journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Development of a daily mortality probability prediction model from Intensive Care Unit patients using a discrete-time event history analysis

Ying Che Huang[a], Kuang Yi Chang[b], Shih Pin Lin[b], Kung Chen[c], Kwok Hon Chan[d], Polun Chang[e,*]

[a] Department of Anesthesiology and Critical Care, Taipei Veteran General Hospital, Institute of Biomedical Informatics, National Yang-Ming University, Taiwan
[b] Department of Anesthesiology and Critical Care, Taipei Veteran General Hospital, Division of Biostatistics/Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taiwan
[c] Department of Computer Science, National Chengchi University, Taiwan
[d] Department of Anesthesiology and Critical Care, Taipei Veteran General Hospital, Taiwan
[e] Institute of Biomedical Informatics, National Yang-Ming University, Taiwan

ARTICLE INFO

Article history:
Received 23 December 2012
Received in revised form
20 March 2013
Accepted 31 March 2013

Keywords:
Mortality probability
Prediction model
Blood cell count
Discrete time event history analysis
Intensive Care Unit
Advisory system

ABSTRACT

As studies have pointed out, severity scores are imperfect at predicting individual clinical chance of survival. The clinical condition and pathophysiological status of these patients in the Intensive Care Unit might differ from or be more complicated than most predictive models account for. In addition, as the pathophysiological status changes over time, the likelihood of survival day by day will vary. Actually, it would decrease over time and a single prediction value cannot address this truth. Clearly, alternative models and refinements are warranted. In this study, we used discrete-time-event models with the changes of clinical variables, including blood cell counts, to predict daily probability of mortality in individual patients from day 3 to day 28 post Intensive Care Unit admission. Both models we built exhibited good discrimination in the training (overall area under ROC curve: 0.80 and 0.79, respectively) and validation cohorts (overall area under ROC curve: 0.78 and 0.76, respectively) to predict daily ICU mortality. The paper describes the methodology, the development process and the content of the models, and discusses the possibility of them to serve as the foundation of a new bedside advisory or alarm system.

1. Introduction

The science of making prognostic estimates for critically ill patients has evolved rapidly since 1985, when the Acute Physiology and Chronic Health Evaluation II (APACHE II) was published [21,48]. Since then, various scoring systems for severity of illness or organ failure and their derivatives have been developed, including the Mortality Probability Model (MPM) score, the Multiple Organ Dysfunction Score (MODS) and the Sequential Organ Failure Assessment score [27,31,45].

The APACHE and MPM scoring systems are global measures of illness severity and outcome prediction; however, the Multiple Organ Dysfunction and Sequential Organ Failure

* *Corresponding author at*: Rm 520, Library Information Building, Institute of Biomedical Informatics, No. 155, Li-Nong St., Sec. 2, Taipei 11221, Taiwan. Tel.: +886 2 28267238; fax: +886 2 28202508.
E-mail addresses: ycsidney@gmail.com (Y.C. Huang), polun@ym.edu.tw, polunchang@gmail.com (P. Chang).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.03.018

Assessment scores were initially designed to describe organ dysfunction or failure in critically ill patients.

2. Background

As various articles [5,12,22,30,46,49,51] have pointed out, severity scores are imperfect at predicting individual clinical chance of survival. Several reasons may explain this deficit [12,33,46,49]. First, the APACHE and MPM scoring systems, in particular, are population-based tools, not designed for use in individual patients or for specific intervention. Second, most scores obtain assessments at a single point, usually upon admission to the Intensive Care Unit (ICU) or within 24 h after admission, although the MPM included scores for 48 and 72 h measurements, respectively [25].

Consequently, many of these assessments do not take into account an important dimension of ICU care: time [33,46,49]. After a period of intensive care, some patients will either die or be discharged on day 1 or 2 post ICU admission [43]. As the ICU stay was less than 3 days for 15–50% of patients in these data sets [18,23,25,36,43,49], most patients (50–85%), after initial intensive treatment, still struggle with their condition in the ICU after day 3. The clinical condition and pathophysiological status of these patients might differ from or be more complicated [25,43] than most predictive models account for; consequently, the treatment guidelines which depend on these models will be less effective [1,7,32]. In addition, as the pathophysiological status changes over time, the likelihood of survival day by day will vary. Actually, it would decrease over time [33] and a single prediction value cannot address this truth. Clearly, alternative models and refinements are warranted [12,46,49].

An alternative model would be based on daily severity score check or daily changes in score values, although it would also have limitations [43]. An example of such a model is the Riyadh ICU Program system developed by Chang et al. [9,10,24]. Timsit et al. [43] were also able to predict hospital mortality in ICU patients hospitalized for more than 72 h, although their model failed to show daily probability of mortality.

Another alternative would be to combine either biomarkers or genetic profiling with other measures (i.e., the measures included in some acute physiology score) to show pathophysiological changes over time [46]. Limitations to such a model include scientific, economic, and organizational barriers [15,44].

3. Design considerations

In this study, we considered that using robust statistical method to build the model used complete blood count and change in its variables to predict daily probability of mortality in individual patients from day 3 to day 28 post ICU admission.

The process should be simple, objective, easy to collect, inexpensive, can be daily checked and provides information about pathophysiological changes over time and may serve as the foundation of a new bedside advisory and alarm system.

To fulfill the purpose, we built two models and then do the comparison to APACHE II score. The first model was a combined model. It used complete blood count and changes in its variables and incorporate with APACHE II score to cover the baseline data of the ICU patients. The second model used complete blood count and change in its variables alone. It was developed to address some important issues [12,13,46,49,51]. First, it could be checked daily and had a robust statistical model. In addition, the burden of data collection and cost were minimal. The complete blood count is an inexpensive and reliable test, often checked daily and easy to collect in the ICU. Second, the more subjective the variables (i.e., calculating the Glasgow Coma Scale in the APACHE II score) in the model, the more dependency on the observer. Several studies have shown that different physicians calculate different APACHE II values from the same data, particularly in patients receiving sedation [8,11,17,29,40]. In the second model, all variables are objective; hence, there is no risk of inter-observer variability. Third, most severity scores were not developed to be used for decisions concerning individuals. One alternative refined model would include daily severity score or daily changes in score values [9,10,24]. The other alternative would combine one or more biomarkers or genetic profiling with other measures as mentioned above. We would like to integrate both features when the model has been developed. Using the absolute complete blood count value and derived changes in variables would establish the relative power of alternative methods. In the second model, the direct daily complete blood count rather than the transformed score would be used.

