Intensive Care Med (2001) 27: 1254–1262
DOI 10.1007/s001340101007

ORIGINAL

Jokin de Irala-Estévez
Diego Martínez-Concha
Carmen Díaz-Molina
Josefa Masa-Calles
Amparo Serrano del Castillo
Rafael Fernández-Crehuet Navajas

Comparison of different methodological approaches to identify risk factors of nosocomial infection in intensive care units

Received: 8 September 2000
Final revision received: 16 January 2001
Accepted: 17 May 2001

J. de Irala-Estévez (✉)
Epidemiology and Public Health Unit,
Faculty of Medicine,
University of Navarre, Irunlarrea s/n,
31008 Pamplona, Navarra, Spain
E-mail: jdeirala@unav.es
Phone: +34-948-425600
Fax: +34-948-425649

D. Martínez-Concha · C. Díaz-Molina ·
J. Masa-Calles · R. Fernández-Crehuet
Navajas
Department of Preventive Medicine,
Reina Sofía University Hospital,
Córdoba, Spain

A. Serrano del Castillo
Preventive Medicine and Public Health.
Faculty of Medicine,
University of Cordoba, Cordoba, Spain

Abstract *Objective*: Comparison of statistical methods and measurement scales to identify nosocomial infection risk factors in intensive care units (ICU).
Design: Prospective study in 558 patients admitted to the ICU of a referral hospital between February and November 1994.
Methods: Analysis using three logistic regression models, three standard Cox regression models, and two Cox regression models with time-dependent extrinsic factors. Different scales were used to measure exposures to risk factors (dichotomous, ordinal, quantitative, and time-dependent variables).
Results: The most appropriate models were those that measured exposure using dichotomous variables. Models using ordinal or quantitative variables estimated biased coefficients and/or failed to comply with the statistical assumptions underlying the analyses. The Cox regression model with quantitative time-dependent variables met all the statistical assumptions, obtained a precise assessment of risk by exposure time, and estimated unbiased coefficients.
Conclusions: The Cox regression analysis with quantitative time-dependent variables is the most valid alternative for assessing the risk of nosocomial infection per day of exposure to an extrinsic risk factor in the ICU.

Keywords Cox proportional hazards · Logistic regression · Intensive care units · Nosocomial infections · Risk factors · Time-dependent variables

Introduction

Differences are often found between studies concerning risk factors for nosocomial infections (NI) in intensive care units (ICU) with regard to the importance which these factors have in inducing infection [1]. The reasons for such inconsistencies include differences between patients selected, diagnostic criteria for NI, and the statistical methods used to analyze findings. The aim of this study was to evaluate the statistical methods and measurement scales for identifying nosocomial infection risk factors. Multivariate analysis techniques are necessary for the valid identification of these risk factors because ICU patients are usually simultaneously exposed to several risk factors [2].

Logistic regression models estimate relative risks (obtaining odds ratios) associated with each risk factor in the model and probabilities of infection for each subject studied. However, they are not appropriate for estimating how quickly a risk factor causes the outcome in question. Although various authors [3, 4, 5, 6, 7, 8, 9] have used this technique to adjust simultaneously for several risk factors, logistic regression assumes the observation time to be similar for all patients studied [10, 11, 12]. This is seldom the case in prospective NI surveillance, where subjects usually have different observation periods (length of hospital stay). Furthermore, in the case of NI it is accepted that the length of hospital stay can be either the cause or the consequence of an infection and therefore must be considered specifically.

In a previous study we suggested that the use of Cox regression instead of logistic regression is more appropriate for the assessment of NI risk factors [13]. Cox regression provides coefficients whose exponentials correspond to adjusted incidence hazard ratios. The time elapsed between an exposure and the incidence of the infection of interest is thereby accounted for. Unfortunately, individual probabilities of infection are more difficult to estimate by this method. Nevertheless, a risk score can be assigned to patients such that higher scores indicate greater risks for early infection [14]. Patients can then be classified into groups based on their risk score to perform risk-adjusted comparisons between them. Also, Cox regression better accounts for different observation times prior to the infection onset.

Risk factors typically implicated in NI (such as therapeutic devices) are time-dependent variables, and exposure time varies with each patient. Patients can also be exposed to the same risk factor for different periods during their stay in the ICU. Several authors point out that these variables should be taken into account when using Cox regression in order to avoid estimating biased coefficients [15, 16, 17, 18]. Other authors claim that results obtained by logistic regression are equally as valid as those obtained by Cox regression when the observation time is short, the effect studied is rare, and the relative risk is moderate [19].

