David S. Warner, M.D., Editor

Risk Stratification Tools for Predicting Morbidity and Mortality in Adult Patients Undergoing Major Surgery

Qualitative Systematic Review

Suneetha Ramani Moonesinghe, F.R.C.A.,* Michael G. Mythen, M.D.,† Priya Das, M.B.B.S.,‡ Kathryn M. Rowan, Ph.D.,§ Michael P. W. Grocott, M.D.‖

ABSTRACT

Risk stratification is essential for both clinical risk prediction and comparative audit. There are a variety of risk stratification tools available for use in major noncardiac surgery, but their discrimination and calibration have not previously been systematically reviewed in heterogeneous patient cohorts.

Embase, MEDLINE, and Web of Science were searched for studies published between January 1, 1980 and August 6, 2011 in adult patients undergoing major noncardiac, nonneurological surgery. Twenty-seven studies evaluating 34 risk stratification tools were identified which met inclusion criteria. The Portsmouth-Physiology and Operative Severity Score for the enUmeration of Mortality and the Surgical Risk Scale were demonstrated to be the most consistently accurate tools that have been validated in multiple studies; however, both have limitations. Future work should focus on further evaluation of these and other parsimonious risk predictors, including validation in international cohorts. There is also a need for studies examining the impact that the use of these tools has on clinical decision making and patient outcome.

ACCURATE prediction of perioperative risk is an important goal—to enable informed consent for patients undergoing surgery and to guide clinical decision making in the perioperative period. In addition, by adjusting for risk, an accurate risk stratification tool enables meaningful comparison of surgical outcomes between providers for service evaluation or clinical audit. Some risk stratification tools have been incorporated into clinical practice, and indeed, have been recommended for these purposes.[1]

Risk stratification tools may be subdivided into risk scores and risk prediction models. Both are usually developed using multivariable analysis of risk factors for a specific outcome.[2] Risk scores assign a weighting to factors identified as independent predictors of an outcome; with the weighting for each factor often determined by the value of the regression coefficient in the multivariable analysis. The sum of the weightings in the risk score then reflects increasing risk. Risk scores have the advantage that they are simple to use in the clinical setting. However, although they may score a patient on a scale on which other patients may be compared, they do not provide an individualized risk prediction of an adverse

* Director, University College London, University College London Hospitals' Surgical Outcomes Research Center, London, United Kingdom; Honorary Senior Lecturer, University College London; and Consultant, Anaesthesia and Critical Care, University College Hospital. † Professor, Smiths Medical Professor of Anaesthesia and Critical Care, University College London; and Honorary Consultant, Anaesthesia, University College Hospital. ‡ Research Fellow, University College London, University College London Hospitals' Surgical Outcomes Research Center, University College Hospital. § Professor and Director, Intensive Care National Audit and Research Center, London, United Kingdom. ‖ Professor of Critical Care Medicine, University of Southampton, Southampton, United Kingdom; Honorary Consultant, Critical Care, Southampton University Hospital; and Director, National Institute for Academic Anaesthesia's Health Services Research Center, London, United Kingdom.

Received from University College London, University College London Hospitals' Surgical Outcomes Research Center, London, United Kingdom. Submitted for publication November 18, 2012. Accepted for publication June 4, 2013. Supported by a grant award from the National Institute for Academic Anaesthesia, London, United Kingdom, and the Frances and Augustus Newman Charitable Foundation, Bristol, United Kingdom (to Dr Moonesinghe). Dr Moonesinghe and Professor Mythen work within the University College London/University College London Hospitals Joint Comprehensive Biomedical Research Center, London, United Kingdom, which receives funding from the Department of Health's National Institute for Health Research, London, United Kingdom. Professor Grocott is funded in part by the University Hospitals Southampton National Health Service Foundation Trust/University of Southampton Respiratory Biomedical Research Unit, Southampton, United Kingdom, which received a portion of its funding from the Department of Health's National Institute for Health Research Biomedical Research Unit funding scheme, London, United Kingdom. The authors declare no competing interests.

Address correspondence to Dr. Moonesinghe: Department of Anaesthesia, University College London Hospitals NHS Foundation Trust, 3rd floor Maples Link Corridor, 235 Euston Road, London NW1 6BU, United Kingdom. rmoonesinghe@gmail.com. Anesthesiology 2013; 119:959–81

outcome.[3] Examples of risk scores are the American Society of Anesthesiologists' Physical Status score (ASA-PS)[4] and the Lee Revised Cardiac Risk Index.[5]

By contrast, risk prediction models estimate an individual probability of risk for a patient by entering the patient's data into the multivariable risk prediction model. Although risk prediction models may be more accurate predictors of an individual patient's risk than risk scores, they are more complex to use in the day-to-day clinical setting.

Despite increasing interest in more sophisticated risk prediction methods, such as the measurement of functional capacity by exercise testing,[6] risk stratification tools remain the most readily accessible option for this purpose. However, clinical experience tells us that they are not commonly used in everyday practice. Lack of use may be due to poor awareness amongst clinicians of the available options and concerns regarding their complexity and accuracy.[7] In other clinical settings, low uptake of risk stratification tools has been ascribed to a lack of clarity on the precision of available tools, resulting from perhaps unnecessary efforts to make minor refinements to existing methods, or to developing novel methods, with the aim of achieving greater predictive accuracy.[8]

With the aim of summarizing the available risk stratification tools in perioperative care, in order to make recommendations about which methods are appropriate for use both in clinical practice and in research, we have undertaken a qualitative systematic review on the available evidence. The specific question we sought to answer was "What is the performance of risk stratification tools, validated for morbidity and/or mortality, in heterogeneous cohort of surgical (noncardiac, nonneurological) patients?" The review had three main objectives as follows: to summarize the available risk prediction methods, to report on their performance, and to comment on their strengths and weaknesses, with particular focus on accuracy and ease of application.

Materials and Methods

Previously published standards for reporting systematic reviews of observational studies were adhered to when undertaking this study.[9] A Preferred Reporting Items for Systematic reviews and Meta-analyses checklist[10] was used in the preparation of this report (appendix 1).

Definitions for the Purposes of This Study

A "risk stratification tool" was defined as a scoring system or model used to predict or adjust for either mortality or morbidity after surgery, and which contained at least two different risk factors. "Major surgery" was defined as a procedure taking place in an operating theatre and conducted by a surgeon; thus, studies of cohorts of patients undergoing endoscopic, angiographic, dental, and interventional radiological procedures were excluded. A "heterogeneous patient cohort" was defined as a cohort of patients including at least two different surgical specialities. Studies of gastrointestinal surgery, which included hepatobiliary surgery, were included. We excluded studies that consisted entirely of cohorts undergoing ambulatory (day case) surgery and cohorts that included cardiac or neurological surgery.

