

The analysis of scoring systems predicting mortality in geriatric emergency abdominal surgery

Murat Özban, M.D., Onur Birsen, M.D., Mahmut Şenel, M.D., Akin Özden, M.D., Burhan Kabay, M.D.

Department of General Surgery, Pamukkale University Faculty of Medicine, Denizli

ABSTRACT

BACKGROUND: Accurate measurement of surgical outcomes, proper evaluation of hospitals and surgeons regardless of case can be performed by mortality prediction models. The aim of this study was to analyze factors affecting mortality, present our clinical experience and patient profile and evaluate different scoring systems in use of these patients.

METHODS: A retrospective review of one hundred and twelve geriatric patients who underwent major abdominal emergency surgery between 2004 and 2008 was performed. APACHE II, ODIN, SAPS II expanded, P-POSSUM, Mannheim peritonitis and Charlson comorbidity index, Goldman and ASA scores were calculated using patient data. Sensitivity, positive predictive value and Odd's ratio were calculated to predict the mortality for these scoring systems.

RESULTS: The overall mortality rate for our patients was found 33.9%. The factors affecting mortality in this study were found to be the duration of initial complaint, requirement of intensive care unit, requirement of mechanical ventilation and its duration, the presence of coexisting disease and peritonitis.

CONCLUSION: According to our study, in this particular group of patients, APACHE II scoring system is more valid and accurate in estimating the mortality risk when compared to other scoring systems.

Key words: Emergency surgery; geriatrics; scoring system.

INTRODUCTION

The definition of 'the aged' changes very rapidly. Even though most authors agree to limit the age of 65, this number is not under control and may vary with longevity and overall health of community.^[1] The elderly has many concomitant diseases limiting functional capacity required for postoperative recovery. In order to ensure the ability and evaluate pathophysiological risks of concomitant diseases and clinical competence, it is essential to provide a safe, beneficial and effective surgical care.^[2]

Elderly patients have more complications after abdominal surgery. Compared to the general population, it is three times

more common in appendix perforation. Similarly, complications of acute cholecystitis is more common in the elderly.^[1] Morrow has determined that 40% of elderly patients with acute cholecystitis also have gallbladder empyema, gangrenous cholecystitis, subphrenic or hepatic abscess or free perforation.^[1] Consequently, in geriatric patients, it is required to select certain indications for reducing the risks of surgery besides an optimal preoperative preparation, the short duration of the operation, the appropriate method of anesthesia and resuscitation and a proper postoperative care.

The purpose of this study was to investigate the factors influencing mortality in this group of patients, which is increasing by number and requiring a multidisciplinary approach, introduce our clinical experience and patient profiles, analyze scoring systems envisioning mortality rates, and determine the most appropriate scoring system for our clinic.

MATERIALS AND METHODS

One hundred and twelve geriatric patients operated as emergency cases were included into the study between October 2004 and December 2008 at Pamukkale University School of Medicine Training and Research Hospital, Department of General Surgery. Ethics committee approval No.4830 was re-

Address for correspondence: Murat Özban, M.D.
Pamukkale Üniversitesi Tıp Fakültesi, Genel Cerrahi Anabilim Dalı,
Denizli, Turkey
Tel: +90 258 - 444 07 28 E-mail: muratozban@yahoo.com

Quick Response Code



Ulus Travma Acil Cerrahi Derg
2015;21(3):182-186
doi: 10.5505/tjtes.2015.05046

Copyright 2015
TJTES

ceived before the study. All patients were selected among the emergency department patients aged 65 and over. Patients were reviewed retrospectively.

Age, sex, length of stay, intensive care unit stay, duration of mechanical ventilation, referral status, social security, complaint, duration of complaint, preoperative diagnosis, postoperative diagnosis, APACHE II score, ASA score, the Goldman cardiac risk index, Charlson comorbidity index, Mannheim Peritonitis Index, ODIN score, the expanded SAPS II score, P-POSSUM score, the risk of venous thromboembolism, the causes of morbidity and mortality were recorded as parameters to data forms. Mortality was analyzed in two groups as compatible and incompatible causes with life, as defined by Seymour and Pringle.^[3] APACHE II, ODIN, the expanded SAPS II and P-POSSUM scores were calculated automatically from www.sfar.com website.

The ability to predict mortality in the validation study was evaluated through observed over expected mortality ratio (O:E ratio) and calibration analysis.

Mortality O:E Ratio

'O' was defined as the number of observed patients who died during the study period. 'E' was defined as the number of operated patients expected to die, multiplied by the mean risk of mortality, expressed as a percentage.

The mean risk of mortality was the mean of estimated risk of mortality given by the 'R' in the logistic regression equation of the P-POSSUM score, calculated in each band risk of mortality.

