CHAPTER III

RESEARCH METHODOLOGY

Research question

Primary research question:-

What are the performances of the Simplified Acute Physiologic Score (SAPS II) and the Mortality Prediction Model (MPM_{24} II) scoring systems in predicting the hospital mortality of critically ill patients in the ICU of Nopparat-Rajathanee Hospital?

Secondary research question:-

What scoring systems perform better in the ICUs?

Research objective

1. To determine the performance of SAPS II and MPM_{24} II scoring systems in predicting the hospital mortality of critically ill patients in the ICUs of Nopparat-Rajathanee Hospital
2. To compared the ability of SAPS II and MPM_{24} II scoring systems in predicting hospital mortality.

Research hypothesis

This study aimed to compare the area under the ROC curve between the two scoring systems. Hence, the null hypothesis was:

The area under the ROC curve from SAPS II is equal to the area under the ROC curve from MPM_{24} II.

The alternative hypothesis was:

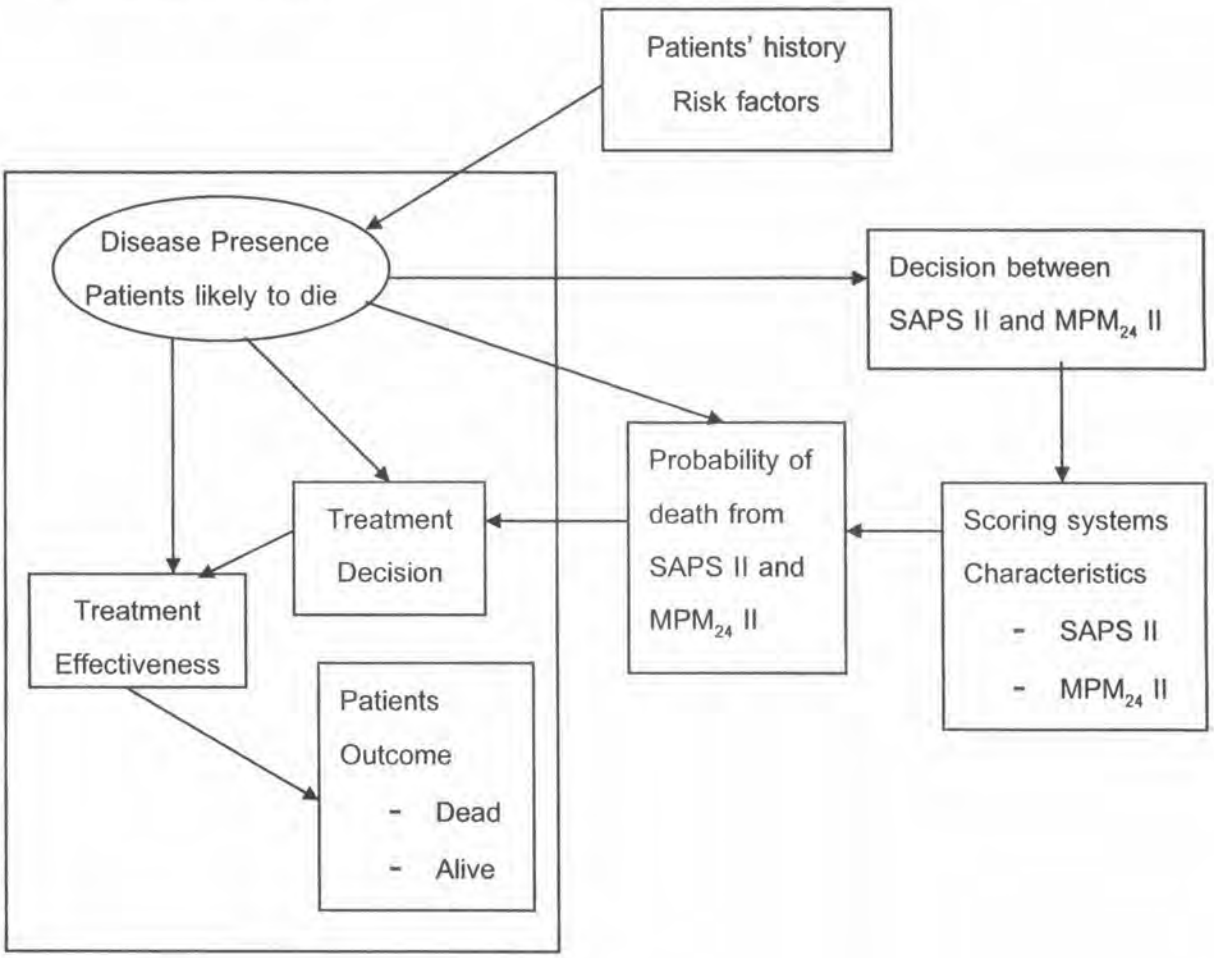
The area under the ROC curve from SAPS II is different from the area under the ROC curve from MPM_{24} II.

The statistic hypothesis was:-

- H_0: ROC AUC SAPS II = ROC AUC MPM_{24} II

  H_1: ROC AUC SAPS II $\neq$ ROC AUC MPM_{24} II

Conceptual framework



Keywords

Severity of illness

Intensive care

Mortality prediction

Simplified Acute Physiologic Score (SAPS II)

Mortality Prediction Model system (MPM II)

Mortality

Operational definitions

1. <u>Definitions for SAPS II data collection</u>

Data are collected during the first 24 hours after ICU admission

- Age:- use the patient's age in years at last birthday

- Heart rate:- use the worst value in 24 hours, either low or high rate; if it varied from cardiac arrest (11pts) to extreme tachycardia (7pts), assign 11points.
- SBP:- use the same method as for hearth rate e.g., if it varied from 60 mmHg to 205 mmHg, assigns 13 Points.
- Body temperature:- use the highest temperature in °C or °F
- PaO2/FiO2 ratio If ventilated or CPAP:- use the lowest value of the ratio.
- Urinary output:- if the patient is in the intensive care unit for less than 24 hours, make the calculation for 24 hours.
- Serum urea or BUN:- use the highest value in mmol// or g/L for serum urea, in mg/dL for the serum urea nitrogen.
- WBC count:- use the worst (high or low) WBC count.
- Serum potassium level:- use the worst (high or low) value.
- Serum Sodium level:- use the worst (high or low) value.
- Serum bicarbonate level:- use the lowest value.
- Bilirubin:- use the highest value in micromol/L or mg/dL
- Glasgow coma score:- use the lowest value. If the patient is sedated, record the estimated Glasgow coma score before sedation.
- AIDS:- yes , if HIV positive with clinical complications as pneumocystis carinii pneumonia, Kaposi's sarcoma, Lymphoma, tuberculosis or toxoplasma infection.
- Hematologic malignancy:- yes, if lymphoma, acute leukemia, or multiple myeloma.
- Metastasis cancer:- yes, if proven metastasis by surgery, CT scan or any other method.

2. Definitions for MPM_{24} II data collection

Data are collected during the first 24 hours after ICU admission.

- Coma or deep stupor at time of ICU admission
  - not due to drug overdosage
  - if patient is on paralyzing muscle relaxant, awakening from anesthesia or heavily sedated, use best judgment of the level of consciousness prior to sedation

- coma: no response to any stimulation, no twitching, no movements in extremities, no response to pain or command, Glasgow coma scale 3
- deep stupor: decorticate or decerebrate posturing; posturing is spontaneous or in response to stimulation or deep pain; posturing is not in response to commands; Glasgow coma scale 4 or 5

- Cirrhosis
  - history of heavy alcohol use with portal hypertension and varices
  - other causes of liver disease with evidence of portal hypertension and varices
  - biopsy confirmation of cirrhosis
- Metastasis malignant neoplasm
  - stage IV carcinomas with distant metastases
  - do not include involvement only of regional lymph nodes
  - include if metastases are obvious by clinical assessment or confirmed by a pathology report
  - do not include if metastases not obvious or if pathology report is not available at the time of ICU admission
  - acute hematologic malignancies are included
  - chronic leukemias are not included unless there are findings attributable to the disease or the patient is under active treatment for the leukemia. Findings include sepsis, anemia, stroke caused by clumping of white blood cells, tumor lysis syndrome with elevated uric acid following chemotherapy, pulmonary edema or lymphangiectatic form of ARDS
- Intracranial mass effect
  - intracranial mass (abscess, tumor, hemorrhage, subdural) as identified by CT scan associated with any of the following: (1) midline shift, (2) obliteration or distortion of cerebral ventricles, (3) gross hemorrhage in cerebral ventricles or subarachnoid space, (4) visible mass > 4 cm or (5) any mass that enhances with contrast media
  - if the mass effect is known within 1 hour of ICU admission, it can be indicated as yes

- CT scanning is not mandated and is only indicated for patients with major neurological insult
- Age in years
  - patient's age at last birthday
- Mechanical ventilation
  - patient is using a ventilator at the time of ICU admission or immediately thereafter
- Medical or unscheduled surgery admission
  - do not include elective surgical patients (surgery scheduled at least 24 hours in advance) or pre-operative Swan-Ganz catheter insertion in elective surgery patients

Research design

Prospective observational study

Research methodology

Population and sample

This study was performed at Nopparat-Rajathanee Hospital, a 580-beds university-affiliated tertiary care referring hospital in Bangkok, Thailand. There are two units in the adult ICU: an eight-bed surgical ICU and an eight-bed mixed medical and coronary care unit. Nurse-to-patient ratio during day and night shifts was approximately 1:1.5, depending on the number of patients. All physiologic monitors were available.

