) and a neurosurgical ICU (ICU-2) of the University Medical Center Utrecht, The Netherlands. This study was approved by the institutional review board. No informed consent was required. ICU-1 has 10 beds, four of which are in separate rooms and ICU-2 has 8 beds, one in a separate room. Nursing and medical staff is not shared between these ICUs. Standard infection control measures were used in both units and these did not change during the period of study.

Microbiological surveillance and genotyping

During an eight-months period, rectal colonization with CRE was determined in all patients admitted to the two ICUs. Rectal swabs were obtained on admission and twice weekly thereafter. Swabs were plated on Chromogenic UTI Agar (Oxoid Limited, Basingstoke, UK) supplemented with 8 μg/ml cefpodoxime (Aventis Pharma, Paris, France) and 6 μg/ml vancomycin. All morphologically different colonies were further processed. Species identification was performed using VITEK II (bioMérieux Benelux B.V., 's Hertogenbosch, Netherlands). Additional susceptibility testing was performed by microdilution according to CLSI guidelines and, subsequently, all isolates not resistant to either cefpodoxime or ceftazidime were excluded from analysis. Two morphologically different isolates per species per patient (if available) were genotyped using Amplified Fragment-Length Polymorphism (AFLP) (16). If more than two isolates of one species were available, first and last isolates were selected. Genetic relatedness was determined on the basis of both visual and computerized interpretation of AFLP patterns of isolates of epidemiological linked patients. A similarity of more than 80 percent, based on similarities in AFLP-patterns among multiple isolates obtained from individual patients, was used as cut-off point.

Colonization with CRE was classified as 'present on admission' when CRE was demonstrated in cultures obtained $<$48 hours after admission and as 'acquired' when demonstrated in cultures obtained $>$48 after admission with a previous negative culture. Two patients in the same ICU were considered to be epidemiologically linked when these patients had either an overlapping period of stay, or, to allow for survival of pathogens in unidentified reservoirs (17), when the time between discharge from the ICU of one of the patients and admission to the ICU of the other patient was at most 7 days. We evaluated the effect of a change in the length of this time window. Possible unidentified reservoirs are health care workers, environmental contamination and other patients, which are not sampled at the site of colonization. Cross transmission was defined as acquired colonization with a CRE that is genetically similar to one previously found in an epidemiologically linked patient. Acquired colonization without epidemiological linkage or genetic relatedness was considered to be endogenous.

Results

Colonization characteristics

In all, 457 patients were studied: 277 admitted to ICU-1 and 180 to ICU-2 and 1,243 rectal swabs were obtained (753 in ICU-1 and 490 in ICU-2) (Table 1). Adherence to the surveillance protocol was close to 100%. Forty-eight patients in ICU-1 and 35 patients in ICU-2 were colonized during their stay. In ICU-1 23 patients were colonized on admission and 23 patients acquired colonization. In ICU-2 ten patients were colonized on admission and 21 patients acquired colonization. Routes of acquisition could not be determined for six patients (2 in ICU-1 and 4 in ICU-2), because first cultures were taken $>$48 hours after admission or because patients had been admitted to the ICU before the start of the study. The mean daily prevalence of colonization with CRE was 26.1 (standard deviation (SD) 15.4) percent in ICU-

1 and 15.1 (13.4) in ICU-2. Acquisition rates were, respectively, 17/1,000 and 18/1,000 patient-days at risk in ICU-1 and ICU-2. The mean time to acquire colonization for patients who acquired colonization was 6 in ICU-1 and 8 days in ICU-2 (SD 8 and 11 days respectively) (Table 1). In total, 174 isolates (107 patients from ICU-1 and 67 patients from ICU-2) were genotyped. Based on AFLP results and epidemiological linkage, five patients in ICU-1 and six patients in ICU-2 acquired colonization via cross transmission. Therefore, five out of 23 (21.7 percent) and six out of 21 (28.6 percent) cases of acquired colonization resulted from cross-colonization in ICU-1 and ICU-2, respectively, representing cross transmission rates of 3.6 and 5.3 per 1,000 patient-days at risk in ICU-1 and ICU-2, respectively. The ratios between endogenous and exogenous acquisition were 3.6:1 for ICU-1 and 2.5:1 for ICU-2. The time interval in the definition of epidemiological linkage hardly influenced the number of acquisitions that were ascribed to cross transmission. Indeed, in both ICUs only one case of cross transmission would be misclassified if the length of the time window would be zero (actual durations between recipients and presumed donor patients were 3 and 4 days). Only if the time window allowed would exceed 21 days, more cases of acquisition would have been considered as cross transmission.

	ICU-1	ICU-2
Admitted pt.	277	180
Rectal swabs	753	490
Pat. colonized (%)	48 (17.3)	35 (19.4)
Pt. with CRE colonization of unknown origin (%)	2 (0.7)	4 (2.2)
Pt. colonized on admission (%)	23 (8.3)	10 (5.6)
Pt. with acquired colonization (%)	23 (8.3)	21 (11.7)
Endemic prevalence, mean (SD)	26.1 (15.4)	15.1 (13.4)
Range (%)	0 - 60	0 - 50
Acquisitions/1000 pt.-days at risk	17	18
Mean time to acquisition (SD)	6 (8)	8 (11)
Length of stay, mean (SD)	8(11)	9 (11)

Table 1: Colonization characteristics of patients admitted to the two ICUs. pt.=patients

	ICU-1		ICU-2	
	Observ.	Model	Observ.	Model
EP	26.1 $\pm$15.4*	27.6	15.1 $\pm$ 13.4*	17.6
% T	21.7	0 (0, 30)†	28.6	0 (0, 25)†

Table 2: Epidemiological variables of cephalosporin-resistant Enterobacteriaceae according to genotyping in combination with epidemiological linkage (=Observ.) and according to model predictions (=Model). * mean $\pm$ SD, † 95% confidence intervals. EP= endemic prevalence, %T =% of acquisitions due to cross transmission.