O:E ratio methodology can be used as a cross sectional audit, as well as for continuous or sequential monitoring of surgical quality performance between subgroups.^[4]

Calibration

Calibration (model fit), assessed by the Hosmer-Lemeshow (HL) statistic is a goodness-of-fit test analogous to a Chi-square statistic where the degrees of freedom (df) equal the number of deciles or the number of risk bands minus. The HL statistics, H² and C², indicate the degree of agreement between the observed and the expected mortality across risk ranges. When p>0.05, the calibration was considered to be 'good'; i.e., the observed mortality was well described, whereas when p≤0.05, calibration was considered to be 'poor'.^[5]

All data were given as the average ± standard error of mean. The differences between the groups were analyzed with Chi-square and Kruskal Wallis test, the differences between two groups were analyzed with Mann-Witney U-test. Statistical analysis was performed by using SPSS software (SPSS Inc., Chicago, IL). Results with %95 confidence interval, significance of p<0.05 were evaluated.

RESULTS

Forty-seven (42%) patients were female and 65 (50%) patients were male. The mean age, duration of symptoms and hospital stay of our patients are given in Table 1. Morbidity was detected in 47 patients and morbidity rate was found to be 63.51%. Thirty-one patients were determined to have pneumonia and atelectasis, 18 patients to have wound infection, 10 patients to have acute renal failure, and 8 patients to have acute coronary syndrome. Mortality was detected in 38 patients (33.9%). Mortality was analyzed in two groups as compatible and incompatible causes with life, as defined by Seymour and Pringle.^[3] The causes of mortality, incompatible with life, are given in Table 2.

Twenty-five (22.3%) patients were detected in the group of compatible with life. Eight (7.1%) of these patients were identified with sepsis, two (1.8%) with anastomosis leakage, two (1.8%) with bleeding, five (4.5%) with pneumonia, seven (6.3%) with myocardial infarction, and one (0.9%) patient with pulmonary embolism were identified. Peritonitis was established in 43 (38.39%) of our cases. Mortality for patients with peritonitis was found to be 41.86% in eighteen patients. Positive predictive values and sensitivity of each scoring systems were given in Table 3.

As determined by Wijesingheve et al, the linear analysis for SAPS II expanded, ODIN, APACHE II scores and both linear and geometric analysis for P-POSSUM score were performed.^[6] The results are given in Table 4.

The presence and severity of peritonitis is a factor affecting mortality in patients. Peritonitis was detected in 38.39% of our patients. Mortality for patients with peritonitis was found to be 41.86%. While 39.53% of patients in the group of

Table 1. The mean age, duration of symptoms and hospitalization

	Age (year)	Duration of symptoms (hour)	Hospitalization (hour)
Mean±SD	74.30±0.5	71.10±5.7	261.47±21.9
Minimum	65	10	8
Maximum	87	360	1680

Table 2. The causes of mortality which incompatible with life

Mortality causes	n	%
Terminal stage of malignancy	5	4.5
Massive bowel infraction	8	7.1
Total	13	11.6

Table 3. The statistical analysis of the scoring systems

	Odds' ratio	%95 CI	p	Sensitivity (%)	Positive predictive value
APACHE II	4.00	2.80-5.19	0.0001	96	86
P-POSSUM	4.43	3.06-5.80	0.0001	87	95
ODIN	2.96	1.93-4.00	0.0001	78	91
SAPS II Expanded	3.2	2.23-4.35	0.0001	74	89
Peritonitis MPI	4.06	1.40-5.77	0.0001	68	94

Table 4. Prediction of mortality

	n	Estimated mortality (n)	Observed mortality (n)	Observed: Estimated mortality ratio
P-POSSUM	112	54	38	0.70
APACHE II	112	35	38	1.08
ODIN	112	32	38	1.18
SAPS II Expanded	112	25	38	1.52

peritonitis score 24 and above died, only 2.32% of patients in the group of peritonitis score less than 24 died. Mortality in patients with Mannheim peritonitis score 24 and above was found to be significantly increased.

DISCUSSION

Basically, all scoring systems aim to predict the risk of death by calculating the value of certain vital, clinical and laboratory findings. A scoring system to predict surgical outcomes via calculating the risk of mortality and morbidity must be a valid and reliable method. A lot of earlier scoring systems were very complex and frequently failed to reach an accurate result in estimating the risk of mortality. There are two objective measures that have been used to evaluate the performance of the prognostic models: calibration and discrimination. Calibration determines the correlation between the probability of mortality with the observed mortality and can be tested using Hosmer-Lemeshow goodness-of-fit statistical analysis. On the other hand, discrimination shows the usefulness of a scoring system in classifying patients properly as survivors or non-survivors and is measured by AUC. While it would be important for a scoring system to have a good discrimination for a specific patient, stratifying patients for clinical trials or comparison of the results and quality of care between different clinics, calibration is preferable. Consequently, we evaluated the performance of selected scoring systems according to calibration analysis in our study.

