Reliability of Different Predictive Systems in Major Gastrointestinal Surgical Patients

Mehmet Özdoğan¹, Riza Halduñ Gündoğdu², Faruk Karateke³, Eren Ersöy², Ebru Menekşe³, Hatice Kaya Özdoğan³, Ali Önder Devay², Adnan Kuvvetli³

¹Medical Park Hospital, Gaziantep, Turkey
²Atatürk Training and Research Hospital, 3. General Surgery Clinics, Ankara, Turkey
³Numune Training and Research Hospital, General Surgery Clinics, Adana, Turkey

Abstract

APACHE II-III, SAPS II, and MPM II are commonly used predictive models. The systems designed for surgical patients include the ASA and P-POSSUM. NRS-2002 score is suggested for screening of nutritional risk. We aimed to assess the performance of prognostic models, and to compare the reliability of NRS-2002 with those in major gastrointestinal surgical patients. APACHE II and III, SAPS II, MPM II and P-POSSUM scores, ASA grading and NRS-2002 scores of the patients underwent major gastrointestinal surgery were collected on admission. Calculations were repeated for APACHE II and III, SAPS II, and MPM II at postoperative 24th hour. Discrimination and calibration characteristics of the scoring systems were evaluated. APACHE II-III, and SAPS II at postoperative 24th hour, and P-POSSUM on admission, had reliable power of discrimination and calibration for mortality prediction in patients undergoing major gastrointestinal surgery. APACHE III, SAPS II and P-POSSUM had the best performance on complication prediction. ASA, MPM II, and NRS-2002 had insufficient discrimination statistics. APACHE II and III, SAPS II, and P-POSSUM are superior to other systems in predicting mortality in gastrointestinal surgical patients. APACHE III, SAPS II, and P-POSSUM are superior in predicting complication.

Keywords: APACHE II, SAPS, MPM, POSSUM, NRS-2002, mortality prediction

Özet


Introduction

There is a number of scoring systems in general use for outcome prediction in hospitalized patients. Common use of most of these systems is illness severity scoring to make comparisons between patient groups or intensive care units (ICU) (1,2). However, they have been used to assess or predict the risk for specific patient groups.

In the context of general surgical patients, there are two groups of scoring system for risk prediction which are widely used: The first group consists of systems originally designed to predict disease severity and probability of hospital mortality and complications in general ICU patients. Acute Physiology and Chronic Health Evaluation II and III (APACHE II and III), Simplified Acute Physiology Score II (SAPS II), and Mortality Probability Model (MPM II) are the most commonly used systems in the first group. There has been a controversy on which predictive system among them is more reliable. The second group includes the systems particularly designed for surgical patients. The American Society of Anesthesiologists (ASA) grading is the simplest subjective system and has been widely used. The most commonly used objective system in this group

Correspondence: Faruk Karateke, Numune Research and Education Hospital, General Surgery Clinic, Adana, Turkey
Tel: +903223550000 karatekefaruk@hotmail.com

Received: 07.08.2014 Accepted: 25.11.2014
ISSN 2148-3132 (print) ISSN 2148-2926 (online)
www.gaziantemedicaljournal.com
DOI: 10.5455/GMJ-30-166055
is Portsmouth variant of Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (P-POSSUM). One of the handicaps of the all of the above-mentioned systems is that they all ignore the nutritional status of the patient.

Malnutrition has been known to be associated with increased rates of complications, length of hospital stay (LOS), and mortality (3-7). Nutritional support decreases complication and infection rates, and reduces LOS in malnourished patients (8,9). Nutritional risk screening 2002 (NRS-2002) score was introduced by the European Society for Clinical Nutrition and Metabolism (ESPEN) as a useful method for screening of nutritional risk in hospitalized patients(10). When the NRS-2002 score is ≥3 patients are accepted as nutritionally under risk (10-13).

True validity of a risk screening or predicting tool can only be discussed in the context of its impact on clinical outcome. The aims of this study were to determine the reliability of systems designed to predict disease severity and probability of mortality and complications, and to compare the reliability of NRS-2002 with different scoring systems in patients undergoing major gastrointestinal surgery. To our knowledge, the present study is the first study in the English literature aiming to compare reliability of a nutritional risk screening index with that of operative scoring systems and severity of illness scores in prediction of surgical outcome.

Materials and Methods
This study was approved by Adana Numune Training and Research Hospital Education, Planning, Coordination and Ethic Committee.

This study was conducted in 175 consecutive patients who underwent major gastrointestinal surgery in Atatürk Training and Research Hospital, 3rd Clinic of General Surgery and Adana Numune Training and Research Hospital, General Surgery Clinic. The major gastrointestinal surgery was defined as operations which were performed due to esophageal, gastric, colorectal, and periampullary malignancies.

The main outcome measures of the study were the reliability of the evaluated scoring systems in predicting the mortality and complication rates during postoperative 30 days. Complications that were taken into consideration included wound-related complications (surgical site infection, evisceration and dehiscence), cardiac complications (arrhythmias, ischemic events), pulmonary complications (atelectasis, respiratory failure, adult respiratory distress syndrome, and pneumonia), gastrointestinal leaks, fistulas and intraabdominal abscesses, and other complications (acute renal failure, acalculous cholecystitis, pressure ulcer). Complications were identified prospectively during the study period.