Model predictions

The MLEs for the parameters α, describing endogenous processes, and β, describing cross transmission, with their 95 percent confidence areas and lines of equal importance of both acquisition routes are depicted in Figure 1. In ICU-1, MLEs for α and β were 0.022 (95 percent confidence interval (CI): 0.013, 0.032) and 0 (95 percent CI: 0.0, 0.035) respectively (Figure 1(a)). In ICU-2, MLEs for α and β were 0.024 (95 percent CI: 0.015, 0.035) and 0 (95 percent CI: 0.0, 0.054), respectively (Figure 1(b)). A re-analysis of the current data with the 'old' model (10) yielded $\alpha = 0.027$ (95 percent CI: 0.016, 0.036) and $\beta = 0$ (95 percent CI: 0, 0.050) for ICU-1 and $\alpha = 0.019$ (95 percent CI: 0.012, 0.027) and $\beta = 0$ (95 percent CI: 0, 0.056) for ICU-2 where the parameter α for the endogenous route also includes admission of colonized patients.

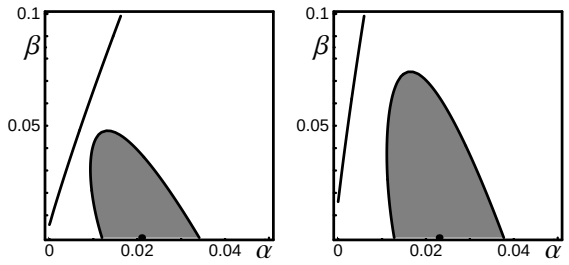
The estimated number of acquisitions was 29.8 (95 percent CI: 28.3, 31.9) and 27.2 (95 percent CI: 26.1, 28.7) for ICU-1 and ICU-2, respectively, which exceeds the observed number of acquisitions by 30 percent. The proportion of acquisitions due to cross transmission was estimated to be 0 percent for both ICUs (95 percent CI: 0, 30 percent and 0, 25 percent for ICU-1 and ICU-2, respectively). The calculated proportions based on epidemiological linkage and genotyping is 21.7 and 28.6 percent for ICU-1 and ICU-2, respectively (Table 2), and so is included in the confidence interval only in case of ICU-1. Using the profile likelihood method, the algorithm established with a confidence level of 99.7 percent and >99.99 percent for ICU-1 and ICU-2, respectively, that less than 50 percent of the acquisitions were due to cross transmission .

The estimated endemic prevalences based on the MLEs for α and β were 27.6 percent and 17.6 percent for ICU-1 and ICU-2, respectively. Both values slightly exceed the observed endemic prevalence (26.1 (SD 15.4) percent for ICU-1 and 15.1 (SD 13.4) percent for ICU-2). Calculated R_N values (expected number of secondary cases through cross transmission generated by a primary case in a pathogen-free ward) were 0 (95 percent CI: 0.0, 0.25) and 0 (95 percent CI: 0, 0.44) for ICU-1 and ICU-2, respectively. A goodness of fit test gave no reason to question our mechanistic transmission model (p=0.29 and p=0.28 for ICU-1 and ICU-2, respectively).

Simulations show that the confidence intervals calculated by our algorithm are conservative when, as with our MLEs, one of the colonization routes is of no importance (see Figure 2). When only the endogenous or the exogenous acquisition route is present, the calculated 95 percent confidence sets in fact represent 97 and 99 percent confidence sets respectively. For other parameter combinations, 95 percent confidence sets will cover the true parameter indeed in 95 percent of the cases when the study period is sufficiently long. In the worst case, for study periods of 6 months, the calculated 95 percent confidence sets still cover the true parameters in 93.5 percent of the simulations.

Discussion

According to the reference standard provided by a combination of genotyping and epidemiological linkage data, the Markov chain model correctly established predominance of endogenous over exogenous acquisition of colonization with cephalosporin-resistant Enterobacteriaceae in two ICUs. The Markov model, therefore, fulfils the need for a reliable tool to evaluate the dynamics of antibiotic resistance and is able to disentangle the relevance of patient-dependent and inde-



(a) ICU-1 (b) ICU-2

Figure 1: **Contour plots of the likelihood of the acquisition parameters α (endogenous acquisition) and β (exogenous acquisition) for cephalosporin-resistant Enterobacteriaceae in ICU-1 and ICU-2.** The shaded area represents the 95% confidence interval. The line represents the parameters for which the endogenous route and the exogenous route are equally important. Supplementary videos demonstrating the development of the confidence intervals in both ICUs in time are available online.

