ORIGINAL ARTICLE

Accuracy of clinical prediction rules in peptic ulcer perforation: an observational study

DAVID LEVARETT BUCK1, MORTEN VESTER-ANDERSEN2 & MORTEN HYLANDER MØLLER3

1Emergency Department, Holbæk Hospital, Denmark, 2Department of Anaesthesiology and Intensive Care Medicine, Copenhagen University Hospital Herlev, Denmark, and 3Department of Anaesthesiology and Intensive Care Medicine, Copenhagen University Hospital Bispebjerg, Denmark

Abstract

Objective. The aim of the present study was to compare the ability of four clinical prediction rules to predict adverse outcome in perforated peptic ulcer (PPU): the Boey score, the American Society of Anesthesiologists (ASA) score, the Acute Physiology and Chronic Health Evaluation (APACHE) II score, and the sepsis score.

Material and methods. Design: an observational multicenter study. Participants and settings: a total of 117 patients surgically treated for PPU between 1 January 2008 and 31 December 2009 in seven gastrointestinal departments in Denmark were included. Pregnant and breastfeeding women, non-surgically treated patients, patients with malignant ulcers, and patients with perforation of other organs were excluded. Primary outcome measure: 30-day mortality rate. Statistical analysis: the ability of four clinical prediction rules to distinguish survivors from non-survivors (discrimination ability) was evaluated by the area under the receiver operating characteristic curve (AUC), positive predictive values (PPVs), negative predictive values (NPVs), and adjusted relative risks.

Results. Median age (range) was 70 years (25–92 years), 51% of the patients were females, and 73% of the patients had at least one co-existing disease. The 30-day mortality proportion was 17% (20/117). The AUCs: the Boey score, 0.63; the sepsis score, 0.69; the ASA score, 0.73; and the APACHE II score, 0.76. Overall, the PPVs of all four prediction rules were low and the NPVs high.

Conclusions. The Boey score, the ASA score, the APACHE II score, and the sepsis score predict mortality poorly in patients with PPU.

Key Words: APACHE II score, ASA score, Boey score, clinical prediction rule, perforated peptic ulcer, prognosis, sepsis

Introduction

Mortality and morbidity following perforated peptic ulcer (PPU) is substantial, and mortality proportions of 25–30% have been reported in population-based studies [1–6].

A large number of prognostic factors for morbidity and mortality following PPU have been characterized [7] and a number of clinical prediction rules proposed [8–12].

The most well-known prediction rule in PPU is the Boey score which seeks to predict mortality based on the presence of major medical illness, preoperative shock, and perforation longer than 24 h [13]. In the original study by Boey et al., the in-hospital mortality proportion increased progressively with the number of prognostic variables, being 0%, 10%, 45.5%, and 100% in patients with none, one, two, or all three variables, respectively.

The most commonly used preoperative clinical prediction rule worldwide is the American Society of Anesthesiologists (ASA) score [14,15]. In the ASA scoring system, the patient’s preoperative health status, independent of the current surgical disease, is graded in five categories: (1) normal health, (2) mild systemic disease, (3) severe systemic disease, (4)
severe systemic disease that is a constant threat to life, and (5) the patient is not expected to survive without surgery [14].

Another commonly used perioperative prediction rule is the Acute Physiology and Chronic Health Evaluation (APACHE) II score, where the predicted mortality risk is based on the sum of points from 12 physiologic measures, age, and health status [16]. The score has been sought to be validated in abdominal surgery in general [17], peritonitis in general [18], and PPU [11,19].

In a recently published paper, 30-day mortality in PPU patients was significantly reduced by implementing a perioperative care protocol, based on diagnostics, monitoring, and treatment according to the Surviving Sepsis Campaign [20]. Sepsis score [21] on admission was included in the protocol.

At present, clinical prediction rules are not routinely used in PPU patients in everyday clinical practice. Accurate and early identification of PPU patients with an increased risk of adverse outcome is needed to plan and target the level of perioperative monitoring and treatment. The aim of the present study was to compare the performance of four clinical prediction rules in PPU: the Boey score, the ASA score, the APACHE II score, and the sepsis score. We hypothesized that existing clinical prediction rules are poor predictors of adverse outcome in PPU.

Material and methods

Design and approval

This observational multicenter study was approved by The Danish Data Protection Agency (no 2007-41-0702) and did not require informed patient consent according to Danish law. The manuscript was prepared according to the STROBE statement [22].