POSSUM scoring system obtained from a heterogeneous population was successfully used as an audit tool, but in low-risk groups, it seemed that it overestimated mortality rates. Predictive value of POSSUM to morbidity was affected by the

type of surgery and the sample size of studies.^[7] Compared with POSSUM, P-POSSUM was more accurate for predicting postoperative mortality.^[8] Therefore, P-POSSUM scoring system designed and used successfully for high risk group of patients.^[9] However, in a study evaluating gastrointestinal surgery in elderly patients, both the POSSUM and P-POSSUM over-predicted the morbidity and mortality.^[10] Similarly, in our study, P-POSSUM scoring system overpredicted the mortality rates when compared with other scoring systems according to both linear analysis and O:E results (Tables 3, 4).

APACHE II is a scoring system in which especially vital signs of the patient besides hematologic and biochemical laboratory findings are evaluated. In a recent study, the discriminative ability of APPACHE II has been found to be excellent in surgical intensive care unit patients.^[11] Moreover, APACHE II also had a better, more appropriate calibration than APACHE III or SAPS II in the same study. APACHE II scores can prognostically stratify acutely ill patients and assist investigators comparing the success of new or differing forms of therapy. Hospital mortality is predicted using the APACHE II score, the principal diagnostic category with which the patient is admitted to ICU and also depending on whether or not the patient required emergency surgery.^[12] In a recent study, APACHE II has also had a better, more appropriate calibration than other widely used scoring systems; so only APACHE II properly predicts mortality risk.^[11] The major limitation of this scoring system is the variability of physiological parameters which are all dynamic and can be influenced by multiple factors, including ongoing resuscitation and treatment, hence, time bias is present. It may explain our findings for APACHE II that although the sensitivity was found to be 96%, positive

predictive value was found to be lower (86%) than others. In spite of these limitations, with calibration analysis, it was found that APACHE II scoring system more accurately predicted the mortality risk when compared to other scoring systems in our patients (Tables 3, 4).

ODIN scoring system includes parameters such as oxygen pressure determined in basic organ functions, arterial pressure, serum creatinine, Glasgow coma score, serum bilirubin level, hematological parameters. The assessment of the presence of infection is the difference of ODIN scoring system from APACHE II. Therefore, although positive predictive value was found to be high (91%), the sensitivity of this system also was found to be quite low (78%). Basically, all scoring systems obtained parameters by calculating vital and laboratory evaluation of patients. While these systems were being developed and modified, the follow-up process and diseases other than these criterias of patients have been added to the evaluated parameters over time. Although SAPS II has similar features, 'SAPS II expanded' has included the duration of hospitalization before intensive care. However, the values of age and gender that it contained were not found as factors affecting mortality in our study. Therefore, the sensitivity (74%) and positive predictive value (89%) were found quite low.

In this study, we investigated the scoring systems that predicted the factors affecting mortality and morbidity rate in geriatric patients, which is increasing by number and requiring a multidisciplinary approach. The mortality rate was found to be 33.9% and the morbidity rate was 63.51%. The factors affecting mortality in our study were found to be the duration of initial complaint, requirement of intensive care unit, requirement of mechanical ventilation and its duration, the presence of coexisting disease and peritonitis. Conversely, poor calibration (goodness-of-fit), especially in subgroup analysis, and underestimation or overestimation of O:E ratios considerably limits the value of P-POSSUM for prediction of mortality in individuals. Therefore, P-POSSUM should not be used in comparison of different clinics and stratifying the patients for in-clinical trials to predict outcomes.^[13]

The present study has some limitations. As a single-centre study, there may be bias with regard to case mix, quality clinical care and policy. In addition, our relatively small sample size is a limiting factor for calibration analysis. Furthermore, APACHE II is based on retrospective data that is available within 24 h of ICU admission; consequently, the sampling rate that is used can influence mortality prediction. In our opinion, a multi-centre study would diminish the concerns over case mix and benefit from a larger sample size.

A better calibration of APACHE II was found besides improving the ability to predict hospital mortality risk in comparison with P-POSSUM, SAPS II expanded or ODIN in geriatric group of patients who underwent emergency abdominal surgery.

Conflict of interest: None declared.