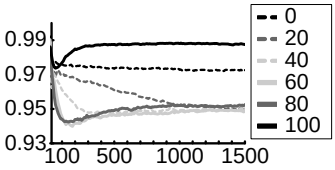


Figure 2: **Fraction of simulations for which the true parameters are contained in the calculated 95% confidence set obtained by our algorithm.** The lines represent different values of the relative importance (in %) of cross transmission. Results are based on 100,000 simulations of an ICU with 10 beds, which are always occupied. The mean prevalence in the ICU was kept constant at 20%, while 5% of the patients are colonized on admission. The length of stay in the ICU was exponentially distributed with a mean of 8 days. Observation of colonization was assumed to be perfect. A perfect method to determine 95% confidence sets would yield the constant 0.95 irrespectively of the duration of the study period.

pendent acquisition routes on the basis of longitudinal data without requiring labour-intensive and costly genotyping procedures.

The Markov algorithm ascribed less cases to cross transmission than found by way of genotyping in combination with epidemiological linkage data. Note, however, that this reference standard does not give a cast-iron answer whether acquisition was exogenous or endogenous. The definition of epidemiological linkage contains an arbitrary time window of 7 days, but more importantly, if two patients are colonized with the same genotype and are epidemiologically linked, this does not necessarily imply that one patient acquired colonization from the other. For instance, a widely spread clone in the extramural population could falsely give the impression that many acquisitions are exogenous. So it may in fact be that the algorithm provides more reliable estimates.

Although the details of the algorithm as presented in the appendix may seem complicated for an audience without a mathematical background, the input to the model consists only of a database with the moments of culturing and the results of these cultures combined with the admission and discharge data from the patients. The output of the algorithm provides results with a clear medical interpretation, e.g., the relative importance of different acquisition routes. Hence, when made user friendly, the software can be a valuable tool which can be used routinely in settings where colonization data are collected.

The framework allows adaptation to alternative Markovian transmission models also and, importantly, individual patient characteristics, such as antibiotic use, the room in which the patient is treated, scores for the severity of illness or multiple sites of colonization, can be incorporated (14). The Markov methodology may also improve the reliability of the interpretation of interventions. Many infection control interventions (such as improving hand hygiene, use of gloves and gowns and antibiotic cycling) have been analyzed in quasi-experimental designs, such as before-after studies (6, 9, 18–21). Results were evaluated by standard statistical tests, such as χ^2 test, Student's t-test and regression analysis that neglect dependence among patients. If cross transmission is relevant, differences between baseline and intervention period, considered to be significant according to these statistical tests, do not necessarily show causality between intervention and outcome (see (22) for a quantitative example of how wrong conclusions can be when dependence is simply ignored). The Markov model provides estimates of confidence intervals for endogenous and exogenous transmission, in itself correcting for autocorrelation when cross transmission is relevant and for chance processes such as a temporarily lower admission rates of colonized patients.

Our model has some limitations. First, only acquisition routes that comply with the Markov property can be incorporated. For instance, environmental contamination, which can persist even after patient discharge (24), does not have the Markov property. However, when environmental contamination depends instantly on the colonization status of a patient, it could be considered as an extension of that patient and the Markov model would still apply. (This is in fact the way we model implicitly temporarily colonized health care workers that act as vectors). Also, the role of persistently colonized health care workers has not been incorporated. Although outbreaks of Enterobacteriaceae caused by health care workers have been reported (25, 26), health care workers are, in general, not considered relevant sources for nosocomial pathogens. But as permanently colonized health care workers would impose a colonization pressure independent of the prevalence of colonized patients, they are mathematically incorporated in the endogenous process. It is possible to explicitly incorporate more complex acquisition routes by adjusting the choice of 'state' and thus retain the Markov property.

Second, colonization is assumed to remain until discharge, which holds true for many but not all antibiotic-resistant nosocomial pathogens. Yet, the possibility of intermittent colonization and eradication, can easily be included. Third, the running time of the algorithm increases with an increasing number of patients with unknown colonization status

(e.g., when incorporating the possibility of false positive and false negative culture results). The actual unit size, on the other hand, can be very large, as long as the number of patients for which the actual colonization status is not known, does not become much larger than 10. If it does, the method can still be used but techniques to approximate the likelihood (e.g., EM algorithm, (27)) have to be used. Fourth, only a limited number of acquisition routes can be incorporated in the model, as otherwise there will be too many parameters that have to be estimated from the (limited amount of) data.

After the initial work of Pelupessy et al. (10), estimation of the relative importance of the different acquisition processes of antibiotic resistance was pursued in two studies. Cooper and Lipsitch (8), building on the work of (10) (and with the same limitations as (10)) proposed a 'hidden Markov model'. Their model uses infection data only and they assume that the Markov model of (10) governs the unobserved dynamics of colonization. Colonized patients then have a constant probability per day to develop an infection, which is observed. As infection rates only represent the tip of the iceberg, long surveillance periods (during which the parameters should remain constant) are needed to derive reliable estimates of the parameters in the underlying transmission process. The counterbalance is that longitudinal data on infection are easier to obtain than data on colonization. The study of Forrester and Pettitt (13), also based on the model of (10), used Monte Carlo Markov Chain methods (MCMC) (28, 29). They estimated transmission rates for MRSA in an ICU where cultures were performed twice weekly. No cultures were performed on admission, all patients were swabbed at the same days of the week (which was required in their analysis) and they assumed that all acquisitions of colonization in between two successive culture moments were independent of each other. Although MCMC is a useful and flexible tool for many situations, a direct calculation of the likelihood avoids typical problems of the MCMC-approach like 1) choosing a prior-distribution of the parameters 2) choosing the burn-in period and 3) bad mixing properties of the Markov-chain.