Participants and settings

A total of 117 consecutive patients surgically treated for PPU between 1 January 2008 and 31 December 2009 in seven gastrointestinal departments in Denmark were included. The participating hospitals were: Copenhagen University Hospital Herlev, Copenhagen University Hospital Bispebjerg, Copenhagen University Hospital Hvidovre, Copenhagen University Hospital Hvidovre, Aarhus University Hospital, Odense University Hospital, and Kolding Hospital. Only patients with intraoperatively confirmed gastric or duodenal PPU were included. Pregnant and breastfeeding women, non-surgically treated patients, patients with malignant ulcers, and patients with perforation of other organs were excluded.

Data extraction and management

The primary data extraction was performed by the local principal investigators at each center using a standardized case report form. Data were subsequently validated and transferred to an electronic database by MHM.

Protocol

Risk stratification according to the four scores (the Boey score, the ASA score, the APACHE II score, and the sepsis score) was done when booking for surgery was performed (preoperatively). Follow-up time was 30 days. The included patients were part of a clinical controlled multicenter intervention trial, recently published [20].

Outcome measures

The primary outcome measure was 30-day mortality. Secondary outcome measures were: (1) development of septic shock (according to the 2001 International Sepsis Conference [21]) within 30 days of surgery and (2) postoperative admission to the intensive care unit (ICU) within 30 days of surgery.

Sample size

Receiver operating characteristics curve: we expected a discriminatory ability (area under the receiver operating characteristic curve; AUC) of 0.75. Based on 80% power to detect a significant association ($p = 0.05$, two-sided) between the clinical prediction rule and the primary outcome, 80 patients were required in the study [23].

Multiple logistic regression: with a binary response variable, nine covariates, $\beta = 0.80$, $\alpha = 0.05$, and an anticipated medium effect size, 113 patients were required in order to detect an association between the variables and endpoints [24,25].

Statistical analysis

Data were analyzed using the statistical software packages SPSS 19.0.

The ability of the four clinical prediction rules to distinguish survivors from non-survivors (discrimination ability) was evaluated by the AUC. The optimal AUC is close to 1 [23].

The risk of dying within 30 days of surgery, given a positive test result (a score above a certain threshold), is reported as the positive predictive value (PPV). Correspondingly, the risk of surviving within
30 days of surgery, given a negative test result (a score below a certain threshold), is reported as the negative predictive value (NPV). The optimal PPV and NPV are close to 100%. The optimal cut-off value of the clinical prediction rules was calculated by means of the Youden index (sensitivity + specificity − 1) [26]. Based on existing literature, and on the Youden indices, the following thresholds were chosen: Boey score ≥ 2, ASA score ≥ 3, APACHE II score ≥ 12, and sepsis score ≥ 3. Finally, adjusted estimates of the four prediction rules’ association to the outcome measures are presented as relative risks (RRs) with 95% confidence intervals (95% CIs). Adjustment for the following confounders was done: age ≥ 65 years, gender, co-morbidity, alcohol abuse, preoperative delay > 24 h, and smoking. Definitions of the variables are found in Table I.

Missing data occurred in less than 5% of the patients. The prevalence and pattern of missing values were evaluated; data were not “missing completely at random” (MCAR). Consequently, multiple imputation for the missing values was performed [27,28]. Five imputations were made, and the five datasets were analyzed using Markov Chain Monte Carlo fully conditional specification [27], and the estimates were pooled for overall estimation. Furthermore, complete-case analysis was performed. The pooled imputed estimates and the corresponding complete-case estimates were compared, and if there were any noticeable differences, both results are presented. Otherwise, only the pooled imputed estimates are presented.

Results

A total of 117 patients were included [20]. Median age (range) was 70 years (25–92 years), 51% of the patients were females, and 73% of the patients had at least one of six co-morbid diseases. Fifty-three percent of the patients were categorized as ASA class ≥ 3. Alcohol abuse was present in 26% of the patients, and 50% of the patients smoked daily (Table I).

Perioperative characteristics of the 117 patients are shown in Table II.

The 30-day mortality proportion was 17% (20/117). In all, 26% (30/117) were diagnosed with septic shock [20], and 49 patients (42%) were admitted to the ICU.

