Original Article

An External Independent Validation of APACHE IV in a Malaysian Intensive Care Unit

Rowena SY Wong, [1]*MSc*, Noor Azina Ismail, [1]*MSc, PhD*, Cheng Cheng Tan, [2]*MBBS, MAnaes*

Abstract

Introduction: Intensive care unit (ICU) prognostic models are predominantly used in more developed nations such as the United States, Europe and Australia. These are not that popular in Southeast Asian countries due to costs and technology considerations. The purpose of this study is to evaluate the suitability of the acute physiology and chronic health evaluation (APACHE) IV model in a single centre Malaysian ICU. Materials and Methods: A prospective study was conducted at the single centre ICU in Hospital Sultanah Aminah (HSA) Malaysia. External validation of APACHE IV involved a cohort of 916 patients who were admitted in 2009. Model performance was assessed through its calibration and discrimination abilities. A first-level customisation using logistic regression approach was also applied to improve model calibration. Results: APACHE IV exhibited good discrimination, with an area under receiver operating characteristic (ROC) curve of 0.78. However, the model's overall fit was observed to be poor, as indicated by the Hosmer-Lemeshow goodness-of-fit test ($\hat{C}$ = 113, P <0.001). Predicted in-ICU mortality rate (28.1%) was significantly higher than the actual in-ICU mortality rate (18.8%). Model calibration was improved after applying first-level customisation ($\hat{C}$ = 6.39, P = 0.78) although discrimination was not affected. Conclusion: APACHE IV is not suitable for application in HSA ICU, without further customisation. The model's lack of fit in the Malaysian study is attributed to differences in the baseline characteristics between HSA ICU and APACHE IV datasets. Other possible factors could be due to differences in clinical practice, quality and services of health care systems between Malaysia and the United States.

Ann Acad Med Singapore 2015;44:127-32

Key words: Mortality, Prognostic models, Severity of illness

Introduction

Over the past 30 years, different versions of severity of illness scoring systems and prognostic models have been developed for prediction of patient outcomes in critical care.[1] These physiological-based systems and models allow patients to be stratified according to their severity of illness and provide prediction of in-hospital mortality.[2] The concept of using intensive care unit (ICU) prognostic models is considered relatively new in Malaysia, despite these models being widely applied in most developed nations. At present, performance assessment of ICUs across the country is managed by the Malaysian Registry of Intensive Care (MRIC), through annual national audits. Assessment and rating of performance in ICUs is primarily based on computation of the simplified acute physiology score (SAPS) II[3] severity of illness scores.[4] A limitation of this assessment is that it considers only the severity of illness scoring component alone, without fully utilising the prediction component. In our opinion, the current assessment approach can be further improved through implementation of a suitable prognostic model that not only provides a severity of illness stratification system, but is also capable of predicting objective mortality risk estimates.

One of the main considerations in this study is to choose a suitable benchmark model among the various available models. We have chosen the acute physiology and chronic health evaluation (APACHE) IV[5] as the reference model in our study after much consideration. An advantage of

[1]Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, Malaysia
[2]Department of Anaesthesia and Intensive Care, Hospital Sultanah Aminah Johor Bahru, Malaysia
Address for Correspondence: Dr Noor Azina Ismail, Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, 50603 Kuala Lumpur, Malaysia.
Email: nazina@um.edu.my

this model is that it is able to provide more information on the physiological conditions of ICU patients since it has the most number of physiological variables compared to its contemporaries, i.e. SAPS 3 admission score[6,7] and the mortality probability model (MPM0-III) admission models.[8] Bearing in mind that the APACHE IV predictive equation was developed using data from multiple institutions, we are aware of the possibility that it may not be applicable in a single centre setting. However, the decision to choose APACHE IV as our reference model is also motivated by the promising potential shown in several recent external validation studies,[9,10] that suggested the possibility of APACHE IV being robust enough for application in other countries and in single centre settings.

Materials and Methods

Data

We conducted a prospective independent validation study at the single centre ICU in Hospital Sultanah Aminah (HSA). This hospital is considered one of the several larger government tertiary referral hospitals in Malaysia. The multidisciplinary ICU in HSA admits general medical, surgical and trauma patients. It is equipped with mechanical ventilation facility and has a current bed size of 16.