Another important issue to address is the extent to which coding schemes used to assess NI risk factors influence the results, especially when using multivariate models. For example, the presence of exposures can be measured using dichotomous variables (present/absent) or quantitative variables (number of days of exposure), the latter providing more information. Measurement scales can influence the statistical assumptions underlying models and thereby alter the validity and reliability of results. Furthermore, regardless of the scale used, risk factors should be considered only if present prior to the onset of the infection of interest. When determining the number of days of exposure to a device, such as a catheter, the only relevant days associated with a higher risk of infection are the days prior to the infection. Studies that use the total number of days of exposure to a device, including the days of use after the onset of infection, fail to account for the days of exposure that are the consequence rather than the cause of the infection [20]. For the same reason, when considering the number of days admitted to the ICU as a risk factor, only days prior to the infection onset should be used for analysis.

This study compared different methods of multivariate analysis and the various measurement scales commonly used to identify NI risk factors (dichotomous, ordinal, quantitative, time-dependent variables). The main objective was to compare the intrinsic validity of each method and scale and to compare other parameters, such as ease of analysis, interpretability, and practical usefulness of the results. Finally, the advantages and drawbacks of each method are discussed.

Methods and materials

The dataset used was obtained from a prospective study of patients admitted to the ICU of a referral hospital between February and November 1994 and had no missing data. The characteristics of the study have been published previously [13]. Patients were monitored daily from the time of their admission to the ICU until 48 h after their discharge [3, 21, 22]. For the purpose of this analysis patients excluded were those in the ICU for less than 24 h, those developing an infection more than 15 days after admission, those with coronary artery disease or decubitus ulcers, and those who had the head of their bed in a horizontal position (< 30°). The final sample consisted of 558 patients. These exclusions were designed solely to obtain a more homogeneous database, keeping in mind that the main objective of this study was methodological (validation of different analytical approaches) rather than the specific identification of risk factors, where a representative sample of ICU patients would be more relevant. During the study period no epidemic outbreaks of NI were detected. The diagnosis of NI was based upon criteria established by the Centers for Disease Control (CDC) [23]. The first three NI were recorded for each patient, including their location and date of onset.

Information concerning variables described in the medical literature as intrinsic risk factors for NI was collected. These variables included age, gender, disease severity at ICU admission (estimated using the Acute Physiology and Chronic Health Evaluation II score [24]) and the primary diagnosis at ICU admission (according to the International Classification of Diseases, 9th Revision Clinical Modification [25]). Other intrinsic factors assessed using dichotomous variables (present/absent) were infection prior to ICU admission, traumatic head injury, coma, renal failure, diabetes, neoplasm, neutropenia, liver cirrhosis, intravenous drug use, and malnutrition.

The presence of extrinsic factors during the ICU stay using dichotomous variables (present/absent) was also assessed. These included anti-H_2 drugs, antacids, sedatives, immunosuppressive agents, bronchoscopy or gastroscopy, antimicrobial treatment prior to the NI, and surgical interventions including those performed prior to the ICU admission. The therapeutic devices or procedures that the CDC considers most important in ICU settings, mechanical ventilation, central intravascular catheter, and urinary catheter [21], were assessed using three different scales of measurement: dichotomous (present/absent), ordinal with three categories (less than 1 day, 1–6 days, and 7 or more days of exposure) and quantitative (number of days of exposure). The presence of a risk factor was only recorded if it preceded the occurrence of the first nosocomial infection to ensure a correct time sequence; likewise, the length of exposure was recorded until that moment.

As part of an initial descriptive analysis, infected and noninfected patients were compared using Student's t test for quantitative variables, and the χ^2 test for qualitative variables. The multivariate analysis consisted of fitting different models using a combination of methods (logistic and Cox regressions) and risk factor scales (dichotomous, ordinal, quantitative, and time-dependent variables). All variables mentioned above known as risk factors for NI were used as possible candidates for the multivariate models evaluated.

Multivariate logistic regression analysis

The dependent variable for the logistic regression analysis was the onset of the first NI during a patient's stay in the ICU. Three analyses were performed using logistic regression, each with a different coding strategy to assess the most important NI risk factors in the ICU (mechanical ventilation, central intravascular catheter, and urethral catheter [21]).

Models were initially constructed using a purposeful selection of variables with a conservative significance level of 0.25. The likelihood ratio test (*G* statistic) was subsequently used to assess the significance of the variables and to obtain the final model with the usual significance level of 0.05. The scales of the quantitative variables were assessed using the Box-Tidwell test. The presence of confounding and interaction was ruled out prior to assessing the model's goodness of fit. When a variable was identified as a confounder, despite presenting borderline significance, it was included in the model to be appropriately adjusted for [26, 27]. Finally, model validity was assessed by estimating its goodness of fit (the extent to which the model reflects the data) using the Hosmer-Lemeshow *C* statistic and by estimating its discrimination (the extent to which the model distinguishes patients who become infected from those that do not) using the area under the receiver operating characteristic curve [26, 27].

