Lifetime Data Analysis, 2, 219–240 (1996)

Estimation of Prolongation of Hospital Stay Attributable to Nosocomial Infections: New Approaches Based on Multistate Models

GABI SCHULGEN gs@imbi.uni-freiburg.de
Institute of Medical Biometry and Medical Informatics, Albert-Ludwigs-University, Stefan-Meier-Str. 26, D-79104 Freiburg, Germany

MARTIN SCHUMACHER
Institute of Medical Biometry and Medical Informatics, Albert-Ludwigs-University, Stefan-Meier-Str. 26, D-79104 Freiburg, Germany

Keywords: Extra hospital stay, Markov processes, Multistate models, Structured nested failure time models, Time-dependent covariates

Received December 8, 1995; accepted June 7, 1996

Abstract. Evaluation of the impact of nosocomial infection on duration of hospital stay usually relies on estimates obtained in prospective cohort studies. However, the statistical methods used to estimate the extra length of stay are usually not adequate. A naive comparison of duration of stay in infected and non-infected patients is not adequate to estimate the extra hospitalisation time due to nosocomial infections. Matching for duration of stay prior to infection can compensate in part for the bias of ad hoc methods. New model-based approaches have been developed to estimate the excess length of stay. It will be demonstrated that statistical models based on multivariate counting processes provide an appropriate framework to analyse the occurrence and impact of nosocomial infections. We will propose and investigate new approaches to estimate the extra time spent in hospitals attributable to nosocomial infections based on functionals of the transition probabilities in multistate models. Additionally, within the class of structural nested failure time models an alternative approach to estimate the extra stay due to nosocomial infections is derived. The methods are illustrated using data from a cohort study on 756 patients admitted to intensive care units at the University Hospital in Freiburg.

1. Introduction

Nosocomial (hospital-acquired) infections are major complications arising in the patients' treatment in hospital. The rate of nosocomial infections is highest in intensive care units (ICU), the most hazardous areas in the hospital setting. Nosocomial infections may cause severe morbidity in patients leading to prolongation of hospital stay and, dependent on the type of infection, may also contribute to an increase in mortality. Therefore, nosocomial infections constitute a substantial medical and socioeconomic problem, apart from the individual consequences for the affected patient (Daschner, 1982, 1984; Haley, 1981). For the United States (US) the total economic burden for the society from nosocomial infections is recently estimated to be about 4.5 billion U.S. $ per year (Bennet & Brachman, 1993). Prolongation of hospital stay due to nosocomial infection is a major issue in cost-benefit analysis of programs for infection control (Haley, 1986). Information about prolongation

of hospital stay due to nosocomial infections is also regarded as a surrogate measure to quantify the medical impact of nosocomial infections.

In estimating the extend to which duration of hospital stay is prolonged by the infection one has to accept that the extra time is not directly observable. Ad-hoc approaches for estimating the extra hospital stay attributable to nosocomial infections applied so far (McGowan, 1982; Brawley et al., 1989; Freeman et al., 1979; Haley et al., 1980; Green et al., 1982; Freeman & McGowan, 1984; Leu et al., 1989; Pittet et al., 1994) suffer the drawback of a necessary retrospective stratification in infected and non-infected patients. These approaches lead to biased results because they treat nosocomial infections like predefined fixed baseline characteristics of the patients neglecting that acquiring infection is a process in time. To assess the consequences of nosocomial infections, the temporal sequence of events and the dynamics of the disease process have to be considered. It will be demonstrated that statistical models based on multivariate counting processes provide an appropriate framework to analyse the occurrence and impact of nosocomial infections. We will propose new approaches to estimate the extra time spent in hospital attributable to nosocomial infections based on functionals of transition probabilities in multistate models. Within the class of structural nested failure time models an additional alternative approach to estimate the extra stay due to nosocomial infections is derived. The methods are illustrated using data from a cohort study on 756 patients admitted to intensive care units at the University Hospital in Freiburg.

The outline of the paper is as follows: Section 2 describes the example data set; section 3 specifies the model based on multivariate counting processes; section 4 deals with non-parametric estimation of the transition intensities and probabilities in multistate models; section 5 describes the different approaches for estimating prolongation of hospital stay attributable to nosocomial infections; section 6 presents the results of an application of the discussed approaches in the example data set. For motivation of the reader it might be worthwhile to skip the more technical section 4 for the first reading and move directly to section 5.

2. Illustrative Example

A cohort study was initiated in 1991 at the University Hospital in Freiburg with prospective assessment of data to evaluate the impact of nosocomial infections. All patients admitted to the anesthesiological (AIT) and medical (MIT) intensive care units during the two year period from July 1991 to June 1993, 18 years of age or older who resided in these ICUs for 48 hours or more entered the study cohort. Overall, 756 patients were included in the cohort and were monitored daily during ICU stay by an independent physician who was not involved in the treatment of the patients. Diagnose of nosocomial infection were made based upon the definitions by the Centers for Disease Control (Garner et al., 1988). Dates recorded are the date of admission to ICU, date(s) of onset of nosocomial infection(s) on ICU, date of discharge from ICU and date of death on ICU, respectively. For every patient the complete follow up information is available, hence, censored event times do not occur.

One-hundred-ninety-seven patients (26.1%) acquired one or more nosocomial infections. The most common type of infection was nosocomial pneumonia which was acquired by

Table 1. Characteristics of the study cohort.

Patient characteristics	AIT (N = 440)		MIT (N = 316)		Total (N = 756)	
Age*	48.5	(21.7)	58.5	(15.2)	52.7	(19.9)
Gender (female)	159	(36.1%)	138	(43.7%)	297	(39.3%)
Admission diagnosis:						
Central nervous system	8	(1.8%)	27	(8.5%)	35	(4.6%)
Cardiopulmonary	36	(8.2%)	151	(47.8%)	187	(24.7%)
Abdominal	89	(20.2%)	13	(4.1%)	102	(13.5%)
Urogenital	8	(1.8%)	3	(0.9%)	11	(1.5%)
Polytrauma/ Head trauma	192	(43.6%)	0	(0%)	192	(25.4%)
Infections	63	(14.3%)	65	(20.6%)	128	(16.9%)
Others	44	(10%)	56	(17.7%)	100	(13.2%)
Operation:						
Elective surgery	106	(24.1%)	10	(3.2%)	116	(15.3%)
Emergency surgery	222	(50.5%)	11	(3.5%)	233	(30.8%)
Infection on admission to ICU	209	(47.6%)	203	(64.2%)	412	(54.6%)
Pneumonia on admission	168	(38.2%)	165	(52.2%)	333	(44%)
Sepsis on admission	44	(10%)	37	(11.7%)	81	(10.7%)

*.mean (standard deviation)

124 patients (16.4%); the second most common nosocomial infection was sepsis which was acquired by 67 patients (8.9%). Overall, 191 patients (25.3%) died on ICU; of the 67 patients who acquired nosocomial sepsis 49 patients (73.1%) died on ICU and of the 124 patients who acquired nosocomial pneumonia without subsequent sepsis 28 patients (22.6%) died on ICU. Median duration of stay was 6 days on AIT and 7 days on MIT. Table 1 describes the patient population by selected baseline variables. Further details of the study are given elsewhere (Kropec et al., 1995).

3. Model Specification Based on Multivariate Counting Processes

In a multistate model for occurrence and impact of nosocomial infections, we regard for example admittance to ICU, nosocomial infections, death and discharge as potential disease states. Acquiring an infection during hospital stay is regarded as a transition to the state "nosocomial infection" out of an initial non-infected state. Occurrence of infections, death and discharge are so called events in the disease course. Figure 1 illustrates the basic potential individual movements in the example presented.

Patients and their event times will be assumed independent of each other. Time will in general be assumed continuous, however, observation of events will in general be in discrete time. The basic time scale in the example is time (days) since admittance to ICU.

