Reliability of Different Predictive Systems in Major Gastrointestinal Surgical Patients

Major Gastrointestinal Cerrahi Hastalarında Farklı Prediktif Sistemlerin Güvenilirliği

Mehmet Özdoğan[1], Rıza Haldun Gündoğdu[2], Faruk Karateke[3], Eren Ersoy[2], Ebru Menekşe[3], Hatice Kaya Özdoğan[3], Ali Önder Devay[2], Adnan Kuvvetli[3]

[1]Medical Park Hospital, Gaziantep, Turkey
[2]Atatürk Training and Research Hospital, 3. General Surgery Clinics, Ankara, Turkey
[3]Numune 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
APACHE II-III, SAPS II ve MPM II sık kullanılan prediktif modellerdir. ASA ve P-POSSUM cerrahi hastalar için dizayn edilmişlerdir. NRS-2002 skorlaması nütrisyonel riskin taranması için önerilmektedir. Bu çalışmada, prognostik modellerin etkinliğini değerlendirmek ve major gastrointestinal cerrahi hastalarda bu modeller ile NRS 2002'nin güvenilirliğini karşılaştırmak amaçlanmıştır. Major gastrointestinal cerrahi uygulanan hastaların başvuru anındaki APACHE II ve III, SAPS II, MPM II ve P-POSSUM skorları, ASA ve NRS-2002 skorları kayıt edildi. Ameliyat sonrası 24.saatte APACHE II ve III, SAPS II,ve MPM II skorları kayıt edildi. Skorlama sistemlerinin farklılığı ve kalibrasyon özellikleri değerlendirildi. Postoperatif 24. saatte APACHE II-III, SAPS II ve hastaneye başvuruda P-POSSUM'un major gastrointestinal cerrahiye giden hastaların mortalite tahminindeki ayırt edici ve değerlendirici gücü güvenilirdi. APACHE III, SAPS II ve P-POSSUM komplikasyon tahmini için en iyi performansı gösteriyordu. ASA, MPM II, ve NRS-2002' nin ise bu hastaları ayırt edici gücü yetersizdi. APACHE II ve III, SAPS II ve P-POSSUM gastrointestinal cerrahi hastalarda mortalite tahmininde diğer hastalara göre üstündür. Komplikasyonların tahmininde ise APACHE III, SAPS II ve P-POSSUM üstündür.
Anahtar kelimeler: APACHE II, SAPS, MPM, POSSUM, NRS-2002, mortalite tahmini

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.gaziantepmedicaljournal.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 have 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 06 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 71 patients with esophago-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,

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

	All	Mortality (-)	Mortality (+)	*P*
Number	175	159	16	
Age (years)	63±1.6	61.7±1.5	67.8±2	NΣ
LOS	19.8±1.8	19.9±2.1	17.7±3.5	NΣ
Patients with complications	72	58	14	**0.032**

LOS: Length of stay
NS: Non-significant

Table 2. The scores of evaluated systems in patients

	All	Mortality (-)	Mortality (+)	*P*
APACHE II_0	7.8±0.6	7.6±0.6	9.7±1.3	NS
APACHE II_{po24}	7.6±0.7	6.8±0.4	12.3±2.0	**0.018**
APACHE III_0	26.2±1.6	25.8±1.8	32.3±6.7	NS
APACHE III_{po24}	24.1±1.9	22.4±2.1	37.3±5.9	**0.029**
SAPS II_0	23.2±0.9	22.8±2.0	26.6±5.2	NS
SAPS II_{po24}	16.7±1.2	15.2±1.3	24.3±4.8	**0.041**
MPM II_0	8.7±0.6	8.6±0.4	9.0±2.1	NS
MPM II_{po24}	4.8±0.4	4.5±0.3	6.4±1.4	NS
NRS-2002	3.3±0.4	3.3±0.3	3.1±0.6	NS
ASA	1.9±0.1	1.9±0.9	2.1±0.3	NS
P-POSSUM$_{Phys}$	21.3±0.7	20.8±0.6	24.8±0.4	**0.024**
P-POSSUM$_{Opsv}$	14.3±0.5	14.2±0.7	15.6±1.5	NS

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

APACHE III, and SAPS II scores on postoperative 24th hour. They also had significantly higher P-POSSUM physiological severity score.

Discrimination and calibration statistics for mortality and complication prediction were presented in table III. None of the systems had perfect or good discrimination power based on the finding that all AUC values are below 0.8. Among the evaluated systems, APACHE II_{po24}, APACHE III_{po24}, SAPS II_{po24}, and P-POSSUM were best performing for mortality prediction. APACHE III_{po24}, SAPS II_{po24}, and P-POSSUM were best performing for complication prediction. Calibration characteristics were statistically adequate for these models.

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

patient 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.