Multivariate Cox regression analysis

The dependent variable for Cox regression was length of ICU stay prior to the onset of the first infection, for the infected subjects, and the time elapsed between ICU admission and discharge from the unit, for the uninfected subjects (uninfected subjects being referred to as censored in the Cox regressions). Three Cox regression models were fit with extrinsic factors as described for logistic regression and measured using the same variables.

The same method of model construction described previously was carried out to develop the models for Cox regression. Final models were also assessed in terms of scales of quantitative variables, confounding effects, and interactions. Their goodness of fit was evaluated using graphic representations of the adjusted values and the Martingale residuals [26, 27, 28]. The proportional hazards assumption was verified by the Phtest procedure of the STATA 4.0 [28] program. Discrimination was assessed by a receiver operating characteristic curve obtained by plotting values of sensitivity and 1–specificity, estimated using each quintile of patient risk scores as cutoff points.

The infection risk scores were estimated using the final NI models and by summing up the products of the coefficients of the models and the corresponding variable values for each patient (risk score = $\beta_1 V_1 + \beta_2 V_2 + ... + \beta_n V_n$; where β_n are the coefficients of the models and V_n are the corresponding model variables).

Multivariate analysis of Cox regression considering the extrinsic factors as time-dependent variables

Two additional Cox regression models were developed, using an automatic procedure of data management [28], in which extrinsic factors were considered time-dependent variables. This procedure takes into account that values of a time-dependent variable can change from one day to the next during the ICU stay of a patient. This is possible by creating one record per day of stay in the ICU and for each patient. Non-time-dependent variables such as gender would be held constant across all records for each patient. When using dichotomous variables, each day of stay (each record thereon) was coded as having or not having that particular factor present on that day and, when using quantitative variables, days of exposure were summed from one day to the next. With this procedure it was possible to consider a subject as exposed or not to a risk factor at any moment in time, to assess the difference between more or less days of exposure, as well as to distinguish between successive episodes of exposures to a same risk factor. The models also differed in the scale used to measure extrinsic factors (quantitative or dichotomous variables). The model selection process followed the same steps discussed previously.

Once the analyses with various methods and scales were performed, we compared their internal validity and other aspects such as the ease of data management and the usefulness of their results. The scoring system used to present results on calibration, discrimination, ease of analysis, interpretability, and practical use of models indicate inadequacy of a model ("poor") or enable the comparison between acceptable values ("moderate") and better ones ("good"). Since the study was an observational study using routinely collected data in a NI surveillance program, the approval by an ethics committee was not warranted.

Patients

The study sample consisted of 558 patients, the majority of whom were men (67.6 %). The mean age was 50.3 ± 17.9 years, with 25 % of the patients over 65 years of age. There was no significant difference in age between men and women. The mean Acute Physiology and Chronic Health Evaluation II score was 11.3 ± 5.9. The most frequent diagnosis was cardiovascular disease, and all patients with this diagnosis had undergone cardiovascular surgery. Approximately 66 % of patients had undergone some type of surgical intervention, and 18.6 % had suffered an infection prior to admission to the ICU.

The mean length of stay in the ICU was 6.6 days, with a statistically significant difference between infected and noninfected patients (16 and 4.3 days respectively; $P < 0.0001$). The mean stay in the ICU prior to the first infection was 4.8 days.

At least one NI was suffered by 109 patients during their stay in the ICU, which represents an annual cumulative incidence of 19.5 infected patients per 100 admissions (95 % CI = 16.3–19.5). However, some of these patients presented more than one infection. The total number of infections was 186, resulting in a ratio of 33.3 infections per 100 admissions. The incidence density of infection was of 50.6 per 1000 patient-days.

Results

Logistic regression analysis

The three logistic regression models fitted are very similar (Table 1). The variable associated with the highest risk of infection is an intrinsic factor, malnutrition (odds ratio greater than 13 in the three models), followed by an extrinsic factor, mechanical ventilation (odds ratio greater than 5 in the models measuring the presence of this risk factors using an ordinal or dichotomous variable). The central intravascular catheter with

Table 1 Factors significantly associated with nosocomial infection in the multivariate logistic regression analysis (OR odds ratio, *CI* confidence interval)