Search Strategy and Study Eligibility

A search for articles published between January 1, 1980 and August 6, 2011 was undertaken using MEDLINE, Embase, and Web of Science. No language restriction was applied. The search strategy and inclusion and exclusion criteria are detailed in appendix 2. Of note, articles reporting development studies were excluded, unless the article included validation in a separate cohort.

Data Extraction and Quality Assessment of Studies

Data extraction was independently undertaken by Drs. Moonesinghe and Das, using standardized tables relating to the study characteristics, quality, and outcomes. Where there was disagreement in the data extraction between these two authors, Dr. Moonesinghe resolved the query by referring again to the original articles. Study characteristics extracted from each article included the number of patients, the country where the study was conducted, the outcome measures and endpoints of each study, and the risk stratification tools being assessed. Data were also extracted regarding the most detailed description of the types of surgery included in each study cohort reported in the articles. We also extracted clinical outcome data (morbidity and mortality) for the cohorts in each study.

Assessment of study quality was based on the framework for assessing the internal validity of articles dealing with prognosis developed by Altman.[11,12] The following criteria were used: the number of patients included in analyses, whether the study was conducted on a single or multiple sites, the timing of data collection (prospective *vs.* retrospective), whether a description of baseline characteristics for the cohort was included (including comorbidities, type of surgery, and demographic data), and selection criteria for patients included in the study (to assess for selection bias). Selection bias was judged to be present if a study restricted the type of patient who could be enrolled based on age, ethnicity, sex, premorbid condition, urgency of surgery, or postoperative destination (*e.g.*, critical care). In addition, we reported the setting of each validation study—*i.e.*, whether the validation was conducted in a split sample of the original development cohort or whether the validation cohort was entirely different from that in which the tool was developed.[13] Finally, as a measure of their clinical usability and reproducibility, we reported whether each risk stratification tool used variables which were objective (*e.g.*, blood results), subjective (*e.g.*, chest radiograph interpretation), or both.[14]

Data Analysis and Statistical Considerations

The performance of each risk stratification tool was evaluated using measures of discrimination and, where appropriate, calibration. Discrimination (how well a model or score correctly identifies a particular outcome) was reported using

either the area under the receiver operating characteristic curve (AUROC) or the concordance (c-) statistic. We considered an AUROC of less than 0.7 to indicate poor performance, 0.7–0.9 to be moderate, and greater than 0.9 to reflect high performance.[15] Calibration is defined as how well the prognostic estimation of a model matches the probability of the event of interest across the full range of outcomes in the population being studied. Where reported, either Hosmer–Lemeshow or Pearson chi-square statistics were extracted as an evaluation of calibration; *P* value of more than 0.05 was taken to indicate that there was no evidence of lack-of-fit.

Results

Search Results

In the initial search, 139,775 articles on MEDLINE and 71,841 on Embase were listed, and the titles and abstracts of these were screened to identify articles which described risk stratification tools used in any adult noncardiac, nonneurological surgery. Seven hundred fifty-one articles then underwent a review. Hand searching of reference lists and citations identified a further 432 studies which were also reviewed in detail.

Three studies were identified that graphically displayed receiver operating characteristic curves in their results but did not report AUROCs.[16–18] The authors of these studies were contacted for additional information; none responded, so these studies were excluded from the analysis. Six foreign language studies, which may have been eligible for inclusion based on review of the abstracts, but for which we were unable to obtain translations, were also omitted from the analysis.[19–24] The flow chart for the review is detailed in figure 1.

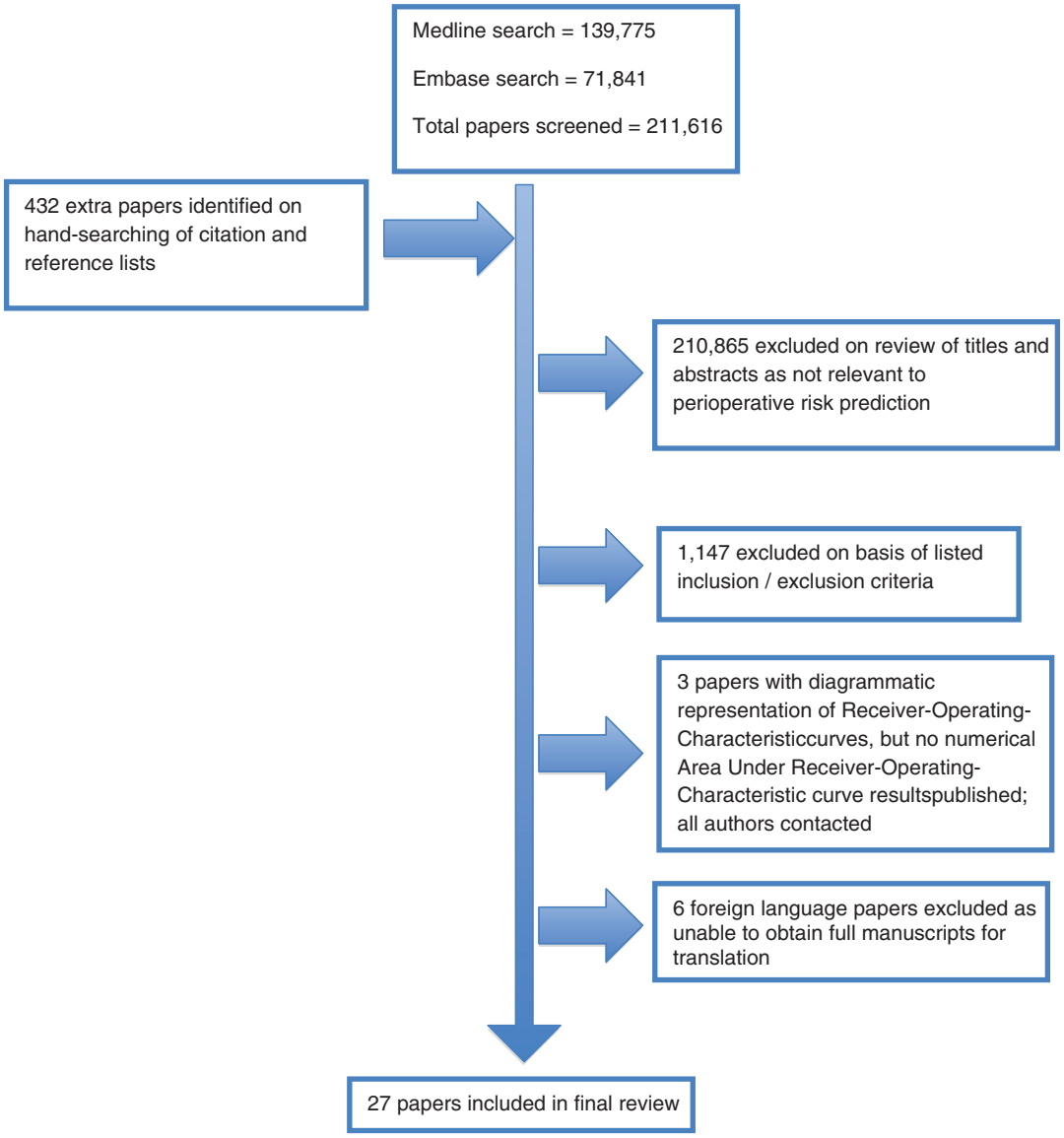

Fig. 1. Flow diagram for the review.

