Clinical risk scores to guide perioperative management

Sarah Barnett,1 Suneetha Ramani Moonesinghe2

ABSTRACT
Perioperative morbidity is associated with reduced long-term survival. Comorbid disease, cardiovascular illness, and functional capacity can predispose patients to adverse surgical outcomes. Accurate risk stratification would facilitate informed patient consent and identify those individuals who may benefit from specific perioperative interventions. The ideal clinical risk scoring system would be objective, accurate, economical, simple to perform, based entirely on information available preoperatively, and suitable for patients undergoing both elective and emergency surgery. The POSSUM (Physiological and Operative Severity Score for the enUmeration of Morbidity and Mortality) scoring systems are the most widely validated perioperative risk predictors currently utilized; however, their inclusion of intra- and postoperative variables precludes validation for preoperative risk prediction. The Charlson Index has the advantage of consisting exclusively of preoperative variables; however, its validity varies in different patient cohorts. Risk models predicting cardiac morbidity have been extensively studied, despite the relatively uncommon occurrence of postoperative cardiac events. Probably the most widely used cardiac risk score is the Lee Revised Cardiac Risk Index, although it has limited validity in some patient populations and for non-cardiac outcomes. Bespoke clinical scoring systems responding to dynamic changes in population characteristics over time, such as those developed by the American College of Surgeons National Surgical Quality Improvement Program, are more precise, but require considerable resources to implement. The combination of objective clinical variables with information from novel techniques such as cardiopulmonary exercise testing and biomarker assays, may improve the predictive precision of clinical risk scores used to guide perioperative management.

INTRODUCTION
Perioperative morbidity and mortality is a significant public health issue, due to its impact on patients’ short and long term survival, and also resource utilisation within the health service. Surgical complications occur in 3–17% of patients1 2; in the general population, surgical mortality is approximately 0.5%, but in elderly patients undergoing emergency surgery the UK mortality rate is >12%.3

As clinicians we strive to find accurate methods to stratify patient risk from undergoing surgery. Clinical risk scores tend to focus on patient history, examination, preoperative investigations, surgical severity, and intraoperative events. Surgical risk may also be assessed using measures of functional capacity, such as the Duke Activity Status Index, and more objectively, using cardiopulmonary exercise testing (CPET).4 5 There is also increasing interest in biological markers of cardiorespiratory health or inflammation which could be used to predict perioperative outcome such as N-terminal brain natriuretic peptide (NT-BNP)6 7 and high sensitivity C reactive protein (hsCRP).8 9 However, while both CPET and serum biomarkers show promise as risk prediction tools, neither are widely available, and both require further validation in large pragmatic multicentre studies.

Scoring systems have the advantage that they are usually cost neutral, and can be performed at any time of the day or night. Patients presenting for emergency surgery are at high perioperative risk,3 and it may be possible to predict the outcomes of such patients using these openly available, relatively quick scoring systems which are not subject to complex interpretation.

The ideal risk prediction model would be one that is simple, reproducible, accurate, objective, and available to all patients. Many hospitals lack the resources to run expensive tests, so ideally it should be cheap, and possible to perform at the bedside. As the purpose of the scoring system is to define an individual patient’s risk before he or she undergoes surgery, a model that is based entirely on preoperative risk factors would be more beneficial than those which include intraoperative and postoperative variables.

In the course of this article we shall describe the different scoring systems that have been developed and are used to predict risk for non-cardiac surgery, highlighting their advantages and disadvantages, and how their accuracy has been shown to vary in a number of different studies.