	Measurement scales of exposures to risk factors					
	Quantitative: days		Ordinal: days		Dichotomous: presence-absence	
Risk factors	OR	95% CI	OR	95% CI	OR	95% CI
Head injury[a]	4.33	1.83–10.26	3.88	1.62–9.28	3.58	1.55–8.29
Coma[a]	1.89	0.84–4.21	1.83	0.83–4.06	1.72	0.79–3.76
Infection prior to ICU admission[a]	3.79	2.05–7.02	2.80	1.41–5.59	2.65	1.35–5.21
Malnutrition[a]	13.13	3.52–48.95	14.95	3.82–58.44	13.90	3.58–53.90
Antacids[a]	2.80	1.08–7.24	2.71	1.01–7.27	–	–
Sedatives[a]	1.92	1.11–3.35	1.81	1.04–3.17	1.77	1.02–3.07
Immunosuppression[a]	3.79	2.05–7.02	3.31	1.77–6.18	3.49	1.89–6.44
Previous antibiotics[a]	0.43	0.21–0.90	0.31	0.14–0.67	0.32	0.15–0.68
Location of surgery[b]						
Limbs, thorax, others	1.40	0.65–3.04	1.04	0.44–2.47	1.41	0.65–3.05
Thorax-abdomen, abdomen	2.97	1.35–6.53	2.41	1.05–2.43	2.76	1.27–5.99
Days with central catheter with a peripheral insertion site	0.89	0.79–0.99	–	–	–	–
Days of mechanical ventilation	1.39	1.21–1.59	–	–	–	–
Days of central catheter[c]						
1–6 days	–	–	1.77	0.87–3.61	–	–
≥7 days	–	–	1.63	0.60–4.40	–	–
Days of mechanical ventilation[c]						
1–6 days	–	–	5.24	2.78–9.88	–	–
≥7 days	–	–	6.02	1.88–19.31	–	–
Mechanical ventilation[a]	–	–	–	–	5.90	3.17–10.97

[a] Reference for dichotomous variables: absence of factor

[b] Reference of locations: head and neck; spinal column; no intervention

[c] Reference of days of factors: less than 24 h of exposure to risk factor

a peripheral insertion site is only identified as a significant variable when it is assessed in days, although as a protective factor. Table 2 compares the internal validity (calibration and discrimination) and other characteristics of these models.

Cox regression analysis

The intrinsic risk factors that presented stronger associations in the univariate analysis (diagnostic categories, malnutrition, and head trauma) were significant in the three multivariate Cox regression analyses (Table 3). Mechanical ventilation was significant regardless of the measurement scale. The use of a central intravascular catheter or a central intravascular catheter with a peripheral insertion site was significantly protective when assessed as either quantitative or ordinal variables. The presence of a urinary catheter, assessed in days, was also identified as a protective factor. Table 4 compares the internal validity and other characteristics of the models.

Cox regression analysis with time-dependent variables

When the exposure to therapeutic devices was assessed using time-dependent variables and the days of exposure prior to the beginning of the first infection was used for analysis, all factors had positive coefficients in the final model. The variable "days of mechanical ventilation" was a significant risk factor for both the first and second episode of device use. The administration of sedative drugs, malnutrition and diagnostic categories were also identified as significant risk factors (Table 5). The model provides additional information about successive episodes of mechanical ventilation, revealing that not only do the adjusted infection rates increase as days of exposure increase, but also this increase is greater during a second episode of mechanical ventilation (after having been extubated from the first episode). The model that assesses the presence of therapeutic devices as time-dependent dichotomous variables is very similar to the one described above (Table 5).

Table 2 Comparison of characteristics of logistic regression models

Comparison criteria	Models with exposures to risk factors measured using different scales		
	Quantitative: days	Ordinal: days	Dichotomous: presence-absence
Calibration[a]	0.4731	0.6711	0.4048
Discrimination[b]	0.8626	0.8699	0.8379
Linear scale[c]	Inadequate	n.a.	n.a.
Ease of analysis	Moderate	Moderate	Good
Interpretability	Moderate	Good	Good
Practical use of results	Good	Good	Moderate

[a] Hosmer-Lemeshow goodness-of-fit: significance level of *C* statistic

[b] Area under the ROC curve

[c] Of quantitative variables

Table 3 Factors associated with nosocomial infection in the multivariate Cox regression (*CI* confidence interval; *IDR* incidence density ratio; *group 2* infectious or parasitic disease, of the digestive system, of the muscular system, skeleton and connective tissue, ill-defined symptoms and signs; *group 3* diseases of the nervous system and sensory organs, of the respiratory system, lesions, neoplasms, endocrine diseases, nutrition and metabolism and immune disorders)

