PLISKA

STUDIA MATHEMATICA
BULGARICA

ПЛИСКА

БЪЛГАРСКИ
МАТЕМАТИЧЕСКИ
СТУДИИ

Pliska Stud. Math. Bulgar. **17** (2005), 221–228

PLISKA
STUDIA MATHEMATICA
BULGARICA

LOGISTIC REGRESSION IN MODELLING DATA FOR CVC – RELATED INFECTION

Krasimira Prodanova, Ionko Stoinov, Dimitar Terziiski

A prospective study of all new central venous catheters (CVC) inserted for patients in intensive care unit in order to identify risk factors for CVC infection and to determine the rate of CVC related infection is undertaken. A catheter-related infection and sepsis was suspected in 62 cases of 118 CVC inserted in intensive care patients. A multiple logistic regression to obtain adjusted estimate of odds ratios and to identify which factors were associated independently with CVC related infection was performed. The variables which entered in the model were those found to be statistically significant ($\alpha \leq 0.5$) on univariate analysis and those which were established risk factors from previous research reports. The dependent variable was the CVC related infection. The independent variables were ten: age, sex, insertion site, number of lumens, duration of catheterization etc. The software package STATISTICA 6.0 was used for analyzing the real data.

1. Introduction

Central Venous Catheters (CVCs) are widely used in critically ill patients in the world. They permit hemodynamic monitoring and allow access for the administration of fluids, blood products, medications, and total parental nutration. Estimates of their use in the United States alone suggest that over five millions CVCs are inserted annually. Although CVCs have significant benefits in many

2000 *Mathematics Subject Classification:* 62J12,62P10
Key words: Central Venious Catheter (CVC) infection, multinomial logistic regression

clinical situations, the increase in their use over the last several years has been associated with at last a doubling of resultant nosocomial infections [1]. A number of factors may contribute to the risk of catheter related infections and sepsis in intensive care patients [2], [3], [4]. Critically ill patients who develop bloodstream infections and sepsis are at a greater risk of death than patients with comparable severity of illness without this complication. Although a common problem, the descriptive epidemiology, pathophysiology, risk factors and best means of diagnosing catheter related infections have not yet been fully elucidated [5].

We undertook a prospective study of all new CVCs inserted into patients in the Department of Anaestheology and Intensive Care Medisine of Military Medical Academy, Sofia, in order to identify the risk factors of CVCs related infections/sepsis and lethal outcome.

A multiple logistic regression is performed to obtain an adjusted estimate of the odds ratios and to identify which factors were associated with these outcomes.

2. Measurements

During the study period (over a two-years) 118 CVCs in 113 patients (female and male) are studied. All 118 catheters are inserted in intensive care units (12 bed) by anaesthetists under strict aseptic techniques. For the insertions could occur two cases: urgent and by plan. Two types of catheters are used - "Selinger" and "Cavafix" with 1 or 2 lumens. Data obtained for each catheter included the patients' APACE (Aqute Phistology Chronic Health Evaluations) scores and primary diagnoses on admission on the first catheter day (categorized as non surgical, trauma, stomac surgical and neuro surgical). Data are also obtained on patients' age, gender, insertion site, clinical and laboratory data pertaining to infections, antibiotic administration, duration of catheterization.

3. Models and Methods

The multinomiale logistic regression model has been applied in a variety of context, including cohort and case-control studies, problems in differential diagnosis and prediction and in the analysis of survey data. Assumed that the categorical outcome of interest Y_j is a multinomiale variable with categories $j = 1, 2, 3$ and that x and β_j is a $(s + 1)$ - dimensional vector of covariates. The polychotomic logistic model specifies a conditional probability of the outcome given x to be of the form

$$p_j(x) = P(\text{outcome falls in} j\text{th category}/x)$$

$$= \frac{\exp(x\beta_j)}{1+\exp(x\beta_j)+\exp(x\beta_2)}, \quad j = 1, 2,$$

and

$$p_3 = 1 - p_1 - p_2,$$

where $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{js}),\ j = 1, 2$ are the regression parameters, s is the number of independent variables in the model and $x = (1, x_1, x_2, \ldots, x_s)^T$ are these independent variables.

Let us consider n observation for every one component of x. Let us denote with $p_{ij} = p_j(x_i),\ i = 1, 2, \ldots, n; j = 1, 2$ and for $Y_j(x_i))$ with $y_{ij} = 1$ when the jth category is chosen and 0 otherwise.