All consecutive ICU admissions were included from 1 May 2008 to 15 March, 2009. Patients' data generated as a result of patient care and recorded in the medical records was collected concurrently for consecutive unselected intensive unit admissions.

Inclusion and exclusion criteria

We prospectively collected data on all patients admit consecutively to the ICUs during the study period. We excluded and did not collect data for patients who admitted for less than 24 hours, patients with burns, patients younger than 18 years of age, and patients missing an acute physiology score on ICU day 1. To avoid counting more than one hospital outcome for the same patient, analysis included only a patient's first ICU

admission. We also excluded from analysis patients admitted from another ICU during the same hospitalization. We did this because extensive life support before ICU admission biased the prognostic implications of first ICU day physiologic measures. These exclusion criteria were in accordance with the original methods used in the development of both scoring systems.

Sample size calculation

For primary research objective, the performances of SAPS II and MPM_{24} II in predicting mortality of ICU patients had to be determined. To estimate the receiver operating characteristic (ROC) curve, the area under the ROC curve and 95% CI for the ROC curve area, the sample size determination was suggested by Xiao-Hua Zhou (Zhou X, Obuchowski, N, McClish D. (2002). Statistic Methods in Diagnostic Medicine, Wiley-Interscience, US.)

A general formula for sample size estimation for constructing a 2-sided CI for single test accuracy was

$$m = \frac{z^2_{1-\alpha/2} V(\hat{\vartheta})}{L^2} \qquad (1)$$

where m is the required patient sample size, and $V(\hat{\vartheta})$ is the variance function (McCullagh and Nelder, 1989) of $\hat{\vartheta}$, and L is the desired width of one-half of the CI. $Z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution, α is the confidence level, We construct 95% CI, in which case α equals 0.05 and $Z_{1-\alpha/2}$ is 1.96.

Here, ϑ is the area under the ROC curve and m is the number of patients with the condition (here is the patients who die) needed for the study. Let k denote the ratio of the number of patients without the condition (here is the patients who survive) to patients with the condition in the study sample. The total number of patients needed for the study is then $m(1+k)$. Since in our hospital the average percentage of hospital death from ICUs is 26.9 %, thus, k = 73.1/26.9 = 2.72 then total number of patients should be 3.72m.

The variance function for the area under the ROC curve used for sample size calculation is (Obuchowski, 1994) as follows:

$$\hat{V}\left(\hat{A}\right) = (0.009 \times e^{-a^2/2}) \times [(5a^2 + 8) + (a^2 + 8)/k] \qquad (2)$$

where $a = \Phi^{-1}(A) \times 1.414$ and Φ^{-1} is the inverse of the cumulative normal distribution function. We expected the area under the ROC curve, A, is 0.85; thus Φ^{-1} =

1.035, then $a = 1.035 \times 1.414 = 1.463$. We want to estimate the area under the ROC curve within $\pm$ 0.05; thus L = 0.05.

From the equation (2),

$$\hat{V}\left(\hat{A}\right) = (0.009 \times e^{-1.463^2/2}) \times [(5 \times 1.463^2 + 8) + (1.463^2 + 8)/2.72] \equiv 0.076$$

By using the formula for sample size estimation (1),

$$m = \frac{(1.96)^2 \times 0.076}{(0.05)^2}$$

$$= 116.78$$

Therefore, for 95% CI for the area under the ROC curve, the required number of patients who die was 117. The total sample size required for the study was $m(1+k)$; thus $117 \times (1 + 2.72) = 117 \times 3.72 = 435.24$.

For secondary research objective, comparison of areas under the ROC curve between SAPS II and MPM_{24} II has to be determined. The Medcalc computer software was used to compute the required sample size for the comparison of the areas under two ROC curves (derived from the same cases) [70]. The sample size took into account the required significance level and power of the test.