A total of 27 studies evaluating 34 risk stratification tools were included in the analysis. All were cohort studies. Eight tools were validated in multiple studies; the most commonly reported were the ASA-PS (four studies, total number of patients, n = 4,014), the Acute Physiology and Chronic Health Evaluation II (APACHE II) scoring system (four studies, n = 5,897), the Physiological and Operative Score for the enUmeration of Mortality and Morbidity (POSSUM; three studies, n = 2,915), the Portsmouth variation of POSSUM (P-POSSUM; five studies, n = 10,648; mortality model only), the Surgical Risk Scale (three studies, n = 5,244; mortality model only), the Surgical Apgar Score (three studies, n = 10,795), the Charlson Comorbidity Index (two studies, n = 2,463,997), and Donati Surgical Risk Score (two studies, n = 7,121). The accuracy of a further 26 tools was evaluated in single-validation studies. A comparison of tools that were validated in multiple studies is detailed in tables 1 and 2. The general characteristics of all included studies are summarized in table 3.

Quality Assessment

The quality assessment of included studies is summarized in table 3. Seven studies were multicenter and 21 were single center. The data collection was prospective in 19 studies, retrospective in 7, and based on administrative data in 2 studies. Sixteen studies used mortality as an outcome measure, four used morbidity, and eight used both. The study endpoints included 30-day outcome in 12 articles, hospital discharge in 15 articles, and 3 articles also included shorter or longer follow-up times ranging from 1 day to 1 yr. Nineteen studies of the total 28 reported baseline patient characteristics of physiology or comorbidity, surgery, and demographics; selection bias was evident in 12 studies.

Outcomes Reporting

Outcomes are summarized in table 4. Surgical mortality at 30 days varied between 1.25 and 12.2% and at hospital discharge between 0.8 and 24.7%.

All but one[25] of the six studies which separately tested the discrimination of stratification tools for morbidity and mortality reported that morbidity prediction was less accurate. There was considerable heterogeneity in the definition of morbidity in the 12 studies that reported this outcome (see appendix 3 for summary), and in keeping with this, there was wide variation in complication rates in different studies (between 6.7[26] and 50.4%).[25]

Calibration

Calibration was poorly reported: 16 studies did not report calibration at all; of the remaining 11 articles, 2 reported only whether the models were of "good fit," without reporting the appropriate statistics. One article did not report calibration in their results, despite stating in the methods that they would calculate it.[27]

Risk Stratification Tools Using Preoperative Data Only

Four entirely preoperative risk stratification tools (ASA-PS, Surgical Risk Scale, Surgical Risk Score, and the Charlson Comorbidity Index) were validated in multiple studies. The Surgical Risk Scale and the Surgical Risk Score both contain the ASA-PS, and the urgency and severity of surgery; both have also been multiply validated. The Surgical Risk Score[28,29] was developed and originally validated in Italy[29] and contains the ASA-PS, a 3-point scale modification of the Johns Hopkins surgical severity criteria and a binary definition of surgical urgency (elective *vs.* emergency). The only published study evaluating the Surgical Risk Score after its initial validation found it to be poorly predictive of inpatient mortality.[28] The Surgical Risk Scale[30–32] uses the ASA-PS alongside United Kingdom definitions of operative urgency (a 4-point scale defined by the United Kingdom National Confidential Enquiry into Postoperative Death and Outcome) and severity (the British United Provident Association classification which is used to rank surgical procedures for the purposes of financial billing in the private sector). Both studies validating this system after its initial development found it to be a moderately discriminant tool (AUROC >0.8).[30,32]

A further 18 different risk stratification tools using solely preoperative data were validated in single publications. Several of these were originally derived and validated for purposes other than the prediction of generic morbidity and mortality: these include cardiac risk prediction scores,[27,32,33] measures of nutritional status,[34] and frailty indices.[27] These tools are described in appendix 4.

Risk Stratification Tools Incorporating Intra- and Postoperative Data

The POSSUM and P-POSSUM scores were the most frequently used tools in heterogeneous surgical cohorts. The POSSUM score was derived by multivariable logistic regression analysis and contains 18 variables, of which 12 were measured preoperatively and 6 at hospital discharge; two separate equations, for morbidity and mortality, were developed and validated.[17,35] After recognition that the POSSUM model overpredicted adverse outcome, the Portsmouth variation (P-POSSUM) was developed to predict mortality, using the same composite variables but a different calculation.[36] P-POSSUM has been used in a larger number of more recent studies[28–30,32,37] than the original POSSUM[25,29,30] and has been found to be of moderate to high discriminant accuracy (AUROC varying between 0.68 and 0.92) with the exception of one Australian study.[37]

Medical Risk Prediction Tools Adapted for Surgical Risk Stratification

Two risk stratification tools, which have been multiply validated, APACHE II[38] and the Charlson Index,[39] were developed for the purposes of risk adjustment and prediction in nonsurgical settings. APACHE II was developed in 1985 as a tool for predicting hospital mortality in patients admitted to