After two models were developed, we carried out the comparison. If the performance of the second model is not inferior to the first one and APACHE II score then we can use the probability of mortality calculated by the second model to form the kernel of the bedside advisory or alarm system. The reason is that the entire second model variables are objective and the data can be collected and calculated automatically by the bedside clinical information system.

4. Description of method

4.1. *Setting*

This retrospective study was performed at a 42 bed adult mixed medical–surgical Intensive Care Unit of a 2700 bed teaching medical center in Taiwan. The Institutional Review Board of the medical center approved the study protocol and waived the requirement for written informed consent from participating subjects or their legal representatives.

4.2. *Study population*

All patients admitted to the ICU from January 2006 to December 2008 were screened. We consecutively allocated first 3/4 of the study patients to the training cohort for the model development and the last 1/4 to the validation cohort. Patients younger than 18 years, readmission to ICU and those whose ICU stay was shorter than 3 days were excluded.

4.3. Data collection

Data were retrieved from electronic medical records stored in a hospital central data warehouse; the following information was collected: age and sex; APACHE II score calculated in the first 24 h in the ICU; International Classification of Disease version 9 code; ICU admission category (i.e., medical or surgical); ICU length of stay; ICU mortality; blood transfusion records; and daily complete blood count (i.e., hemoglobin [HGB], white blood cell [WBC] and platelet count measured as daily frequency). Day 1 was defined as the interval from ICU admission to 12:00 AM on the next day; all other days were calendar days from 12:00 AM to 11:59 PM.

4.4. Data retrieval and database quality

We retrieved data from the data warehouse using Structured Query Language [16]. The resulting tables were imported into Access and Excel 2007 software (Microsoft Inc., Redmond, WA), then both the time-independent and time-dependent variables were calculated. Data quality was controlled by having a biostatistician from a non-participating department check a 2% random sample of the study data.

4.5. Variables settings

Variables were defined as time-dependent or time-independent. Time-dependent variables were platelet transfusion records (red cell transfusion records were not included [47]) daily HGB, WBC and platelet count and change in blood count over 2 or 3 days (expressed as Δ21, Δ32 or Δ31 and the term). Two expressions of change in blood count were developed. The first is the ratio type, i.e., the mathematical increase or decline in the ratio of the blood count over a 1-day or 2-day interval. For instance, with three consecutive daily platelet counts, P1, P2, and P3, we could calculate two 1-day interval change ratios, [i.e., (P2 – P1)/P1 and (P3 – P2)/P2] and present those as the variables Δ21PLATE and Δ32PLATE. Similarly, we could also calculate the 2-day interval (P3 – P1)/P1 and present that as the variable Δ31PLATE. The second type of change in blood count is the variation type, which refers to the calculation of the "pulse pressure variation" [19,38]. For example, we can define var21PLATE by calculating (P2 – P1)/[(P2 + P1)/2]. Using the same method, we can define var32PLATE and var31PLATE. These two expressions of change in blood count can be applied to WBC, HGB and platelets.

The time-independent variables were all other variables collected.

All the time-dependent and time-independent variables were analyzed using statistical models.

4.6. Statistical analysis

We compared categorical and continuous variables between survivors and non-survivors by using chi square test and Students *t*-test, respectively. Since the daily HGB, WBC and platelet count fluctuated over time, all of these and related derived changes in variables were considered time-dependent covariates in the analysis. However, patients who stayed in the ICU only 2 days would lack third day data and we would not be able to calculate the relative change in variables. Accordingly, patients who stayed in the ICU for less than 3 days were excluded in the following analysis. Since the time recorded in this study was discrete, a discrete time event history model was applied to analyses of risk of ICU mortality when ICU stay was more than 3 days [3,42,50]. The model was constructed as follows. Let P_{it} be the conditional event probability or hazard that patient i has an event at time t, given that an event has not happened yet to that person. A logistic regression used to model the relationship between P_{it} and covariates would be:

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = \alpha_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk}$$

where t is the discrete time point and α_t is the baseline hazard at time t. The details of statistical technique on discrete time event history model refer to the previous statistical literature [3,42,50]. Note that, unlike a Cox regression model, the baseline hazard must be specified in a discrete-time hazard model. To minimize the impact of outliers, the time t was truncated at day 28, as previously done [44,53].

To evaluate the net effect of daily complete blood count or its derived variables, the model was adjusted to include initial APACHE II score and receipt of platelet transfusion, however, red cell transfusion records were not included, since previous study [47] did not support the view that red cell transfusions were associated with mortality. The platelet transfusion record was also considered a time-dependent variable and coded as 1 for receiving transfusion and 0 for not. To evaluate effects of collected variables on ICU mortality, odds ratios (OR) and their 95% confidence interval (CI) were calculated. All reported p values were two-sided, and p values less than 0.05 were considered statistically significant. The potential colinearity of the variables in the final model was checked and r values less than 0.2 were considered acceptable [2].

To assess the model performance, the goodness-of-fit as evaluated by the Hosmer–Lemeshow statistic was used with higher p values (>0.05) indicating better fit. We also use the area under the curve (AUC) of the receiver-operating characteristic (ROC) curve to check the discrimination ability of the model to separate survivors and non-survivors. Comparisons of AUC of ROC were conducted using the method proposed by DeLong et al. [14]. For the validation phase of the study, values for predictive variables measured in the validation cohort were entered into the discrete time event history model using the corresponding regression coefficients estimated from the discrete time event history model analysis on the training cohort. Validation of the prediction model was established by comparing the area under the receiver operating characteristic curve of the study models in the training and validation data sets. Calibration of the validation model was also assessed using goodness-of-fit as evaluated by the Hosmer–Lemeshow statistic. All analyses were performed with SAS, version 9.1 (SAS Institute Inc., Cary, NC) and SPSS, version 17 (SPSS Inc., Chicago, IL).

5. Results

From January 2006 to December 2008, of a total of 3135 patients admitted to this mixed Intensive Care Unit, 2488 met our

Table 1 – Patient characteristics.

Variables	Training cohort ($n = 1624$; 75.3%)	Validation cohort ($n = 534$; 24.7%)	p-Value
Age (year)	70.5 (17.379)	69.98 (17.866)	0.553
Gender (male/female)			
Male	1157 (71.2%)	375 (70.2%)	0.653
ICU admission category (medical/surgical)			
Surgical	741 (45.6%)	255 (47.8%)	0.393
APACHE II	22.98 (7.586)	22.33 (7.627)	0.087
ICU length of stay	13.08 (13.330)	11.34 (9.832)	0.001
ICD 9 diagnosis main organ dysfunction at ICU admission			
Cardiovascular	80 (4.9%)	35 (6.6%)	
Respiratory	220 (13.5%)	73 (13.7%)	
Neurology	18 (1.1%)	5 (0.9%)	
Genitourinary	179 (11%)	51 (9.6%)	0.097
Digestive system	328 (20.2%)	96 (18%)	
Hematology	4 (0.2%)	6(1.1%)	
Others (neoplasm, monitoring, etc.)	795 (49%)	268 (50.2%)	

ICU, Intensive Care Unit; APACHE II, Acute Physiology and Chronic Evaluation II; ICD-9, International Classification of Disease version 9.

Table 2 – Variables associated with death of training cohort in the ICU by multivariable analysis.

	β	SE	p	OR	95% CI
WBC1	.037	.004	<0.001	1.037	1.029–1.046
PLT	−.097	.010	<0.001	.907	.889–.926
HGB	−.186	.041	<0.001	.830	.766–.899
APACHE2	.063	.009	<0.001	1.066	1.047–1.084
var32WBC	.317	.155	.040	1.373	1.014–1.859
var21PLATE	−.391	.128	.002	.677	.527–.869
var32PLATE	−.339	.122	.006	.712	.560–.905
Multivariable analysis without APACHE II					
WBC1	.035	.004	<0.001	1.036	1.028–1.045
PLT	−.087	.009	<0.001	.917	.900–.933
HGB	−.159	.038	<0.001	.853	.793–.918
var32WBC	.396	.150	.008	1.486	1.107–1.995
var21PLATE	−.494	.125	<0.001	.610	.477–.780
var32PLATE	−.386	.121	.001	.680	.537–.861

ICU, Intensive Care Unit; β, estimated regression coefficient; SE, standard error of β; OR, odds ratio, which is equal to e^{β}; CI, confidence interval; APACHE II, Acute Physiology and Chronic Evaluation II; WBC, white blood cell count divided by 1000; PLT, platelet count divided by 10,000; HGB, hemoglobin; var21PLATE, var32PLATE and var31PLATE: expressions of change in blood count over 2 or 3 days applied to platelets (variation type are expressed as var21, var32 or var31 and the term).

inclusion criteria. Of these, 2158 patients with no missing data and with a total 27,211 patient-days formed the database for this study. Among these 2158 patients, 1624 patients with 20,409 patient-days data were consecutively enrolled from January 2006 to December 2007 for the training cohort of the prognostic model and 534 patients with 6802 patient-days data were consecutively enrolled from December 2007 to December 2008 to form the model's validation cohort. Patient demographics for the training and validation cohort are presented in Table 1. The groups were mostly similar, but patients in the validation cohort had shorter ICU stay.

To reduce the complexity of the model and take into account the risk of co-linearity, the variables associated with ICU death at p value less than 0.001 instead of 0.05 by the univariate analysis were selected and introduced in the multivariate model to improve model deviance [6,37]. The results are presented in Table 2.

The effects of APACHE II score, daily HGB, WBC and platelet count on ICU mortality remained significant in the multivariate analysis. When derived changes in variables were checked, only var32WBC, var21PLATE and var32PLATE remained significantly associated with ICU mortality. The final first combined model of mortality probability prediction by complete blood count variability (BCV) with APACHE II score is presented below:

$$P=\frac{1}{1+\exp[-(\alpha(t)-0.063\text{APACHE II}-0.037(\text{WBC}/1000)+0.097(\text{PLATE}/10,000)+0.186\text{HGB}-0.317\text{var32WBC}+0.391\text{var21PLATE}+0.339\text{var32PLATE})]},$$

where "P" is the probability of ICU mortality on day (t) and $\alpha(t)$ represents the baseline hazard on day (t), which is specified in Table 3. Another competitive model without the APACHE II

Table 3 – Baseline hazard estimates from day 3 to day 28 for the training cohort.

	β	SE	p
Day(3)	.157	.611	.797
Day(4)	−.141	.621	.821
Day(5)	−.066	.624	.916
Day(6)	−.180	.635	.777
Day(7)	.055	.630	.930
Day(8)	.279	.627	.657
Day(9)	.007	.642	.991
Day(10)	.556	.627	.376
Day(11)	.194	.648	.765
Day(12)	.295	.649	.649
Day(13)	.726	.634	.252
Day(14)	−.104	.690	.880
Day(15)	.330	.666	.620
Day(16)	.012	.703	.986
Day(17)	.341	.680	.617
Day(18)	−.226	.744	.761
Day(19)	−.310	.776	.690
Day(20)	−.288	.777	.711
Day(21)	−.235	.778	.763
Day(22)	.062	.745	.934
Day(23)	.306	.722	.672
Day(24)	−.275	.830	.741
Day(25)	.513	.722	.478
Day(26)	−.535	.926	.563
Day(27)	−.007	.831	.993
Day(28)[a]	−3.311	.457	.000

β, estimated regression coefficient; SE, standard error of β.
[a] Day(28) is set as the reference group for the rest of ICU days.