All patients start in the transient, non-recurrent state "alive, free of nosocomial infection" (:= state 0) with admittance to the intensive care unit (ICU) at $t = 0$. Infections, which are acquired within the first 48 hours after admission are not regarded as nosocomial. We are mainly interested in the occurrence of nosocomial pneumonia and sepsis, the most frequent

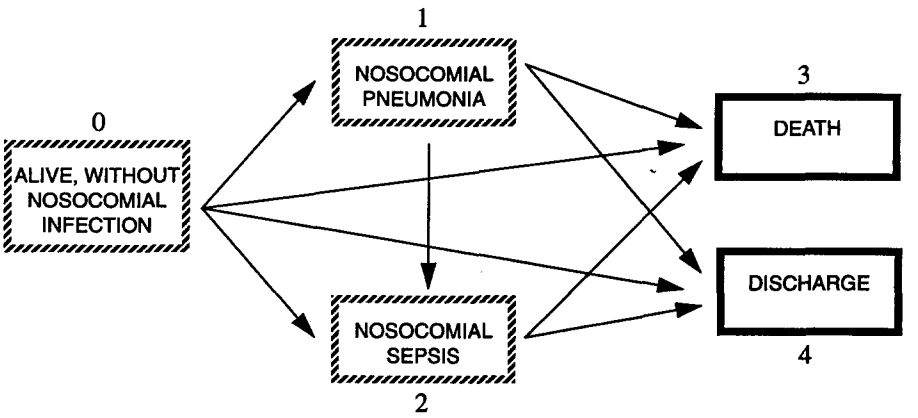

Figure 1. Potential states and transitions for occurrence and impact of nosocomial infections.

types of nosocomial infections and most serious complications arising in the study population. The states 1 and 2 represent the occurrence of nosocomial pneumonia and nosocomial sepsis, respectively, both are regarded as transient states. Both states are non-recurrent because in the available database, there is essentially no information about termination of an infection. For the same reason multiple infections of the same type are not modelled. Finally, patients may be discharged to a normal care unit within hospital or they may die while still on ICU. Naturally, death (:= state 3) is an absorbing state. Discharge (:= state 4) is also regarded as an absorbing state because observation terminates at the time of discharge and no further events have been registered. We will not model transitions from state 2 to state 1 because nosocomial sepsis is an almost always fatal event in the patients course at ICU and a model for pneumonia following sepsis is clinically irrelevant.

We assume a non-homogeneous Markov model for the movements of the individuals among the several states (Andersen et al., 1993). Let $(X(t), t \in T)$ denote the Markov process with state space $\{0, 1, 2, 3, 4\}$ and the initial distribution degenerated at state 0. Then $N = (N_{jh};\ j, h = 0, 1, 2, 3, 4;\ j \neq h)$ denotes the multivariate counting process, counting the number of direct transitions from j to h, i.e. the number of infections, deaths and discharges with the time scale being "time since admittance". Let N be adapted to the filtration $F_t = N_t \cup F_0$, where N_t denotes the self-exiting filtration generated by N and F_0 represents the information fixed at time 0 which may be given by the covariates.

Figure 2 illustrates the processes counting the number of pneumonia cases, sepsis cases and deaths observed in the first 40 days after admittance to the ICU.

4. Estimation of the Integrated Transition Intensities and Transition Probabilities

Estimation in the multiplicative intensity model centers around estimators for the integrated transition intensities and the transition probabilities, which are important in judging the

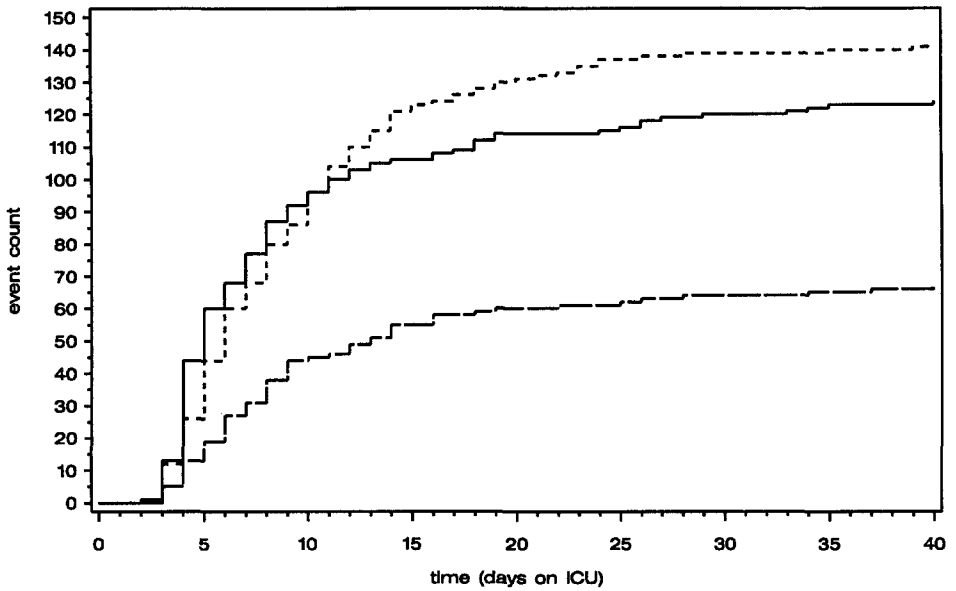

Figure 2. Illustration of observed counting processes for transition from the initial state to nosocomial pneumonia (____), sepsis (_ _) and death (_ _ _) for the first 40 days on ICU.

prognosis of a patient. Estimates of the event specific hazard functions can be obtained by application of smoothing techniques based on kernel functions (Ramlau-Hansen, 1983). However, we will concentrate here on the Nelson-Aalen estimator for the cumulative hazard function and the Aalen-Johansen estimator for the transition probabilities. The book by Andersen, Borgan, Gill and Keiding (1993) provides a thorough overview of methods and applications and will serve as the basic reference and source of notation.

Let A denote the event-specific cumulative hazard function with

$$A_{jh}(t) = \int_0^t \alpha_{jh}(s)d(s),$$

where the force of transition from state j to h, $j \neq h$, is given by the event-specific hazard function $\alpha_{jh}(t, \theta)$ (with some $\alpha_{jh}(t) = 0$).

An estimator for $A_{jh}(t)$ is given by the Nelson-Aalen estimator (Nelson, 1972; Aalen, 1978):

$$\hat{A}_{jh}(t) = \int_0^t \frac{J_j(s)}{Y_j(s)} dN_{jh}(s),$$

where $Y_j(t)$ denotes the indicator for X being in state j just prior to time t and $J(t) = I(Y(t) > 0)$ and $J(t)/Y(t) = 0$ if $Y(t) = 0$. Here, $dN(t)$ denotes the increments of the counting process N in the time interval $[t, t + dt)$: $dN(t) = N((t + dt)-) - N(t-)$. A proof of consistency and asymptotic normality of the estimator is given in Aalen (1978).

The transition probabilities can be determined from the transition intensities by product-integrating the intensity measure. Using the product-integral notation (Gill & Johansen, 1990) the matrix of transition probabilities $P(s, t)$ with elements $\{P_{jh}(s, t)\}$ for $s < t$ is given by

$$\mathbf{P}(s, t) = \prod_{u \in (s,t)} (\mathbf{I} + d\mathbf{A}(u)),$$

with A being the matrix of the transition intensity measure with elements $\{A_{jh}\}$. The product sign over an interval denotes the product integral, a continuous version of the ordinary finite product.

Let $\hat{A}_{jh}(t)$ denote the Nelson-Aalen estimator of the cumulative hazard function A_{jh} for direct transitions from j to $h (j \neq h)$ and let $\hat{A}_{jj}(t) = -\sum_{j \neq h} \hat{A}_{jh}(t)$. The Aalen-Johansen estimator for the matrix of transition probabilities in non-homogeneous Markov processes (Aalen & Johansen, 1978) is given by

$$\hat{\mathbf{P}}(s, t) = \prod_{u \in (s,t)} (\mathbf{I} + d\hat{\mathbf{A}}(u)),$$

with $\hat{A} = \{\hat{A}_{jh}\}$. Because the Nelson-Aalen estimator is a step function with a finite number of jumps in (s, t) the Aalen-Johansen estimator is a finite product of matrices. The estimator was proposed by Aalen and Johansen (1978) as a product-limit estimator generalising the Kaplan-Meier estimator for the survival function in the two-state model (Kaplan & Meier, 1958). The Aalen-Johansen estimator is almost unbiased and may be interpreted as a nonparametric maximum likelihood estimator (Andersen et al., 1993). In case of no censoring before the end of follow up the Aalen-Johansen estimator for the probability of being in state h at time t is just the fraction of sample paths observed in state h at time t.