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 for scoring so that standardized comparisons can be performed between patient groups and between ICUs. However, to some extent, they can be used to predict risk or to assess a probability of hospital death for general ICU patient groups(1,2). There have been controversies in the use of these systems in general surgical patients, since they do not consider the extent of the surgical procedure. However, attempts have been made to apply and evaluate the performances of these scoring systems in patients with cancer(20), trauma(21), postoperative abdominal complications (22), emergency surgery (23), and elective surgery (16,24,25). In a recent critical review, den Boer et al. demonstrated that APACHE II, SAPS II, and MPM II had poor to good discrimination performances, and generally underestimated the mortality risk in solid organ cancer patients (20). The most widely used of these systems, APACHE II, has been shown to be able to predict mortality (25), and even the increasing levels of both local and systemic complication rates in elective surgery (16,24,25). Our results show that APACHE II is useful in mortality prediction, but not in complication prediction, in surgical patients with gastrointestinal malignancy. The present study confirms that the general severity of illness scores should be used postoperatively in surgical patients,

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

1. Boyd O, Jackson N. How is risk defined in high-risk surgical patient management? Crit Care 2005;9:390-396.
2. Le Gall JR. The use of severity scores in the intensive care unit. Intensive Care Med 2005;31:1618-1623.
3. Correia MI, Waitzberg DL. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. Clin Nutr 2003;22:235-239.
4. Reilly JJ Jr, Hull SF, Albert N, et al. Economic impact of malnutrition: a model system for hospitalized patients. J Parenter Enteral Nutr 1988;12:371-376.
5. Kuzu MA, Terzioğlu H, Genç V, et al. Preoperative nutritional risk assessment in predicting postoperative outcome in patients undergoing major surgery. World J Surg 2006;30:378-390.
6. Schneider SM, Veyres P, Pivot X, et al. Malnutrition is an independent factor associated with nosocomial infections. Br J Nutr 2004;92:105-111.
7. Correia MI, Waitzberg DL. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. Clin Nutr 2003;22:235-239.
8. Waitzberg DL, Saito H, Plank LD, et al. Postsurgical infections are reduced with specialized nutrition support. World J Surg 2006;30:1592-1604.
9. Xu J, Zhong Y, Jing D, et al. Preoperative enteral immunonutrition improves postoperative outcome in patients with gastrointestinal cancer. World J Surg 2006; 30:1284-1289.
10. Kondrup J, Allison SP, Elia M, et al. Educational and Clinical Practice Committee, European Society of Parenteral and Enteral Nutrition (ESPEN). ESPEN guidelines for nutrition screening 2002. Clin Nutr 2003;22:415-421.
11. Johansen N, Kondrup J, Plum LM, et al. Effect of nutritional support on clinical outcome in patients at nutritional risk. Clin Nutr 2004;23:539-550.
12. Kondrup J, Rasmussen HH, Hamberg O, et al. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. Clin Nutr 2003;22:321-336.
13. Sorensen J, Kondrup J, Prokopowicz J, et al; EuroOOPS study group. EuroOOPS: An international, multicentre study to implement nutritional risk screening and evaluate clinical outcome. Clin Nutr 2008;27:340-349.
14. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.

15. Lemeshow S, Hosmer DW Jr. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115:92-106.
16. Ribeiro KC, Kowalski LP. APACHE II, POSSUM and ASA scores and the risk of perioperative complications in patients with oral or oropharyngeal cancer. Arch Otolaryngol Head Neck Surg 2003;129:739-745.
17. Jones HJ, de Cossart L. Risk scoring in surgical patients. Br J Surg 1999;86:149-157.
18. Prytherch DR, Whiteley MS, Higgins B, et al. POSSUM and Portsmouth POSSUM for predicting mortality.Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. Br J Surg 1998;85:1217-1220.
19. Tekkis PP, Kocher HM, Bentley AJ, et al. Operative mortality rates among surgeons: comparison of POSSUM and p-POSSUM scoring systems in gastrointestinal surgery. Dis Colon Rectum 2000;43:1528-1532.
20. den Boer S, de Keizer NF, de Jonge E. Performance of prognostic models in critically ill cancer patients - a review. Crit Care 2005;9:R458-463.
21. Rutledge R, Fakhry S, Rutherford E, et al. Comparison of APACHE II, Trauma Score, and Injury Severity Score as predictors of outcome in critically injured trauma patients. Am J Surg 1993;166:244-247.
22. Altomare DF, Serio G, Pannarale OC, et al. Prediction of mortality by logistic regression analysis in patients with postoperative enterocutaneous fistulae. Br J Surg 1990;77:450-453.
23. Koperna T, Semmler D, Marian F. Risk stratification in emergency surgical patients: is the APACHE II score a reliable marker of physiological impairment? Arch Surg 2001;136:55-59.
24. Grant CA, Dempsey GA, Lowe D, et al. APACHE II scoring for the prediction of immediate surgical complications in head and neck cancer patients. Plast.Reconstr. Surg, 2007;119:1751-1758.
25. Goffi L, Saba V, Ghiselli R, et al. Preoperative APACHE II and ASA scores in patients having major general surgical operations: prognostic value and potential clinical applications. Eur J Surg 1999;165:730-735.
26. Kilic YA, Dogrul AB, Karakoc D. Impact of organ failure on mortality prediction in a Turkish surgical intensive care unit. Int Care Med 2005;31(suppl):S47.
27. Arabi Y, Al Shirawi N, Memish Z, et al. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study. Crit Care 2003;7:R116-122.
28. Schiesser M, Müller S, Kirchhoff P, et al. Assessment of a novel screening score for nutritional risk in predicting complications in gastro-intestinal surgery. Clin Nutr 2008;27:565-570.
29. Kyle UG, Kossovsky MP, Karsegard VL, et al. Comparison of tools for nutritional assessment and screening at hospital admission: a population study. Clin Nutr 2006;25:409-417.

How to cite:

Özdoğan M, Gündoğdu RH, Karateke F, Ersoy E, Menekşe E, Özdoğan H K, Devay A Ö, Kuvvetli A. Reliability of Different Predictive Systems in Major Gastrointestinal Surgical Patients. Gaziantep Med J 2015; 21 (1): 21-25.