Risk factors	Measurement scales of exposures to risk factors					
	Quantitative: days		Ordinal: days		Dichotomous: presence-absence	
	IDR	95% CI	IDR	95% CI	IDR	95% CI
Head injury[a]	1.97	1.13–2.82	1.87	1.14–3.05	1.86	1.14–3.05
Coma[a]	1.59	0.89–2.87	–	–	–	–
Malnutrition[a]	2.38	1.40–4.03	2.62	1.57–4.39	2.12	1.23–3.68
Sedatives[a]	1.79	1.13–2.82	–	–	–	–
Diagnosis						
Group 2	2.50	1.17–5.36	2.63	1.43–4.82	2.63	1.46–4.75
Group 3	2.28	1.17–4.44	2.25	1.25–4.04	2.33	1.32–4.11
Location of surgery[b]						
Limbs, thorax, others	1.16	0.58–2.33	–	–	–	–
Thorax-abdomen, abdomen	1.85	1.05–3.26	–	–	–	–
Days with central catheter	0.91	0.83–1.00	–	–	–	–
Days with central catheter with a peripheral insertion site	0.84	0.76–0.92	–	–	–	–
Days with urinary catheter	0.81	0.71–0.92	–	–	–	–
Days of mechanical ventilation	1.24	1.10–1.38	–	–	–	–
Days of central catheter[c]						
1–6 days	–	–	0.98	0.61–1.58	–	–
≥7 days	–	–	0.28	0.14–0.59	–	–
Days of central catheter with a peripheral insertion site[b]						
1–6 days	–	–	0.75	0.47–1.20	–	–
≥7 days	–	–	0.18	0.09–0.38	–	–
Days of mechanical ventilation[b]						
1–6 days	–	–	3.81	2.25–6.46	–	–
≥7 days	–	–	4.81	1.97–11.72	–	–
Mechanical ventilation[a]	–	–	–	–	2.83	1.49–5.37
Previous antibiotics[a]	–	–	–	–	0.58	0.36–0.94

[a] Reference category for dichotomous variables: absence of the factor

[b] Reference of locations: head and neck; spinal column; no intervention

[c] Reference of days of factors: less than 24 h of exposure to risk factor

Discussion

We analyzed a database from a prospective epidemiological surveillance with concurrent, daily visits to the ICU. This type of surveillance presents the greatest sensitivity and specificity for the identification of NI [29, 30]. As our objective in this paper was not the identification of NI risk factors, we were not exhaustive in considering extrinsic risk factors, especially therapeutic devices, as possible candidates in our analyses. Instead, we fo-

Table 4 Comparison of characteristics of Cox regression models

Comparison criteria	Models with different measurement scales Quantitative: days	Ordinal: days	Dichotomous: presence-absence
Calibration[a]	Good	Good	Good
Discrimination	Moderate	Good	Good
Proportional risks assumption[b]	No	Complies	Complies
Ease of analysis	Moderate	Moderate	Good
Interpretability	Moderate	Good	Good
Practical use of results	Good	Good	Moderate

[a] Goodness of fit with graphs of adjusted values and Martingale residuals

[b] STATA Ph test

cused on three of them (mechanical ventilation, central intravascular catheter, and urinary catheter) and concentrated our efforts on the description of alternative analytical and variable coding approaches.

It is not easy to interpret why the "days of exposure to a central catheter with a peripheral insertion site" was identified as protective because one would expect the contrary. It is possible that this variable had a confounding association with a protective variable that we have not studied. However, the most plausible explanation is that this is an artifact due to the scale used to measure the variable.

The days of exposure to a risk factor refer to days prior to the onset of a NI in infected patients, and prior to discharge in noninfected patients. Should the factor studied not increase the risk of infection, noninfected patients would contribute to more days of exposure than infected patients, and this survival-bias-like error could mislead us to believe that the factor is protective when it actually has no effect on the outcome. This is especially true for early infections. In an attempt to correct this artifact, exposures can be estimated as ratios, that is, dividing the number of days of exposure to a device or risk factor by the days at risk for infection. This is analogous to what some authors do when calculating the "device utilization ratio" proposed by the CDC [31]. However, this ratio does not resolve the problem. As Fig. 1 illustrates, with early infections, and assuming the factors studied are not associated with a higher risk of infection, a greater value for the "utilization ratio" would result for infected patients (the denominator being smaller) despite the fact that both infected and noninfected patients (patients A and B) present the same level of risk (two devices used during 7 days). Singh-Naz et al. [31] have suggested that when using logistic regression analysis, the "utilization ratio" for devices is an adequate estimator of infection risk but, unlike in our study, the infections they studied had a late onset

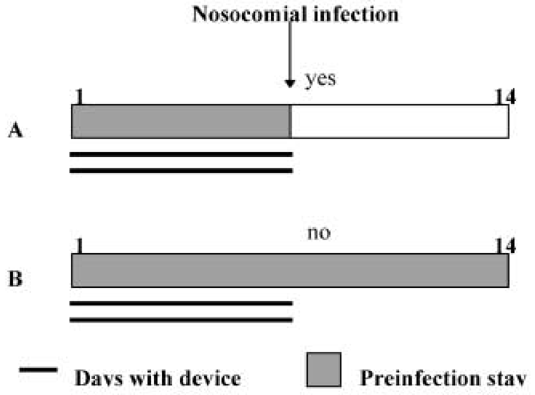
Table 5 Factors significantly associated with nosocomial infection in the Cox regression multivariate analysis, considering the extrinsic CDC factors as time-dependent variables (*CI* confidence interval; *IDR* incidence density ratio; *group 2* infectious or parasitic disease, of the digestive system, of the muscular system, skeleton and connective tissue, ill-defined symptoms and signs; *group 3* diseases of the nervous system and sensory organs, of the respiratory system, lesions, neoplasms, endocrine diseases, nutrition and metabolism and immune disorders)