5. Estimation of Prolongation of Hospital Stay Attributable to Nosocomial Infections

Attempts have been made to estimate the additional time spent in hospital by "expert rating". An independent observer examines (retrospectively) the patients' files to determine the extra length of hospital stay attributable to an acquired infection. The resulting estimate is regarded as being not very reliable due to its subjective nature and is believed to underestimate the extra hospitalization time (McGowan, 1982). In consequence, we have to rely on statistical methods to estimate the extra stay in hospital realising that the additional time is not observable.

For each patient the time of admission to the hospital and the time of discharge is observable. For patients who acquire an infection, the time can be observed at which the infection becomes manifest. For patients who develop an infection, however, the duration of stay "had they not acquired the infection" is not observable. Similarly, for patients who were discharged without having acquired infection during hospital stay, the duration of stay "had they acquired an infection" is not observable.

We now describe two approaches which, in an ad-hoc manner, have been applied to estimate the extra hospitalization time on an empirical basis. Two non-parametric estimators

for the effect of nosocomial infection will then be proposed which are based on functionals of the transition probabilities in multistate models. They will be illustrated and compared to the ad-hoc approaches for the example data. Finally, a fifth approach based on G-estimation in structural nested failure time models will be described and compared to the others using the same example data set.

5.1. The "Naive" Approach

With the simplest method, we like to call it "naive" approach, patients are retrospectively stratified into two groups according to whether they have acquired nosocomial infection(s) or not. The mean duration of hospitalisation in the two groups of infected and non-infected patients is to be compared (see for example McGowan, 1982 or Brawley et al., 1989). This method disregards the time infected patients are "ahead" compared to patients who are discharged without having acquired an infection. Patients who are discharged early and who therefore have a lower chance to acquire an infection are classified as non-infected. This naive method will lead to biased results in that it overestimates the effect of nosocomial infection on duration of hospitalisation (Freeman et al., 1979; Haley et al., 1980; Green et al., 1982; Freeman & McGowan, 1984).

The same effect is known from phase II and phase III cancer clinical trials evaluating efficacy of chemotherapeutic agents where the effect of response to therapy on survival time is grossly overestimated when a simple comparison of responder and non-responder is performed and, in addition, a treatment comparison based on duration of response in the responders would give misleading results (Morgan, 1988; Weiss et al., 1983; Simon & Makuch, 1984; Temkin, 1978; Begg & Larson, 1982). Similarly, in studies on the effect of heart transplantation on survival time using patients who are accepted for transplantation but still waiting for a suitable donor as control group for transplanted patients in a simple two-group comparison would lead to an overestimate of the transplantation effect because early deaths would automatically be counted in the control group (Gail, 1972; Turnbull et al., 1974; Mantel & Byar, 1974; Aitkin et al., 1983).

5.2. The Matching Approach

A matching procedure is frequently applied to estimate the extra stay attributable to nosocomial infections (Leu et al., 1989; Kappstein et al., 1992a, 1992b; Pittet et al., 1994). For each infected patient (case) one or several control patients are selected from the pool of patients who are discharged from hospital without having acquired an infection and who are as similar as possible to the case with respect to specified risk factors for duration of hospitalisation. The matching approach consists of an individual comparison of length of hospital stay of patients who acquire an infection at some time during hospital stay (cases) with those who are discharged (alive) without having acquired an infection (controls). To account for factors which influence the duration of hospital stay, those factors should be considered as matching criteria. Furthermore, a suitable control patient should still be "at risk" for developing an infection until the time at which the infection becomes manifest

in the corresponding case. This requirement corresponds to what is known as "incidence-density sampling" in epidemiological case-control studies (Clayton & Hills, 1993). Hereby, some part of the positive bias in the naive method should have been eliminated. In contrast to the classical matched case-control study, where past exposure is of interest, the aim is to compare cases and controls with respect to a quantity (time until discharge) which can only be assessed future to development of disease. For this reason, in contrast to the nested case-control study within a cohort (Clayton & Hills, 1993), it is not sensible to allow cases to be used as controls at a time prior to their infection. In consequence, a suitable control has to fulfil the condition that (s)he will not become a case in the future. This necessary condition violates the basic principle in the counting process framework, that conditioning is only allowed on past events. Therefore an induction of bias by applying the matching approach is unavoidable.

5.3. Nosocomial Infection as Time-Dependent Covariate

One approach deserves to be mentioned which can be used to judge whether there is an effect of infection on time to discharge at all. However, this approach does not lead directly to an estimate of the extra days in hospital but should be applied before trying to estimate the extra time in hospital due to nosocomial infection by more elaborated methods. The approach avoids the drawback of the ad-hoc approaches, i.e. the retrospective stratification in infected and non-infected individuals and can be applied using widely available standard statistical software. The approach consists in fitting a proportional hazards model (Cox, 1972) to the discharge hazard with nosocomial infection included as time-dependent covariate (Andersen, 1986). The effect of this covariate can be tested by a test for the corresponding regression coefficient to be zero (Wald test; Kalbfleisch & Prentice, 1980). The size of the estimated regression coefficient gives a first impression of the effect that nosocomial infection has on prolongation of hospital stay. However, it does not lead to an estimate of the extra time because the model is formulated in terms of the hazard and not the actual time to discharge.

5.4. Nonparametric Estimation of the Extra Stay in Hospital Based on Functionals of the Transition Probabilities in Multistate Models: Approach A

In the following we will propose an estimator which is based on functionals of the transition probabilities in multistate models. In addition to the model assumptions mentioned in section 4 (independence of individual event times; non-homogeneous Markov model) this approach requires the assumption of independence of the risk of infection and the risk of death or discharge or, at least, independence conditional on measured covariates.

Let $\pi_0(0, t)$ denote the probability of an individual admitted to hospital at time 0 to be alive and in hospital (or ICU) from time 0 to time t. This probability is a composition of transition probabilities in the multistate model:

$$\pi_0(0, t) = P_{00}(0, t) + P_{01}(0, t) + P_{02}(0, t)$$

$$= \exp\left\{-\int_0^t (\alpha_{01}(u) + \alpha_{02}(u) + \alpha_{03}(u) + \alpha_{04}(u))du\right\}$$
$$+ \int_0^t (P_{00}(0,u)\alpha_{01}(u)P_{11}(u,t))du$$
$$+ \int_0^t (P_{00}(0,u)\alpha_{02}(u)P_{22}(u,t))du.$$

$\pi_0(0,t)$ denotes the probability that an individual (i) either stays in the initial state until time t without acquiring an infection and without being discharged or dieing before t, or, (ii) that the individual acquires pneumonia or sepsis at some time before t and stays in the infection state until t without being discharged or dieing before t. $\pi_0(0,t)$ can be estimated by inserting the Aalen-Johansen estimator for each of the component probabilities. $\pi_0(0,t)$ integrated over all times t is then the expected (or average) duration of hospital stay of a patient admitted to hospital at time 0.