RISK STRATIFICATION SCORING SYSTEMS
Risk stratification scoring systems can be classified into those estimating population risk, such as the American Society of Anaesthesiologists Physical Status score (ASA-PS), and those estimating individual risk. Risk scores developed to calculate individual risk can be subdivided into those which are designed to predict cardiac morbidity and mortality such as the Lee Revised Cardiac Risk Index (RCRI), and those which predict generic morbidity and mortality. These include scores which use solely preoperative risk factors—for example, Charlson Index—and those looking at a combination of preoperative, intraoperative, and postoperative factors—for example, Physiological and Operative Severity Score for the enUmeration of Morbidity and Mortality (POSSUM). We shall now consider each of these in turn.
The ASA-PS uses preoperative physical fitness to categorise patients subjectively into five subgroups. The score was devised in 1941 originally as a statistical tool for retrospective analysis of hospital records. It has been revised on several occasions and in 1965 the five point scale most commonly known was first reported. There is now an additional category for patients whose organs are being removed for donor purposes (table 1).

Underlying fitness is an important predictor of survival from surgery, and the ASA-PS score has been shown to correlate with outcome in a number of different settings. It is simple, easy to understand, and commonly used as part of the preoperative assessment. It is also an easy and useful tool to assist descriptions of workload and ‘anaesthetic risk’ for audit and research purposes.

However, the ASA-PS system has a number of limitations. It does not consider the preoperative optimisation of the patient, the surgery planned, or the proposed postoperative care—that is, critical care unit admission—and it makes no adjustments for age, sex, weight, or pregnancy. In this sense it does not give a prediction of risk for an individual patient or operation. Wolters’ study found a population mortality of 0.1% for ASA I patients, and an 18.3% mortality for patients graded ASA IV. However, more recent work demonstrated that the ASA-PS was poorly predictive of individual patient outcome: while it correctly predicted an uncomplicated course in 96% of patients, complications were correctly predicted in only 16% of patients in whom they occurred (positive predictive value 57%, negative predictive value 80%). Similar results were found in a more recent study which looked at infective complications and morbidity in patients undergoing total knee replacement.

### Table 1 American Society of Anaesthesiologists Physical Status Score

<table>
<thead>
<tr>
<th>ASA grade</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A normal healthy patient</td>
</tr>
<tr>
<td>II</td>
<td>A patient with mild systemic disease</td>
</tr>
<tr>
<td>III</td>
<td>A patient with severe systemic disease</td>
</tr>
<tr>
<td>IV</td>
<td>A patient with severe systemic disease that is a constant threat to life</td>
</tr>
<tr>
<td>V</td>
<td>A moribund patient who is not expected to survive without the operation</td>
</tr>
<tr>
<td>VI</td>
<td>A declared brain dead patient whose organs are being removed for donor purposes</td>
</tr>
</tbody>
</table>

**Table 1 Revised Cardiac Risk Index (Lee et al 1999)**

1. High risk type of surgery
2. Ischaemic heart disease (includes any of the following: history of myocardial infarction, history of a positive exercise test, current complaint of chest pain, considered to be secondary to myocardial ischaemia, use of nitrate therapy, or electrocardiography with pathologic Q waves)
3. Congestive heart failure
4. History of cerebrovascular disease
5. Preoperative treatment with insulin
6. Preoperative serum creatinine >2.0 mg/dl

There is ongoing work on adapting and improving the predictive accuracy of the RCRI. A recent prospective, single centre study assessed the predictive power of NT-proBNP, CRP, and RCRI in combination for the risk of a perioperative major cardiovascular event, and found that the predictive power of the RCRI could be improved significantly by the addition of CRP and NT-proBNP to RCRI (adjusted relative risk (RR) 4.6) (p<0.001).

While cardiac morbidity is important, it is relatively uncommon compared with other types of perioperative complication, which may have significant implications for patients both in the short and long term. There is therefore a requirement for a clinical risk score which may predict generic morbidity and mortality with accuracy. Existing systems which are used for this purpose are now discussed.
following colorectal surgery, and it has helped to calculate mortality following cardiac surgery. The CACI has also, in certain patient groups, been found to predict ‘long term’ mortality.

Following head and neck surgery, and radical prostatectomy, the Charlson Index has a predictive ability comparable to that of the ASA-PS when ascertaining morbidity and mortality. However, the ASA-PS was more accurate in its calculation than the Charlson Index in a cohort of patients undergoing liver resection.