Data related to calculations of APACHE II and III, SAPS II, MPM II and P-POSSUM scores, ASA grading, and NRS-2002 score were collected on admission. Calculations were repeated for APACHE II and III, SAPS II, and MPM II at postoperative 24th hour. Calculations were performed by using calculators of the Muavenet Intensive Care Information System (http://www.icu.hacettepe.edu.tr/micis.html). Informed consents of all patients were taken.

Statistical Analysis
Discrimination and calibration define the overall predictive power of a model. “Discrimination” refers to ability of a model to distinguish patients who experienced an event from those who did not. Discrimination was measured by the receiver operating characteristic (ROC) curves. The area under the curve (AUC) represents the probability that a patient who experienced the event had a higher predicted probability of having that event than a patient who did not(14). The higher the true-positive rate is relative to the false-positive rate, the greater is the AUC. An AUC of 0.5 indicates that the model does not predict better than chance. The discrimination power of a model is considered perfect if AUC=1, good if AUC>0.8, moderate if AUC is between 0.6 and 0.8, and poor if AUC<0.6.

“Calibration” refers to the agreement between the “predicted probabilities” and the “true probabilities”. Calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test and the corresponding calibration curves. P value less than 0.05, which means significant difference exists between observed and predicted event, indicates a lack of fit of the model (15).

Continuous variables were presented as means ±SEM and were compared using Mann-Whitney U test. Categorical values were analyzed using Chi-square test. P<0.05 was considered as statistically significant. Statistical evaluation was performed by using STATA 8.0 and SPSS 11.0 statistical packages.

Results
During the study period 175 patients underwent surgery due to gastrointestinal malignancies (Table I). There were 77 patients with esophageal-gastric malignancies, 92 patients with colo-rectal malignancies, and 12 patients with periampullary malignancies. A total of 16 patients died. The causes of mortality were cardiopulmonary problems in ten patients and intra-abdominal sepsis in six patients. Significantly increased percentage of non-survivors had complication.

Table II shows overall scores and the differences between hospital survivors and non-survivors. Non-survivors had significantly higher APACHE II,
Many illness severity scoring systems have been produced but a few are currently in clinical use. The most commonly used systems include APACHE II and III, SAPS II and MPM II(2). These systems consider objective physiological criteria, age of the patient or the nature of the intended operation (1). The scoring system specifically designed for surgical patients is POSSUM, and it is accepted as the most appropriate score to predict the operative risk (16,17). It consists of a physiological score and an operative severity score, and considers both for predicting operative mortality. Variations of POSSUM scoring, especially P-POSSUM(18), has been suggested to be more reliable in gastrointestinal surgery(19).

Furthermore, it can be used as a part of risk stratification to identify patients who might benefit from postoperative ICU care(1). The present study is in accordance with the literature, demonstrating the reliability of the P-POSSUM system in gastrointestinal surgery patients. On the other hand, widely used ASA score was not found to be reliable for mortality and complication prediction.

### Table 1.
The study population (Data on Age and LOS are presented as mean ±SD).

<table>
<thead>
<tr>
<th>Number</th>
<th>All</th>
<th>Mortality (-)</th>
<th>Mortality (+)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>175</td>
<td>159</td>
<td>16</td>
<td>NS</td>
</tr>
<tr>
<td>LOS</td>
<td>62±1.6</td>
<td>61.7±1.5</td>
<td>67.8±2</td>
<td>NS</td>
</tr>
<tr>
<td>Patients with complications</td>
<td>19.8±1.8</td>
<td>19.9±2.1</td>
<td>17.7±3.5</td>
<td>NS</td>
</tr>
<tr>
<td>72</td>
<td>58</td>
<td>14</td>
<td>0.032</td>
<td></td>
</tr>
</tbody>
</table>