Consider now the (hypothetical) situation that by introducing infection control programs or by establishing a certain vaccination procedure, one type of infection, e.g. nosocomial pneumonia, could be prevented. Dependent on the efficacy of the intervention, the hazard for nosocomial pneumonia may ideally be completely eliminated or may at least be reduced by a certain proportion. Let $\pi_0^\theta(0,t)$ denote the probability of an individual admitted to hospital at time 0 to be alive and in hospital (or ICU) from time 0 to time t in this hypothetical situation, where $\theta (0 \leq \theta \leq 1)$ denotes the reduction in the hazard of nosocomial pneumonia by the prophylactic measure. Then let $\pi_0^\theta(0,t)$ be given by:

$$\pi_0^\theta(0,t) = P_{00}^\theta(0,t) + P_{01}^\theta(0,t) + P_{02}^\theta(0,t)$$
$$= \exp\left\{-\int_0^t (\theta \times \alpha_{01}(u) + \alpha_{02}(u) + \alpha_{03}(u) + \alpha_{04}(u))du\right\}$$
$$+ \int_0^t (P_{00}^\theta(0,u) \times \theta \times \alpha_{01}(u) \times P_{11}(u,t))du$$
$$+ \int_0^t (P_{00}^\theta(0,u) \times \alpha_{02}(u) \times P_{22}(u,t))du.$$

$\pi_0^\theta(0,t)$ denotes a hypothetical probability for hospital stay up to time t in the situation where the intervention only affects the hazard of nosocomial pneumonia. This approach fits into a competing risk framework where one would ask the question of how many days in hospital can be saved when, as an extreme case, nosocomial infections could completely be eliminated ($\theta = 0$). This "partial" probability (Andersen et al., 1993) is only interpretable under the assumption that the risk of infection and the "risk" of discharge act independently from each other, and that by eliminating nosocomial infection, the intensity of the other events are not altered. This assumption would be violated if, for example, special prophylactic measures would prevent nosocomial infections, but side effects of this prophylactic treatment would prolong hospitalisation. Another scenario would be, that only less severe infections can be prevented and the more severe infections with a possibly higher intensity of death still occur; this would also violate the above assumption.

Then the hypothetical expected (or average) duration of hospital stay of a patient admitted to hospital at time 0 when the hazard of nosocomial pneumonia is reduced by a certain

factor θ is $\pi_0^\theta(0,t)$ integrated over all times t. Also $\pi_0^\theta(0,t)$ can be estimated by the Aalen-Johansen estimator for the component "θ-partial" probabilities. The Aalen-Johansen estimator $\hat{P}^\theta(0,t)$ for the matrix of the "θ-partial" transition probabilities is given by

$$\hat{P}^\theta(0,t) = \prod_{\upsilon=1}^{m} (I + \Delta\hat{A}^\theta(T_\upsilon)),$$

with

$$\mathbf{I} + \Delta\hat{\mathbf{A}}^\theta(T_\upsilon) = \begin{bmatrix} 1 - \frac{\Delta N_0^\theta(T_\upsilon)}{Y_0(T_\upsilon)} & \frac{\theta \times \Delta N_{01}(T_\upsilon)}{Y_0(T_\upsilon)} & \frac{\Delta N_{02}(T_\upsilon)}{Y_0(T_\upsilon)} & \frac{\Delta N_{03}(T_\upsilon)}{Y_0(T_\upsilon)} & \frac{\Delta N_{04}(T_\upsilon)}{Y_0(T_\upsilon)} \\ 0 & 1 - \frac{\Delta N_1(T_\upsilon)}{Y_1(T_\upsilon)} & \frac{\Delta N_{12}(T_\upsilon)}{Y_1(T_\upsilon)} & \frac{\Delta N_{13}(T_\upsilon)}{Y_1(T_\upsilon)} & \frac{\Delta N_{14}(T_\upsilon)}{Y_1(T_\upsilon)} \\ 0 & 0 & 1 - \frac{\Delta N_2(T_\upsilon)}{Y_2(T_\upsilon)} & \frac{\Delta N_{23}(T_\upsilon)}{Y_2(T_\upsilon)} & \frac{\Delta N_{24}(T_\upsilon)}{Y_2(T_\upsilon)} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

where

$$\Delta N_0^\theta(T_\upsilon) = \theta \times \Delta N_{01}(T_\upsilon) + \Delta N_{02}(T_\upsilon) + \Delta N_{03}(T_\upsilon) + \Delta N_{04}(T_\upsilon)$$

and $Y_j(T_\upsilon)$ denotes the number of patients at risk in state j at time T_υ, where T_υ denote the observed event times.

In the computations, only the terms including the number of individuals moving from state 0 to state 1 are affected by the reduction factor. All the other transition intensities remain unchanged according to the above assumptions.

An estimator for the extra stay in hospital attributable to nosocomial infection (here: pneumonia) can now be constructed based on the difference between the "observed" average duration of hospital stay and the "hypothetical" average duration of hospital stay dependent on θ. An estimator of the extra hospitalisation timed per patient of the cohort is given by:

$$\hat{E}_C^\theta(\delta) = \int_0^\tau \hat{\pi}_0(0,t)\,dt - \int_0^\tau \hat{\pi}_0^\theta(0,t)\,dt,$$

where τ denotes the last observation time. Note, that when the largest observation time is censored the estimate of the mean is biased.

In cost-benefit analysis of infection control programs for a postulated reduction of nosocomial infections, the number of hospital days saved per 1000 admissions, for example, and the corresponding reduction in cost can be calculated by this approach. Now $\hat{E}_C^\theta(\delta)$ is a population measure for the effect of nosocomial infections on length of hospital stay. It can be transformed into a measure for the individual infected patient. Dividing the estimated extra time per patient of the cohort by the rate of infections that hypothetically is prevented one obtains an estimate of the extra length of stay for the individual infected patient.

5.5. *Nonparametric Estimation of the Extra Stay in Hospital Based on Functionals of the Transition Probabilities in Multistate Models: Approach B*

While Approach A used a hypothetical quantity to estimate the extra time attributable to nosocomial infections, which is only valid under additional assumptions, Approach B relies

on observable quantities only and requires only the assumptions mentioned in section 4 (independence of individual event times; non-homogeneous Markov model).

Let $\pi_0(s,t)$ denote, analogously to Approach A, the probability of an individual to be alive and in hospital (or ICU) from time s to time t. $\pi_0(s,t)$ is a conditional probability, namely the probability of hospital stay at t given that the individual is still alive, in hospital and without nosocomial infection at time s, i.e. conditional on the individual is still in state 0. For each time s we can then define the expected "residual" duration of hospital stay given the individual has not moved out of state 0 until time s but may well acquire nosocomial infection later on:

$$r_0(s) = \int_s^\tau \pi_0(s,t)dt.$$

Let $\pi_1(s,t)$ denote the probability of an individual who acquired pneumonia at or prior to s and who is still alive and in hospital at time s, to be still alive and in hospital at time t and possibly having acquired sepsis prior to t in addition to nosocomial pneumonia. Then, $\pi_1(s,t)$ can be expressed as:

$$\begin{aligned}\pi_1(s,t) &= P_{11}(s,t) + P_{12}(s,t)\\ &= \exp\left\{-\int_s^t (\alpha_{12}(u) + \alpha_{13}(u) + \alpha_{14}(u))du\right\}\\ &\quad + \int_s^t P_{11}(s,u)\alpha_{12}(u)P_{22}(u,t)du.\end{aligned}$$

The expected "residual" duration of hospital stay given the individual is in the nosocomial pneumonia state at time s is then defined as

$$r_1(s) = \int_s^\tau \pi_1(s,t)dt.$$

We can now compare the expected duration of hospital stay of an individual who acquired pneumonia at time s (or prior to s) with the expected duration of stay of an individual who is free of infection at s, but—different from the matching approach—will possibly acquire pneumonia future to s. An alternative estimator of the expected extra hospitalisation time δ of an infected individual dependent on time s may therefore be constructed as:

$$\hat{E}_1(\delta, s) = \int_s^\tau \hat{\pi}_1(s,t)dt - \int_s^\tau \hat{\pi}_0(s,t)dt = \hat{r}_1(s) - \hat{r}_0(s),$$

where estimates are obtained by inserting the corresponding Aalen-Johansen estimator for the component transition probabilities. In the computations the state of the system is fixed at time s, i.e. the number of patients "at risk" in the various states as it is observed at time s is kept, but all transitions which occurred prior to time s are ignored. To obtain a summary measure we compute a weighted average of the extra hospital stay over all times s, weighted by the subdistribution function for the time of infection among those infected.

5.6. *Confidence Intervals for Approach A and B*

To derive confidence intervals for the extra time spent in hospital due to nosocomial infection we would need an estimate of the asymptotic variance-covariance matrix for the matrix of the transition probabilities. Computation of this matrix would require extensive computational work to be implemented. Therefore we choose to use bootstrap confidence intervals (Efron & Tibshirani, 1986) based on percentiles of the bootstrap distribution of $\hat{E}_C^{\theta}(\delta)$ and $\hat{E}_I^{\theta}(\delta)$, respectively. By means of a Monte Carlo algorithm we draw 2000 bootstrap samples (random sample of size n drawn with replacement from the actual sample of size n) to approximate the bootstrap distribution using the random number generator CALL RANUNI in SAS. The percentile method interval is the interval between the $100 \times (\alpha/2)$ and $100 \times (1 - \alpha/2)$ percentiles of the bootstrap distribution.