In this paper we focussed on the methodology and the data analysis, rather than on the clinical effects of candidate infection control measures in the considered units or on risk factors for colonization. So one should not apply our findings concerning the unimportance of cross transmission too readily to other settings with different patient populations, infrastructure, ecology, antibiotic use, infection control adherence, patient-staff ratio and colonization pressure. Our aim here has simply been to introduce a reliable method for obtaining clear conclusions from data that are not too difficult to collect. We hope that the method will be fruitfully applied to investigate obscure details of acquisition for many other nosocomial pathogens in a variety of hospitals/ICUs.

Acknowledgement: We thank Richard Gill for very helpful advice and Hans Metz for a useful conversation. Financial support was provided by Nederlandse Organisatie voor Wetenschappelijk Onderzoek Grants ZonMW 2100.0051 and CLS 635.100.002.

References

1. M.K. Kollef, V.J. Fraser. Antibiotic resistance in the intensive care unit. Ann Intern Med 2001;134:298–314.
2. M.J.M. Bonten, R.A. Weinstein. The role of colonization in the pathogenesis of nosocomial infections. Infect Control Hosp Epidemiol 1996;17:193–200.
3. M.J.M. Bonten, D.J. Austin, M. Lipsitch. Understanding the spread of antibiotic resistant pathogens in hospitals: mathematical models as tools for control. Clin Infect Dis 2001;33:1739–46.
4. M.J.M. Bonten, S. Slaughter, A.W. Ambergen, et al. The role of colonization pressure in the spread of vancomycin-resistant enterococci. an important infection control variable. Arch Intern Med 1998;158:1127–32.
5. J Merrer, F Santoli, C Appéré-De Vecchi, et al. Colonization pressure and risk of acquisition of methicillin-resistant *Staphylococcus aureus* in a medical intensive care unit. Infect Control Hosp Epidemiol 2000;21:718–23.
6. L.A. Puzniak, T. Leet, J. Mayfield, et al. To gown or not to gown: The effect on acquisition of vancomycin-resistant enterococci. Clin Infect Dis 2002;35:18–25.
7. P. de Man, E. van der Veeke, M. Leemreijze, et al. Enterobacter species in a pediatric hospital: horizontal transfer or selection in individual patients? J Infect Dis 2001;184;211–4.
8. B. Cooper, M. Lipsitch. The analysis of hospital infection data using hidden markov models. Biostatistics 2004;5:223–37.
9. A.D. Harris, D.D. Bradham, M. Baumgarten, et al. The use and interpretation of quasi-experimental studies in infectious diseases. Clin Infect Dis 2004;38:1586–91.
10. I. Pelupessy, M.J.M. Bonten, O. Diekmann. How to assess the relative importance of different colonization routes of pathogens within hospital settings. Proc Natl Acad Sci USA 2002;99:5601–5.
11. J. Schlesinger, S. Navon-Venezia, I. Chmelnitsky et al. Extended-Spectrum Beta-Lactamase among Enterobacter isolates obtained in Tel Aviv, Israel. Antimicrob Agents Chemother 2005; 49:1150-6.
12. A. Paauw, A.C. Fluit, J. Verhoef et al. Enterobacter cloacae Outbreak and Emergence of Quinolone Resistance Gene in Dutch Hospital. Emerg Infect Dis. 2006;12:807-12.
13. M. Forrester, A.N. Pettitt. Use of stochastic epidemic modeling to quantify rates of colonization with methicillin-resistant *Staphylococcus aureus* in an intensive care unit. Infect Control Hosp Epidemiol 2005;26:598–606.
14. M.C.J. Bootsma. Mathematical studies of the dynamics of antibiotic resistance. PhD thesis, Utrecht University, 2005. http://igitur-archive.library.uu.nl/dissertations/2005-0526-201427/.
15. D.R. Cox, D.V. Hinkley. Theoretical Statistics. Chapman & Hall/CRC, 1979.
16. R.J.L. Willems, J. Top, N. van den Braak, et al. Host specificity of vancomycin-resistant enterococcus faecium. J Infect Dis 2000;182:816–23.
17. H. Grundmann, S. Barwolff, A. Tami, et al. How many infections are caused by patient-to-patient transmission in intensive care units? Crit Care Med 2005;33:946–51.
18. D. Pittet, P. Mourouga, T.V. Perneger. Compliance with handwashing in a teaching hospital. Ann Intern Med 1999;130:126–30.
19. S. Slaughter, M.K. Hayden, C. Nathan, et al. A comparison of the effect of universal use of gloves and gowns with that of glove use alone on acquisition of vancomycin-resistant enterococci in a medical intensive care unit. Ann Intern Med 1996;125:448–56.
20. E.A. Dominguez, T.L. Smith, E. Reed, et al. A pilot study of antibiotic cycling in a hematology-oncology unit. Infect Control Hosp Epidemiol 2000;21:S4–S8.
21. D.P. Raymond, S.J. Pelletier, T.D. Crabtree, et al. Impact of a rotating empiric antibiotic schedule on infectious mortality in an intensive care unit. Crit Care Med 2001;29:1101–8.
22. S. Nijssen, M.C.J. Bootsma, M.J.M. Bonten. Potential confounding in evaluating infection control interventions in hospital settings: changing antibiotic prescription. Clin Infect Dis 2006;43:616–23.
23. O. Diekmann, J.A.P. Heesterbeek. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley, Chichester, 2000.
24. S.S. Huang, R. Datta, R. Platt. Risk of acquiring antibiotic-resistant bacteria from prior room occupants. Arch Intern Med. 2006;166:1945–51.
25. W.H. Van Nierop, A.G. Duse, R.G. Stewart, et al. Molecular epidemiology of an outbreak of enterobacter cloacae in the neonatal intensive care