Table 1. Mortality Models Validated in Multiple Studies

Model	No. of Variables	Pre-, Intra-, or Postoperative Data Used	Original Derivation Cohort and Outcome	Studies (n)	Author	Patients (n)	Type of Surgery	Surgical Urgency	Endpoint	AUROC (CI)
APACHE II	16	Postoperative	Critical care patients; all diagnoses (not just surgical); hospital mortality[38]	3	Jones[25]	117	Gastrointestinal, vascular, renal, and urology	All	30 d	HDU admission score: 0.539 (+/–0.083)
					Osler[63]	5,322	Noncardiac	All	Hospital discharge	ICU admission score: 0.806
					Stachon[40]	271	Ortho, spinal, trauma, visceral surgery, limb surgery	All	Hospital discharge	First 24-h worst score: 0.777
ASA-PS	1	Preoperative	General surgical patients[4]	2	Sutton[31]	1,946	Gastrointestinal, vascular, trauma	All	Hospital discharge	0.93 (0.90–0.97)
					Donati[29]	1,849	Abdominal, vascular, orthopedics, urology, endocrine, otolaryngology, neurological, gynecology, eye, thoracic, other	All	Hospital discharge	0.810 (0.792–0.828)
Charlson	17	Preoperative	Medical patients; 10-yr mortality[39]	2	Atherly[42]	2,167	General, vascular		30 d	0.52
					Sundararajan[43]	2,461,830	All inpatient surgery	All	Hospital discharge	0.85–0.87*
POSSUM	18	Pre- and intraoperative	General surgery; 30-d mortality[17]	3	Jones[25]	117	Gastrointestinal, vascular, renal, and urology	All	30 d	0.753 (+/–0.081)
					Donati[29]	1,849	Abdominal, vascular, orthopedics, urology, endocrine, otolaryngology, neurological, gynecology, eye, thoracic, other	All	Hospital discharge	0.915 (0.884–0.947)
					Brooks[30]	949	General, colorectal, upper gastrointestinal, urology, head, and neck	All	30 d	0.92 (0.90–0.95)
P-POSSUM	18	Pre- and intraoperative	General surgery; 30-d mortality[36,64]	5	Organ[37]	229	General, vascular, otolaryngology, plastics, thoracic, urology, other	All	30 d	0.68 (0.57–0.78)
					Donati[29]	1,849	Abdominal, vascular, orthopedics, urology, endocrine, otolaryngology, neurosurgery, gynecology, eye, thoracic, other	All	Hospital discharge	0.912 (0.898–0.924)
					Brooks[30]	949	General, colorectal, upper gastrointestinal, urology, head, and neck	All	30 d	0.92 (0.90–0.95)
					Neary[32]	2,349	General, vascular, otolaryngology, urology, orthopedics, other	Emergent and urgent	30 d	0.90 (0.87–0.93)
									1 yr	0.90 (0.8–1.0)
					Haga[28]	5,272	Gastrointestinal and hepatobiliary	Elective	30 d	0.74 (0.63–0.86)
									Hospital discharge	0.81 (0.75–0.88)
Surgical Apgar	3	Intraoperative	Colorectal; 30-d mortality[65]	2	Regenbogen[66]	4,119	General and vascular	All	30 d	0.81
					Haynes[67]	5,909	Any noncardiac	All	Inpatient	0.77
Surgical Risk Scale	3	Preoperative	General surgery; inpatient mortality[31]	3	Sutton[31]	1,946	Gastrointestinal, vascular, trauma	All	Hospital discharge	0.95 (0.93–0.97)
					Brooks[30]	949	General, colorectal, upper gastrointestinal, urology, head, and neck	All	30 d	0.89 (0.86–0.93)
					Neary[32]	2,349	General, vascular, otolaryngology, urology, orthopedics, other	Emergent and urgent	30 d	30 d: 0.85 (0.82–0.89)
									1 yr	1 yr: 0.84 (0.75–0.94)
Surgical Risk Score (Donati)	3	Preoperative	General surgery; inpatient mortality[29]	2	Donati[29]	1,849	Abdominal, vascular, orthopedics, urology, endocrine, otolaryngology, neurosurgery, gynecology, eye, thoracic, other	All	Hospital discharge	0.888 (0.838–0.937)
					Haga[28]	5,272	Gastrointestinal, hepatobiliary	Elective	Hospital discharge	0.73 (0.63–0.83)

* Varied between year and whether ICD-9 or ICD-10 administrative data used (table 3).

APACHE II = Acute Physiology and Chronic Health Evaluation II; ASA-PS = American Society of Anesthesiologists' Physical Status score; AUROC = area under receiver operating characteristic curve; HDU = high dependency unit; ICD = International Classification of Diseases; ICU = intensive care unit; (P)-POSSUM: (Portsmouth)-Physiological and Operative Severity Score for the enUmeration of Morbidity and morbidity.

Table 2. Morbidity Models Validated in Multiple Studies

Model	Number of Validation Studies	Pre-, Intra-, or Postoperative Variables	Original Derivation Cohort and Outcome	Author	N	Type of Surgery and Urgency	Surgical Urgency	Endpoint	AUROC for Outcome
ASA-PS	3	Preoperative	General surgery[4]	Goffi[68]	187	General	All	30 d (mortality and morbidity combined)	0.777
				Hightower[69]	32	Major abdominal (gastrointestinal, urology)	Elective	7 d	0.688 (0.523–0.851)
				Makary[27]	594	Unselected inpatient	All	Hospital discharge	0.626
APACHE II	1	Postoperative	Critical care patients; any diagnosis (not just surgical); hospital mortality[38]	Goffi[68]	187	General	All	30 d (mortality and morbidity combined)	Hospital admission score: 0.866 Preoperative score: 0.894*
POSSUM	2	Pre- and intraoperative	General surgery; 30-d morbidity[17]	Jones[25]	117	Gastrointestinal, vascular, renal, and urology	All	30 d	0.82
				Brooks[30]	949	General, colorectal, upper gastrointestinal, urology, head, and neck	All	30 d	0.92
Surgical Apgar	3	Intraoperative	Colorectal; 30-d mortality[65]	Gawande[65]	767	General and vascular	All	30 d (mortality and morbidity combined)	0.72
				Regenbogen[66]	4,119	General and vascular	All	30 d	0.73
				Haynes[67]	5,909	Any noncardiac	All	Inpatient	0.70

* See table 4 for AUROC on subgroup cohort.

APACHE II = Acute Physiology and Chronic Health Evaluation II; ASA-PS = American Society of Anesthesiologists' Physical Status score; AUROC = area under receiver operating characteristic curve; POSSUM = Physiological and Operative Severity Score for the enUmeration of Morbidity and Mortality.

critical care; the score consists of 12 physiological variables and an assessment of chronic health status. This approach has face validity, as APACHE II is a summary measure of acute physiology and chronic health, both of which may influence surgical outcome. Only one of the four studies reporting the APACHE II score's predictive accuracy used it in the way originally intended: by incorporating the most deranged physiological results within 24 h of critical care admission.[40]

The Charlson comorbidity score was developed to predict 10-yr mortality in medical patients.[39] A combined age-comorbidity score was subsequently validated for the prediction of long-term mortality in a population of patients who had essential hypertension or diabetes and were undergoing elective surgery.[41] It is the original Charlson score, however, which is used in two studies identified in our search to stratify risk of short-term outcome.[42,43] These two studies reported very different predictive accuracy for the Charlson score; however, the largest single study included in this entire review found the Charlson score (measured using administrative data) to be a moderately accurate tool.[44]

Discussion

The purpose of this systematic review was to identify all risk stratification tools, which have been validated in heterogeneous patient cohorts, and to report and summarize their discrimination and calibration. We have found a plethora of instruments that have been developed and validated in single studies, which unfortunately limits any assessment of their usefulness and generalizability. A smaller number of tools have been multiply validated which could be used universally for perioperative risk prediction; of these, the P-POSSUM and Surgical Risk Scale have been demonstrated to be the most consistently accurate systems.