5.7. *G-estimation in Nested Structural Failure Time Models: A Counterfactual Approach*

One approach to estimate prolongation of hospital stay attributable to nosocomial infections which uses a completely different reasoning compared to the previously described ones, is based on a class of estimators referred to as G-estimators of parameters of a new class of causal models, the structural nested failure time models (SNFTM) (Robins et al., 1992).

Let T_i^- and $T_i^+(v)$ be defined as counterfactual variables. Let T_i^- be the duration of hospital stay of subject i if, possibly contrary to fact, no infection is acquired during hospital stay and let $T_i^+(v)$ be the time to discharge of subject i if nosocomial infection becomes manifest at time v. On the individual level, $T_i^+(v) - T_i^-$ then is the extra hospital stay due to infection.

A simple SNFTM formulated in terms of the counterfactual variables assumes that residual duration of hospital stay after having acquired nosocomial infection is expanded or contracted by a certain factor $\exp(-\eta)$, where η is unknown and has to be estimated:

$$(T_i^- - v) \times \exp(-\eta) = T_i^+(v) - v.$$

If $\eta < 0$ nosocomial infection prolongs residual hospital stay, otherwise if $\eta > 0$ residual duration of hospital stay would be decreased by nosocomial infection. The factor $\exp(-\eta)$ is called the expansion factor and $\exp(-\eta) - 1$ is the fractional increase in residual hospital stay due to nosocomial infection.

This model implies that if V_i is the observed time of nosocomial infection of subject i (if infection is acquired) and T_i is the observed duration of hospital stay then

$$\begin{aligned} T_i^- &= V_i + (T_i - V_i) \times \exp(\eta) && \text{if } (V_i < T_i) \\ T_i^- &= T_i && \text{otherwise.} \end{aligned}$$

For a series of values of η we can compute a new variable $L_i(\eta) = V_i + (T_i - V_i) \times \exp(\eta)$ for patients who acquire infection and $L_i(\eta) = T_i$ for patients who were discharged without having acquired nosocomial infection. Robins et al. (1992) propose a G-test of the hypothesis that the "true causal parameter" η equals some particular value η^* by fitting a

proportional hazards model to the hazard for nosocomial infection with $L(\eta)$ as time fixed covariate,

$$a_{inf}(t \mid F_{t-}, T_i) = a_{0,inf}(t) \exp(\beta \times L(\eta)),$$

and testing for the parameter $\beta = 0$. An asymptotic 95%-confidence interval $(\hat{\eta}_{low}, \hat{\eta}_{high})$ for η^* is given by those values of η which are consistent with the null hypothesis $\beta = 0$ at the 5% level; an unbiased and asymptotically normal distributed point G-estimate for the value of η is given by that particular η which leads to the score test statistic being equal to zero (Robins et al., 1992).

Having obtained an estimate of η, we can estimate the counterfactual duration of stay T_i^- for each individual who acquired infection as $\hat{T}_i^- = V_i + (T_i - V_i) \times \exp(\hat{\eta})$. An estimate of the extra hospital stay per infected patient attributable to nosocomial infection is then given by the average of the individual differences between the observed duration of stay $T_i(v)$ and the estimated counterfactual duration of stay $\hat{T}_i^-$. Constructing a 95%-confidence interval for the extra stay by using the upper and lower limits of the 95%-confidence interval for η^* would not take all sources of sampling variation into account (as noted by referee I). Therefore, we choose to use bootstrap confidence intervals for the extra stay as described in section 5.6.

6. Application of the Approaches to the Example Data Set for Nosocomial Pneumonia

6.1. The "Naive" Approach

Patients without acquiring nosocomial pneumonia during hospital stay are discharged alive from ICU on average 11.9 (± 19.2) days after admission, while patients who acquired nosocomial pneumonia have an average duration of ICU stay of 26.4 (± 21.4) days. The estimated extra stay due to nosocomial pneumonia would therefore be (over) estimated as being 14.4 days using the "naive" method. By applying a logrank test one would conclude that nosocomial pneumonia significantly prolongs ICU stay ($p < 0.001$).

6.2. The Matching Approach

The matching approach uses as matching criteria age (within ten years) and sex of the patients, and time to manifestation of nosocomial pneumonia in the case. The maximum number of controls per patient is restricted to five. For all but one case, control patients are available; in total 450 patients serve as controls. Duration of hospital stay is estimated to be prolonged by nosocomial pneumonia for 8.2 (± 13.1) days which is about one week less than the corresponding estimate obtained by the naive approach.

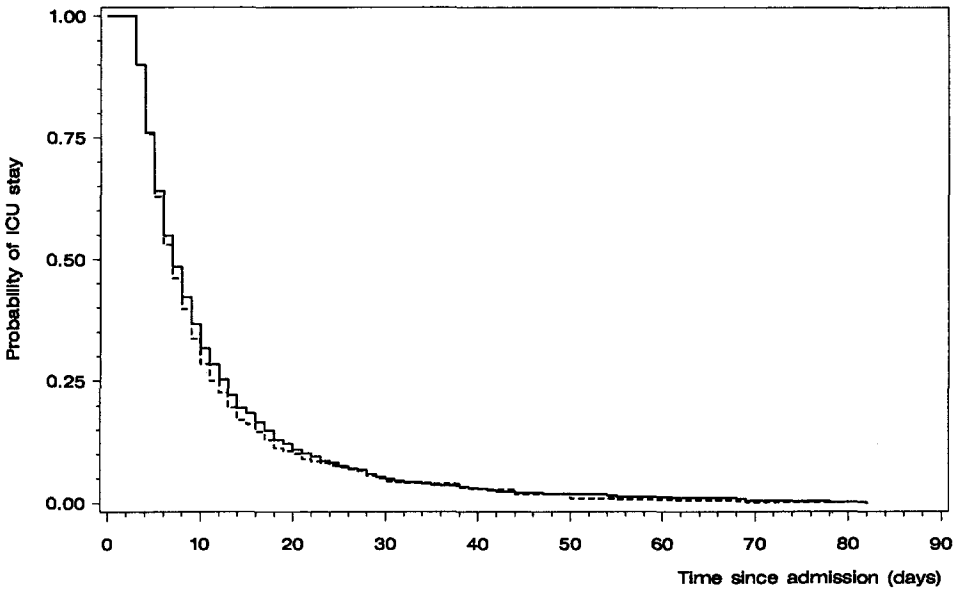

Figure 3. Probability of hospitalisation a.) with risk of nosocomial pneumonia acting (upper solid curve) and b.) with risk of nosocomial pneumonia removed completely (lower broken curve). The extra time due to nosocomial pneumonia per patient of the cohort is the area between the curves (Approach A).

6.3. Time-Dependent Covariate

Fitting a proportional hazards model to the discharge hazard with a time dependent covariate for nosocomial pneumonia which is coded "0" as long as no infection is acquired by the patient and "1" at the time the infection becomes manifest yields the following result: The effect of the time dependent covariate is highly significant ($p < 0.0001$). Nosocomial pneumonia significantly reduces the discharge hazard [RR = 0.6; 95%-CI: (0.4–0.7)], i.e. prolongs ICU stay.

6.4. Approach A

Figure 3 illustrates the Aalen-Johansen estimates for $\pi_0(0, t)$, the probability of being hospitalised at time t in the situation where the risk of nosocomial pneumonia is acting (upper solid curve), and for the "θ-partial" probability $\pi_0^0(o, t)$, the probability of being hospitalised at time t in the hypothetical situation where the risk of nosocomial pneumonia is removed completely, i.e. $\theta = 0$ (lower broken curve).

At first sight, the difference between the two curves does not look very impressive. However, the extra time per patient of the cohort attributable to nosocomial infection is estimated to be 0.54 days for $\theta = 0$, corresponding to the area between the curves in Figure 3.