unit of a provincial hospital in gauteng, south africa. J Clin Microbiol 1998;36:3085–7.
26. W.L. Yu, H.S. Cheng, H.C. Lin, et al. Outbreak investigation of nosocomial enterobacter cloacae bacteraemia in a neonatal intensive care unit. Scand J Inf Dis 2000;32:293–8.
27. L. Wasserman. All of Statistics. Springer Verlag, New York, 2004.
28. G.J. Gibson, E. Renshaw. Likelihood estimation for stochastic compartmental models using markov chain methods. Statistics and Computing 2001;11:347–58.
29. P.D. O'Neill, G.O. Roberts. Bayesian inference for partially observed stochastic epidemics. J. Royal Statistical Soc. C 1999;162:121–9.
30. P.K. Andersen, Ø. Borgan, R.D. Gill et al. Statistical Models Based on Counting Processes. Springer-Verlag, New York, 1993.

Appendix

Definitions

We divide the period of stay of a patient retrospectively into (at most) 3 periods based on the results of the culturing :

1. patient is known to be uncolonized $t_0 \leq t \leq t_1$
2. patient may or may not be colonized $t_1 < t < t_2$
3. patient is known to be colonized $t_2 \leq t \leq t_e$

Per day, we have three categories of patients, uncolonized patients, patients whose colonization status is uncertain and colonized patients. We label these three categories $\mathcal{U}$, $\mathcal{Q}$ (for 'questionable') and $\mathcal{C}$ respectively. The number of patients in the categories are represented by, respectively, u, q and c. For later convenience, for every time t, we order the patients in category $\mathcal{Q}$ in increasing order of the time they are already in category $\mathcal{Q}$, i.e., a patient entering category Q will be the first one in the ordering.

By definition of the categories, we are certain about the colonization status of the patients in $\mathcal{U}$ and $\mathcal{C}$. For each day, the number of patients in these two categories can be determined from the data directly. Hence, u and c are treated as (time-dependent) parameters and, for computational reasons, patients in $\mathcal{U}$ and $\mathcal{C}$ are not included in our definition of the state space. As each of the patients in $\mathcal{Q}$ can be colonized or not, the total number of possible states for the ICU is $(\mathbb{Z}_2)^q = \{0,1\}^q$. Note that the dimension of the state space, 2^q, changes over time as q may change from day to day.

Each ICU state is denoted by a vector $\mathbf{v} = (v_1, v_2, \ldots, v_q)$, with $v_k \in \{0,1\}$, where v_k denotes whether the k^{th} patient in $\mathcal{Q}$ is colonized or not. The state $(v_1, v_2, \ldots, v_q)$ can also be represented by the binary number $v_1 v_2 \ldots v_q$ and therefore we have a natural labeling j of each of the 2^q states ($0 \leq j \leq 2^q - 1$).

For notational convenience, we would like to be able to switch back and forth between a state represented as a finite sequence of 0's and 1's, i.e., as an element of $(\mathbb{Z}_2)^q$, and its corresponding number. Therefore we introduce the numbering function defined as:

$$N : (\mathbb{Z}_2)^q \to \mathbb{Z}_{2^q} \subset \mathbb{N}; \quad (v_1, v_2, \ldots, v_q) \mapsto \textstyle\sum_{i=1}^{q} v_i 2^{(q-i)} \qquad \mathbf{1}$$

The inverse of the numbering function, N^{-1}, relates a state number m ($0 \leq m \leq 2^q - 1$) to the colonization status of the individuals in $\mathcal{Q}$. Specifically, the component $N^{-1}(m)_k$ shows whether in ICU state m individual k ($1 \leq k \leq q$) is colonized or not. We also introduce an ordering on $(\mathbb{Z}_2)^q$.

$$\mathbf{v} \geq \mathbf{w} \Leftrightarrow v_i \geq w_i \quad \forall\, 1 \leq i \leq q \qquad \mathbf{2}$$

and the l^1-norm: $|v| = \sum_{i=1}^{q} v_i$.

As the state is actually uncertain, we want to employ a stochastic description and assign to each state a probability that it is the actual (unknown) state. So we introduce the probability vector $\mathbf{p}(t) = \{p_0(t), p_1(t), \ldots, p_{2^q-1}(t)\}$ of length 2^q in which $p_j(t)$ denotes the likelihood that the ICU is in state j.

We consider a period of observation from time 0 till time T. The observations can be divided into two parts: those before or at time t and those after time t. We define the 'forward' vector $\mathbf{v_f}$, a column vector, based on the observations until time t. The forward vector has as its components the unnormalized (i.e., relative) probabilities that the system is in a specific state at time t if we take into account only the observations until time t. However, the best estimate for $\mathbf{p}(t)$ is no longer $\mathbf{v_f}(t)$ when results of cultures performed after time t become available. To take such additional information into account, we define the 'backward' vector $\mathbf{v_b}(t)$, a row vector, for which the i^{th} component is the unnormalized probability that, given that the ICU is at time t in the state numbered i, the ICU will develop in a way that is compatible with all observations after time t. With these definitions, the i^{th} component of the probability vector $\mathbf{p}(t)$ that takes all observations into account is

$$\mathbf{p}_i(t) = \mathbf{v_b}_i(t)\mathbf{v_f}_i(t) / \sum_{j=0}^{2^q-1} \mathbf{v_b}_j(t)\mathbf{v_f}_j(t). \qquad \mathbf{3}$$

All observations before or at the start of the study period are incorporated in $\mathbf{v_f}(0)$. In the case that all patients are cultured at $t = 0$, we know the ICU state at time $t = 0$ with certainty and the component of $\mathbf{v_f}(0)$ corresponding to this state will be one while all other components of $\mathbf{v_f}(0)$ are zero. When $t = T$, all states are compatible with the observations after time T as there are no such observations, and hence $\mathbf{v_b}(T) = \mathbf{1}$, with $\mathbf{1}$ the row vector with all elements equal to one.

We now construct an algorithm to calculate the forward and the backward vector for all $0 \leq t \leq T$.

The forward process

To calculate the time evolution of the forward vector $\mathbf{v_f}$, we need the mechanistic model. The mechanistic model gives probabilities A_{mn}, ($0 \leq m, n \leq 2^q - 1$), which describe how likely state m is at time $t + 1$ just before culturing, discharge and admission, given that the system was in state n at time t just after the culturing, discharge and admission. At this point we do not yet express in the notation that A_{mn} depends on t, simply since q does; note that it depends on q what the numbers m and n tell us about the ICU state. The evolution can then be defined in terms of matrix multiplication:

$$A: \quad \mathbb{R}^{2^q} \to \mathbb{R}^{2^q}; \quad \mathbf{w} \mapsto A\mathbf{w} \quad \text{with } A = (A_{mn}) \qquad \mathbf{4}$$

The matrix A has a special structure. Let $\pi(k)$ be the probability that an uncolonized patient acquires colonization during a day, given that there are k colonized patients in the ward. Each transition probability in column m is either zero, when the transition to state m is not allowed by the mechanistic model, or it can be written as a product of powers of $\pi(c+j)$ and $(1-\pi(c+j))$ with c the number of colonized patients in $\mathcal{C}$ and j the number of patients in $\mathcal{Q}$ that are colonized when the system is in state n. Explicitly,

$$\begin{cases} A_{mn} = & 0 & \text{if } N^{-1}(m) \not\geq N^{-1}(n) \\ A_{mn} = & (1-\pi(k))^u \pi(k)^l (1-\pi(k))^{q-l} & \text{if } N^{-1}(m) \geq N^{-1}(n) \\ & \text{with } l = |N^{-1}(m)| - |N^{-1}(n)| & \text{and } k = c + |N^{-1}(n)| \end{cases} \quad \mathbf{5}$$

For instance, in the case that there are 2, c and u patients in $\mathcal{Q}$, $\mathcal{C}$ and $\mathcal{U}$ respectively, with $\rho(k) = 1-\pi(k)$ the matrix A becomes:

$$\begin{pmatrix} \rho(c)^{2+u} & 0 & 0 & 0 \\ \rho(c)^{1+u}\pi(c) & \rho(c+1)^{1+u} & 0 & 0 \\ \rho(c)^{1+u}\pi(c) & 0 & \rho(c+1)^{1+u} & 0 \\ \rho(c)^{u}\pi(c)^2 & \rho(c+1)^{u}\pi(c+1) & \rho(c+1)^{u}\pi(c+1) & \rho(c+2)^{u} \end{pmatrix}$$

Note that, when $u \neq 0$, the matrix A does not preserve the norm of the vector on which it acts. (This is due to the fact that we leave out of consideration all transitions that could in principle have happened to the $\mathcal{U}$ category.)

We now will use the culture results, the discharge data and the admission data. Suppose that the k^{th} patient in $\mathcal{Q}$ is cultured. By the definition of the category $\mathcal{Q}$, this culture will be positive. Therefore only the states m, $0 \leq m \leq 2^q - 1$, with $N^{-1}(m)_k = 1$ are allowed by the data and the other states have zero a posteriori probability. Mathematically, culturing of patient k in $\mathcal{Q}$ amounts to projecting the vector $A\mathbf{w}$ on a linear subspace isomorphic to $\mathbb{R}^{2^{q-1}}$. The diagonal matrix

$$C_k \colon \mathbb{R}^{2^q} \to \mathbb{R}^{2^q}; \quad \mathbf{w} \mapsto C_k \mathbf{w} \quad \text{with } C_k = (c_{mn}) \qquad \mathbf{6}$$

is given by: $c_{mn} = \begin{cases} 0 & \text{if } n \neq m \text{ or } N^{-1}(n)_k = 0 \\ 1 & \text{if } n = m \text{ and } N^{-1}(n)_k = 1 \end{cases}$.

Example: In the case that $q = 3$ and we culture the second patient in $\mathcal{Q}$ and before the culturing the state vector is $\mathbf{w} = (w_0, w_1, \ldots, w_7)$, then after the culturing, the vector will be $(0, 0, w2, w3, 0, 0, w6, w7)$.