Risk Stratification Tools in Practice: Complexity versus *Parsimony*

There are two key considerations when assessing the clinical utility of the various risk stratification tools reviewed in our study. First, what level of predictive accuracy is fit for the purposes of risk stratification? Second, what is the likelihood that each of the described instruments may be used in everyday practice by clinicians? Although the answer to the first question may be to aim as "high" (accurate) as possible, this must also be balanced against the issues raised by the second question. Risk models incorporating over 30 variables may be highly accurate but are less likely to be routinely incorporated into preoperative assessment processes than scores of similar performance that use only a few data points. Furthermore, clinical experience tells us that the clinician is less likely to use complex mathematical formulae, as opposed to additive scores, when attempting to risk stratify patients at the bedside or in the preoperative clinic.[1]

P-POSSUM

The P-POSSUM model was developed in the United Kingdom and has since been validated in Japan, Australia, and Italy. Although this is the most frequently and widely validated model identified by our study, it has some limitations. First, it includes both preoperative and intraoperative variables, and therefore cannot be used for preoperative risk prediction. Second, several of the variables are subjective (*e.g.*, chest radiograph interpretation), carrying the risk of measurement error. Third, in common with the original POSSUM, the P-POSSUM tends to overestimate risk in low-risk patients. Fourth, it contains 18 variables, which must be entered into a regression equation to obtain a predicted percentage risk value, and clinicians may not wish to use such a complex system. Finally, the inclusion of intraoperative variables, particularly blood loss, which may be influenced by surgical technique, runs the risk of concealing poor surgical performance, therefore, jeopardizing its face validity as a risk adjustment model for comparative audit of surgeons or institutions.

Surgical Risk Scale

The Surgical Risk Scale consists entirely of variables that are available before surgery, making it a useful tool for preoperative risk stratification for the purposes of clinical decision making. However, there are also some limitations. First, it incorporates the ASA-PS, which may be subject to interobserver variability and therefore measurement error.[44–46] Second, the surgical severity coding is not intuitive, and some familiarity with the British United Provident Association system would be required for bedside estimation, unless a reference manual was available. Finally, it has only been validated in single-center studies within the United Kingdom; therefore, its generalizability to patient populations in the United States and worldwide is unknown.

Other Options

The ASA-PS is widely used as an indicator of whether or not a patient falls into a high-, medium-, or low-risk population, but it was not originally intended to be used for the prediction of adverse outcome in individual subjects.[4] It is perhaps surprising that the ASA-PS was reported as having good discrimination for predicting postoperative mortality, as it is a very simple scoring system, which has been demonstrated to have only moderate to poor interrater reliability.[44–47] Nevertheless, the ASA-PS has face validity as an assessment of functional capacity, which is increasingly thought to be a significant predictor of patient outcome, as demonstrated by more sophisticated techniques such as cardiopulmonary exercise testing.[48] Although it is possible that this provides some explanation for the high discriminant accuracy for ASA-PS found in this systematic review, it is possible that publication bias, favoring studies with "positive" results, may also be a factor.

The Biochemistry and Hematology Outcome Model is a parsimonious version of POSSUM, which omits the subjective variables such as chest radiography and electrocardiogram results. It also has the advantage of consisting of variables which are all available preoperatively, with the exception of operative severity. Given the Biochemistry and Hematology

Table 3. Characteristics of All Included Studies

First Author	Region	N	No. of Centers	Data Acquisition	Selection Bias	Subject Description	Type of Surgery	Surgical Urgency	Models Used	Validation Cohort: Internal *vs.* External *vs.* Temporal*	Outcome	Endpoint
Atherly[42]	United States	2,167	M	Administrative (ICD-9 codes)	N	N	General, vascular	All	Charlson Comorbidity I ndex based on ICD-9 codes	External	Mortality	30 d
Brooks[30]	United Kingdom	949	S	Prospective	N	N	General, colorectal, upper GI, urology, head, and neck	All	POSSUM, P-POSSUM, Surgical Risk Scale	Temporal	Mortality	30 d
Dasgupta[33]	Canada	125	S	Prospective	Y: >70 yr only	Y	General, abdominal, orthopedic, neurosurgery, carotid surgery	Elective	Detsky Index Edmonton Frail Scale	External	Morbidity	Hospital discharge
Davenport[26]	United States	5,878	S	Prospective	N	Y	General, neurosurgery, orthopedic, plastic, thoracic, vascular	All	ASA-PS	External	Morbidity, mortality	30 d
Donati[29]	Italy	1,849	M	Prospective	N	Y	Abdominal, vascular, ortho, urology, endocrine, otolaryngology, neuro, gynecology, eye, thoracic, other	All	Surgical Risk Score, POSSUM, P-POSSUM	Temporal	Mortality	Hospital discharge
Gawande[65]	United States	767	S	Retrospective	N	Y	General and vascular	All	Surgical Apgar Score	Temporal	Major complications or mortality (combined endpoint)	30 d
Goffi[68]	Italy	187	S	Prospective	N	N	General	All	ASA-PS, APACHE II on hospital admission, APACHE II immediately preoperative	External	Combined endpoint: mortality, morbidity	30 d
Hadjianastassiou[70]	United Kingdom	4,494	S	Retrospective	N	Y	Maxillofacial, general, orthopedic, renal, urology, neuro	All	Surgical Mortality Score	Internal	Mortality	Hospital discharge
Haga[28]	Japan	5,272	M	Prospective	N	Y	Gastrointestinal, hepatobiliary	Elective	E-PASS, mE-PASS, P-POSSUM, Surgical Risk Score (Donati)	External	Mortality	Hospital discharge
Haynes[67]	International	5,909	M	Prospective	N	Y	Any noncardiac	All	Surgical Apgar	External	Mortality, morbidity	Hospital discharge
Hightower[69]	United States	32	S	Prospective	Y: major abdominal and fit enough for CPET	Y	Major abdominal (gastrointestinal, urology)	Elective	ASA-PS	External	Morbidity	7 d
Hobson[71]	United Kingdom	163	S	Prospective	Y: emergent surgery only	N	General, gynecology, renal, urology, vascular	Emergent	POSSUM, P-POSSUM	External	Mortality	30 d
Jones[25]	United Kingdom	117	S	Prospective	Y: HDU admissions only	N	Gastrointestinal, vascular, renal, and urology	All	POSSUM, APACHE II	External	Morbidity, mortality	30 d
Kuzu[34]	Turkey	460	S	Prospective	N	Y	Gastrointestinal, vascular, hepatobiliary, gynecology	Elective	Nutritional Risk Index, Maastricht Index, Subjective Global Assessment, Mini Nutritional Assessment	External	Mortality, morbidity	Hospital discharge or 30 d (whichever later)