	Measurement scales of exposures to risk factors			
	Quantitative: days		Dichotomous: presence/absence	
Risk factors	IDR	95% CI	IDR	95% CI
Malnutrition[a]	2.56	1.55–4.23	2.39	1.44–3.96
Diagnosis				
Group 2[b]	2.28	1.25–4.16	2.14	1.18–3.88
Group 3[b]	2.66	1.60–4.42	2.32	1.38–3.90
Sedation[a]	1.61	1.03–2.51	1.51	0.97–2.36
Days of mechanical ventilation				
First episode	1.17	1.08–1.27	–	–
Second episode	1.24	1.01–1.52	–	–
Mechanical ventilation[c]	–	–	2.53	1.66–3.85

[a] Reference category for dichotomous variables: absence of the factor

[b] Reference category: circulatory disease, genitourinary disease, skin and subcutaneous tissue disease, poisoning, complications of pregnancy, childbirth and puerperium

[c] Reference of locations: head and neck, spinal column, no intervention



$$\textbf{Device utilization ratio} = \frac{\sum(\textbf{days with device})}{\textbf{Preinfection stay}}$$

$$\textbf{Device utilization ratio (A)} = \frac{7+7}{7} = 2$$

$$\textbf{Device utilization ratio (B)} = \frac{7+7}{14} = 1$$

Fig. 1 Device utilization ratio

Table 6 Comparison of characteristics of Cox regression models with extrinsic CDC factors as time-dependent variables

Comparison criteria	Measurement scales of exposures to risk factors	
	Quantitative: days	Dichotomous: presence-absence
Calibration[a]	Good	Good
Discrimination	Moderate	Moderate
Proportional risks assumption[b]	Complies	No
Ease of analysis	Moderate	Good
Interpretability	Moderate	Moderate
Practical use of results	Good	Moderate

[a] Goodness of fit with graphs and Martingale residuals
[b] STATA Ph test

(mean preinfection stay of 9.1 days and mean stay in non-infected patients of 4 days). Also, the "utilization ratio" included, in addition to days with an intravascular catheter, days with a urinary catheter and days on a ventilator. In any case, when we applied the Box-Tidwell test to the quantitative variables in our logistic regression analysis, the underlying statistical assumptions were not met.

The three logistic regression models presented in Table 2 suggest that the model assessing extrinsic factors in days is inadequate because assumptions of linearity are not met, and the assessment of days prior to infection biases the estimation of coefficients. Conversely, the logistic regression with days of exposure coded as ordinal variables presents good calibration and discrimination parameters (slightly better than the model which uses a dichotomous coding). The ordinal scale is more useful than the dichotomous one. It enables the evaluation of different levels of risk among exposed subjects even though dichotomous variables are easier for data collection and apparently estimate unbiased coefficients.

The explanation for the negative coefficients in the Cox regression for some of the variables that measure days of exposure to a device is similar to that mentioned earlier for the logistic models. The Cox regression does not solve the problem of negative coefficients because, if a variable is not associated with NI, using "days exposed to a device" would "penalize" the infected subjects who would always have less days of exposure. This falsely suggests that having fewer days of exposure is associated with a higher risk of NI when in fact the exposure has no effect. Similarly, when using ordinal variables, the infected subjects would always have a greater probability of being included in categories with fewer days of exposure.

There are a few differences between the three Cox regression analyses (Table 4). The model assessing extrinsic factors in days is not valid in terms of discrimination, it does not fulfill the proportional hazards assumption, and the assessment of the days prior to infection leads to biased coefficients. The models with ordinal or dichotomous variables present similar goodness of fit and discrimination values and comply with the proportional hazards assumption. However, the use of ordinal variables does not eliminate possible errors in the estimation of coefficients and could result in falsely identifying risk factors as protective. The dichotomous variable is easier to manage in the data collection phase of a study, thus suggesting that it might be the preferred scale when Cox regression is used.

After comparing the two Cox regression models with the extrinsic factors measured as time-dependent variables, the quantitative scale appeared as the most valid coding alternative for the identification of NI risk factors (Table 6). This model presents good discrimination and goodness of fit parameters and meets all the statistical assumptions. It enables a comprehensive assessment of the risk implied by each day of exposure and, when there is a need for distinguishing between different episodes of exposure as in our case, it allows an adequate assessment of each episode independently. The time-dependent coding of variables eliminated the biases of coefficient estimation described in the previous analyses (Tables 1, 3). The Cox regression model with time-dependent dichotomous variables does not comply with the proportional hazards assumption.