If several category $\mathcal{Q}$ patients are cultured at the same time, the operator $C(t)$ consists of a product of C_k's. When a category $\mathcal{Q}$ patient 'leaves' $\mathcal{Q}$, either because he/she was cultured or because he/she leaves the unit without being cultured, the number of possible states is reduced by a factor 2.

For $1 \leq k \leq q$ we can define the operator R_k that removes the k^{th} patient in Q via:

$$R_k \colon \quad \mathbb{R}^{2^q} \to \mathbb{R}^{2^{q-1}}; \quad \mathbf{w} \mapsto \mathbf{w}' \qquad \mathbf{7}$$

where the components of $\mathbf{w}'$ are defined by: $\mathbf{w}'_{N(v_1,\ldots,v_{q-1})} = \sum_{i\in\{0,1\}} \mathbf{w}_{N(v_1,\ldots,v_{k-1},i,v_k,\ldots,v_{q-1})}$ This operator R_k adds the probabilities of the two states for which the colonization status of the remaining $q-1$ patients is identical.

If several category $\mathcal{Q}$ patients 'leave' $\mathcal{Q}$ at the same time, the operator $R(t)$ consists of a product of R_k's. To avoid confusion about which of the patients in $\mathcal{Q}$ 'leaves' $\mathcal{Q}$, we should use some convention, for instance, order the operators such that we do the removal in decreasing order of the patient number in $\mathcal{Q}$.

Suppose now that l patients enter category $\mathcal{Q}$ at a certain time t. By the definition of the category $\mathcal{Q}$, patients enter category $\mathcal{Q}$ directly after their last negative culture, so we know that these patients enter category $\mathcal{Q}$ uncolonized. As we ordered the patients in category $\mathcal{Q}$ according to the day they entered this category, these l patients correspond to the first l digits in the binary expansion. Due to this ordering, the function I_l that deals with the admission of l new patients to $\mathcal{Q}$ is defined by:

$$I_l \colon \mathbb{R}^{2^q} \to \mathbb{R}^{2^{q+l}}; \quad \mathbf{w} \mapsto \mathbf{w}' \qquad \mathbf{8}$$

where the elements in the vector $\mathbf{w}'$ are given by $(0 \leq k \leq 2^{q+l} - 1)$:

$$\mathbf{w}'_k = \begin{cases} 0 & \text{if } k \geq 2^q \\ \mathbf{w}_k & \text{if } k < 2^q \end{cases} \qquad \mathbf{9}$$

Note that $R(t)$ and I_l involve a change of the dimension of the state space. Indeed, we 'glue' together state spaces of different size according to the need as exposed by observed events.

With the previous definition of the operators, the 'forward vector' can be written as:

$$\mathbf{v_f}(t) = \prod_{\tau=0}^{t-1} I(\tau+1)R(\tau+1)C(\tau+1)A(\tau)\mathbf{v_f}(0) \qquad \mathbf{10}$$

or derived iteratively from the recursion:

$$\mathbf{v_f}(t) = I(t)R(t)C(t)A(t-1)\mathbf{v_f}(t-1). \qquad \mathbf{11}$$

The likelihood of the observed events during one day is the norm of the final state vector (assuming that the initial state vector had norm 1). More precisely, the likelihood is given by $|CA\mathbf{v_f}|/|\mathbf{v_f}|$. The likelihood of the observed events over several days is the product of the relevant 1 day likelihoods and the overall likelihood is given by $|\mathbf{v_f}(T)|$. This likelihood function leads to maximum likelihood estimates of parameters and to confidence regions (15).

Note that for calculation of the likelihood of the observations, it suffices to consider only the forward vector. When during a certain period the dynamics of acquisition in an ICU is followed, the new observations that become available each day can be processed by the algorithm to improve the estimates of the transmission parameters as more data become available.

The backward process

The 'backward vector' can be written as:

$$\mathbf{v_b}(t) = \mathbf{1} \prod_{\tau=t}^{T-1} I(\tau+1)R(\tau+1)C(\tau+1)A(\tau) \qquad \mathbf{12}$$

or derived iteratively from the backward recursion:

$$\mathbf{v_b}(t) = \mathbf{v_b}(t+1)I(t+1)R(t+1)C(t+1)A(t) \qquad \mathbf{13}$$

To explain equation 13, we consider the special case that both $R(t+1)$ and $I(t+1)$ are the identity operator or, in words, the

case that neither discharge nor admission occurs at $t+1$. (The general case differs from the special case in bookkeeping aspects, all truly probabilistic considerations are incorporated in $A(t)$ and $C(t+1)$.) Let $S(t)$ denote the state at time t and let $\mathcal{O}_{\tau_1}^{\tau_2}$ denote the observations for $\tau_1 \leq t \leq \tau_2$. We can write:

$$\begin{gathered}\mathbf{v_b}_i(t) = P(\mathcal{O}_{t+1}^T | S(t) = i) = \\ = \sum_j P(\mathcal{O}_{t+2}^T | S(t+1) = j) P(S(t+1) = j, \mathcal{O}_{t+1}^{t+1} | S(t) = i) = \\ \sum_j \mathbf{v_b}_j(t+1) C_{jj}(t+1) A_{ji}(t)\end{gathered}$$

which is exactly the i^{th} component of the identity 13. Here we have used in particular that when we look at times $\geq t+2$ and condition on the state being j at $t+1$, knowledge about the state at t is irrelevant.