LOS: Length of stay
NS: Non-significant

### Table 2.
The scores of evaluated systems in patients

<table>
<thead>
<tr>
<th>All</th>
<th>Mortality (-)</th>
<th>Mortality (+)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>APACHE II</td>
<td>7.6±0.6</td>
<td>9.7±1.3</td>
<td>NS</td>
</tr>
<tr>
<td>APACHE IIpo24</td>
<td>7.6±0.7</td>
<td>12.3±2.0</td>
<td>0.018</td>
</tr>
<tr>
<td>APACHE III</td>
<td>26.2±1.6</td>
<td>32.3±6.7</td>
<td>NS</td>
</tr>
<tr>
<td>APACHE IIIpo24</td>
<td>24.1±1.9</td>
<td>37.3±5.9</td>
<td>NS</td>
</tr>
<tr>
<td>SAPS II</td>
<td>23.2±0.9</td>
<td>26.5±5.2</td>
<td>NS</td>
</tr>
<tr>
<td>SAPS IIpo24</td>
<td>16.7±1.2</td>
<td>24.3±4.8</td>
<td>0.041</td>
</tr>
<tr>
<td>MPM II</td>
<td>8.7±0.6</td>
<td>9.0±2.1</td>
<td>NS</td>
</tr>
<tr>
<td>MPM IIpo24</td>
<td>8.5±0.4</td>
<td>6.4±1.4</td>
<td>NS</td>
</tr>
<tr>
<td>NRS-2002</td>
<td>3.3±0.4</td>
<td>3.1±0.6</td>
<td>NS</td>
</tr>
<tr>
<td>ASA</td>
<td>1.9±0.1</td>
<td>2.1±0.3</td>
<td>NS</td>
</tr>
<tr>
<td>P-POSSUM</td>
<td>21.3±0.7</td>
<td>24.8±0.4</td>
<td>NS</td>
</tr>
<tr>
<td>P-POSSUMIps</td>
<td>14.3±0.5</td>
<td>15.6±1.5</td>
<td>NS</td>
</tr>
</tbody>
</table>

APACHE: Acute Physiology and Chronic Health Evaluation
SAPS: Simplified Acute Physiology Score
MPM: Mortality Probability Model
NRS-2002: Nutritional risk screening 2002
ASA: The American Society of Anesthesiologists

**Discussion**

Patients undergoing a major surgical procedure are at higher risk of developing complications and mortality. The magnitude of this risk is determined by the patient and surgery related factors. Several scoring systems have been used for the assessment of operative risk. The simplest and most widely used method has been ASA scoring, however it does not consider objective physiological criteria, age of the patient or the nature of the intended operation (1). The scoring system specifically designed for surgical patients is POSSUM, and it is accepted as the most appropriate score to predict the operative risk (16,17). It consists of a physiological score and an operative severity score, and considers both for predicting operative mortality. Variations of POSSUM scoring, especially P-POSSUM(18), has been suggested to be more reliable in gastrointestinal surgery(19).
based on the findings that the scores at postoperative 24th hour had moderate to good discrimination power. Beginning from the admission, prospective scoring using APACHE II and III, and SAPS II that can reveal continuous improvement or deterioration of scores may lead to a reliable prediction of mortality in this patient group. APACHE III and SAPS II can also be used for postoperative complication prediction.

Although Kilic et al. and Arabi et al. have previously shown that MPM II had best discrimination for prediction of mortality in general surgical ICU patients (26) and in patients with severe sepsis and septic shock (27), the present study demonstrated insufficient discrimination and calibration statistics for MPM II in gastrointestinal surgical patients. A possible explanation for this result is that MPM II system has steep cut off points for perfusion parameters: 90 mmHg for systolic pressure and 150/min for pulse rate. Most of our patients did not have these extreme hemodynamic values postoperatively, and so did not need postoperative ICU stay.

One of the major drawbacks of all above-mentioned prediction systems is they do not specifically take the nutritional status of the patient into account. Numerous tools for the screening of malnutrition and nutritional risk have been proposed. The NRS-2002, endorsed by ESPEN, consists of a nutritional score, a disease severity score, and an age adjustment for patients aged > 70 years. Total score is calculated and patients are classified as at no risk to high risk (10,12). Although NRS-2002 has not been specifically constructed for use in surgical patients, it was found as a sensitive screening tool in patients undergoing elective surgery (13,28). A high NRS-2002 score is significantly associated with increased complication rate and prolonged LOS (11,13,28,29). The purpose of nutritional screening is to predict the probability of a better or worse outcome due to nutritional factors, and whether nutritional treatment is likely to influence it (10). Therefore, an ideal nutritional screening tool is supposed to predict postoperative mortality and complication rates in surgical patients, so that nutritional intervention can be provided in high risk patients. To our knowledge, this is the first study in the English literature comparing reliability of a nutritional risk screening index with that of operative scoring systems and severity of illness scores. The present data indicate an insufficient discrimination statistics for NRS-2002 for postoperative mortality and complication prediction in major gastrointestinal surgical patients. Our results contradict with that of Schiesser et al., the only study in the literature which evaluated the value of NRS-2002 score in predicting the complications specifically in gastrointestinal surgery (28). They found that the patients with increased nutritional risk according to the NRS-2002 score had significantly more severe complications. They compared complication rates within groups with and without increased risk by using Pearson chi-square test. However, we could not demonstrate a sufficient discrimination power for NRS-2002 score by ROC analysis. Although NRS-2002 considers disease severity, it does not take into account the type and extent of the surgical procedure, and this might be responsible for this result.

Conclusion
Overall, the preoperative risk formulas are, in their current form, useful mainly for 1) clinical research to quantify the degree of risk in a study sample or to evaluate whether two study groups are comparable; and 2) quality improvement programs. The present study, in conclusion, showed that APACHE II and III, SAPS II, and P-POSSUM are superior to other scoring systems in predicting mortality in major gastrointestinal surgical patients. APACHE III, SAPS II, and P-POSSUM are superior in predicting complication. Reliabilities of these models are superior when evaluated postoperatively.

References


How to cite: