# 77369_Auto_Edited - check.docx

*Retrospective Study*

**Scoring systems in critically ill: Which one to use in cancer patients?**

Scoring systems in oncology ICUs

Anisha Beniwal, Deven Juneja, Omender Singh, Amit Goel, Akhilesh Singh, Hemant Kumar Beniwal

**Abstract**

BACKGROUND

Scoring systems have not been evaluated in oncology patients. We aimed to assess the performance of Acute Physiology and Chronic Health Evaluation (APACHE) II, APACHE III, APACHE IV, Simplified Acute Physiology Score (SAPS) II, SAPS III, Mortality Probability Model (MPM) $II_0$ and Sequential Organ Failure Assessment (SOFA) score in critically ill oncology patients.

AIM

To compare the efficacy of seven commonly employed scoring systems to predict outcomes of critically ill cancer patients.

METHODS

We conducted a retrospective analysis of 400 consecutive cancer patients admitted in medical intensive care unit over a two-year period. Primary outcome was hospital mortality in and secondary outcome measure was comparison of various scoring systems in predicting hospital mortality.

RESULTS

In our study, the overall intensive care unit (ICU) and hospital mortality was 43.5% and 57.8%, respectively. All the seven tested scores underestimated mortality. The mortality as predicted by MPM $II_0$ predicted death rate (PDR), was nearest to actual mortality followed by that predicted by APACHE II, with a standardized mortality rate (SMR) of 1.305 and 1.547, respectively. The best calibration was shown by the APACHE III score ($\chi 2$=4.704, $P$ = 0.788). On the other hand, SOFA score ($\chi 2$=15.966, $P$ = 0.025) had the worst calibration, although the difference was not statistically significant. All the seven scores had acceptable discrimination with good efficacy, however SAPS III PDR and MPM $II_0$ PDR (AUROC=0.762), had better performance as compared to others. The correlation between the different scoring systems, was significant (P<0.001).

CONCLUSIONS

All the severity scores tested under-predicted mortality, in the present study. As the difference in efficacy and performance was not statistically significant, the choice of scoring system used may depend on the ease of use and local preferences.

Key Words: APACHE score; Intensive care unit; Medical oncology; SOFA score; Scoring systems; Severity of illness index.

Beniwal A, Juneja D, Singh O, Goel A, Singh A, Beniwal HK. Scoring systems in critically ill: Which one to use in cancer patients? *World J Crit Care Med* 2022; In press

Core Tip: Scoring systems are important for patient triaging, benchmarking ICU performance, comparing different ICUs and may also help in patient prognostication, selecting treatment options and resource utilization. However, validity and utility of these scores may be questionable in the patient population apart from where they were developed. Hence, these scores need to be tested and validated in different patient populations, in different geographical areas and over different time periods. There is a lack of an ideal score for prognostication of critically ill cancer patients. In our retrospective study, analyzing data from 400 patients and comparing seven commonly employed critical illness scores, we observed that all the scores had similar efficacy and under-predicted mortality. Therefore, the selection of severity of illness score should depend on the ease of use and local preferences.

INTRODUCTION

The application of prognosticating scoring systems is considered as an important phase in intensive care units (ICUs) since these severity scoring systems estimate the probability of mortality for patients. These scores help the physicians to facilitate resource utilization or continuous quality improvement and to stratify the patients for clinical research[1,2].

ICU scoring systems can help both patients as well as their attendants to select from further treatment options. Further, the scores calculated by these scoring systems help in evaluating the impact of newer treatment modalities and organizational changes which in turn contributes towards the development of treatment standards. In addition to the above, the scoring systems' outcomes also help in benchmarking ICU performance and comparing the scores secured by different ICU patient populations so as to find out the differences in mortality. However, these systems are unreliable in predicting the clinical outcomes of an individual though it has proven efficacy in predicting mortality for a particular patient cohort[3].

Acute Physiology and Chronic Health Evaluation (APACHE) II and Simplified Acute Physiology Score (SAPS) II are arguably the two most-commonly used and validated tools used in the prediction of ICU patient outcomes [4,5]. These scoring systems were developed in 1980s and have become outdated due to technological and clinical advancements in critical care management of patients in recent years. Hence, there is a need to develop new scoring systems that include APACHE IV, SAPS III and Mortality Probability Model (MPM) II0 [6-9]. Such newly-created systems encompass a large number of variables and are highly complicated to compute.