Combining both processes

Note first of all that the inner product $\mathbf{v_b}(t)\mathbf{v_f}(t)$ is independent of t and hence is equal to $\mathbf{1}\mathbf{v_f}(T) = |\mathbf{v_f}(T)|$, the overall likelihood of the observations. Once we know $\mathbf{v_f}(t)$ and $\mathbf{v_b}(t)$, and hence, given all the information that we possess, the true probabilities for each state at all times, we can calculate the expected prevalence by averaging over all the states. Calculation of the number of acquisitions per day and the subdivision of this number according to the routes requires one additional step.

Fix t. Let $P(j,i)$ be the probability that the ICU is in state j at day t and in state i at day $t+1$. With K_m the projection operator on the m^{th} component, we can write:

$$P(j,i) = \frac{\mathbf{v_b}(t+1) K_i I(t+1) R(t+1) C(t+1) A(t) K_j \mathbf{v_f}(t)}{|\mathbf{v_f}(T)|} \quad \textbf{14}$$

To explain equation 14, we again focus on the case in which there is neither discharge nor admission.

$$\begin{aligned}P(j,i) &= P\left(S(t) = j, S(t+1) = i | \mathcal{O}_0^T\right) = \\ &\frac{P\left(S(t)=j, S(t+1)=i, \mathcal{O}_0^T\right)}{P\left(\mathcal{O}_0^T\right)} = \frac{P\left(S(t)=j, S(t+1)=i, \mathcal{O}_0^T\right)}{|\mathbf{v_f}(T)|}\end{aligned} \quad \textbf{15}$$

We can write:

$$\begin{gathered}P\left(S(t) = j, S(t+1) = i, \mathcal{O}_0^T\right) = \\ P\left(S(t) = j, S(t+1) = i, \mathcal{O}_0^{t+1}\right) P\left(\mathcal{O}_{t+2}^T | S(t) = j, S(t+1) = i, \mathcal{O}_0^{t+1}\right) \\ = P\left(S(t) = j, S(t+1) = i, \mathcal{O}_0^{t+1}\right) P\left(\mathcal{O}_{t+2}^T | S(t+1) = i\right) = \\ P\left(S(t) = j, S(t+1) = i, \mathcal{O}_0^{t+1}\right) \mathbf{v_b}_i(t+1)\end{gathered} \quad \textbf{16}$$

and

$$\begin{gathered}P\left(S(t) = j, S(t+1) = i, \mathcal{O}_0^{t+1}\right) = \\ P\left(S(t+1) = i, \mathcal{O}_{t+1}^{t+1} | S(t) = j, \mathcal{O}_0^t\right) P\left(S(t) = j, \mathcal{O}_0^t\right) = \\ P\left(S(t+1) = i, \mathcal{O}_{t+1}^{t+1} | S(t) = j\right) P\left(S(t) = j, \mathcal{O}_0^t\right) = \\ P\left(S(t+1) = i, \mathcal{O}_{t+1}^{t+1} | S(t) = j\right) \mathbf{v_f}_j(t) = \\ C_{ii}(t+1) A_{ij}(t) \mathbf{v_f}_j(t)\end{gathered} \quad \textbf{17}$$

Combining the identities 15, 16 and 17 we obtain equation 14. (Strictly speaking the derivation above only applies when $P(j,i) > 0$.)

The expected number of acquisition during day t is: $\sum_{i,j} P(j,i) g_{ij}$ with g_{ij} the expected number of acquisitions during the transition from ICU state j to i. Clearly, g_{ij} depends on whether patients who were discharged at day $t+1$ without being cultured at discharge, acquired colonization during day t or not. Therefore, it is convenient to count the number of acquisitions before performing the bookkeeping operations. So we adapt the evolution matrix and define the matrix $m(t)$ by:

$$m_{ij}(t) = a_{ij}(t) f_{ij}(t) \quad \textbf{18}$$

with $a_{ij}(t)$ the components of the evolution matrix $A(t)$ and $f_{ij}(t)$ the expected number of acquisitions by a route or combination of routes in case of a transition from state j to state i. Note carefully that several choices of f_{ij} are relevant; also note that these numbers depend on t simply because the state space, and hence the precise meaning of i and j changes with time. In case we are interested in the number of acquisitions by the endogenous route, we define:

$$f_{ij}(t) = \left(|N^{-1}(i)| - |N^{-1}(j)|\right) \frac{\alpha}{\alpha + \beta\left(c(t) + |N^{-1}(j)|\right)/n} \quad \textbf{19}$$

The expected number of acquisitions $\theta(t)$ during day t by the route(s) under consideration is given by the following inner product:

$$\frac{\mathbf{v_b}(t+1) I(t+1) R(t+1) C(t+1) m(t) \mathbf{v_f}(t)}{|\mathbf{v_f}(T)|} \quad \textbf{20}$$

When we sum over all days, we obtain the expected total number of acquisitions per route during the study period. We use these expressions to calculate the relative importance of each route. By comparing the maximum likelihood with the maximum likelihood constrained to the parameter space for which the relative importance of cross transmission is less than 50 percent, we can establish a confidence level that less than 50 percent of the acquisitions were due to cross transmission.