Data collection involved consecutive admissions to HSA ICU between 1 January 2009 and 31 December 2009. The study followed the eligibility criteria defined in APACHE IV. For this study, we excluded patients below 16 years old, with less than 4 hours stay in the ICU, transferred from another ICU, and with incomplete clinical information. We have also excluded burns, transplant and postcardiac surgical patients from the study since they are treated in separate units. As for readmission cases, we only considered data from the first admission for analysis.

All data from patients are analysed anonymously. We did not obtain informed consent from all participants since this requirement is waived by the Medical Research and Ethics Committee, Ministry of Health, Malaysia, which is also responsible for approving this study. Data collection was manually performed by HSA ICU nurses and transferred to an online web-based database by medical officers. Details of the data collected are shown in Table 1. The admission diagnoses for each patient are determined by the ICU specialist on duty and subsequently verified by the intensivist.

Statistical Analysis

We performed manual calculation of the acute physiology score (APS) for each patient using Microsoft® Excel 2007, by combining the points for the worst physiological variables over the first day of stay in the ICU. Patients with incomplete first day APS information were excluded from the analysis. Missing physiological values were assumed normal and assigned zero points.[11] We validated APACHE IV by fitting in data that were obtained from HSA ICU patients into the APACHE IV multiple logistic regression equation in order to generate individual mortality estimates.

Model accuracy was evaluated through 3 measures, i.e. the model's discrimination, calibration and standardised mortality ratio (SMR). The model's ability to discriminate between surviving and non-surviving patients was tested through area under a receiver operating characteristic (ROC) curve. Analysis of model discrimination was performed using MedCalc version 10.4, in which estimation of area under ROC was based on a non-parametric approach.[12] Calibration was used to measure the degree of correspondence between predicted and actual mortality risk. The model's overall calibration was assessed through a calibration curve and

Table 1. Types of Data Collected in HSA ICU

Type of Data	Details	Frequency of Collection
Demographic	age, gender, ethnic group	once (ICU admission)
ICU admission	date and time, source prior to admission, principal admission diagnosis, operative status, pre-ICU length of stay	once (ICU admission)
Chronic health	AIDS, hepatic failure, lymphoma, metastatic cancer, leukaemia/multiple myeloma, immunosuppression, cirrhosis, diabetes	once (ICU admission)
Physiological and laboratory variables	pulse, mean blood pressure, temperature, respiratory rate, ABGs (PaO_2, FiO_2, A-aDO_2, $PaCO_2$), hematocrit, white blood cell count, creatinine, urine output, blood urea nitrogen, sodium, albumin, bilirubin, glucose, arterial pH, GCS score, mechanical ventilation status	physiological (hourly) laboratory (approximately twice per day) neurological (between 4 to 6 times/day)
ICU discharge	vital outcome status, discharge location	once (ICU discharge)

HSA: Hospital Sultanah Aminah; ICU: Intensive care unit; ABG: Arterial blood gas; GCS: Glasgow Coma Scale; AIDS: Acquired immunodeficiency syndrome

the Hosmer-Lemeshow goodness-of-fit test,[13] where *P* value <0.05 was used to imply a model's overall lack of fit. The model's SMR was calculated by taking the ratio of the mean observed deaths over the mean predicted deaths for the duration of study.

We applied a first level customisation strategy[14] to improve calibration of the APACHE IV model. The approach involves fitting a simple logistic regression model with the observed in-ICU mortality rate being the dependent variable, and the original logit term in the APACHE IV model being the independent variable. The new estimated probability of death for each patient is then calculated from the customised model and calibration is re-evaluated for the customised model. We performed all statistical analysis using SPSS 17.0 for Windows.

Results

Baseline Characteristics of Study Population

A total of 1084 patients were admitted to HSA ICU from 1 January 2009 to 31 December 2009. After applying APACHE IV exclusion criteria, 916 eligible admissions were considered for analysis. Postoperative coronary artery bypass graft (CABG) patients were treated in a separate unit and were excluded from the study. A comparison of the differences in demographic and clinical characteristics between HSA ICU and APACHE IV (developmental) datasets is shown in Table 2. Admissions to HSA ICU recorded a higher percentage of male patients (60.6%) compared to the APACHE IV developmental sample (54.2%). The ethnic composition between Malaysia and the United States is different, where ethnic groups in HSA ICU were divided into 4 categories (Malay, Chinese, Indian and Others).