In addition, both validity and utility of the existing scoring systems may be questionable in terms of current patient population compared to the patient population during when they were developed. These scores are widely used and the scoring systems have been validated for a long known time to predict the outcome in general medical or surgical procedures conducted upon critically ill patients. However, whether these systems can predict the mortality accurately among cancer patients remain unknown[10]. There is a dearth of studies that compare different generations of scoring systems especially the ones used upon cancer patients admitted in medical oncology ICUs. Only a few studies has assessed their usefulness in cancer patients with conflicting results. Moreover, geographic variations in patient populations and the types of cancer necessitate that these scores should be evaluated for different populations[11].

Therefore, the current study is aimed at analysing the efficacy of seven commonly-used scoring systems to predict the mortality amongst patients admitted in oncology ICUs.

## MATERIALS AND METHODS

A retrospective observational cohort study was carried out at multi-disciplinary onco-medical ICU of a tertiary care centre in India. We have an advanced ICU setup and 24-hours intensivist coverage with state-of-the-art facilities. Approval for the study and a consent waiver from institutional ethics committee was obtained.

The data from the records of adult patients who were admitted between January 2018 and February 2020 i.e., two years, was collected and analysed. If the patient was readmitted to ICU more than once during his/her hospital stay, only the first admission was included in the study. Patients who had ICU stay of less than 12 h, post-operative patients and those admitted from or discharged to another ICU were excluded from the study. Patients fulfilling inclusion criteria were serially recruited. The researchers collected the following data; baseline patient characteristics, indication for ICU admission, type of malignancy, presence of metastasis, need for vasopressor, renal and Mechanical Ventilation (MV), length of ICU and hospital stay, and ICU and hospital mortality. The data, required to compute various scores, was collected and calculated specified by the procedures.

### Statistical analysis

The collected data was then transformed into variables, coded, entered in Microsoft Excel. Then, it was statistically analysed using SPSS software (version. PC-25). Quantitative data was expressed in mean ± SD or median with interquartile range. Normality distribution difference between two comparable groups was measured using student's t-test or Mann Whitney 'U' test. Qualitative data was expressed in percentage whereas the statistical differences between the proportions were tested using chi square or Fisher's exact test, as appropriate.

Standardized Mortality Ratio (SMR) was computed by dividing the observed 28 days' mortality by predicted hospital mortality based on different scores. Further, 95% Confidence Interval (CI) was calculated for SMR by considering the observed mortality as a Poisson variable and then dividing its 95% CI by predicted mortality.

The calibration of the scores was executed using Hosmer-Lemeshow goodness-of-fit statistics which divides the subjects into deciles based on the predicted probabilities of death. Afterwards, it computes a Chi-square value from the observed and expected frequencies. Low Chi-square values and high P values (P > 0.5) correspond to better fit.

The ability of the scores, to predict ICU mortality, was explored and discrimination was tested using Area Under Receiver Operating Characteristic (AUROC) curves. If AUROC curves are more than 0.8, it denotes excellent outcome while 0.6-0.8 are considered to be acceptable. The cut-off values were calculated for different scores using Youden's index based on which sensitivity and specificity of the scores were calculated.

Clinically-relevant variables that produced P <0.05 during univariate analyses and are easily accessible on admission were also entered into multiple logistic regression models as the outcome variable of interest. Odds Ratio (OR) was calculated along with 95% CI. P value <0.05 was considered to be statistically significant.

**Sample size calculation**

The sample size calculation was done for estimation of the AUROC curve for APACHE 2 score, using the following formula:

$$n \geq Z^2_{\alpha/2} \, V \, (AUC) \div d^2$$

Where, $V(AUC) = 0.0099 \times e^{-a2/2} \times (6a^2 + 16)$, $a = \phi^{-1} (AUC) \times 1.414$ and $\phi^{-1}$ is the inverse of standard cumulative normal distribution for AUC.

For 95% level of confidence $Z_{\alpha/2 = 1.96}$; d = 0.05 which is the margin of error in estimation and AUC was obtained from a similar study conducted by Schellongowski et al,[12] who reported an AUC of 0.776 for APACHE II score.