The estimates of the $\beta_j,\ j = 1, 2$ are obtained by maximizing the likelihood function with respect to β_j. The likelihood function and the logarithm of this function are given [6] by:

$$L = \prod_{i=1}^{n} p_{i1}^{y_{i1}} p_{i2}^{y_{i2}} p_{i3}^{y_{i3}}$$

$$\ln L = \sum_{i=1}^{n} (\sum_{j=1}^{2} y_{ij}(x_j\beta_j) - \sum_{j=1}^{3} y_{ij} \ln(1 + \sum_{j=1}^{2} \exp(x_i\beta_j))),$$

where $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{js})$ and $x_i = (1, x_{i1}, \ldots, x_{is})$. The maximum likelihood estimator for the regression parameters $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{js})$ is obtained from the system:

$$\frac{\partial lnL}{\partial \beta_{jk}} = \sum_{i=1}^{n} (y_{ij} - p_{ij}) x_{ik} = 0, \quad k = 0, \ldots, s$$

Newton-Raphson procedure is used to approximate the maximum likelihood estimator $\hat{\beta}$.

Since the statistical software usually allows only two outcomes for the dependent variable (binary case),we follow the suggestion given in [6] and [7]: When the outcomes are three an approximate solution can be obtained by fitting two separate binary (dichotomous) logistic regressions.The first should exclude all respondents in the unrestricted category, and the second should exclude all respondent in the prohibited category.

The dichotomous logistic regression model assumes that the logit function (logarithm of odds ratio)

$$\ln \frac{p}{1-p},$$

where p is the probability that the outcome belongs to the group coded with "1", can be modeled as a linear function of a set of s explanatory variables.

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \ldots, \beta_s x_s$$

The likelihood function has the form

$$L = \prod_{i=0}^{n} p^{y_i} (1-p)^{1-y_i},$$

where $y_i = 0$ or 1 depending on in which group the i-th observation belongs.For given random sample of n observations the likelihood estimator $\hat{\beta}$ can be obtained.

Remark: In comparison to the multiple linear regression model, the coefficient vector $\hat{\beta}$ in this case must be iterpreted differently. A marginal one unit increase of $x_k,\ k = 1, 2, \ldots, s$ brings about an increase in $\ln \frac{p}{1-p}$ of the amount $\hat{\beta}_k$.

Let us consider two binary logistic models.

The first logit compares category $j = 1$ to the remaining categories using $\ln \frac{p_1}{p_2+p_3}$.Thus the first partition consists of the first category in the first set and the remaining categories in the second set.

The second logit then compares the category $j = 2$ to the remaining category from the second set of the first partition. Thus the second logit is given by $\ln \frac{p(2)}{p(3)}$. The diagram below illustrates the pattern for a total three categories:

Logit 1: 1 versus [2,3].

Logit 2: 2 versus 3.

It is important to note that only the first model equation uses the entire data set for estimation. The second equation uses only the observations corresponding to categories 2 and 3, and therefore, the probabilities in the second logit function are more correctly represented by the conditional probabilities $p(2) = p(2/[2, 3])$ and $p(3) = p(3/[2, 3])$. Having determined $p([2, 3])$ from the first binary logistic model, the second model can be used to determine p_2 and p_3 and hence $p_2 = p([2, 3])p(2)$ and $p_3 = p([2, 3])p(2)$.

Tests concerning statistically significant parameters of the regression can be conducted using Wald test with test statistic:

$$\beta^T [Var(\beta)]^{-1} \beta,$$

where $Var(\beta)$ is the estimated covariance submatrix for the corresponding parameters. If the null hypothesis is that the q - dimensional vector $\beta = 0$, then this statistic has approximately χ^2 distribution with q degrees of freedom.

4. Results and Discussions.

For our data the dependent variable Y is with three outcomes: the first - the patient has no infection or sepsis (normal condition), the second - the patient has infection and the last outcome - the patient has the sepsis. The independent variables in the model are 10. Three of them are quantitative and 7 are qualitative. The list of minimal and maximal values of the qualitative variables and the codes of the categories for the qualitative variables are presented bellow:

- The age $(x_1) : \min = 18,\ \max = 80$;
- The gender (x_2): male - "1", female - "0";
- APACHE scores $(x_3) : \min = 2,\ \max = 29$;
- The diagnostic group (x_4): non surgical - "0", trauma - "1", st.surgical - "2", neuro surgical - "3";
- The changes in the type of antibiotic during the treatment (x_5): more than two - "1", less or two - "0";
- The type of CVC (x_6): "Selinger" - "0", "Cavafix" - "1";
- The insertion site (x_7): v.Femoralis - "-1", v.Subclavia - "0", v.Jugularis - "1";
- The duration of catherization (in days)(x_8):$\min = 4,\ \max 45$;
- The number of CVC - lumens (x_9): one - "1",more than one - "0";
- Type of catetherization (x_{10}): planed - "0", urgent - "1".

Let us denote by p_1 the probability, that the patient is in normal condition, by p_2 - that he has infection and by p_3 - the probability that the patient has sepsis. As it is mentioned above one have to estimate the three probabilities $p_j,\ j = 1, 2, 3$ in two stages.