An analysis using the ordinal time-dependent coding was omitted because it was considered less advantageous than the quantitative coding. Also, it would be more complicated to transform the quantitative variables into categories and then use the time-dependent analysis than merely to use quantitative variables.

The problem of the length of ICU stay

Several authors have found that the risk of infection is proportional to the length of stay, and they therefore highlight the need to control for this factor [3, 32, 33, 34]. However, the infection itself prolongs the patient's stay in the hospital, hence the length of stay may be a cause and/or a consequence of infection. This is why "days prior to the beginning of the first infection" is frequently used to estimate the risk attributable to the length of stay rather than the "the total length of the stay" [35, 36, 37]. The logistic regression analysis should accordingly control for the length of stay prior to infection. However, biased coefficients might be estimated for some therapeutic devices and could be identified as protective factors, in the case of early infections, or as risk factors, in the case of late infections, when they in fact have no real effect on the outcome. This could be the reason why Singh-Naz et al. [30] reported an association between length of stay and NI in their logistic regression analysis; their mean preinfection stay in infected subjects was double the total stay in uninfected subjects. In their case the uninfected patients may have been somewhat penalized by this artifact.

Finally, the most valid alternative for the assessment of the risk per day of exposure to extrinsic risk factors is Cox regression with quantitative time-dependent variables. Cox regression has the advantage of obtaining incidence density ratios, more appropriate for the analysis of the length of stay in the ICU. However, this method implies knowing the exact starting and ending dates of the exposures which requires a prospective surveillance system that consumes more resources. In addition, it requires a more elaborated data management process prior to the data analysis.

When the objective is the mere detection of risk factors without necessarily needing the estimate of the risk per day of exposure, a dichotomous scale is preferable and can be applied in both logistic and Cox regressions. Logistic regression is preferable when the range of days in the ICU is not very large and when there is an interest in obtaining infection probabilities, provided that we control for preinfection stay. However, Cox regression with dichotomous variables has the advantage of accounting for preinfection stay. The decision about which method to choose and what coding scheme to use depends upon the research goals of a study, and the resources available to achieve them should be considered.

Acknowledgements We thank the staff of the Intensive Care Unit of the Reina Sofía Hospital in Cordoba and to J.P. Glutting and Marta Pankovich for translating this paper into English.

References

1. Chevret S, Hemmer M, Carlet J, Langer M and the European Cooperative Group on Nosocomial Pneumonia (1993) Incidence and risk factors of pneumonia acquired in intensive care units. Results from a multicenter prospective study on 996 patients. Intensive Care Med 19: 256–264
2. Freeman J, McGowan JE Jr (1978) Risk factors for nosocomial infection. J Infect Dis 138: 811–819
3. Craven DE, Kunches LM, Lichtenberg DA, Kollisch NR, Barry A, Heeren TC, et al (1988) Nosocomial infections and fatality in medical and surgical intensive care unit patients. Arch Intern Med 148: 1161–1168
4. Celis R, Torres A, Gatell JM, Almela M, Rodríguez-Roisin R, Agustí-Vidal A (1988) Nosocomial pneumonia: a multivariate analysis of risk and prognosis. Chest 93: 318–324
5. Harkness GA, Bentley DW, Roghmann KJ (1990) Risk factors for nosocomial pneumonia in the elderly. Am J Med 89: 457–463
6. Hanson LC, Weber DJ, Rutala WA (1992) Risk factors for nosocomial pneumonia in the elderly. Am J Med 92: 161–166
7. Moro ML, Viganó EF, Cozzi Lepri A, The Central Venous Catheter-Related Infections Study Group (1994) Risk factors for central venous catheter-related infections in surgical and intensive care units. Infect Control Hosp Epidemiol 15: 253–264
8. Fridkin SK, Pear SM, Williamson TH, Galgiani JN, Jarvis WR (1996) The role of understaffing in central venous catheter-associated bloodstream infections. Infect Control Hosp Epidemiol 17: 150–158
9. Torres A, Aznar R, Gatell JM, Jiménez P, González J, Ferrer A, et al (1990) Incidence, risk, and prognosis factors of nosocomial pneumonia in mechanically ventilated patients. Am Rev Respir Dis 142: 523–528
10. Myers MH, Hankey BF, Mantel N (1973) A logistic-exponential model for use with response-time data involving regression variables. Biometrics 29: 257–269
11. Elandt-Johnson RC (1980) Time dependent logistic models in follow-up studies and clinical trials. I. Binary data. Institute of Statistic mimeo series no 1309. University of North Carolina, Chapel Hill
12. Friedman SR, Des Jarlais DC (1993) The harm reduction approach and risk factors for human immunodeficiency virus (HIV) seroconversion in injecting drug users, Amsterdam. Am J Epidemiol 138: 768–769
13. Fernández-Crehuet R, Díaz-Molina C, De Irala J, Martínez-Concha D, Salcedo-Leal I, Masa-Calles J (1997) Nosocomial infection in an intensive care unit: identification of risk factors. Infect Control Hosp Epidemiol 18: 825–830
14. Cox DR (1988) Regression models and life tables. J R Stat Soc B 34: 187–220
15. Rozman C, Montserrat E, Rozman M, Bladé J, Cervantes F (1989) Análisis de la respuesta al tratamiento. Consideraciones sobre las variables pronósticas evolutivas (tiempo-dependientes). Med Clin (Barc) 92: 52–55