Substituting these values in the above formula gives n ≥ 196. As our study was retrospective in nature, we included 400 patients.

## RESULTS

During the study period, the data from 400 patients who fulfilled the inclusion criteria were included in the final analysis. Thirty eight patients were excluded, because 31 were admitted from or discharged to another ICU, five were post-operative patients and two had ICU stay less than 12 h. Their baseline characteristics are given in Table 1 and the comparison between various scores is given in Table 2.

### Predicted mortality

All the scoring systems tested in current study, underestimated the mortality (table 3). The mortality, predicted by MPM $II_0$ PDR, was nearest to the actual mortality with an SMR of 1.305, followed by APACHE II (1.547) and SAPS II (1.74).

### Calibration

Using the Lemeshow-Hosmer goodness-of fit test, APACHE III (4.704) achieved the best calibration with $P = 0.788$ whereas SOFA score (15.966) was the worst with $P = 0.025$ (Table 4). The least statistically significant discrepancy between the predicted and observed mortality was shown by the APACHE III score.

### Discrimination

The efficacy of various scores is given in figure 1. All the scores tested in current study, exhibited good efficacy, even though there was no statistically significant difference between AUROCs and SAPS III PDR. On the other hand, MPM $II_0$ PDR (AUROC=0.762) yielded the best performance (table 5).

### Correlation between various scoring systems

As shown in table 6, there was a significant correlation found among various scoring systems ($P<0.001$) as assessed by linear regression analysis.

### Factors associated with hospital mortality

Five factors that showed significance in univariate analysis such as hypertension, surgery for cancer, use of MV, vasopressors and renal support were used in multivariate analysis too. Out of the five, one two factors i.e., need for MV (OR 2.437, 95%CI=1.315-4.515, $P = 0.005$) and vasopressor support (OR 10.465, 95%CI=5.901-18.557, $P = 0.000$) were statistically associated with hospital mortality.

## DISCUSSION

The current study compared various mortality prediction scoring systems and found that all the scores under-predicted the mortality in critically-ill cancer patients. Amongst the scoring systems considered, mortality predicted by MPM PDR was the closest to that of the actual mortality with an SMR of 1.305. AUROC values showed that all the seven scoring sytems, had good efficacy and acceptable discrimination. MPM PDR and SAPS III PDR achieved the best discrimination. We found the best sensitivity in SAPS II score (76.2%) and best specificity in SAPS III PDR score (92%). The Lemeshow-Hosmer goodness-of fit tests showed that APACHE III score had the best calibration although there was no statistically significant difference.

In current study, all the scores were significantly higher among non-survivors (p value< 0.001) as reported in the literature [13-18]. However, all the scores tested in this study underestimated the mortality (SMR >1), alike previous studies [14,15,19,20].

Discrimination is the ability to determine the patients who may die and who will survive. Measures of discrimination include sensitivity, specificity and AUROC curve. But no single scoring system excelled in all the three areas. SAPS III PDR and MPM $II_0$ PDR (AUROC=0.762) had the best AUROC values whereas sensitivity was at its best for SAPS II and specificity was at its best for SAPS III PDR. However, these differences were not statistically significant. In current study, AUROC outcomes showed that discrimination is acceptable in all the scoring systems tested in current study, as reported in the literature [14-16,20-22]. All the severity illness scores showed good efficacy with no statistically significant difference in AUROCs.

Calibration evaluates the accuracy of the degree of correspondence between the estimated probability of mortality and the observed actual mortality. Calibration is good, if the predicted mortality is close to observed mortality. APACHE III (4.704) had the best calibration with $P$ = 0.788. This infers that it had the least statistically significant discrepancy between the predicted and observed mortality. Good calibration of these scores have also been reported by other authors [14-16,20].

A significant correlation was found among various scoring systems (P<0.001) as per linear regression analysis. This correlation may be attributed to the overlap of multiple variables, considered for calculating the scores. Sculier *et al* too reported an excellent correlation between APACHE II and SAPS II in their study on oncology patients [21].

ICU mortality rate, among cancer patients, was reportedly high and in the range of 30 to 77% [23-26]. The overall ICU mortality in current study was 43.5%. Even though it is higher, the ICU mortality of the current cohort does not differ from the mortality reported in similar studies conducted earlier [23,24]. The hospital mortality in current study was 57.8% which is again similar as reported earlier [27,28].