This series of papers represents a relatively large and diverse experience with the Charlson Index, mainly confirming its predictive validity. However, disadvantages of this scoring system are the lack of information regarding the surgical procedure, and the subjectivity of the patients’ comorbidity assessment, which may potentially lead to error.

**GENERIC MORBIDITY AND MORTALITY: PERIOPERATIVE RISK FACTORS**

The Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (POSSUM) was developed by Copeland et al in 1991 as a scoring system for surgical audit, where they used multivariate logistic regression to identify the most significant of 48 physiological variables and 12 operative and postoperative variables for the prediction of the 30 day morbidity and mortality rates. The final POSSUM system incorporates 12 variables to assess physiological status and six ‘surgical’ variables, resulting in an 18 component score. In order for perioperative risk to be calculated, the sum of the physiological and surgical variables are entered into two mathematical equations which are used to calculate the risk of morbidity and mortality (table 3).

A potential drawback in using this scoring system to assess risk is the time at which the operative variables are obtained. The physiological variables are documented before the commencement of surgery, and include the patient’s clinical symptoms, signs elicited, haematology and biochemistry results, and an assessment of the electrocardiogram. When it comes to collecting the surgical data, this can be delayed for a prolonged period of time, as included in the score are the number of subsequent operations within 30 days, and presence of malignancy. This causes obvious difficulties in using this as the sole tool to make decisions on the appropriateness of surgery, as a number of variables will not be available until the surgery is complete.

POSSUM was originally developed and validated in a single UK district general hospital, in a study which included emergency and elective procedures in patients undergoing urological, gastrointestinal, vascular, and hepatobiliary surgery. Subsequently, this system has also been used in comparisons of surgeons, resource utilisation, and to compare surgical outcomes in different countries. While some studies have found confirmatory evidence that POSSUM predicts individual patient risk of morbidity and mortality, others have found a significant overestimation of mortality using this score. This is the result of using logistic regression to predict risk, as the lowest possible mortality risk is 1.08%.

Using alternative risk equations, but the same physiological and surgical variables, a new risk model was developed in Portsmouth (P-POSSUM), and validated in a large single centre cohort, and since then has been shown to predict in-hospital mortality more accurately than POSSUM. P-POSSUM has no morbidity prediction equation, as a result of the original authors’ lack of confidence in the reporting of perioperative complications. Subsequent studies have shown P-POSSUM to both over-predict and under-predict mortality, in different settings.

Large cohorts of patients have been used to validate variations of POSSUM for specific surgical groups, including Cr-POSSUM (colorectal surgery), V-POSSUM (vascular surgery), and O-POSSUM (vascular surgery), which are more sensitive and specific for predicting patient outcome following these operations. However, a recent systematic review of the use of different POSSUM models in oesophageal–gastric surgery found that the P-POSSUM was superior to both the original POSSUM and the surgery specific O-POSSUM model for the prediction of postoperative mortality. These data highlight the considerable variation in the predictive precision of these models in different surgical cohorts. One of the reasons for this variation may be the inclusion of a number of subjective variables such as jugular venous pressure measurement and chest radiograph interpretation, which may introduce error in risk score calculation. Nevertheless, POSSUM and its subsequent variations remain the most internationally validated risk scoring system for predicting individual patient risk.

**BIOCHEMICAL AND HAEMATOLOGICAL OUTCOME MODELS**

The Biochemistry and Haematology Outcome Model (BHOM) was developed by Prytherch et al in 2003, and data collected includes age, sex, mode of admission (eg, emergency or elective), physiological parameters (haemoglobin, white cell count, sodium, potassium, urea), BUPA operative severity score, and 30 day mortality. The BUPA operative severity score is used in the private sector (UK) as an indicator of surgical complexity and workload. The BHOM uses information which is easily obtainable from most hospital information technology systems to calculate predictive mortalities. As previously stated, a risk assessment scoring system should ideally have few variables, which are simple to collate and therefore make the score

<table>
<thead>
<tr>
<th>Table 2 Charlson Index (Charlson et al 1987)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>Myocardial infarction</td>
</tr>
<tr>
<td>Congestive heart failure</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
</tr>
<tr>
<td>Cerebrovascular disease</td>
</tr>
<tr>
<td>Dementia</td>
</tr>
<tr>
<td>Chronic pulmonary disease</td>
</tr>
<tr>
<td>Connective tissue disease</td>
</tr>
<tr>
<td>Ulcer disease</td>
</tr>
<tr>
<td>Mild liver disease</td>
</tr>
<tr>
<td>Diabetes</td>
</tr>
</tbody>
</table>

The table above lists the various comorbidities that are assessed by the Charlson Index, with the specific conditions that are included in each category. The index is used to calculate the risk of mortality following surgery, and can be used to predict patient outcomes in different surgical and medical settings.
easy to implement in daily practice. It should be available for every patient and have limited potential inter-observer variability. BHOM uses objective data, which are easily available from a single blood sample. It is therefore feasible to collect this for all patients as part of routine care. However, BHOM is one of the less widespread outcome models used presently, and is yet to be validated in large multicentre cohorts including patients undergoing different types of surgery.

THE SURGICAL APGAR SCORE

The Apgar score was originally developed as a powerful tool in obstetrics to rapidly assess and gain feedback on a newborn’s condition. The surgical Apgar score was piloted using general and vascular patients where the score was significantly associated with the occurrence of major complications or death within 50 days of surgery (p<0.001). The 10 point score is calculated from the estimated blood loss, lowest heart rate, and lowest mean arterial pressure during an operation. It uses routinely available data, and has been shown to identify immediately and effectively those patients at higher and lower risk of complications and death postoperatively. Its limitations include its lack of validity in other cohorts of patients, lack of comparison between different institutions, and the potential for imprecision resulting from estimating blood loss. As a consequence, it may be considered one of the less applicable scoring systems in current clinical practice.

AN AMERICAN APPROACH: THE NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM

The American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) is the first nationally validated, risk adjusted, outcomes based programme used to measure and improve the quality of surgical care in the USA. A prospective, validated database is used to quantify 30 day risk adjusted surgical outcomes. A comparison of outcomes can then be made for all the hospitals within the programme. ACS NSQIP collects data on 156 variables for patients undergoing major surgical procedures, including preoperative risk factors, intraoperative variables, and 30 day postoperative morbidity and mortality. The data are collected, validated, and submitted by a trained surgical clinical reviewer at each site. Using logistic regression, risk adjusted 30 day morbidity and mortality outcomes are made for each participating hospital on an annual basis. Both generic and surgery specific models are generated, and outcomes are reported as observed versus expected (O/E) ratios. The ACS NSQIP database therefore allows hospital specific adjustments to surgical risk to be incorporated, by allowing the categorisation of hospitals performing better or worse than expected.

The unique aspect of this initiative is the generation of bespoke models which are responsive to changes in population characteristics over time, and therefore are more precise. This was demonstrated in a recent study which compared an ACS-NSQIP model for colorectal surgery with existing validated models including the CR-POSSUM and the Association Française de Chirurgie colorectal model. The ACS NSQIP colorectal risk calculator demonstrated far higher predictive precision than existing models in both development and validation cohorts. Despite the advantages of such initiatives, their application internationally is limited by the resources required to collect and record multiple variables prospectively and accurately.

CONCLUSIONS AND FUTURE WORK

Clinicians within the perioperative team utilise risk prediction scores to help make important, informed decisions to optimise
Main messages

- An ideal clinical risk score would consist of objective preoperative variables, be economical and simple to implement, and suitable for patients undergoing elective and emergency surgery.
- The POSSUM scoring systems are the most widely validated, internationally available clinical risk scores currently available for the prediction of generic perioperative morbidity and mortality; however, they are not validated for preoperative risk stratification, as they incorporate intraoperative variables. The Lee RCRI is the most widely validated system for the prediction of adverse cardiac outcomes, although it should be recognised that cardiac events account for only a very small fraction of postoperative morbidity, and the RCRI is poorly predictive of all cause morbidity and mortality.
- Initiatives such as the ACS-NSQIP programme in the USA allow the development of bespoke risk prediction models which are responsive to changes in population characteristics and patient management; however, considerable resources would be required to implement such programmes internationally.