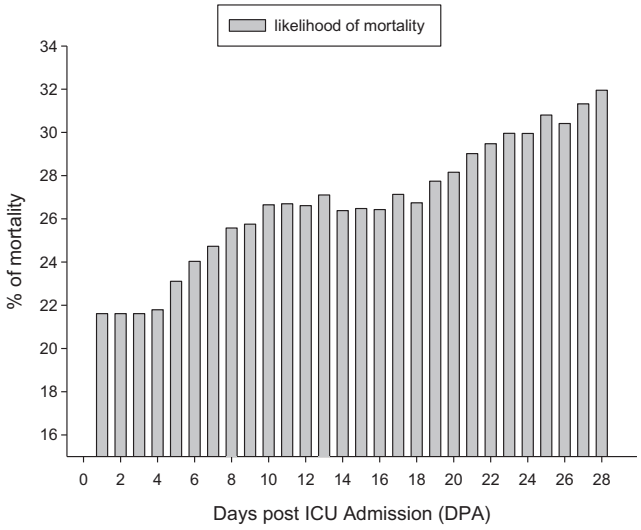

Fig. 1 – Likelihood of mortality by days post ICU admission (DPA). Length of stay of all patients was more than 3 days. The x-axis of figure begins from DPA 3, as patients who died on DPA 1 and DPA 2 were excluded.

score, the pure BCV model, was modified from the previous model as follows:

$$P = \frac{1}{1 + \exp[-(\alpha'(t) - 0.035(\mathrm{WBC}/1000) + 0.087(\mathrm{PLATE}/10,000) + 0.159\mathrm{HGB} - 0.396\mathrm{var32WBC} + 0.494\mathrm{var21PLATE} + 0.386\mathrm{var32PLATE})]}$$

In terms of calibration and discrimination, we first validated the BCV with APACHE II model to predict the ICU mortality. The final model exhibited good calibration (Hosmer–Lemeshow chi-square, 6.163; $p = 0.629$) and good discrimination (AUC-ROC curve, 0.804, 95% confidence interval [CI] [0.780, 0.828]). For comparison, the pure BCV model also showed good calibration (Hosmer–Lemeshow [HL] chi-square, 8.713; $p = 0.367$) but less significant discrimination (AUC-ROC curve, 0.787, 95% CI [0.761, 0.813]). Fig. 2A shows the ROC curves of the BCV incorporated with the APACHE II model and the pure BCV model compared to the APACHE II score, in which the AUC was 0.682 (95% CI [0.65, 0.714]) in our training cohort.

For internal validation, we used values of predictive variables measured in patients from the validation cohort. The same two models and the APACHE II score had AUC-ROC curves showed in Fig. 2B. Performance was slightly lower in the validation cohort but still exhibited good calibration and discrimination in the BCV with APACHE II score model (Hosmer–Lemeshow chi-square, 5.391; $p = 0.715$, AUC-ROC curve, 0.778, 95% CI [0.729, 0.826]) and the pure BCV model (Hosmer–Lemeshow chi-square, 11.133; $p = 0.194$, AUC-ROC curve, 0.758, 95% CI [0.706, 0.811]). The AUC of APACHE II score was still low in validation cohort (AUC-ROC curve, 0.671, 95% CI [0.618, 0.723]).

Moreover, we grouped the 20,409 patient-days in the training cohort by their days post ICU admission (DPA) from the third to the 28th day (DPA 3 to DPA 28; note that we use DPA instead of length of stay). Data for DPA after day 28 were truncated. The AUC_ROC (for mortality prediction) of two models (the BCV with APACHE II, the BCV alone) and the APACHE II score were computed by DPA and the results compared (Fig. 3). The APACHE II score had significantly lower AUC value from DPA 3 to DPA 28 and the lowest one was on DPA 19 (0.531, 95% CI [0.320, 0.741]).

From August 2009, our ICU built up the bedside clinical information system using the software contains HL7 messaging information for the Philips® Intellivue Clinical Information Portfolio (hereinafter called the ICIP system, or ICIP). ICIP was designed using standard Microsoft® tools. This product was formerly known as CareVue Chart. We incorporated the model formula as mentioned above into ICIP and combined the CBC data-triggered calculation rules by using configuration tool of ICIP. The pilot advisory or alarm system (we named it as Critical Stage Index, or CSI) was developed and the bedside monitor snap shot showed as Fig. 4.

6. Lessons learned

Of all patients admitted to the ICU during the whole study period, 20.6% (647/3135) were discharged from the ICU or died on DPA 1 and DPA 2. This means that as many as 80% of patients remained in the ICU three or more days, an amount compatible with previous ICU studies [18,23,25,36,43,49]. For this group, the likelihood of mortality varied and kept increasing from 21.6% to 31.9% over time (Fig. 1), a finding similar to

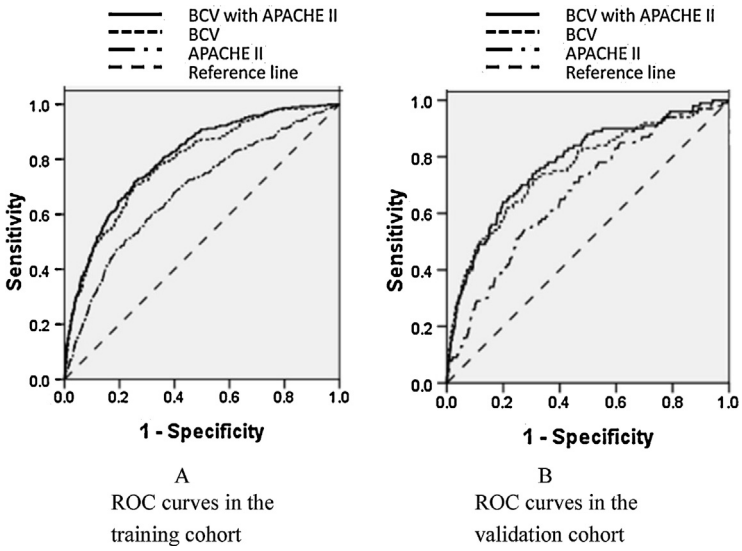

Fig. 2 – ROC curves of BCV with APACHE II model and pure BCV model compared to APACHE II severity score in the training and validation cohort. (A) AUC-ROC in the training cohort are 0.804 (95% CI [0.780, 0.828]), 0.787 (95% CI [0.761, 0.813]) and 0.682 (95% CI [0.65, 0.714]), respectively. (B) AUC-ROC in the validation cohort are 0.778 (95% CI [0.729, 0.826]), 0.758 (95% CI [0.706, 0.811]) and 0.671 (95% CI [0.618, 0.723]), respectively. APACHE II, Acute Physiology and Chronic Evaluation II; BCV, complete blood count variability; AUC-ROC, area under the curve of the receiver-operating characteristic (ROC) curve.

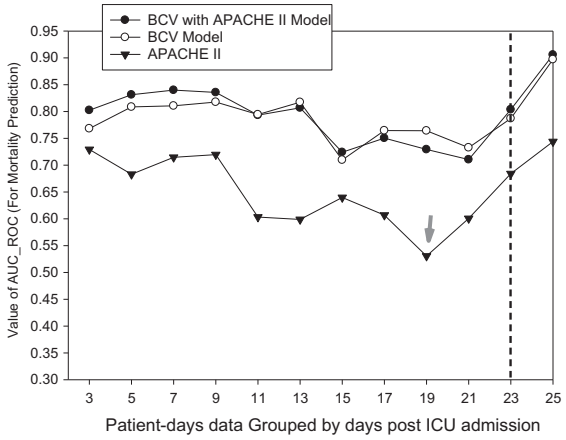

Fig. 3 – AUC_ROC (for mortality prediction) of two models (BCV with APACHE II, BCV alone) and APACHE II according to the day post admission (DPA) to the Intensive Care Unit. This figure demonstrate that, after DPA 3 (we grouped patient-days data by DPA and computed the value of each AUC_ROC), the predictive accuracy of the APACHE II score model tended to decline until DPA 21; the lowest AUC-ROC was 0.531 ± 0.108 on DPA 19 (dark gray arrow). The larger variation and elevation after DPA 23 (short dash line) could be explained by the decreasing number of deaths as DPA increased APACHE II, Acute Physiology and Chronic Evaluation II; BCV, complete Blood Count Variability; AUC-ROC, Area Under the Curve of the Receiver-Operating Characteristic (ROC) curve. Note that *x*-axis begins from DPA 3, since the length of stay for all patients was more than 3 days, DPA 1 and DPA 2 had no mortality data.

that presented recently by Meadow et al. [33]. This observation suggests that using a one-time prediction, such as the APACHE II score, to estimate patient mortality over an entire stay in the ICU is not appropriate; a measure which takes into account daily changes in condition would likely better predict a patient's daily likelihood of mortality. However, we believe it is still appropriate to include the APACHE II score on the day of ICU admission in our first model for the following reasons. First, the variables in APACHE II score and the time-dependent variables in the BCV model were not recorded at the same time, meaning the HGB and WBC counts computed were different values. Second, the APACHE II score can be considered the baseline condition when the patient was admitted to the ICU and omitting it might decrease the information in the final model. Third, despite the statistical method used, no significant co-linearity was seen between variables in our model which incorporated the APACHE II score (all r values <0.2).

However, in Fig. 3, we demonstrate that, after DPA 3, the predictive accuracy of the APACHE II score tended to decline until DPA 21; the lowest AUC-ROC was 0.531 ± 0.108 on DPA 19 (Fig. 3, dark gray arrow). The larger variation and elevation after DPA 23 (Fig. 3, short dash line) could be explained by the decreasing number of deaths as DPA increased. On DPA 26, only two cases died and the small sample size likely increased the statistical variation. Since the APACHE II score was developed using a population-based model and developed to integrate both chronic and acute conditions within 24 h of admission, this decline illustrates the fact that the longer a patient stays in ICU, the less predictive accuracy the APACHE II score has.

As indicated in Fig. 3, we showed the ability of our two models to fit the daily prediction. Using a sound statistical method (i.e., discrete time event history model instead of common

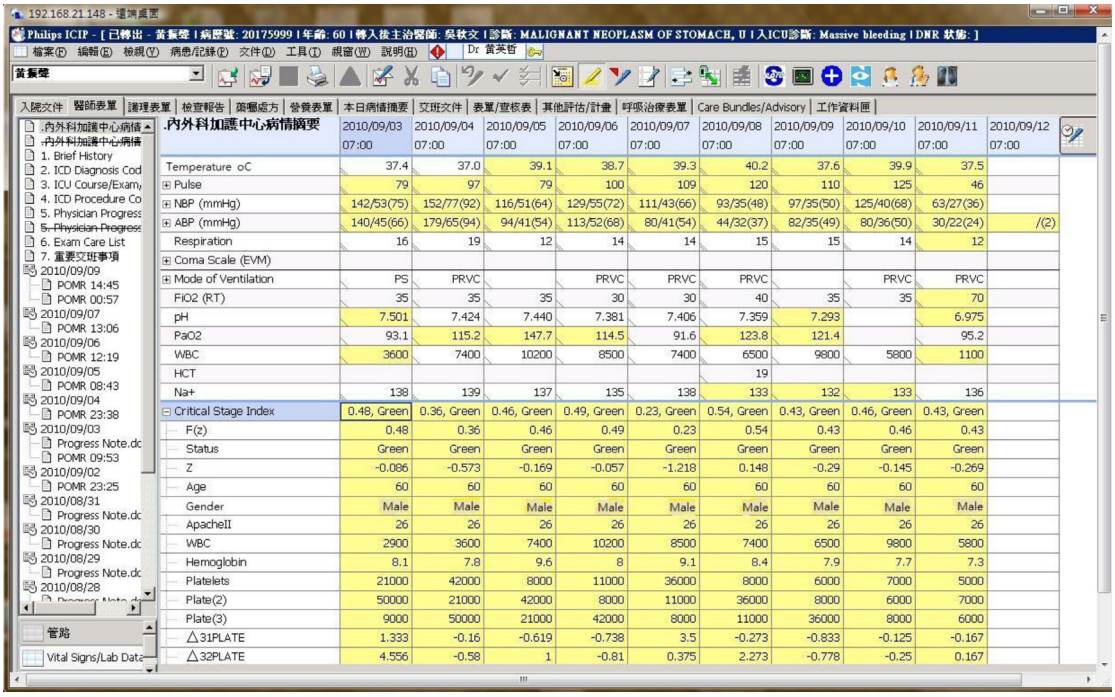

Fig. 4 – Bedside monitor snap shot showed the pilot advisory or alarm system (we named it as Critical Stage Index, or CSI, see the blue area in the middle of the figure, part of Chinese labels were translated to English ones).

logit or COX model), we used the value of β on specific DPA (Table 3) and other time-dependent variables to develop our BCV with APACHE II model and the pure BCV model. These two models allowed for daily prediction of mortality probability and we checked these predictions from DPA 3 to DPA 28. The BCV with APACHE II model and the pure BCV model had higher AUC-ROC values than APACHE II alone and showed good discrimination ability to separate survivors and non-survivors (Figs. 2 and 3).

The pure BCV model is not inferior to BCV with APACHE II model and can fulfill the purpose of our design. Nevertheless, we did not compare our models to models other than the APACHE II, such as the APACHE III/IV, Sequential Organ Failure Assessment, MPM II/III, SAPS II/III or their derivatives. For this cohort, we had only the APACHE II score for disease severity prediction. However, these other scores would face the same issues we tried to address. Since our models focuses on DPA 3 to DPA 28 mortality prediction, we suggest using the APACHE II or MPM II/III model [25,26,44] to predict DPA 1 and DPA 2 mortality if early mortality is the concern.

For the underlining mechanism which the proposed measurement increased the accuracy of the prediction, we can explain it from two perspectives. From the statistic point of view, the BCV model with miscellaneous time-dependent covariates are generated from arithmetical manipulation of CBC measures and then the discrete-time event history analysis was applied to select factors which made significant contributions to the accuracy of daily mortality prediction with or without the combination of APACHEII score for critically ill patients. With the aid of dynamic (time-dependent) information (in the BCV model which means $\alpha'(t)$, var32WBC, var21PLATE, and var32PLATE data), the accuracy of the prediction will increase. From the clinical point of view, the BCV model uses the platelet data which is not included in the APACHEII score and it was proved both a low nadir platelet count and a large fall in platelet count imply a poor outcome in adult ICU patients [20,35,41]. After the model selection using the discrete-time event history analysis, the platelet count and its derived time-dependent variables in combination did improve the predictive power of the selected model.

In conclusion, our two models use complete blood count and change in its variables to predict daily probability of mortality in individual patients from DPA 3 to 28. The process is simple, objective, easy to collect, inexpensive, can be daily checked and provides information about pathophysiological changes over time. The pure BCV model which is no dependency on the observer may serve as the foundation of a new bedside advisory or alarm system, since daily mortality probability could be collected and calculated automatically using a hospital's existing clinical information system. In the recent study [28], Lenvin et al. used the same concept and collected the naturally generated provided orders to develop the system for real-time length of stay prediction and patient flow management. However, its benefit to decision making and the subsequent improvement in mortality rate should be confirmed in a prospective study.

7. Future plans

Our study has several future plans. First, we will conduct a multi-center prospective study and do the external validation.

The original study was a retrospective design and the data were collected in a single medical center. Since the study unit is a 2700 bed teaching medical center and many of the patients had been transferred from different regional hospitals, the 2158 patients and 27,211 patient-days data collected over a 2 year period were enough to develop the model. Nevertheless, like other similar models [4,24,28], the results should be confirmed in a prospective study and such models should only be used by those familiar with their limitations. Although internal validation showed significant results, we did not evaluate the performance of models in a different patient population or institution, that is, did not do the external validation. This is one of the limitations for our study. Our measures of performance may therefore over estimate actual predictive ability when new patients, who were not included in the developmental data set, are considered. Consequently, whenever the prediction accuracy deteriorates, the BCV model should be revised and updated [34,52]. Second, we would evaluate the cost-effectiveness of the pilot system. Due to the complexity of calculating the discrete time event history model, this model is ideally suited to a computerized environment. Although computerized clinical decision support systems have become increasingly more common [39], the cost of setting up our clinical information system is more than one million and a half US dollars; we should carefully evaluate the cost-effectiveness in the near future.

Conflict of interest

The authors declare that they have no financial or ethical conflicts of interest in publishing this study.

Funding

The authors declare that they have no funding received for the work from any organizations.

Acknowledgements

We acknowledge the data transformation and quality checking by Bo Wei Lin, Department of Computer Science, National Chengchi University, Taipei, Taiwan/ROC and Chien Kun Ting, Department of Anaesthesiology and Critical Care, Taipei Veteran General Hospital, Taipei, Taiwan/ROC.

Appendix.

For better understanding, we provided a clinical example for demonstrating how to use the model to predict the daily mortality rate of the patient.

A 65-year-old man is admitted to the ICU because of multiple traumas caused by a traffic accident. The table below listed the CBC data after he admitted to ICU for 3 days:

	DPA 1	DPA 2	DPA 3
WBC (/μl)	6800	8700	15,400
Hemoglobin (g/dl)	7.6	9.2	9.8
Platelets (/mm^3)	126,000	85,000	57,000

And his derived parameters on DPA3 were calculated as follow:

	var32WBC	var21PLATE	var32PLATE	$\alpha'(t)$
Calculation	(15,400 − 8700)/((15,400 + 8700)/2)	(85,000 − 126,000)/((85,000 + 126,000)/2)	(57,000 − 85,000)/((57,000 + 85,000)/2)	1.905–0.157
Results	0.556	−0.389	−0.394	1.748

Our pure BCV model was listed as below:

$$P = \frac{1}{1 + \exp[-(\alpha'(t) - 0.035(\mathrm{WBC}/1000) + 0.087(\mathrm{PLATE}/10,000) + 0.159\mathrm{HGB} - 0.396\mathrm{var32WBC} + 0.494\mathrm{var21PLATE} + 0.386\mathrm{var32PLATE})]}$$

In this equation, "P" is the probability of ICU mortality on day (t) and $\alpha'(t)$ represents the baseline hazard on day (t), which is specified in Table 3. In this case, $\alpha'(t)$ is equal to 1.748(1.905 − 0.157).

The probability of ICU mortality on the post admission day 3 would be:

$$P = \frac{1}{1 + \exp[-(1.748 - 0.035(15,400/1000) + 0.087(57,000/10,000) + 0.159(9.8) - 0.396(0.556) + 0.494(-0.389) + 0.386(-0.394))]} = 0.063$$

On the day 3 post ICU admission, the mortality probability of this patient would be 6.3%.

REFERENCES

[1] E. Abraham, P.F. Laterre, R. Garg, H. Levy, D. Talwar, B.L. Trzaskoma, et al., Drotrecogin alfa (activated) for adults with severe sepsis and a low risk of death, New England Journal of Medicine 353 (2005) 1332–1341.

[2] C. Adrie, A. Francais, A. Alvarez-Gonzalez, R. Mounier, E. Azoulay, J.R. Zahar, et al., Model for predicting short-term mortality of severe sepsis, Critical Care 13 (2009) R72.

[3] P.D. Allison, Event History Analysis: Regression for Longitudinal Event Data, Sage Publications, Newbury Park, CA, 1984.

[4] M.W. Angstwurm, C.E. Dempfle, M. Spannagl, New disseminated intravascular coagulation score: a useful tool to predict mortality in comparison with Acute Physiology and Chronic Health Evaluation II and Logistic Organ Dysfunction scores, Critical Care Medicine 34 (2006) 314–320 (quiz 28).

[5] R. Barrera, S. Nygard, H. Sogoloff, J. Groeger, R. Wilson, Accuracy of predictions of survival at admission to the intensive care unit, Journal of Critical Care 16 (2001) 32–35.

[6] P. Bauer, B. Potscher, P. Hackl, Model selection by multiple test procedures, Statistics 19 (1988) 39–44.

[7] G.R. Bernard, J.L. Vincent, P.F. Laterre, S.P. LaRosa, J.F. Dhainaut, A. Lopez-Rodriguez, et al., Efficacy and safety of recombinant human activated protein C for severe sepsis, New England Journal of Medicine 344 (2001) 699–709.

[8] F.V. Booth, M. Short, A.F. Shorr, N. Arkins, B. Bates, R.L. Qualy, et al., Application of a population-based severity scoring system to individual patients results in frequent misclassification, Critical Care 9 (2005) R522–R529.

[9] R.W. Chang, Individual outcome prediction models for intensive care units, Lancet 2 (1989) 143–146.

[10] R.W. Chang, D.J. Bihari, Outcome prediction for the individual patient in the ICU, Unfallchirurg 97 (1994) 199–204.

[11] L.M. Chen, C.M. Martin, T.L. Morrison, W.J. Sibbald, Interobserver variability in data collection of the APACHE II score in teaching and community hospitals, Critical Care Medicine 27 (1999) 1999–2004.

[12] N.H. Cohen, The real reasons not to rely on severity scores, Critical Care Medicine 38 (2010) 334–335.

[13] D.J. Cullen, B. Chernow, Predicting outcome in critically ill patients, Critical Care Medicine 22 (1994) 1345–1348.

[14] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.

[15] P.A. Deverka, T. Doksum, R.J. Carlson, Integrating molecular medicine into the US health-care system: opportunities, barriers, and policy challenges, Clinical Pharmacology and Therapeutics 82 (2007) 427–434.

[16] D. Donald, R.F.B. Chamberlin, Implementation of a structured English query language, in: SIGFIDET'74 Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control, 1974, pp. 249–264.

[17] M.R. Gill, D.G. Reiley, S.M. Green, Interrater reliability of Glasgow Coma Scale scores in the emergency department, Annals of Emergency Medicine 43 (2004) 215–223.

[18] D.R. Goldhill, A. Sumner, Outcome of intensive care patients in a group of British intensive care units, Critical Care Medicine 26 (1998) 1337–1345.

[19] S.R. Gunn, M.R. Pinsky, Implications of arterial pressure variation in patients in the intensive care unit, Current Opinion in Critical Care 7 (2001) 212–217.

[20] P. Hui, D.J. Cook, W. Lim, G.A. Fraser, D.M. Arnold, The frequency and clinical significance of thrombocytopenia complicating critical illness, Chest 139 (2011) 271–278.

[21] W.A. Knaus, E.A. Draper, D.P. Wagner, J.E. Zimmerman, APACHE II: a severity of disease classification system, Critical Care Medicine 13 (1985) 818–829.

[22] J.A. Kruse, M.C. Thill-Baharozian, R.W. Carlson, Comparison of clinical assessment with APACHE II for predicting mortality risk in patients admitted to a medical intensive care unit, Journal of American Medical Association 260 (1988) 1739–1742.

[23] J.R. Le Gall, S. Lemeshow, F. Saulnier, A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study, Journal of American Medical Association 270 (1993) 2957–2963.

[24] R. Lefering, B. Wolfrum, H. Wauer, E.A. Neugebauer, Limitations of score-based daily outcome predictions in the individual intensive care patient. An example of the RIAHDH algorithm, Inflammation Research 53 (Suppl. 2) (2004) S169–S174.

[25] S. Lemeshow, J. Klar, D. Teres, J.S. Avrunin, S.H. Gehlbach, J. Rapoport, et al., Mortality probability models for patients in the intensive care unit for 48 or 72 h: a prospective, multicenter study, Critical Care Medicine 22 (1994) 1351–1358.

[26] S. Lemeshow, J.R. Le Gall, Modeling the severity of illness of ICU patients. A systems update, Journal of American Medical Association 272 (1994) 1049–1055.

[27] S. Lemeshow, D. Teres, J. Klar, J.S. Avrunin, S.H. Gehlbach, J. Rapoport, Mortality probability models (MPM II) based on an international cohort of intensive care unit patients, Journal of American Medical Association 270 (1993) 2478–2486.