Admissions to HSA ICU for the period of study were almost equally divided between non-operative and postoperative patients, whereas approximately 70% of admissions used in the development of APACHE IV were non-operative. Difference in types of admissions between HSA ICU and APACHE IV was also reflected through a higher percentage of emergency surgery patients in HSA ICU (36.6%), compared to APACHE IV (5.7%). The principal diagnostic categories according to major organ-related functions for both operative and non-operative admissions to HSA ICU for the period of study are shown in Figure 1. The majority of postoperative patients were admitted due to trauma, whereas most of the non-operative patients were those with cardiovascular-related diseases.

The mean age of patients in HSA ICU at 43.44 years was notably lower compared to the corresponding mean of 61.51 years in APACHE IV. Younger patients formed the majority of HSA ICU admissions, with nearly 30% below the age of 30 years. These patients were mostly admitted to HSA ICU for trauma-related illnesses (Fig. 2). Patients in the middle age groups were mostly admitted due to cardiovascular and respiratory diseases, whereas those who

Table 2. Demographic and Clinical Characteristics of Admissions to HSA ICU (January 2009 to December 2009) and APACHE IV

Characteristics	HSA ICU (n = 916)	APACHE IV (n = 66,270)
Gender (% male)	60.6	54.2
Ethnic group		
White	NA	69.3
Malay	56.4	NA
Chinese	24.1	NA
Indian	10.7	NA
Others	8.7	NA
Mean age (years)	43.44 ± 0.58	61.51 ± 0.07
Mean APS	69.59 ± 1.06	38.83 ± 0.10
Mean pre-ICU length of stay (square root days)	0.794 ± 0.023	0.786 ± 0.004
Died in ICU (%)	18.8	13.6
With comorbidities (%)	3.7	10.3
Emergency surgery (%)	36.6	5.7
Postoperative patient (%)	49.2	30.9
Ventilated on ICU Day 1 (%)	83.0	35.1
Unable to assess GCS (%)	23.1	8.0

HSA: Hospital Sultanah Aminah; APACHE: Acute physiology and chronic health evaluation; NA: Not applicable; APS: Acute physiology score; ICU: Intensive care unit; GCS: Glasgow Coma Scale

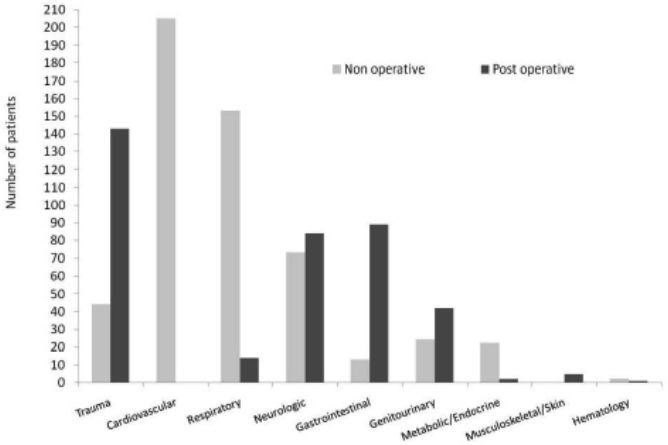

Fig. 1. Graph showing breakdown of principal diagnostic categories according to major organ functions for postoperative and non-operative HSA ICU admissions in year 2009.

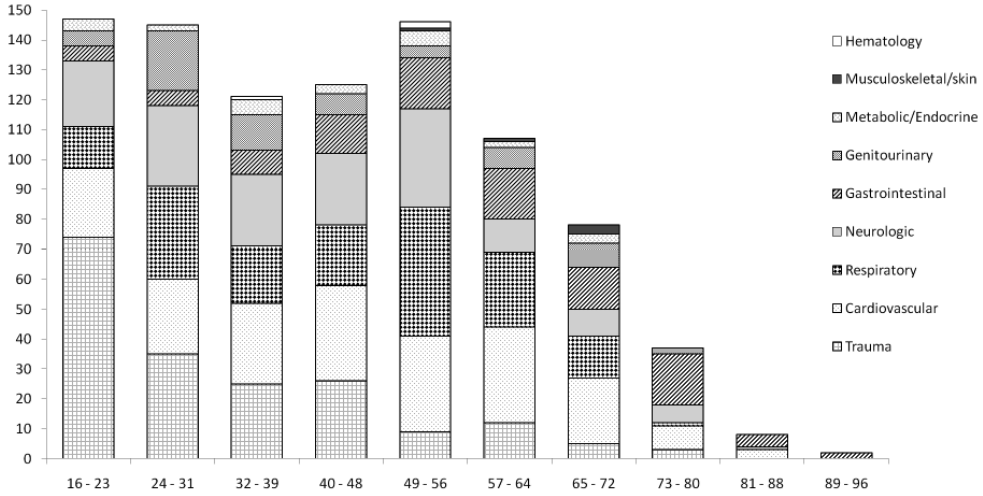

Fig. 2. Graph showing the detailed principal diagnostic categories according to age group for HSA ICU admissions in year 2009.

were older than 70 years old were mostly admitted due to gastrointestinal problems. A small percentage of HSA ICU patients (3.7%) disclosed that they have at least 1 of the 7 comorbidities defined in APACHE IV, with one-third having immunodeficiency disorders. It is possible that some of the patients who were admitted to HSA ICU could have been deliberately withholding important information about their underlying conditions, or could not provide accurate information due to lack of awareness of their previous medical condition.

The first day APS values for patients who were admitted to HSA ICU for the period of study varied between 11 and 171, with the majority having values between 41 and 50. There were also isolated cases of patients having APS values that were greater than 140. The mean APS value for ICU Day 1 admissions to HSA ICU (69.59) is significantly higher than the mean APS value for APACHE IV (38.83). This value implies that the severity of illness of patients who were admitted to HSA ICU was much greater compared to the APACHE IV cohort. The overall number of deaths in HSA ICU for the period of study was 172 (18.8%). Application of APACHE IV in HSA ICU produced a significantly higher overall predicted in-ICU mortality rate of 28.11% and an overall SMR value of 0.67.

Validation of APACHE IV Model

On the whole, APACHE IV appeared to exhibit good discrimination when applied to HSA ICU, with an area under ROC of 0.78 (Fig. 3). Despite having good discrimination, the model's calibration in HSA ICU was found to be poor, as reflected in the calibration curve (Fig. 4) and results obtained from the Hosmer-Lemeshow goodness-of-fit tests (Table 3). Model fit is considered to be perfect when the observed values lie exactly on the diagonal line of a calibration curve. From the calibration curve, model fit appeared to be quite acceptable for the first 3 risk deciles. However, predictions were inaccurate starting from the fourth decile onwards, where the observed outcomes were much lower than predicted outcomes. The majority of patients in HSA ICU were in the first 3 groups and were associated with lower mortality risk. A significant improvement was observed in calibration after applying first level customisation on the APACHE IV equation (Table 3). The overall fit for the customised model was found to be good although there was no improvement in discrimination.

Discussion

There has been considerable debate over the use of SMR as

Table 3. Performance Indicators of APACHE IV and the First-level Customised Model

	SMR	Area under ROC Curve	Hosmer-Lemeshow X^2	$\hat{C}$ Statistic *P* Value
APACHE IV	0.67	0.78	113	< 0.0001
Customised model	1.00	0.78	6.39	0.78

SMR: Standardised mortality ratio; ROC: Receiver operating characteristic

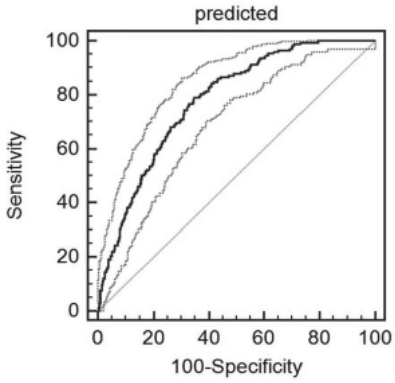

Fig. 3. Area under receiver operating characteristic curve for HSA ICU admissions in year 2009.