Future research

- Identification of risk factors which are common to the various validated clinical risk scores and use these to develop a simple objective dataset.
- Develop models which incorporate clinical data and information from novel techniques such as cardiopulmonary exercise testing and biomarker assays, to assess if these improve predictive precision.
- Develop large multicentre databases which allow the development and validation of bespoke risk prediction models which may be modified on an annual basis and therefore be responsive to changes in clinical management, patient characteristics, and healthcare delivery.

Key references


If this is then combined with biomarker assays and assessment of functional capacity, improved performance may be possible.

Further work is required to improve the identification of patients at high risk of perioperative morbidity. Such work should include both research required to identify objective risk factors which are significant on multivariate analysis and may be combined in a scoring system with high predictive precision, and also investment in resources enabling both objective and subjective variables to be recorded and collected accurately. As medical and surgical practice changes with time, it is likely that the predictive precision of risk scoring systems may also change: the implications of comorbid illness may not be the same now as during the 1990s when systems such as the RCRI and POSSUM score were first validated. Furthermore, standards of healthcare and perioperative practice will differ between institutions and countries, with the consequence that a ‘one size fits all’ solution may not be possible. Therefore, the ideal risk prediction system, in addition to consisting of predominantly preoperative variables, should also be responsive to changes in clinical management over time. If subjective variables are used, then these should be assessed by staff trained in such assessment. The ACS-NSQIP initiative approaches these ideals; however, for the rest of the world, while existing systems such as the RCRI and POSSUM scores are helpful, further work aimed at evaluating and refining them in different populations and healthcare systems is required.

### MULTIPLE CHOICE QUESTIONS (TRUE (T)/FALSE (F); ANSWERS AFTER THE REFERENCES)

1. **The ideal risk prediction model would be**
   - Simple
   - Subjective
   - Expensive
   - Possible to perform at the bedside
   - Purely for elective cases

2. **The American Society of Anaesthesiologists’ physical status score**
   - Is commonly used as part of the preoperative assessment
   - Considers the preoperative optimisation of the patient
   - Makes no adjustment for the age of the patient
4. The Physiological and Operative Severity Score for the enumeration of Mortality and Morbidity (POSSUM)

A. Was first reported to assist comparative audit among anaesthetic services
B. Incorporates 12 physiological variables and six surgical variables
C. Uses purely preoperative variables
D. Contains no subjective variables
E. Is the most internationally validated risk scoring system for predicting individual patient risk

5. Clinical risk scores
A. Inter-observer variation in data collection can influence predictive accuracy
B. Adaptation to changes in patient population over time may improve predictive precision
C. The Biochemistry and Haematology Outcome Model (BHOM) has been validated in multicentre cohorts
D. The surgical Appgar score is a 10 point score calculated from intraoperative heart rate, blood pressure, and estimated blood loss
E. The Charlson-Comorbidity Index uses no information regarding the surgical procedure

Competing interests SRM works within the University College London/University College London Hospitals’ Joint Comprehensive Biomedical Research Centre, which received a proportion of funding from the UK Department of Health’s National Institute for Health Research Biomedical Research Centres’ funding scheme.

Provenance and peer review Commissioned; externally peer reviewed.

REFERENCES
ANSWERS

1. A (T); B (F); C (F); D (T); E (F)
2. A (T); B (F); C (F); D (F); E (T)
3. A (T); B (F); C (F); D (F); E (T)
4. A (F); B (T); C (F); D (F); E (T)
5. A (T); B (F); C (F); D (F); E (T)