Table 2. Result of Approach A for the data on nosocomial pneumonia: values of the hypothetical reduction factor θ (column 1) between 1 (no reduction) and 0 (complete removal of risk) in steps of 0.1; estimated rate of nosocomial pneumonia (column 2); estimated hypothetical rate of nosocomial pneumonia (column 3); estimated expected duration of ICU stay (column 4); estimated hypothetical duration of ICU stay (column 5).

θ	$\hat{P}_{inf}$	$\hat{P}^{\theta}_{inf}$	$\int \hat{\pi}_0$	$\int \hat{\pi}^{\theta}_0$	Extra days per patient of cohort	Days saved per 1000 patients	Extra days per infected patient
1.0	16.4%	16.4%	10.85	10.85	0.000	0	-
0.9	16.4%	15.1%	10.85	10.80	0.046	46	3.41
0.8	16.4%	13.7%	10.85	10.75	0.093	93	3.39
0.7	16.4%	12.2%	10.85	10.71	0.141	141	3.36
0.6	16.4%	10.7%	10.85	10.66	0.191	191	3.34
0.5	16.4%	9.1%	10.85	10.61	0.242	242	3.33
0.4	16.4%	7.5%	10.85	10.55	0.295	295	3.31
0.3	16.4%	5.8%	10.85	10.50	0.351	351	3.30
0.2	16.4%	3.9%	10.85	10.44	0.410	410	3.29
0.1	16.4%	2.0%	10.85	10.38	0.472	472	3.28
0.0	16.4%	0.0%	10.85	10.31	0.538	538	3.28

In the situation that the risk of nosocomial pneumonia can completely be removed one could save 538 ICU days per 1000 patients (ICU admissions). If the risk can be reduced by 50% ($\theta = 0.5$) one could save 242 days of ICU stay per 1000 patients. Table 2 summarises the results for the data on nosocomial pneumonia for values of θ (first column) between 1 (no reduction) and 0 (complete removal of risk) in steps of 0.1.

The estimated expected duration of ICU stay in the cohort is 10.85 days (column 4); the rate of nosocomial pneumonia is 16.4% (column 2). For θ e.g. equal to 0.6, that is, a reduction of the hazard of nosocomial pneumonia of 40%, one would expect the average duration of ICU stay to be 10.66 days (column 5) and the rate of infection to be 10.7% (column 3). A reduction of 40% would therefore result in saving of 191 hospital days per 1000 patients (column 7). In the last column the estimated extra time spent on ICU for an infected patient is given. We observe a slight dependency of $\hat{E}^{\theta}_{I}(\delta)$ on θ. The estimate varies between 3.41 and 3.28 extra days of ICU stay attributable to nosocomial pneumonia per infected patient. For $\theta = 0$ we obtain a 95%-confidence interval for the extra stay per infected patient due to nosocomial pneumonia of (0.8 days, 6.0 days). For the cohort effect our data are consistent with nosocomial pneumonia causing 137 to 969 extra days per 1000 ICU admissions at the $\alpha = 0.05$ level.

6.5. Approach B

As a result of Approach B Figure 4 illustrates the estimated expected duration of ICU stay conditional on being alive, hospitalised and infected at time s (solid line) and not infected at time s (broken line), respectively, up to day 50 on ICU (for later points in time, the risk

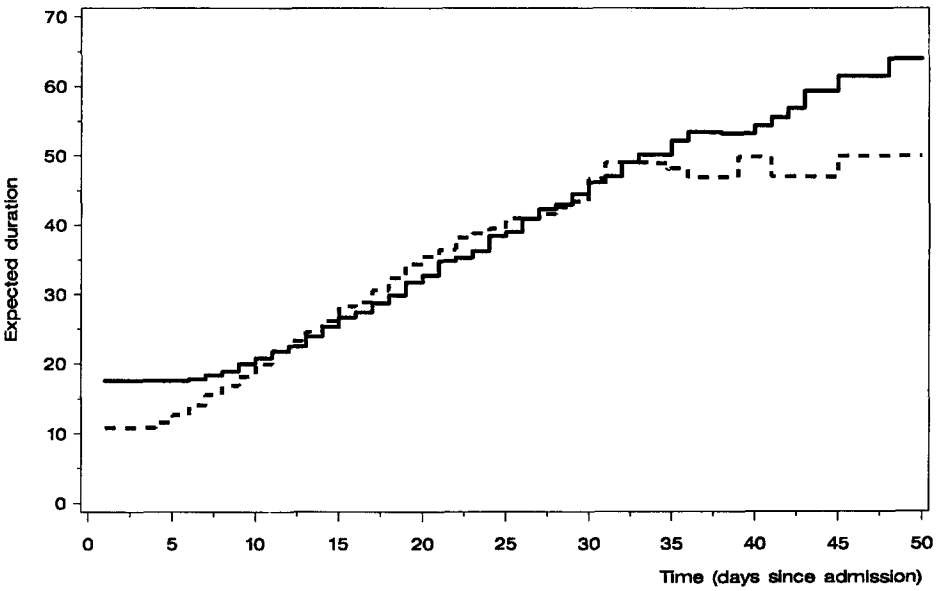

Figure 4. Estimates of the expected duration of ICU stay (days) given nosocomial pneumonia is acquired at or prior to time s and patient is alive and in hospital (solid line) and the expected duration of hospital stay (days) given the patient is alive, in hospital and free of nosocomial pneumonia at time s (broken line).

set in state 0 is empty). We observe that the two curves cross. The crossing of curves might have occurred by chance but may possibly also indicate a selection effect.

Weighting the difference between the curves in Figure 4 by the subdistribution function for the time of infection among those infected, which is illustrated in Figure 5, we obtain an estimate for the extra hospital stay per infected patient attributable to nosocomial pneumonia of 3.44 days, which is of the same order of magnitude as the estimate obtained with Approach A. We obtain a 95%-confidence interval (based on 2000 bootstrap replications) for the extra hospital stay of (1.44, 5.40).

6.6. The Counterfactual Approach

A proportional hazards regression analysis for the time to manifestation of nosocomial pneumonia including time to discharge as time fixed covariate gives a Score statistic of 8.92 so that the hypothesis of "no causal effect of nosocomial infection on duration of ICU stay" can be rejected ($p < 0.0028$) under the assumption of no unmeasured confounders. Table 3 summarises the results obtained in the SNFTM.

A subject's residual hospital stay after having acquired pneumonia is estimated to be prolonged by the expansion factor 1.48, resulting in absolute terms in an estimated extra hospital stay of 4 days attributable to nosocomial pneumonia.

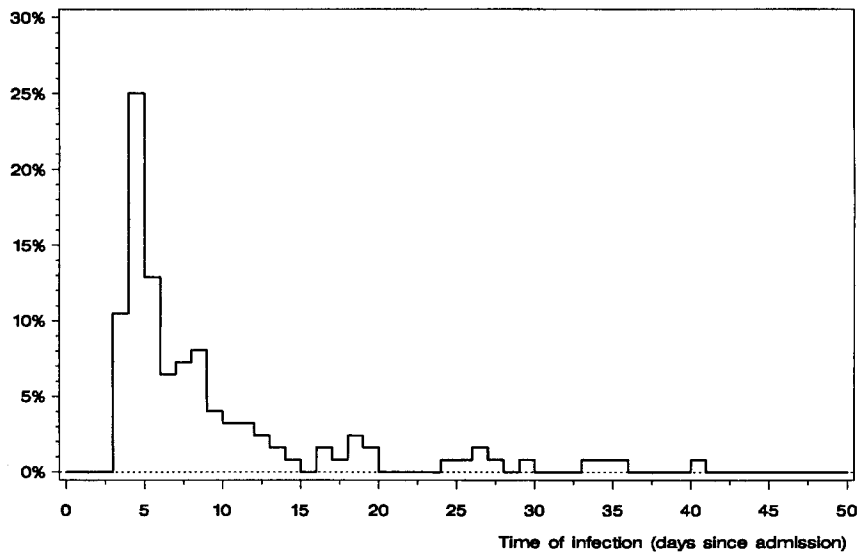

Figure 5. Distribution of the time to nosocomial pneumonia among those infected.