The first binary regression with outcomes "normal condition of the patient with CVC" (coded by "0") and "the patient has infection or sepsis" (coded by "1") is considered. By use of STATISTICA 6.0 and all 118 data the following dichotomous logistic regression model is obtained:

$$\ln \frac{\overline{p}}{1-\overline{p}} = -2.38 + 0.01x_1 - 0.51x_2 + 0.34x_3 + 0.54x_4 - 0.07x_5 + 0.8x_6 + 1.01x_7 + 0.03x_8 - 0.72x_9 + 0.05x_{10}.$$

Thus we estimate $p_1 = 1 - \overline{p}$.

The remaining 62 data for the patients with infection and sepsis are coded: "sepsis" by "1", "infection" by "0" and from the second binary regression we get:

$$\ln \frac{\hat{p}}{1-\hat{p}} = -2 - 0.06x_2 + 0.12x_3 + 0.15x_4 - 0.49x_5 - 0.73x_6 + 0.66x_7 +$$

$$0.08x_8 - 0.33x_9 + 0.8x_{10},$$

where the probability $\hat{p}$ is the probability that the patient from this group of 62 patients has sepsis.

Using the conditional probabilities $\hat{p}$ and $1-\hat{p}$ one can estimate: $p_3 = \overline{p}\hat{p}$ and $p_2 = \overline{p}(1-\hat{p})$.

The Wald test statistic give us that in the first binary model statistically significant parameters ($p-$ level < 0.05) are: $\beta_0(p = 0.01)$, $\beta_4(p = 0.001)$ and $\beta_7(p = 0.008)$ and in the second - only the variable x_3 - APACE-scores with $p = 0.023$. From these results and according to the Remark: in previous section we can infer that the probability for normal condition of the CVC-patient is higher if:

- The APACE-scores (x_3) tends to the minimum;
- The patient doesn't have surgical intervention (x_4);
- The insertion site (x_7) is v.Femoralis;
- If the APACE-scores increases it is more likely to have sepsis than infection.

An additional dichotomous logistic regression model was considered in which dependent variable has two categories:the patient has survived or not (lethal outcome)was considered. The independent variables are 11: the ten mentioned above plus the independent variable of the first binary model - x_{11} - normal condition with code "0" and infection or sepsis with code "1". Based on all 118 data the following dichotomous logistic regression model is obtained:

$$\ln \frac{p}{1-p} = -6.12 + 0.01x_1 - 1.84x_2 + 0.28x_3 + 0.37x_4 - 0.85x_5 - 0.54x_6 + 0.05x_7 +$$

$$0.07x_8 + 1.552x_9 - 1.09x_{10} + 0.58x_{11}.$$

The significant variables we obtain are x_2 - gender ($p = 0.003$), x_3 - APACHE($p = 0.000$), x_8 - the duration of catheterization ($p = 0.049$) and x_9 - the number of CVC's lumens ($p = 0.042$). From these results it follows: if the patient is male the lethal outcome is less likely;the probability for the lethal outcome decreases with decrease in APACHE scores, number of lumens and duration of catheterization.

By use of logistic regression the factors, which were associated with infections, sepsis and lethal outcome for the patients with central venous catheters (CVC) are successfully identify.

REFERENCES

[1] Atanasov D., K. Prodanova, D. Terziiski Risk Factors of CVC Related Infections in Intensive Care Patients, *Pliska Stud. Math Bulgar.* **14** (2003) 15–22.

[2] Jobson, J.D. *Applied Multivariate Data Analysis* Vol.**II**, Springer Verlag,1991.

[3] McKinly S. et al. Incidence and Prediction of CVC Related Infections in Intensive Care Patients, *Anaesth Intensive Care* **27** (1999) 164–169.

[4] O'Grady N.P. et al. Guidelines for the prevention of intravascular catheter-related infections. Catheter for Disease Control and Prevention, *MMWR Recomm Rep.* **51** (2002) 1–29.

[5] Oncu S. et al. CVC Related Infections: Risk Factors and the Effect of Glycopeptide Antibiotics, *Annals of Clinical Microbiology and Antimicrobials* **2** (2003) 1–6.

[6] Tacconelli, E et al. CVC-Related Sepsis in a Cohort of 366 Hospitalised Patients, *Eur.J.Clin.Microbiol.Infect.Dis.* **16** (1997) 203–209.

[7] In: Case studies in biometry ISBN 0-471-58885-7, John Wiley & Sons,Inc., 1994, 249–271.

Krassimira Prodanova
Technical University of Sofia
Faculty of Applied Mathematics
and Informatics
e-mail: kprod@tu-sofia.bg

Ionko Stoinov
Technical University of Sofia
Faculty of Applied Mathematics
and Informatics

Dimitar Terziiski
Military Medical Academy,
Department of Anaesthesiology
and Intensive Care Medicine,
Sofia