Use of MV and vasopressor support have a direct association with hospital mortality. Similar studies conducted earlier have also reported the need for organ support in the form of MV. At times, vasopressor use is directly associated with increased mortality among cancer patients[29].

An ideal scoring system is the need of the hour. This system should be well calibrated, easy to compute, able to have high levels of discrimination and predict mortality rate with high accuracy based on the easily-available patient parameters. Additionally, an ideal score also needs to be dynamic, reflecting the change in management and case mix over time. In this search for an ideal scoring system, newer scoring systems have been developed. However, these systems are highly complex in nature, demand huge set of patient data and need computer assistance to calculate the scores. Hence, the development of an ideal scoring system is a long way to go.

The accuracy of scoring systems may differ over a period of time and may produce varied results in different countries due to differences in ethnicity, patient population, healthcare systems, ICU structure and organisation. So, its accuracy cannot be generalized and all such models need external validation in independent patient populations to prove its reproducibility. Therefore, it becomes imperative to compare and test the validity of scoring systems under different geographical areas and upon different patient populations. The current study is one of the few studies conducted in Indian

subcontinent, and the researchers have compared a huge number of scoring systems developed for cancer patients in a significantly large cohort of patients.

The current study has a limitation to address i.e., being a single centre retrospective study, concerns may arise in terms of generalizing the conclusions arrived in this study. The missing data may have also led to information bias. Nonetheless, the study has several salient features such as the comparison of seven scoring systems, fairly large sample size, well-defined study protocol, and the inclusion of only medical oncology patients.

## CONCLUSION

The current study concludes that all the scoring systems considered for this study cohort under-predicted the mortality. However, APACHE III score had the least discrepancy between the predicted and observed mortality. There was no statistically significant difference in efficacy and all the scores tested had good calibration and acceptable discrimination. Hence, the choice of scoring system in critically-ill oncology patients should not only be based on the performance of the score, but also on other factors such as ease of use and local preferences.

# 77369_Auto_Edited - check.docx

ORIGINALITY REPORT

# 12%
SIMILARITY INDEX

PRIMARY SOURCES

| | | | |
|---|---|---|---|
| 1 | cyberleninka.org<br>Internet | 75 words — | 3% |
| 2 | www.ijccm.org<br>Internet | 53 words — | 2% |
| 3 | ajcc.aacnjournals.org<br>Internet | 34 words — | 1% |
| 4 | f6publishing.blob.core.windows.net<br>Internet | 31 words — | 1% |
| 5 | Mona Mansour, Iman Galal, Enas Kassem. "Study of the role of different severity scores in respiratory ICU", Egyptian Journal of Bronchology, 2013<br>Crossref | 27 words — | 1% |
| 6 | www.ncbi.nlm.nih.gov<br>Internet | 22 words — | 1% |
| 7 | core.ac.uk<br>Internet | 18 words — | 1% |
| 8 | bmcinfectdis.biomedcentral.com<br>Internet | 17 words — | 1% |
| 9 | bmccardiovascdisord.biomedcentral.com<br>Internet | 15 words — | 1% |

**10** Jun Ho Lee, Young Hwan Lee, Yong Hwan Park, Yong Hwan Kim, Chong Kun Hong, Kwang Won Cho, Seong Youn Hwang. "The difference in C-reactive protein value between initial and 24 hours follow-up (D-CRP) data as a predictor of mortality in organophosphate poisoned patients", Clinical Toxicology, 2012
Crossref

14 words — < 1%

**11** "ESICM 2010 TUESDAY SESSIONS 12 October 2010", Intensive Care Medicine, 2010
Crossref

12 words — < 1%

**12** J. P. Thompson. "Severity scoring systems in the critically ill", Continuing Education in Anaesthesia Critical Care & Pain, 08/22/2008
Crossref

12 words — < 1%

**13** ccforum.com
Internet

12 words — < 1%

| EXCLUDE QUOTES | ON | EXCLUDE SOURCES | < 12 WORDS |
|---|---|---|---|
| EXCLUDE BIBLIOGRAPHY | ON | EXCLUDE MATCHES | < 12 WORDS |