(*Continued*)

Table 3. (*Continued*)

First Author	Region	N	No. of Centers	Data Acquisition	Selection Bias	Subject Description	Type of Surgery	Surgical Urgency	Models Used	Validation Cohort: Internal *vs.* External *vs.* Temporal*	Outcome	Endpoint
Liebman[49]	The Netherlands	33,224	S	Prospective	N	Y	General and trauma	Emergent	Identification of Risk In Surgical patients	Internal	Mortality, morbidity	Hospital discharge
Makary[27]	United States	594	S	Prospective	Y: elective only	Y	Unselected inpatient	Elective	ASA-PS, Lee RCRI, and Eagle Scores alone and in combination with Frailty Index	External	Morbidity	Hospital discharge
Nathanson[72]	United States	13,417	M	Retrospective	Y: postoperative ICU admission only	Y	All excluding cardiac, neurosurgery, and trauma	Elective and emergent in separate cohorts	MPM_0-III	External	Mortality	Hospital discharge
Neary[32]	United Kingdom	2,349	S	Prospective	Y: emergent and urgent only	N	General, vascular, otolaryngology, urology, orthopedic, other	Emergent and urgent	RCRI, P-POSSUM, Surgical Risk Scale, BHOM	External	Mortality	30 d and 1 yr
Organ[37]	Australia	229	S	Retrospective	Y: ICU only	N	General, vascular, otolaryngology, plastics, thoracic, urology, other	All	P-POSSUM	External	Mortality	30 d
Osler[63]	United States	5,322	S	Retrospective	Y: ICU only	N	Noncardiac	All	APACHE II, ICISS	External	Mortality	Hospital discharge
Pillai[73]	New Zealand	6,492	M	Retrospective	N	Y	GI, breast, endocrine, vascular, gynecology, orthopedic, hepatobiliary	All	Otago Surgical Audit Score	External	Morbidity	Hospital discharge
Regenbogen[66]	United States	4,119	S	Prospective	N	Y	General and vascular	All	Surgical Apgar Score	External	Mortality, morbidity	30 d
Stachon[40]	Germany	271	S	Prospective	Y: ICU only	Y	Orthopedic, spinal, trauma, visceral surgery, limb surgery	All	APACHE II, SAPS II, APACHEN, SAPSN	External	Mortality	Hospital discharge
Stachon[74]	Germany	283	S	Prospective	Y: ICU only	Y	Orthopedic, spinal, trauma, visceral surgery, limb surgery	All	DELAWARE, APACHE II, SAPS II	Temporal/ external	Mortality	Hospital discharge
Story[75]	Australia	256	S	Retrospective	Y: >70 y only	Y	General, colorectal, orthopedic, plastics, urology, vascular, other	All	Perioperative Mortality Risk Score	Internal	Mortality	30 d
Sundararajan[43]	Australia	2,461,830	M	Administrative (ICD-9 and -10 codes)	N	Y	All inpatient surgery	All	Charlson Comorbidity Index using administrative data (ICD-9 and ICD-10 coding)	External	Mortality	Hospital discharge
Sutton[31]	United Kingdom	1,946	S	Prospective	N	N	Gastrointestinal, vascular, trauma	All	Surgical Risk Scale; ASA-PS	Temporal	Mortality	Hospital discharge

* Definitions of validation cohorts: External = validation in new cohort unrelated to derivation study; Internal = validation in split sample of same study population as derivation cohort; Temporal = validation in new cohort from derivation study but same institution(s).

APACHE II = Acute Physiology and Chronic Health Evaluation II; APACHEN = Acute Physiology and Chronic Health Evaluation-Nucleated; ASA-PS = American Society of Anesthesiologists' Physical Status score; BHOM = Biochemistry and Hematology Outcome Model; CPET = cardiopulmonary exercise testing; DELAWARE = Dense Laboratory Whole Blood Applied Risk Estimation; E-PASS = Estimation of Physiologic Ability and Surgical Stress; GI = gastrointestinal; HDU = high dependency unit; ICD = International Classification of Diseases; ICISS = International Classification of Disease Illness Severity Score; ICU = intensive care unit; M = multicenter; (m)E-PASS = (modified) Estimation of Physiologic Ability and Surgical Stress; MPM_0 = Mortality Prediction Model; N = no; (P)-POSSUM = (Portsmouth)-Physiology and Operative Severity Score for the enUmeration of Morbidity and Mortality; RCRI = Revised Cardiac Risk Index; SAPS = Simplified Acute Physiology Score; SAPSN = Simplified Acute Physiology Score-Nucleated; S = Single center; Y = Yes.