The AUCs of the four clinical prediction rules – with respect to the primary outcome 30-day mortality – were as follows: the Boey score 0.63; the sepsis score 0.69; the ASA score 0.73; and the APACHE II score 0.76 (Figure 1). With respect to the secondary endpoint septic shock, the APACHE II score and the sepsis score had the highest AUC (Figure 2), whereas in relation to ICU admission, the ASA score and the APACHE II score had the highest AUC (Figure 3).

Overall, the PPVs of all four prediction rules were low: 30-day mortality (24–41%), septic shock (35–72%), and ICU admission (49–80%) (Table III). The NPVs with respect to 30-day mortality were on the other hand high: ASA score ≥ 3, 93%; APACHE II score ≥ 12, 97%; Boey score ≥ 2, 87%; and sepsis score ≥ 3, 90% (Table III). For septic shock and ICU admission, the NPVs were lower (Table III).

For 30-day mortality, the APACHE II score and the ASA score had the highest RRs, 31.6 (1.8–545.2) and 21.5 (3.1–150.0), respectively. The sepsis score had the strongest association with septic shock: RR 14.6 (4.2–50.2). Only ASA score and sepsis score were significantly associated with ICU admission (Table III).

Discussion

In this observational multicenter study in patients surgically treated for PPU, we found poor discriminatory abilities of all four evaluated clinical prediction rules: the Boey score, the ASA score, the APACHE II score, and the sepsis score. This was the case for all outcome measures.

During the last decade, fast-track surgery and evidence based in-hospital care have been sought to be implemented in various fields of surgery, leading to a reduction in morbidity and mortality [29]. In patients with PPU, mortality has been reduced considerably by the implementation of a standardized evidence based in-hospital care protocol [20]. The limited number of ICU and high dependency unit beds emphasizes the importance of individual risk

---

Table I. Demographic characteristics.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n = 117</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, median (range)</td>
<td>70 (25–92)</td>
</tr>
<tr>
<td>Men, no (%)</td>
<td>57 (49)</td>
</tr>
<tr>
<td>Body mass index, median (range)</td>
<td>24 (15–65)</td>
</tr>
<tr>
<td>Co-morbidity †, no (%)</td>
<td>85 (73)</td>
</tr>
<tr>
<td>ASA class ‡, no (%)</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>12 (10)</td>
</tr>
<tr>
<td>II</td>
<td>43 (37)</td>
</tr>
<tr>
<td>III</td>
<td>45 (38)</td>
</tr>
<tr>
<td>IV</td>
<td>15 (13)</td>
</tr>
<tr>
<td>V</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Daily smoking, no (%)</td>
<td>59 (50)</td>
</tr>
<tr>
<td>Alcohol abuse †, no (%)</td>
<td>30 (26)</td>
</tr>
</tbody>
</table>

† One or more of the following diseases: diabetes, chronic obstructive pulmonary disease, heart disease, liver cirrhosis, active cancer disease, AIDS or other chronically treated disease.
‡ American Society of Anesthesiologists class.
§ More than 36 g/day (men) or 24 g/day (women).
stratification [30]. Early and accurate identification of patients with increased risk of adverse outcome is needed to plan and target the level of perioperative monitoring and treatment. Thus, a clinical scoring system should be able to predict adverse outcome with a high degree of precision. Furthermore, the score should be easy to calculate, preferably bedside. These characteristics have proven difficult to realize [31].

In the present study, the Boey score had the poorest discriminatory ability of survival, with an AUC of 0.63. Furthermore, the PPV and NPV were low, and the RR was statistically insignificant. In the Boey score, in-hospital mortality increases progressively with the presence of major medical illness, preoperative shock, and perforation longer than 24 h [13]. The Boey score has been re-evaluated in a number of single center studies [3,10,32–35], though without being able to fully replicate the convincing results found by Boey et al. This is both the case for the PPV, the NPV, and the discriminatory function (AUC). When we apply the same cut-off value as used in the present study (Boey score $\leq 2$) to the original work of Boey et al. [13] this results in a substantially lower PPV and NPV, comparable to the

<table>
<thead>
<tr>
<th>Table II. Perioperative characteristics.</th>
<th>($n = 117$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time to admission, median (range), hours</td>
<td>10 (1–336)</td>
</tr>
<tr>
<td>Time from admission to surgery, median (range), hours</td>
<td>5 (1–72)</td>
</tr>
<tr>
<td>Duration of surgery, median (range), minutes</td>
<td>83 (30–240)</td>
</tr>
<tr>
<td>Surgical procedure, no (%)</td>
<td>96 (83)</td>
</tr>
<tr>
<td>Simple open closure ± omental patch</td>
<td>12 (10)</td>
</tr>
<tr>
<td>Laparoscopic closure</td>
<td>8 (7)</td>
</tr>
<tr>
<td>Resection</td>
<td>42 (36)</td>
</tr>
<tr>
<td>Gastric site of perforation, no (%)</td>
<td>56 (48)</td>
</tr>
<tr>
<td>Preoperative metabolic acidosis, no (%)</td>
<td>10 (8)</td>
</tr>
<tr>
<td>Preoperative hemoglobin &lt;6.0 mmol/L, no (%)</td>
<td>33 (28)</td>
</tr>
<tr>
<td>Preoperative creatinine &gt;110/130 μmol/L, no (%)</td>
<td>56 (48)</td>
</tr>
<tr>
<td>Preoperative albumin &lt;550 μmol/L, no (%)</td>
<td>44 (38)</td>
</tr>
<tr>
<td>Use of NSAIDs at admission, no (%)</td>
<td>36 (31)</td>
</tr>
<tr>
<td>Use of aspirin at admission, no (%)</td>
<td>17 (15)</td>
</tr>
<tr>
<td>Use of steroids at admission, no (%)</td>
<td>7 (6)</td>
</tr>
<tr>
<td>Use of anticoagulants at admission, no (%)</td>
<td>28 (24)</td>
</tr>
</tbody>
</table>

Abbreviation: NSAID = non-steroidal anti-inflammatory drug.
*Blood pressure $<90$ mmHg and heart rate $>100$ per min.

Figure 1. Receiver operating characteristics (ROC) curves with area under the curve (AUC). 30-day mortality.

The ASA score (AUC 73%)
The APACHE II score (AUC 76%)
The sepsis score (AUC 69%)
The Boey score (AUC 63%)
The reference line

1.0 The ASA score (AUC 73%)
0.8 The APACHE II score (AUC 76%)
0.6 The sepsis score (AUC 69%)
0.4 The Boey score (AUC 63%)
0.2 Reference line

0.0 0.2 0.4 0.6 0.8 1.0
1 – Specificity

0.0 0.2 0.4 0.6 0.8 1.0
Sensitivity
previously mentioned studies [32–35]. The Boey score is crude, consisting of only three parameters. Consequently, it does not include many of the other existing and well-examined prognostic factors for adverse outcome in PPU, e.g. old age, tachycardia, and acute renal failure [7]. This might explain the inferior performance in the present study. On the other hand, it is simple to calculate and was created specifically for patients with PPU.

The ASA score has been criticized for its subjectivity and the wide inter-observer variability [36]. Nevertheless, it predicts postoperative mortality well [37]. In the present study, the AUC of ASA class was 0.73 (mortality), 0.67 (septic shock), and 0.69 (ICU). The only other study examining the discriminatory ability of ASA score in PPU patients found a somewhat higher AUC of 0.91 [34]. This discrepancy is most likely caused by differences in

Figure 2. Receiver operating characteristics (ROC) curves with area under the curve (AUC). Septic shock.

Figure 3. Receiver operating characteristics (ROC) curves with area under the curve (AUC). Intensive care unit (ICU) admission.
demographic characteristics and in the mortality rate. The PPVs and NPVs of other studies evaluating ASA score are of similar strengths as in our study [35,38–40]. The explanation for the inferior performance of the ASA score could be the fact that ASA score is not developed within PPU patients and only accounts for the pre-morbid status. However, in our experience it is difficult to completely ignore the current clinical condition of the patient when assessing the ASA score. Consequently, the performance of the ASA score could in fact be overestimated.