16. Mantel N (1982) An uncontrolled clinical trial-treatment response or spontaneous improvement? Controlled Clin Trials 3: 369–370
17. Weiss GB, Bunce III H, Hokanson JA (1983) Comparing survival of responders and non-responders after treatment: a potential source of confusion in interpreting cancer clinical trials. Controlled Clin Trials 4: 43–52
18. Anderson JR, Cain KC, Gelber RD (1983) Analysis of survival by tumor response. J Clin Oncol 1: 710–719
19. Green MS, Symons MJ (1983) A comparison of the logistic risk function and the proportional hazards model in prospective epidemiological studies. J Chronic Dis 36: 715–723
20. National Nosocomial Infections Surveillance (1995) Semiannual report, May 1995. Am J Infect Control 23: 377–385
21. Jarvis WR, Edwards JR, Culver DH, Hughes JM, Horan T, Emori TG, et al (1991) Nosocomial infection rates in adult and pediatric Intensive Care Units in the United States. Am J Med 91 (supl 3B):185S–91S
22. Joshi N, Russell Localio A, Hamory BH (1992) A predictive risk index for nosocomial pneumonia in the Intensive Care Unit. Am J Med 93: 135–142
23. Garner JS, Jarvis WR, Emori TG, Horan TC, Hughes JM (1988) CDC definitions for nosocomial infections 1988. Am J Infect Control 16: 128–140
24. Knaus WA, Wagner DP, Draper EA, Zimmerman JE (1985) APACHE II: a severity of disease classification system. Crit Care Med 13: 818–829
25. Instituto Nacional de la Salud (1988) Clasificación Internacional de Enfermedades 9ª Revisión, Modificación Clínica, vol 1. Ministerio de Sanidad y Consumo. INSALUD, Madrid
26. Lee ET (1992) Statistical methods for survival data analysis, 2nd edn. Wiley, New York
27. Hosmer DW, Lemeshow SA (1989) Applied logistic regression. Wiley, New York
28. Anonymous (1995) Stata statistical software: release 4.0. Stata Corporation, College Station
29. García Martín M, Fernández-Crehuet Navajas M, Gálvez Vargas R (1993) Sistemas de vigilancia de la infección nosocomial. In: Gálvez Vargas R, Delgado Rodríguez M, Guillén Solvas JF (eds) Infección hospitalaria. Universidad de Granada, Granada, pp 299–311
30. Glenister H, Taylor L, Bartlett C, Cooke M, Sedgwick J, Leigh T (1991) An assessment of selective surveillance methods for detecting hospital acquired infections. Am J Med 91 [Suppl 3b]:121–124
31. Singh-Naz N, Sprague BM, Patel KM, Pollack MM (1996) Risk factors for nosocomial infection in critically ill children: a prospective cohort study. Crit Care Med 24: 875–878
32. Constantini M, Donisi PM, Turrin MG, Diana L (1987) Hospital acquired infections surveillance and control in intensive care services. Results of an incidence study. Eur J Epidemiol 3: 347–355
33. Daschner FD, Frey P, Wolff G, Baumann PC, Suter P (1982) Nosocomial infections in intensive care wards: a multicenter prospective study. Intensive Care Med 8: 5–9
34. Nystrom B, Fredereci H, Von Euler C (1988) Bacterial colonization and infection in an intensive care unit. Intensive Care Med 14: 34–48
35. Haley RW, Schaberg KB, Crossley DR, Von Allmen SD, McGowan JE Jr (1981) Extra charges and prolongation of stay attributable to nosocomial infections: a prospective interhospital comparison. Am J Med 70: 51–58
36. Peña C, Pujol M, Pallarés R, Corbella X, Vidal T, Tortras N, Ariza J, Gudiol F (1996) Estimación del coste atribuible a la infección nosocomial: Prolongación de la estancia hospitalaria y cálculo de costes alternativos. Med Clin (Barc) 106: 441–444
37. Jarvis WR, Cookson ST, Robles MB (1996) Prevention of nosocomial bloodstream infections: a national and international priority. Infect Control Hosp Epidemiol 17: 272–275