Table 4. Outcomes, Discrimination, and Calibration

Author	Models Used	Endpoint	Morbidity (%)	AUROC Morbidity (95% CI)	Mortality (%)	AUROC Mortality (95% CI)	Calibration (*P* Value for Hosmer–Lemeshow Statistic Unless Otherwise Stated)
Atherly[42]	Charlson Comorbidity Index using ICD-9 coding	30 d	NR	NR	1.3	0.47	NR
Brooks[30]	POSSUM	30 d	NR	NR	8.4	POSSUM: 0.92	NR
	P-POSSUM					P-POSSUM: 0.92	NR
	Surgical Risk Scale					Surgical Risk Scale: 0.89	NR
Dasgupta[33]	Detsky Index	Hospital discharge	25	Detsky: 0.51 (0.39–0.63)	0.8	NR	NR
	Edmonton Frail Scale			Edmonton Frail Scale: 0.69 (0.58–0.79)			NR
Davenport[26]	NSQIP	30 d	6.7	NSQIP: 0.769	1.5	NSQIP: 0.958	NR
	ASA-PS			ASA-PS: 0.722		ASA-PS: 0.889	NR
	ASA-PS and NSQIP combined			NSQIP with ASA-PS: 0.782		NSQIP with ASA-PS: 0.960	NR
Donati[29]	Surgical Risk Score	Hospital discharge	NR	NR	1.9	Surgical Risk Score: 0.888 (0.838–0.937)	0.744
	POSSUM					POSSUM: 0.915 (0.884–0.947)	0.0004
	P-POSSUM					P-POSSUM: 0.912 (0.898–0.924)	0.1528
	ASA-PS					ASA-PS: 0.810 (0.792–0.828)	NR
Gawande[65]	Surgical Apgar Score	30 d	9.1	NR	1.4	Combined outcome of mortality and morbidity: 0.72	Pearson goodness-of-fit: 0.57
Goffi[68]	ASA-PS, Preoperative APACHE II	30 d	Overall: 26.7, Elective: 15.9, Emergent: 57.1	NR	Overall: 8.6, Elective: 4.3, Emergent: 20.4	Combined outcome of mortality and morbidity:	NR
						ASA-PS: 0.777	NR
						Hospital Admission APACHE II: 0.866	NR
						Immediate preoperative APACHE II: overall: 0.894, elective surgery: 0.826, emergent surgery: 0.873, cancer surgery: 0.915, noncancer surgery: 0.869	NR
Hadjianastassiou[70]	Surgical Mortality Score	Hospital discharge	NR	NR	4.1	0.82 (0.78–0.85)	0.10
Haga[28]	E-PASS, mE-PASS, P-POSSUM, Surgical Risk Score (Donati)	Hospital discharge 30 d	NR	NR	NR	Hospital discharge / 30 d	NR
						E-PASS: 0.86 (0.79–0.93) / 0.82 (0.69–0.95)	
						mE-PASS: 0.86 (0.79–0.92) / 0.81 (0.66–0.96)	
						P-POSSUM: 0.81 (0.75–0.88) / 0.74 (0.63–0.86)	
						Surgical Risk Score: 0.73 (0.63–0.83) / —	
Haynes[67]	Surgical Apgar	Hospital discharge	9.2 (major)	0.70	1.4	0.77	NR
Hightower[69]	ASA-PS	7 d	50	0.688 (0.523–0.851)	NR	NR	NR
Hobson[71]	POSSUM, P-POSSUM	30 d	NR	NR	30 d: 9.2	30 d:POSSUM: 0.946, P-POSSUM: 0.940	NR
		Hospital discharge			Hospital discharge: 12.9		
Jones[25]	POSSUM	30 d	50.4	POSSUM: 0.82	11.1	POSSUM: 0.75	NR
	APACHE II					APACHE II: 0.54	
Kuzu[34]	Subjective Global Assessment	Hospital discharge or 30 d (whichever later)	28.47	Subjective Global Assessment: 0.669	4.34	Subjective Global Assessment: 0.687	NR
	Nutritional Risk Index			Nutritional Risk Index: 0.659		Nutritional Risk Index:0.797	
	Maastricht Index			Maastricht Index: 0.671		Maastricht Index: 0.743	

(*Continued*)

Table 4. (*Continued*)

Author	Models Used	Endpoint	Morbidity (%)	AUROC Morbidity (95% CI)	Mortality (%)	AUROC Mortality (95% CI)	Calibration (*P* Value for Hosmer–Lemeshow Statistic Unless Otherwise Stated)
Liebman[49]	Identification of Risk In Surgical patients	Hospital discharge	13.3	0.77	2.2	0.90	NR
Makary[27]	ASA-PS, Lee, and Eagle with and without Frailty Index added	Hospital discharge	Not stated for entire cohort	ASA-PS: 0.626 ASA-PS + Frailty: 0.699 Lee: 0.618 Lee + Frailty: 0.669 Eagle: 0.678 Eagle + Frailty: 0.714	NR	NR	NR (but reported that this would be calculated in methods)
Nathanson[72]	MPM_0-III	Hospital discharge	NR	NR	Elective: 5.3 Emergent: 14.4	Elective: 0.79 Emergency: 0.79	Good fit
Neary[32]	RCRI	30 d and 1 yr	NR	NR	30 d: 6.0, 1 yr: 10.8	RCRI:30 d: 0.731 yr: 0.71	NR
	P-POSSUM					P-POSSUM:30 d: 0.901 yr: 0.90	Good fit
	Surgical Risk Scale					Surgical Risk Scale:30 d: 0.851 yr: 0.84	Good fit
	BHOM					BHOM:30 d: 0.841 yr: 0.86	Good fit
Organ[37]	P-POSSUM	30 d	NR	NR	12.2	0.68	<0.001
Osler[63]	APACHE II	Hospital discharge	NR	NR	13.9	APACHE II: 0.806	0.002
	ICISS					ICISS: 0.892	0.15
	APACHE and ICISS combined					Combined: 0.903	0.038
Pillai[73]	Otago Surgical Audit Score	Hospital discharge	NR for validation cohort	0.86	NR	NR	Good fit
Regenbogen[66]	Surgical Apgar Score	30 d	14.1	0.73	2.3	0.81	NR
Stachon[40]	APACHE II	Hospital discharge	NR	NR	24.7	APACHE II: 0.777	NR
	SAPS II					SAPS II: 0.785	
	APACHEN					APACHEN: 0.829	
	SAPSN					SAPSN: 0.823	
Stachon[74]	DELAWARE	Hospital discharge	NR	NR	23.3	DELAWARE: 0.813	0.44
	APACHE II					0.777	NR
	SAPS II					0.785	NR
Story[75]	Perioperative Mortality Risk Score	30 d	NR	NR	6.0	0.79	0.35
Sundararajan[43]	Charlson Comorbidity Index using administrative data (ICD-9 and ICD-10 coding)	Hospital discharge	NR	NR	Overall mortality not reported	ICD-9 1996–1997: 0.87ICD-9 1997–1998: 0.86ICD-10 1998–1999: 0.85ICD-10 1999–2000: 0.86ICD-10 2000–2001: 0.86ICD-10 2001–2002: 0.85	NR
Sutton[31]	Surgical Risk Scale	Hospital discharge	NR	NR	2.41	Surgical Risk Scale: 0.95	0.65
	ASA-PS					ASA-PS: 0.93	NR

APACHE II = Acute Physiology and Chronic Health Evaluation II; APACHEN = Acute Physiology and Chronic Health Evaluation-Nucleated; ASA-PS = American Society of Anesthesiologists' Physical Status score; AUROC = area under receiver operating characteristic curve; BHOM = Biochemistry and Hematology Outcome Model; DELAWARE = Dense Laboratory Whole Blood Applied Risk Estimation; ICD = International Classification of Diseases; ICISS = International Classification of Disease Illness Severity Score; (m)E-PASS = (modified) Estimation of Physiologic Ability and Surgical Stress; MPM_0 = Mortality Prediction Model; NR = not reported; NSQIP = National Surgical Quality Improvement Program; (P)-POSSUM = (Portsmouth)-Physiology and Operative Severity Score for the enUmeration of Morbidity and Mortality; RCRI = Revised Cardiac Risk Index; SAPS = Simplified Acute Physiology Score; SAPSN = Simplified Acute Physiology Score-Nucleated.