[28] S.R. Levin, E.T. Harley, J.C. Fackler, C.U. Lehmann, J.W. Custer, D. France, et al., Real-time forecasting of pediatric intensive care unit length of stay using computerized provider orders, Critical Care Medicine 40 (2012) 3058–3064.

[29] B.M. Livingston, S.J. Mackenzie, F.N. MacKirdy, J.C. Howie, Should the pre-sedation Glasgow Coma Scale value be used when calculating Acute Physiology and Chronic Health Evaluation scores for sedated patients? Scottish Intensive Care Society Audit Group, Critical Care Medicine 28 (2000) 389–394.

[30] J.P. Marcin, M.M. Pollack, K.M. Patel, U.E. Ruttimann, Decision support issues using a physiology based score, Intensive Care Medicine 24 (1998) 1299–1304.
[31] J.C. Marshall, D.J. Cook, N.V. Christou, G.R. Bernard, C.L. Sprung, W.J. Sibbald, Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome, Critical Care Medicine 23 (1995) 1638–1652.
[32] D.K. McClish, S.H. Powell, How well can physicians estimate mortality in a medical intensive care unit? Medical Decision Making 9 (1989) 125–132.
[33] W. Meadow, A. Pohlman, L. Frain, Y. Ren, J.P. Kress, W. Teuteberg, et al., Power and limitations of daily prognostications of death in the medical intensive care unit, Critical Care Medicine 39 (2011) 474–479.
[34] L. Minne, S. Eslami, N. de Keizer, E. de Jonge, S.E. de Rooij, A. Abu-Hanna, Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment, Intensive Care Medicine 38 (2012) 40–46.
[35] D. Moreau, J.F. Timsit, A. Vesin, M. Garrouste-Orgeas, A. de Lassence, J.R. Zahar, et al., Platelet count decline: an early prognostic marker in critically ill patients with prolonged ICU stays, Chest 131 (2007) 1735–1741.
[36] R. Moreno, J.L. Vincent, R. Matos, A. Mendonca, F. Cantraine, L. Thijs, et al., The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM, Intensive Care Medicine 25 (1999) 686–696.
[37] R.P. Moreno, P.G. Metnitz, E. Almeida, B. Jordan, P. Bauer, R.A. Campos, et al., SAPS 3 – from evaluation of the patient to evaluation of the intensive care unit. Part 2: development of a prognostic model for hospital mortality at ICU admission, Intensive Care Medicine 31 (2005) 1345–1355.
[38] G. Natalini, A. Rosano, M.E. Franceschetti, P. Facchetti, A. Bernardini, Variations in arterial blood pressure and photoplethysmography during mechanical ventilation, Anesthesia and Analgesia 103 (2006) 1182–1188.
[39] J.A. Osheroff, J.M. Teich, B. Middleton, E.B. Steen, A. Wright, D.E. Detmer, A roadmap for national action on clinical decision support, Journal of the American Medical Informatics Association 14 (2007) 141–145.
[40] K.H. Polderman, E.M. Jorna, A.R. Girbes, Inter-observer variability in APACHE II scoring: effect of strict guidelines and training, Intensive Care Medicine 27 (2001) 1365–1369.
[41] B. Sharma, M. Sharma, M. Majumder, W. Steier, A. Sangal, M. Kalawar, Thrombocytopenia in septic shock patients – a prospective observational study of incidence, risk factors and correlation with clinical outcome, Anaesthesia and Intensive Care 35 (2007) 874–880.
[42] J.D. Singer, J.B. Willett, It's about time: using discrete-time survival analysis to study duration and the timing of events, Journal of Educational and Behavioral Statistics 18 (1993) 155–195.
[43] J.F. Timsit, J.P. Fosse, G. Troche, A. De Lassence, C. Alberti, M. Garrouste-Orgeas, et al., Accuracy of a composite score using daily SAPS II and LOD scores for predicting hospital mortality in ICU patients hospitalized for more than 72 h, Intensive Care Medicine 27 (2001) 1012–1021.
[44] E.E. Vasilevskis, M.W. Kuzniewicz, B.A. Cason, R.K. Lane, M.L. Dean, T. Clay, et al., Mortality probability model III and simplified acute physiology score. II: assessing their value in predicting length of stay and comparison to APACHE IV, Chest 136 (2009) 89–101.
[45] J.L. Vincent, A. de Mendonca, F. Cantraine, R. Moreno, J. Takala, P.M. Suter, et al., Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. Working group on "sepsis-related problems" of the European Society of Intensive Care Medicine, Critical Care Medicine 26 (1998) 1793–1800.
[46] J.L. Vincent, S.M. Opal, J.C. Marshall, Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions, Critical Care Medicine 38 (2010) 283–287.
[47] J.L. Vincent, Y. Sakr, C. Sprung, S. Harboe, P. Damas, Are blood transfusions associated with greater mortality rates? Results of the Sepsis Occurrence in Acutely Ill Patients study, Anesthesiology 108 (2008) 31–39.
[48] D.P. Wagner, E.A. Draper, R. Abizanda Campos, P. Nikki, J.R. Le Gall, P. Loirat, et al., Initial international use of APACHE. An acute severity of disease measure, Medical Decision Making 4 (1984) 297–313.
[49] D.P. Wagner, W.A. Knaus, F.E. Harrell, J.E. Zimmerman, C. Watts, Daily prognostic estimates for critically ill adults in intensive care units: results from a prospective, multicenter, inception cohort analysis, Critical Care Medicine 22 (1994) 1359–1372.
[50] K. Yamaguchi, Event History Analysis, Sage Publications, London, 1991.
[51] J.G. Zijlstra, L.M. Dijkema, J.J. Ligtenberg, A.R. Girbes, Severity scoring abuse, Critical Care Medicine 38 (2010) 1502.
[52] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients, Critical Care Medicine 34 (2006) 1297–1310.
[53] J.E. Zimmerman, A.A. Kramer, D.S. McNair, F.M. Malila, V.L. Shaffer, Intensive care unit length of stay: benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV, Critical Care Medicine 34 (2006) 2517–2529.