Of the four clinical prediction rules evaluated, the APACHE II score performed best overall with AUCs of 0.76 (mortality), 0.78 (septic shock), and 0.72 (ICU admission). An APACHE II score ≥12 seems to imply a nearly 32 times increased risk of 30-day mortality. However, only 24% of the patients predicted to die by the APACHE II score died (PPV 24%). On the other hand, all but four patients predicted to survive (APACHE II score <12) survived (NPV 97%). Use of APACHE II score as a clinical prediction rule in PPU patients has been sparsely examined [8,10,11]. Overall, PPVs and NPVs are of similar strengths as in the present study. The only other study examining the discriminatory ability of APACHE II score found a somewhat higher AUC of 0.87 [11]. This discrepancy is most likely caused by differences in demographic characteristics and in the mortality rate. The APACHE II score takes into account age, co-existing diseases, the current clinical status, and current biochemical findings: in total, 14 variables [16]. This is most likely the main reason why the APACHE II score performs best among the four clinical prediction rules evaluated in the present study. One possible explanation for the somewhat inadequate performance of the APACHE II score in general could be the fact that the score is not developed within a PPU cohort. Thus, well-established important prognostic factors for mortality in PPU patients are not included in the APACHE II score: preoperative delay, co-existing use of medications, and low serum albumin blood level [7].

It is well known that sepsis (systemic inflammatory response syndrome and proven or highly suspected infection) [41] is associated with adverse outcome in PPU patients [7]. To our knowledge, the discriminatory ability of the sepsis score in PPU has not been examined previously. In the present study, the sepsis score had low AUCs: 30-day mortality 0.69, septic shock 0.74, and ICU admission 0.64. Furthermore, we found low PPVs and NPVs. This might be explained by the fact that the sepsis score is not specific for PPU patients, it does not include co-existing diseases, and it only takes into account few of the other known prognostic factors for PPU [7].

### Limitations

This study has limitations. The cohort was part of an interventional study [20], which potentially could affect the external validity of the study. Hence, external or internal (boot-strapping) validation of the present results is warranted. The risk of confounding was sought to be minimized by the use of logistic regression, with adjustment for previously identified prognostic factors and suspected confounding factors (multivariate technique). However, one major problem is that confounding variables are not always known or measurable (residual confounding). The calculated 95% CIs are wide, indicating low statistical precision. Choosing other definitions and categorizations of the demographic variables and confounders, and choosing other cut-off values of the clinical prediction rules could lead to other results. Despite using the preferred method of handling missing values that are not MCAR (multiple imputation) [27,28], missing values could affect the validity of the results. The PPU patients in the present cohort have one particular type of case mix. The results of this study should not be applied

### Table III. Positive predictive values (PPVs), negative predictive values (NPVs), and relative risks (RRs) with 95% confidence intervals (CIs) of the four clinical prediction rules (n = 117).

<table>
<thead>
<tr>
<th>Score</th>
<th>30-day mortality</th>
<th>Septic shock</th>
<th>ICU admission</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PPV</td>
<td>NPV</td>
<td>RR* (95% CI)</td>
</tr>
<tr>
<td>ASA score ≥3</td>
<td>26%</td>
<td>93%</td>
<td>21.5 (3.1–150.0)</td>
</tr>
<tr>
<td>APACHE II score ≥12</td>
<td>24%</td>
<td>97%</td>
<td>31.6 (1.8–345.2)</td>
</tr>
<tr>
<td>Boey score ≥2</td>
<td>24%</td>
<td>87%</td>
<td>64 (1.2–34.7)</td>
</tr>
<tr>
<td>Sepsis score ≥3</td>
<td>41%</td>
<td>90%</td>
<td>7.2 (2.1–28.0)</td>
</tr>
</tbody>
</table>

Abbreviations: APACHE = Acute Physiology and Chronic Health Evaluation; ASA = American Society of Anesthesiologists; ICU = intensive care unit.

*Adjusted for age ≥65 years, yes/no; male gender, yes/no; co-morbidity, yes/no; alcohol abuse, yes/no; preoperative delay >24 h, yes/no; and smoking, yes/no.
uncritically to other categories of patients and/or to PPU patients with different characteristics.

**Conclusion, clinical implications and perspectives**

In patients surgically treated for PPU, the Boey score, the ASA score, the APACHE II score, and the sepsis score predict mortality, septic shock, and admission to ICU poorly. The AUCs and PPVs were overall low, while the NPRVs were 90–95%. Thus, in the clinical setting, the scores can rule out mortality within 30 days of surgery with a high degree of precision in a PPU cohort with a similar case mix.

In order to be able to precisely predict adverse outcome in PPU patients, we suggest that a score developed within PPU patients and including both pre-morbid objective measures and current objective measures is used.

**Declaration of interest:** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

**References**