The required input were type I error (the probability of rejecting the null hypothesis when in fact it is true), type II error (the probability of accepting the null hypothesis when in fact it is false), the hypothesized area for the first and second ROC curves, and the hypothesized rank correlation coefficient between the two assays in the positive (death cases) and in the negative (survival cases) groups. Usually, these values were obtained from previous studies. But from literature reviews, we could not find the reported correlation coefficient in the positive and negative groups. Therefore, we used these values from the preliminary study to answer our primary research question. With 0.05 Type I error, and 0.2 type II error to detect the difference between area under ROC curve of 0.89 and 0.92. The correlation coefficient in the positive groups was 0.90 and the correlation coefficient in the negative groups was 0.84. The required number of patients who die has to be 170. Since in 2007 the average percentage of in-hospital death of ICUs patients was 26.9 %, therefore the total sample size is 170/0.27 = 632. From May 1, 2008 to March 15, 2009, there were 218 in-hospital deaths among the 639 ICU patients. Hence, we collected data during this period.

Intervention

None

Outcome measurement

The main outcome was survival status on discharge from the hospital, including deaths in the ICU and hospital wards after discharge from the ICU.

Data collection

Data collection was last from May 1, 2008 to March 15, 2009. All data was collected concurrently for consecutive ICU admissions.

The following data was collected:

- Basic demographic characteristics included sex, age, and principal diagnostic category leading to ICU admission.
- Type of patient status was defined classically as medicine, scheduled surgery, or unscheduled surgery.
- ICU length of stay (LOS)
- Patients were followed up until ICU and hospital discharge in order to registrar their survival status.

Calculating SAPS scores and predicted mortality

The worst values on SAPS II variables during 24 h following ICU admission was collected from the patient's charts and clinical flow sheets, using the variable definitions reported in the literature [3]. The worse value was defined as the value that would have been assigned the greatest number of points. The resulting SAPS II score was then entered into a published mathematical formula whose solution gave the numerical value of the predicted hospital mortality.

Calculating MPM_{24} II predicted mortality

MPM_{24} II data was obtained during 24 h following ICU admission in accordance with the original methodology [4]. In sedated patients, the Glasgow Coma Score (GCS) was determined from medical records before sedation. Except for age, all these values were dichotomous; in other words, values were either present or absent. As an example, a blood pressure below 90 mmHg was worth one point; all other values were assigned a zero. Each of these variables was entered into a published mathematical formula whose solution provided the predicted mortality.

Data analysis

Descriptive statistics were used to report age, gender, type of admission, length of stay and mortality. All continuous variables were presented as mean (SD) or medians with the range Categorical variables were expressed as numbers and percentages All statistical tests were two-sided, with $P < 0.05$ considered statistically significant. The Statistical Package for Social Sciences (SPSS), version 15.0 was used for data analysis. The outcome measurement was hospital mortality, defined as death occurring before hospital discharge. Probabilities of hospital death were compared with the actual outcome.

The 2 × 2 classification table at the predicted risk of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 was used to compare predicted and observed outcomes, record sensitivity (true positive), specificity (true negative), positive predictive value (1-false positive), negative predictive value (1-false negative) and positive likelihood ratios.

Sensitivity is the proportion of patients who die that is predicted correctly by SAPS II and MPM_{24} II.

Specificity is the proportion of patients who do not die that is predicted correctly by SAPS II and MPM_{24} II.

Accuracy or overall correct classification rate is the proportion of true results (both the proportion of patients who die and do not die that is predicted correctly by SAPS II and MPM_{24} II) in the population.

- Positive predictive value is the proportion of the patients who actually die who are predicted to die by SAPS II and MPM_{24} II.

Negative predictive value is the proportion of the patients who do not actually die who are predicted not to die by SAPS II and MPM_{24} II.

Positive likelihood ratio is the ratio of the probability of the predicted mortality that is predicted to die by SAPS II and MPM_{24} II in the patients who die to the probability of the predicted mortality that is predicted to die by SAPS II and MPM_{24} II in the patients who survive.

In analyzing the performance of hospital mortality predictions of scoring systems, two assessments are especially important: discrimination and calibration.

Discrimination

Discrimination refers to a scoring system's ability to distinguish survivors from non-survivors. The area under the receiver operating characteristic (ROC) curve was measured to test discriminative power. A ROC curve was constructed from the patients' predicted outcomes and observed outcomes using 10 % stepwise increments in predicted mortality. The area under the ROC curve represents the probability that a patient who died had a higher predicted probability of dying than a patient who survived. An AUC of 0.5 indicates that the scoring system does not predict better than chance. The discrimination of a prognostic scoring system is considered perfect if AUC = 1, good if AUC >0.8, moderate if AUC is 0.6 to 0.8, and poor if AUC <0.6 [71]. A comparison of the areas under the ROC curves of the two scoring systems was done using formulas suggested by Hanley and McNeil (Radiology; 1983 148: 839-843) [70]

Calibration

The AUC of a scoring system gives no indication of how close the predicted probabilities are to the observed outcome. To take this aspect of a scoring system's performance into account, we have to look at the calibration of the prognostic scoring systems. Calibration refers to the agreement between predicted probabilities and the 'true probabilities'. Of course, the true probability of a patient's outcome is not known; otherwise there would be no need to develop prognostic scoring systems (we only know whether a patient died or not). However, the true probabilities can be approximated by taking the mean of the observed outcomes within predefined groups of patients. The Hosmer-Lemeshow goodness-of-fit statistic [72] was used to evaluate calibration. It compares the observed mortality in a group with the predicted mortality of that group. The subjects were divided into approximately 10 groups of roughly the same size based on the percentiles of the estimated probabilities. This goodness-of-fit statistic had a chi-square distribution and the discrepancies between the observed and expected number of observations in these groups were summarized by the Pearson chi-square statistic. A high Hosmer-Lemeshow chi-square statistics relates to a small *p* value, implying significant difference between observed and predicted mortality, and thus indicates a lack of fit of the scoring system.

Calibration curves were drawn by plotting predicted against actual mortality for groups of the patient population stratified by 10% increments of predicted mortality. This graph gives a visual representation of the agreement or calibration at each of levels of risk.

Evaluation of overall scoring system prediction

Standardized mortality ratios (SMR) is an overall summary statistic providing information about how the overall mortality rate agrees with the mortality prediction for the sample. It compares the observed number of deaths to the predicted number of deaths. Large departures will indicate failure of the scoring systems to predict the probability of patient death in that context. It is the ratio of observed deaths to predicted deaths or observed mortality rate to predicted mortality rate.

$$SMR = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} \pi_i}$$

There are n patients indexed by i, π is the estimate of the probability of death provided by the scoring system. For a patient who dies, the outcome, $Y_i = 1$ or $Y_i = 0$ if the patient survives.

The estimate of the 95% CI of the SMR is:

$$SMR \pm 1.96 \frac{\sqrt{\sum_{i=1}^{n} \hat{\pi}_i.(1-\hat{\pi}_i)}}{\sum_{i=1}^{n} \hat{\pi}_i}$$

Ethic considerations

Before conducting this study, the protocol was reviewed and approved by the ethics committee of the Bureau of Medical Technical Development, Department of medical services, Ministry of Public Health and the Institutional Review Board of Nopparat-rajathanee Hospital.

Since, data used for this study had already been collected for routine clinical purposes; informed consent of the patients was waived with the permission of the ethical review committees. Access to the data was granted by the ethic committees and the database management committee of the hospital. All data was used for study purpose only and was confidential. The research was conducted on de-identified patient data

Limitation

This study had some possible limitations to the generalizability. As a single center study, the results therefore reflected the outcome of patients in a university-affiliated tertiary care referring hospital and may not be applied to all hospitals in the country. However, the study gave some insight into this issue, at least from a tertiary care perspective.

Moreover, critical illness was a dynamic process and therefore severity based on a single time point such as ICU admission, did not consider changes in patients' clinical status over time and their response to treatment. Serial predictions over a period of time, may improve prediction accuracy and clinical utilities, although acquiring these data continuously was difficult in practice.

Expected benefit

Before the clinical application of any of scoring systems, they must be validated on the population under evaluation. If the SAPS II and MPM_{24} II can accurately predict the outcomes of our patients, we intended to select the appropriate one to apply it for routine use in our ICUs.

The proposed roles of scoring systems in intensive care can be divided into three main areas:

1. Comparative audit:

Comparing the actual outcomes with the expected outcomes for groups of patients, calculated using a scoring system, has been proposed as the basis for initial exploratory control comparisons of different providers. The use of case-mix-adjusted outcomes as a measure of the clinical effectiveness of intensive care assumes that an SMR greater than 1.0 may reflect poor care and, conversely, an SMR less than 1.0 may reflect good care.

2. Evaluative research:

Accurate objective estimates of the probabilities of hospital death, when translated into expected hospital death rates for groups of patients, have been proposed as the basis for research studies to identify those components of intensive care structure and process that are linked to improved patient outcome.

3. Clinical management of individual patients.

Although the early scoring systems were proposed only as a means for comparing observed and expected outcomes for groups of intensive care patients, some of the subsequent second- and third-generation methods are promoted as methods to guide the clinical care of individual patients. It is proposed that an accurate objective estimate of the risk of hospital death can provide additional information to help make clinical decisions about treatment aims for individual patients. Such decisions might include when to withdraw treatment or when to discharge a patient.