Table 3. Results of counterfactual approach for nosocomial pneumonia based on the structural nested failure time model (SNFTM): point estimates and 95% confidence intervals (CI).

Estimated parameter of SNFTM		Expansion factor		Estimated extra hospital stay	
$\hat{\eta}$	95%-CI	$exp(-\hat{\eta})$	95% CI	mean	95% CI
−0.39	(−0.69, −0.13)	1.48	(1.14, 1.99)	3.98 days	(1.23, 6.71)

6.7. *Summary of Results*

Table 4 summarizes the results obtained for the example data set for all approaches.

The "naive" approach leads to the highest estimates of the extra hospital stay attributable to nosocomial infections. The matching approach results in an estimate which is considerably below the "naive" one. Including nosocomial infection as time dependent covariate in a proportional hazards model for the discharge hazard results in a significant effect of this time dependent covariate, indicating that nosocomial infection prolongs hospital stay. The approaches based on functionals of the (partial and influenced) transition probabilities result in estimates which are again considerably below the matching estimate and are very close to the estimate based on the counterfactual approach.

The same relative behaviour of the approaches has been observed in two other applications concerning the effect of postoperative wound infections and ventilator-associated

Table 4. Summary of the results of the discussed approaches for the data on nosocomial pneumonia: estimated extra days on ICU with 95% confidence interval; p value, estimated relative risk (RR) and 95% confidence interval for nosocomial pneumonia as time dependent covariate in a Cox model for the discharge hazard.

Approach	Estimated extra stay (days)	95% - confidence interval
"Naive" Approach	14.4	(10.7, 18.2)
Matching Approach	8.2	(5.9, 10.5)
Time-dependent covariate	$p < 0.0001$, $RR = 0.6$	(0.4, 0.7)*
Multistate Approach A	3.4	(0.8, 6.0)
Multistate Approach B	3.4	(1.4, 5.4)
Counterfactual Approach	4.0	(1.5, 6.1)

*.for RR

nosocomial pneumonia on prolongation of hospital stay (Schulgen, 1995), although the absolute effect was quite different in the two studies (Kappstein et al., 1992a, 1992b).

7. Discussion

In the early seventies a publication of a study on the effect of heart transplantation on survival time (Clark et al., 1971) used the "naive" approach to estimate the effect of transplantation on prolongation of life. In the sequel, starting with the paper by Gail (1972), a discussion about the adequate statistical analysis of the transplantation data was set off in the statistical community. In the eighties, the discussion was resumed with respect to the problem of evaluating the relationship between survival and attainment of tumour response in cancer clinical trials. In this context, Simon & Makuch (1984) suggested an approach for graphical display of the relation between survival and occurrence of response which is some kind of mixture between our multistate approaches A and B.

Published literature as well as the results of this work suggests that the use of the "naive" and the matching approaches are not adequate to estimate the extra hospitalisation time due to nosocomial infections. The new approaches with "working titles" A and B are based on functionals of the transition probabilities. They are embedded in a sound methodology, estimators of their component parts exist with properties that are already known. Approach A compares the observed duration of hospitalisation of the study cohort with the hypothetical duration that would result from partial prevention of nosocomial infections. This approach avoids the drawback of retrospective stratification and yields population based estimates that can be used in cost-benefit analyses. Approach B compares the residual hospital stay of a currently infected patient with the residual hospital stay of a currently non-infected patient. This approach also avoids the drawback of retrospective stratification and uses observable quantities for the estimation of the extra stay on a per patient basis. The approach fits into the concept of dynamic predictive causality in longitudinal studies (Arjas & Eerola, 1993). Klein et al. (1994) applied this concept in a multicenter study on bone marrow transplantation in acute leukaemia patients. Within the class of structural nested failure

time models an additional approach to estimate the extra stay due to nosocomial infections has been developed. This approach is based on the counterfactual principle in causality and relates model based hypothetical to observed discharge times (see Eerola (1994) for a summary and discussion of philosophical theories in statistical analysis of causality).

In presenting methods for estimating prolongation of hospital stay attributable to nosocomial infections we have implicitly assumed that infections acquired in hospital are a causal factor for the excess stay. However, there may exist measured or unmeasured 'internal' covariates which may in part explain the occurrence of nosocomial infection, like the status of the immune system or the requirement for ventilation. These covariates may also be predictive for duration of hospital stay. Infection would then only be a marker of a predisposing condition and prolongation of hospital stay would be caused by these internal covariates. In this paper we have not considered any covariates (except for the matching approach) and we have presented only crude results. However, it is possible to include covariates in the multistate approaches by inserting the so called 'Breslow estimator' for the cumulative hazard function in the estimator for the transition probabilities (Andersen et al., 1993; Klein et al., 1994). Also for the counterfactual approach it is possible to let η depend on covariates (Robins et al., 1992). Adjustment of the estimates for measured influential covariates as well as application of methods which take unobserved heterogeneity (frailty models) into account have to be left to future work. It should also be noted that the new approaches allow for censored observations.

Application of the proposed methods rely on complete observation for each individual under study of the time at which transitions between different states occur. In the study used as examples this requirement is fulfilled. Another basic assumption is that the individuals and the observed sample paths are independent from each other. Dealing with infectious diseases this assumption seems to be somewhat unrealistic. However, nosocomial infections are in their majority caused by microorganisms of the patients' normal flora due to the weakening of their immunesystem by severe underlying illness or due to invasive diagnostic or therapeutic measures. Therefore an infected patient usually does not carry the infection to other patients. If an epidemic cannot be ruled out, then patients hospitalised in the same ward at the same time clearly can not be assumed independent.

Another crucial point is related to the assumption of an underlying non-homogeneous Markov process, that is, the assumption that the transition intensities and probabilities depend only on the time elapsed since the defined starting point and on the currently occupied state. However, transitions from states other than the initial one may also depend on the time elapsed since entry into the current state which would be a violation of the Markov assumption. The existence of such a violation can be checked by including the time elapsed since entry into the current state into a regression model based on time since entry into the system (Farewell & Cox, 1979). If transitions depend only on the time elapsed since entry into the current state the underlying process is semi-Markov (or a Markov renewal model). Voelkel & Crowley (1984) showed that the class of hierarchical semi-Markov models fits into the multiplicative intensity model of counting processes. These models certainly deserve consideration in future work.

The estimates for the extra hospital stay may become negative if infection results in a quick death of the patients. However, if this is the case or if infection contributes to a substantial

increase in mortality, like for nosocomial sepsis, analysing the impact on mortality should be the main issue of a statistical analysis. Analysing prolongation of hospital stay would then be pointless.

Which of the multistate approaches should be selected for the estimation of the extra stay due to nosocomial infections is still an open issue and may also depend on the question to be answered. The additional assumptions underlying approach A allow to assess the potential benefit achieved by the reduction of infection rates on the overall duration of hospitalisation within the population. This information can be easily incorporated into cost-benefit analyses. Approach B directly addresses the prediction of the extra stay of an infected patient without making the additional assumptions required for approach A.

In summary, multistate models based on multivariate counting processes were found to offer adequate tools and pave the way for new approaches to explore statistical problems arising in the dynamic process underlying occurrence and impact of nosocomial infections. In addition, the new class of structural nested failure time models seems to be a promising alternative. The relationship between approaches A and B based on the multistate framework and the counterfactual approach based on G-estimation in structural nested failure time models are not yet examined and further research is required.

Acknowledgments

We are grateful to Per Kragh Andersen for valuable suggestions and encouraging discussions. We also like to acknowledge the helpful comments of referee I. This work was supported by grant SCHU756/2-1 from the Deutsche Forschungsgemeinschaft (German National Research Council).

References

O. O. Aalen, "Nonparametric inference for a family of counting processes," *Annals of Statistics* vol. 6 pp. 701–726, 1978.

O. O. Aalen, and S. Johansen, "An empirical transition matrix for non-homogeneous Markov chains based on censored observations," *Scandinavian Journal of Statistics* vol. 5 pp. 141–150, 1978.

M. Aitkin, N. Laird, and B. Francis, "A reanalysis of the Stanford heart transplant data," *Journal of the American Statistical Association* vol. 78 pp. 264–292, 1983.

P. K. Andersen, "Time-dependent covariates and Markov processes," *Modern Statistical Methods in Chronic Disease Epidemiology* (S. Moolgavkar and R. L. Prentice, eds.), Wiley: New York, 1986, pp. 82–103.

P. K. Andersen, O. Borgan, R. D. Gill and N. Keiding, *Statistical Models Based on Counting Processes*, Springer: New York, 1993.

E. Arjas and M. Eerola, "On predictive causality in longitudinal studies," *Journal of Statistical Planning and Inference* vol. 34 pp. 361–386, 1993.

C. B. Begg and M. Larson, "A study of the use of the probability-of-being-in-response-function as a summary of tumor response data," *Biometrics* vol. 38 pp. 59–66, 1982.

J. V. Bennet and P. S. Brachman (eds.), *Nosocomial Infections*, Little, Brown and Company, 1993.

L. R. Brawley, D. J. Weber, G. P. Samsa and W. A. Rutala, "Multiple nosocomial infections. An incidence study," *American Journal of Epidemiology* vol. 130 pp. 769–780, 1989.

D. A. Clark, E. B. Stinson, R. B. Griepp, J. S. Schroeder, N. E. Shumway and D. C. Harrison, "Cardiac transplantation in man. Prognosis of patients selected for cardiac transplantation," *Annals of Internal Medicine* vol. 75 pp. 15–21, 1971.

D. Clayton and M. Hills, *Statistical Models in Epidemiology*, Oxford University Press Inc.: New York, 1993.

D. R. Cox, "Regression analysis and life tables" (with discussion), *Journal of the Royal Statistical Society* vol. B-34 pp. 187–220, 1972.

F. D. Daschner, "Economic aspects of hospital infections," *Journal of Hospital Infection* vol. 3 pp. 1–4, 1982.

F. D. Daschner, "The cost of hospital-acquired infection," *Journal of Hospital Infection* vol. 5 pp. 27–33, 1984.

B. Efron and R. Tibshirani, "Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy,"*Statistical Science* vol. 1 pp. 54–77, 1986.

M. Eerola, "Probabilistic Causality in Longitudinal Studies," *Lecture Notes in Statistics 92*, Springer: New York, 1994.

V. T. Farewell and D. R. Cox, "A note on multiple time scales in life testing," *Applied Statistics* vol. 28 pp. 73–75, 1979.

J. Freeman, B. A. Rosner, J. E. McGowan Jr., "Adverse effects of nosocomial infections," *Journal of Infectious Diseases* vol. 140 pp. 732–740, 1979.

J. Freeman and J. E. McGowan Jr., "Methodologic issues in hospital epidemiology. III. Investigating the modifying effects of time and severity of underlying illness on estimates of cost of nosocomial infection, *Reviews of Infectious Diseases* vol. 6 pp. 285–300, 1984.

M. H. Gail, "Does cardiac transplantation prolong life? A reassessment," *Annals of Internal Medicine* vol. 76 pp. 815–817, 1972.

J. S. Garner, W. R. Jarvis, T. G. Emori, T. C. Horan and J. M. Hughes, "CDC definitions for nosocomial infections," *American Journal of Infection Control* vol. 16 pp. 128–140, 1988.

R. D. Gill and S. Johansen, "A survey of product-integration with a view towards application in survival analysis," *Annals of Statistics* vol. 18 pp. 1501–1555, 1990.

M. S. Green, E. Rubinstein and P. Amit, "Estimating the effects of nosocomial infections on the length of hospitalization," *Journal of Infectious Diseases* vol. 145 pp. 667–672, 1982.

R. W. Haley, *Managing Hospital Infection Control for Cost-Effectiveness*, American Hospital Publishing: Chicago, 1986.

R. W. Haley, D. R. Schaberg, K. B. Crossley, S. D. Von Allmen and J. E. McGowan Jr., "Extra charges and prolongation of stay attributable to nosocomial infections: a prospective interhospital comparison," *American Journal of Medicine* vol. 70 pp. 51–58, 1981.

R. W. Haley, D. R. Schaberg, S. D. Von Allmen and J. E. McGowan Jr., "Estimating the extra charges and prolongation of hospitalization due to nosocomial infections: a comparison of methods," *Journal of Infectious Diseases* vol. 141 pp. 248–257, 1980.

J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley: New York, 1980.

E. L. Kaplan and P. Meier, "Non-parametric estimation from incomplete observations," *Journal of the American Statistical Association* vol. 53 pp. 457–481, 562–563, 1958.

I. Kappstein, G. Schulgen, U. Beyer, K. Geiger, M. Schumacher and F. D. Daschner, "Prolongation of stay and extra charges due to ventilator-associated pneumonia in an anesthesiological intensive care unit," *European Journal of Clinical Microbiology and Infectious Diseases* vol. 11 pp. 504–508, 1992a.

I. Kappstein, G. Schulgen, G. Fraedrich, V. Schlosser, M. Schumacher and F. D. Daschner, "Added hospital stay due to wound infections following cardiac surgery," *The Thoracic and Cardiovascular Surgeon* vol. 40 pp. 148–151, 1992b.

J. P. Klein, N. Keiding and E. A. Copelan, "Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients," *Statistics in Medicine* vol. 13 pp. 2315–2332, 1994.

A. Kropec, G. Schulgen, H. J. Just, K. Geiger, M. Schumacher, F. D. Daschner, "A scoring system for nosocomial pneumonia in intensive care units," *Intensive Care Medicine*, 1995 (in press).

H. S. Leu, D. L. Kaiser, M. Mori, R. F. Woolson and R. P. Wenzel, "Hospital-acquired pneumonia. Attributable mortality and morbidity," *American Journal of Epidemiology* vol. 129 pp. 1258–1267, 1989.

N. Mantel and D. P. Byar, "Evaluation of response-time data involving transient states: An illustration using heart-transplant data," *Journal of the American Statistical Association* vol. 69 pp. 81–86, 1974.

J. E. McGowan Jr., "The cost of hospital acquired infection," *Hospital Infection and its Control*, (S. Sabri and J. R. Tittensor, eds.), Barker Publication Ltd.: Richmond (UK), 1982.

T. M. Morgan, "Analysis of duration of response: a problem of oncology trials," *Controlled Clinical Trials* vol. 9 pp. 11–18, 1988.

W. Nelson, "Theory and applications of hazard plotting for censored failure data," *Technometrics* vol. 14 pp. 265–275, 1975.

D. Pittet, D. Tarara and R. P. Wenzel RP, "Nosocomial bloodstream infection in critically ill patients: excess length of stay, extra costs, and attributable mortality," *Journal of the American Medical Association* vol. 271 pp. 1598–1601, 1994.

H. Ramlau-Hansen, "Smoothing counting process intensities by means of kernel functions," *Annals of Statistics* vol. 11 pp. 453–466, 1983.

J. M. Robins, D. Blevins, G. Ritter and M. Wulfson, "G-estimation of the effect of prophylaxis therapy for pneumocystis carinii pneumonia on the survival of AIDS patients," *Epidemiology* vol. 3 pp. 319–336, 1992.

G. Schulgen, *Multistate Models for Event-History Data: Modelling Occurrènce and Impact of Nosocomial Infections*, PhD-Thesis, University of Dortmund, 1995.

R. Simon and R. W. Makuch, "A non-parametric graphical representation of the relationship between survival and the occurrence of an event: application to responder versus non-responder bias," *Statistics in Medicine* vol. 3 pp. 35–44, 1984.

N. R. Temkin, 'An analysis for transient states with application to tumor shrinkage," *Biometrics* vol. 34 pp. 571–580, 1978.

B. W. Turnbull, B. W. Brown Jr. and M. Hu, "Survivorship analysis of heart transplant data," *Journal of the American Statistical Association* vol. 69 pp. 74–80, 1974.

J. G. Voelkel and J. Crowley, "Nonparametric inference for a class of semi-Markov processes with censored observations," *The Annals of Statistics* vol. 60 pp. 142–160, 1984.

G. B. Weiss, Ill H. Bunce and J. A. Hokanson, "Comparing survival of responders and nonresponders after treatment: a potential source of confusion in interpreting cancer clinical trials," *Controlled Clinical Trials* vol. 4 pp. 43–52, 1983.