REFERENCES

- Hendrickson M, Naparst TR. Abdominal surgical emergencies in the elderly. *Emerg Med Clin North Am* 2003;21:937-69. [CrossRef](#)
- Halpern LR, Feldman S. Perioperative risk assessment in the surgical care of geriatric patients. *Oral Maxillofac Surg Clin North Am* 2006;18:19-34. [CrossRef](#)
- Seymour DG, Pringle R. A new method of auditing surgical mortality rates: application to a group of elderly general surgical patients. *Br Med J (Clin Res Ed)* 1982;284:1539-42. [CrossRef](#)
- Tekkis PP, McCulloch P, Poloniecki JD, Prytherch DR, Kessaris N, Steger AC. Risk-adjusted prediction of operative mortality in oesophagogastric surgery with O-POSSUM. *Br J Surg* 2004;91:288-95. [CrossRef](#)
- Hosmer DW, Lemeshow S. *Applied logistic regression*. Wiley, New York.
- Wijesinghe LD, Mahmood T, Scott DJ, Berridge DC, Kent PJ, Kester RC. Comparison of POSSUM and the Portsmouth predictor equation for predicting death following vascular surgery. *Br J Surg* 1998;85:209-12. [CrossRef](#)
- Markus PM, Martell J, Leister I, Horstmann O, Brinker J, Becker H. Predicting postoperative morbidity by clinical assessment. *Br J Surg* 2005;92:101-6. [CrossRef](#)
- Chen T, Wang H, Wang H, Song Y, Li X, Wang J. POSSUM and P-POSSUM as predictors of postoperative morbidity and mortality in patients undergoing hepato-biliary-pancreatic surgery: a meta-analysis. *Ann Surg Oncol* 2013;20:2501-10. [CrossRef](#)
- Pelavski AD, Lacasta A, de Miguel M, Rochera MI, Roca M. Mortality and surgical risk assessment among the extreme old undergoing emergency surgery. *Am J Surg* 2013;205:58-63. [CrossRef](#)
- Wakabayashi H, Sano T, Yachida S, Okano K, Izuishi K, Suzuki Y. Validation of risk assessment scoring systems for an audit of elective surgery for gastrointestinal cancer in elderly patients: an audit. *Int J Surg* 2007;5:323-7. [CrossRef](#)
- Gilani MT, Razavi M, Azad AM. A comparison of Simplified Acute Physiology Score II, Acute Physiology and Chronic Health Evaluation II and Acute Physiology and Chronic Health Evaluation III scoring system in predicting mortality and length of stay at surgical intensive care unit. *Niger Med J* 2014;55:144-7. [CrossRef](#)
- Rapsang AG, Shyam DC. Scoring systems in the intensive care unit: A compendium. *Indian J Crit Care Med* 2014;18:220-8. [CrossRef](#)
- Merad F, Baron G, Pasquet B, Henner H, Kohlmann G, Wärlin F, et al. Prospective evaluation of in-hospital mortality with the P-POSSUM scoring system in patients undergoing major digestive surgery. *World J Surg* 2012;36:2320-7. [CrossRef](#)

ORIJİNAL ÇALIŞMA - ÖZET

Geriatrik abdominal acil cerrahide mortaliteyi öngörmede skorlama sistemlerinin analizi

Dr. Murat Özban, Dr. Onur Birsen, Dr. Mahmut Şenel, Dr. Akın Özden, Dr. Burhan Kabay

Pamukkale Üniversitesi Tıp Fakültesi, Genel Cerrahi Anabilim Dalı, Denizli

AMAÇ: Cerrahi sonuçların doğru ölçümü, hastane ve cerrahın doğru değerlendirilmesi, olgudan bağımsız olarak mortalite tahmin yöntemleri tarafından yapılır. Bu çalışmanın amacı, mortaliteyi etkileyen faktörleri analiz etmek, klinik deneyimimizi ve hasta profilimizi sunmak ve bu hastalarda farklı skorlama sistemlerinin kullanımını araştırmaktır.

GEREÇ VE YÖNTEM: 2004-2008 yılları arasında majör abdominal acil cerrahi geçirmiş 112 yaşlı hastanın dosyası geriye dönük olarak incelendi. APACHE II, ODIN, SAPS II genişletilmiş, P-POSSUM, Mannheim peritonit skoru, Charlson komorbidite indeksi, Goldman ve ASA skorları hesaplandı. Bu skorlama sistemlerinin mortaliteyi öngörme açısından duyarlılık, pozitif prediktif değer ve odds oranları hesaplandı.

BULGULAR: Hastalarımız için tüm mortalite oranı %33.9 oldu. Bu çalışmada mortaliteyi etkileyen faktörler başlangıç şikayeti, yoğun bakım ihtiyacı, mekanik ventilasyon ihtiyacı ve süresi, peritonit varlığı ve eşlik eden hastalık durumu olarak belirlendi.

TARTIŞMA: Buna göre, bizim çalışmamızda APACHE II skorlama sisteminin bu grup hastalarda mortalite riskini öngörmede daha gerçekçi sonuçlar verdiği saptanmıştır.

Anahtar sözcükler: Acil cerrahi; geriatri; skorlama sistemi.

Ulus Travma Acil Cerrahi Derg 2015;21(3):182-186 doi: 10.5505/tjtes.2015.05046