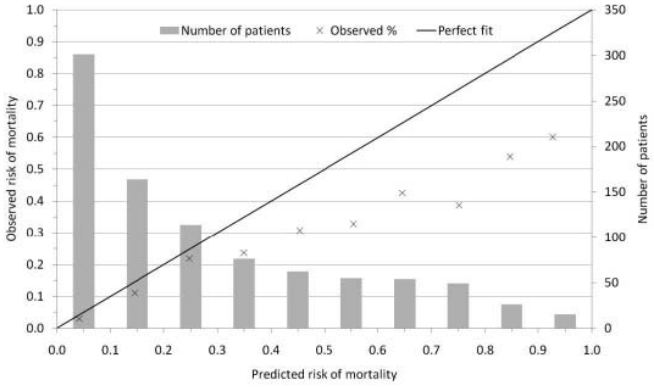

Fig. 4. Calibration curve to compare observed and predicted in-ICU mortality rates across 10% intervals of predicted risk.

a valid indicator of quality of care that is being provided in an ICU.[15-17] We acknowledge that SMR may not be a perfect indicator to evaluate model performance and interpretation of this index should be done with caution. However, SMR is still a universal standard approach for reporting of hospital mortality and has been applied to evaluate model performance in many well established hospital prognostic systems, including APACHE IV. In this study, external validation of APACHE IV in HSA ICU yielded a low SMR value of 0.67. There are 2 possible interpretations of a low SMR value, i.e. it could indicate that HSA ICU performed well with very ill patients, or it could suggest poor calibration of APACHE IV in the Malaysian ICU. We do not think that our ICU did an excellent job with very ill patients as Malaysia is still a developing country and the ICU was seriously short of trained staff at the time the study was conducted. We believe that significant differences in case mix (age distribution, types of admission, admission diagnoses) between HSA ICU and APACHE IV contributed towards the lower SMR.

APACHE IV exhibits good discrimination but tends to overestimate in-ICU mortality, especially for mid- to high-risk ranges in this study. Our study reveals that APACHE IV equation for prediction of in-ICU mortality risk does not fit well and is not suitable for application in HSA ICU. The model's lack of fit is likely due to differences in patient characteristics between APACHE IV and HSA ICU datasets, especially in terms of age, types of disease and admission types. While most of the patients involved in the development of APACHE IV were elderly, a significant number of HSA ICU patients were from the younger age groups with trauma-related illnesses. These patients were mostly victims of road accidents, who were transferred from the Accident & Emergency (A&E) unit.

Discrepancy in the mean APS between HSA ICU and APACHE IV is probably another influential factor that affected APACHE IV's prediction accuracy. The mean APS in HSA ICU is considerably higher compared to APACHE IV. Differences in quality and health care systems could probably account for the discrepancy in APS between HSA ICU and APACHE IV admissions. The physiological components of a patient are also affected by factors such as genetics, lifestyle and cultural habits. Moreover, the calculation approach in APACHE IV considers points that are assigned to the worst values of each physiological variable. The choice of worst values is highly dependent on data variation. It is highly possible that variability in data is greater for HSA ICU patients, resulting in extreme values being chosen as worst values and thus contributing to higher APS values. In our study, frequency of data collection followed the current practice in HSA ICU, where data collection intervals were not equal time-spaced for all physiological variables. Some variables were monitored on an hourly basis, whereas other variables were measured less frequently. Differences in the data collection intervals play an important role in influencing the choice of worst values for physiological variables. We believe that the choice of worst values for infrequently measured variables may be affected by detection bias[18] where variables are only detected if they are measured and unmeasured variables are assumed normal.

The majority of patients who were admitted to HSA ICU reported that they do not have existing chronic health conditions. Although APACHE IV (non-CABG model) did not list diabetes as one of the chronic health conditions, this information was collected in our study. Interestingly, 184 patients (20.1%) revealed that they were diabetic patients. This figure is consistent with the findings in a recent study[19] that reported the overall prevalence of diabetes in Malaysia (11.6%) to be higher than other regions in the world. In view

of the high percentage of diabetic patients in HSA ICU, there is a possibility that this variable may potentially be a significant predictor of in-ICU mortality risk in the context of our Malaysian study. In addition, although ethnicity was not included as one of the main predictors in APACHE IV, the importance of this variable merits further investigation. Further work is required to evaluate whether differences in cultural and dietary habits among various ethnicities in Malaysia play an important role in predicting in-ICU mortality outcomes in the local context.

There are several limitations to our study. First, due to the single centre nature of this study, findings that are obtained in HSA ICU may not be representative of other ICUs in Malaysia. Potential differences in case mix, clinical practice, discharge policies and quality of care between HSA ICU and other ICUs may limit the generalisation of our findings. Second, the sample size in our single centre study is considered relatively small and our case mix may not be sufficiently diverse compared to APACHE IV. The small sample size is a limiting factor in the analysis of uniformity of fit among different subgroups in our study. Although a multi centre study will probably mitigate issues regarding sample size and case mix, we were unfortunately not able to obtain cooperation from other ICUs in the country due to limited manpower and funding resources.

Conclusion

In this study, we have shown that APACHE IV has poor calibration and acceptable discriminatory power when applied to a Malaysian ICU cohort. There is great potential in using prognostic models to enhance the quality of critical care in Malaysia. A good prognostic model will definitely serve as a clinical decision support tool that is beneficial to medical practitioners in the long run. To the best of our knowledge, this is the first study on external validation of APACHE IV in Malaysia. This study is beneficial in the sense that it provides an insight on the characteristics of patients who were admitted to HSA ICU.

Acknowledgements

This study was funded by University of Malaya grant RG138/09HTM Assessment of Performance of the Scoring Systems at an interdisciplinary ICU. The authors would like to thank two researchers, Dr Azim Mohd Yunos and Rafidah Atan, from Monash University Malaysia, for their support in initiating the design of this study.

REFERENCES

1. Keegan MT, Gajic O, Afessa B. Severity of illness scoring systems in the intensive care unit. Crit Care Med 2011;39:163-9.
2. Bouch DC, Thompson JP. Severity scoring systems in the critically ill. Contin Educ Anaesth Crit Care Pain 2008;8:181-5.
3. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993;270:2957-63.
4. Tong JMG, Tai LL, Tan CC, Othman AS, Shukor AA, Lim CH. Malaysian Registry of Intensive Care Report for 2012. Available at: www.mric.org.my. Accessed 6 August 2013.
5. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. Crit Care Med 2006;34:1297-310.
6. Metnitz PG, Moreno RP, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 - From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 2005;31:1336-44.
7. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, 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 Med 2005;31:1345-55.
8. Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III). Crit Care Med 2007;35:827-35.
9. Kherallah M, Hazza M, Dahhan T, Tantawy T, Jamil MG, Al-Tarifi A. Performance of the acute physiology and chronic health evaluation IV at a tertiary Saudi hospital. Chest 2008;134:p112003.
10. Bhattacharyya M, Todi S. APACHE IV: benchmarking in an Indian ICU. Crit Care 2009;13:P510.
11. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100:1619-36.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.
13. Hosmer DW, Lemeshow S. A goodness-of-fit test for the multiple logistic regression model. Commun Stats 1980;A10:1043-69.
14. Bakhshi-Raiez F, Peek N, Bosman RJ, de Jonge E, de Keizer NF. The impact of different prognostic models and their customization on institutional comparison of intensive care units. Crit Care Med 2007;35:2553-60.
15. Goldman DA, Brender JD. Are standardized mortality ratios valid for public health data analysis? Stat Med 2000;19:1081-8.
16. Jarman B. In defence of the hospital standardized mortality ratio. Healthc Pap 2008;8:37-42;discussion 69-75.
17. van Gestel YR, Lemmens VE, Lingsma HF, de Hingh IH, Rutten HJ, Coebergh JW. The hospital standardized mortality ratio fallacy: a narrative review. Med Care 2012;50:662-7.
18. Holmes CL, Gregoire G, Russell JA. Assessment of severity of illness. In: Hall JB, Schmidt GA, Wood LDH, editors. Principles of critical care. 3rd ed. Blacklick: McGraw-Hill Professional Publishing; 2005. p. 63-78.
19. Letchumanan GR, Wan Nazaimoon WM, Wan Mohamad WB, Chandran LR, Tee GH, Jamaiyah Y, et al. Prevalence of diabetes in the Malaysian National Health Morbidity Survey III 2006. Med J Malaysia 2010;65:173-9.