Outcome Model's similarity in predictive accuracy to P-POSSUM in the one study, we identified which made a direct comparison,[32] this system warrants further evaluation. Finally, the Identification of Risk In Surgical patients score was developed in The Netherlands and consists of four variables (age, acuity of admission, acuity of surgery, and severity of surgery). In the study, which developed and validated it on separate cohorts, the validation AUROC was 0.92.[49] Again, further investigation of this simple system would be useful.

Generalizability of Findings

Clinical and Methodological Heterogeneity. Clinical heterogeneity (both within- and between-cohort patient heterogeneity) and methodological heterogeneity (between-study differences in the outcome measures used) are both likely to have had a significant influence on some of our findings. For example, between-cohort heterogeneity, and variation in how morbidity is defined (appendix 2), may explain the wide range of morbidity rates reported in different studies. Heterogeneity of morbidity definitions may also in part explain the lower accuracy of models for predicting morbidity compared with mortality. On a different note, our study included all populations of patients who were determined to be heterogeneous, using the definitions described in our methods. However, the degree of heterogeneity varied among studies, including whether or not patients of all surgical urgency categories were included, and this may have affected the predictive accuracy of models in different studies.

Objective *versus* Subjective Variables and Issues Surrounding Data Collection Methodology. The variables included in risk stratification tools may be classified as objective (*e.g.*, biochemistry and hematology assays), subjective (*e.g.*, interpretation of chest radiographs), and patient-reported (*e.g.*, smoking history). In some clinical settings, the reliability of nonobjective data may be questionable; for example, previous reports have demonstrated significant interrater variation in the interpretation of both chest radiographs[50] and electrocardiograms.[51] Patients may also under- or overestimate various elements of their clinical or social history when questioned in the hospital setting. Despite these concerns, the discrimination of predictors incorporating patient-reported and patient-subjective variables was high in the studies included. This may be due to publication bias; it may also be explained by the fact that in all of these studies, data were collected prospectively by trained staff. Previous work has demonstrated an association between interobserver variability in the recording of risk and outcome measures, and the level of training that data collection staff have received.[52] These caveats are important when considering the generalizability of our findings to the everyday clinical setting, where data reporting and interpretation may be conducted by different types and grades of clinical staff. Finally, concerns have also been raised over the clinical accuracy of administrative data used for case-mix adjustment purposes.[53,54] However, one large study included in our review[43] showed high discriminant performance when using International Classification of Diseases 9 and 10 administrative coding data to define the Charlson Index variables.

Limitations of This Study

This study has limitations in a number of factors. First, the focus was on studies that measured the discrimination and/or calibration of risk stratification tools in cohorts that were heterogeneous in terms of surgical specialities; therefore, a large number of single-speciality cohort studies identified in the search were excluded from the analysis.

Second, although the inclusion criteria for our review ensured that a standard measure of discrimination was reported (AUROC or c-statistic), many studies did not report measures of calibration. However, in a systematic review such as this, calibration may be seen to be a less important measure of goodness-of-fit than discrimination for a number of reasons. Calibration can only be used as a measure of performance for models that generate an individualized predicted percentage risk of an outcome (*e.g.*, the POSSUM systems) as opposed to summative scores, which use an ordinal scale to indicate increasing risk (*e.g.*, the ASA-PS). Calibration drift is likely to occur over time and will be affected by changes in healthcare delivery; good calibration in a study over 30 yr ago may be unlikely to correspond to good calibration today.[55,56] Although such calibration drift may affect the usefulness of a model for predicting an individual patient's risk of outcome, poorly calibrated but highly discriminant models will still be of value for risk adjustment in comparative audit. Finally, the probability of the Hosmer–Lemeshow statistic being significant (thereby indicating poor calibration) increases with the size of the population being studied.[57] This may explain why many of the large high-quality studies we evaluated did not report calibration or reported that calibration was poor.

Third, by using the AUROC as the sole measure of discrimination, a number of studies were excluded, particularly earlier articles that used correlation coefficients between risk scores and postoperative outcomes. This was felt to be necessary, as a uniform outcome measure provides clarity to the reader. Fourth, publication bias, where studies are preferentially submitted and accepted for publication if the results are positive, is likely to be a particular problem in cohort studies. Finally, despite an extensive literature search, it is possible that some studies which would have been eligible for inclusion may have been missed. Multiple strategies have been used to prevent this; however, in a review of this size, it is possible that a small number of appropriate articles may have been omitted.

Future Directions

Undertaking clinical risk prediction should be a key tenet of safe high-quality patient care, it facilitates informed consent and enables the perioperative team to plan their clinical management appropriately. Equally, accurate risk adjustment is required to enable meaningful comparative audit between teams and institutions, to facilitate quality improvement for patients and providers. Although we identified dozens of scores and models

which have been used to predict or adjust for risk, very few of these achieved the aspiration of being derived from entirely preoperative data, and of being accurate, parsimonious, and simple to implement. The Surgical Risk Scale is the system that comes closest to achieving these goals; the P-POSSUM score is more accurate, but its value is limited by the fact that some of the variables are only available after surgery has been completed. Future work which might be of value would include further comparison of the Surgical Risk Scale, P-POSSUM, and objective models such as the Biochemistry and Hematology Outcome Model in international multicenter cohorts and further investigation of models which combine novel variables such as measures of functional capacity, nutritional status, and frailty.

There is another possible approach. The American College of Surgeons' National Surgical Quality Improvement Program was created in the 1990s to facilitate risk-adjusted surgical outcomes reporting in Veterans' Affairs hospitals, and now also includes a number of private sector institutions. Risk adjustment models are produced annually and observed that the expected ratios of surgical outcomes are reported back to institutions and surgical teams to facilitate quality improvement. This organization has published a number of risk calculators to help clinicians to provide informed consent and plan perioperative care. However, none of these calculators have been included in our review, as they have all been developed and validated for use in either specific types of surgery (*e.g.*, pancreatectomy,[58] bariatric,[59,60] or colorectal[60] surgery) or for specific outcomes (*e.g.*, cardiac morbidity and mortality).[61] A parsimonious, entirely preoperative National Surgical Quality Improvement Program model for predicting mortality in heterogeneous cohorts would be of value in the United States; its validation in international multicenter studies would also be a worthwhile endeavor.

Finally, although there are multiple studies aimed at developing and validating risk stratification tools, we do not know how widely such tools are used. Use of mobile technology, such as apps to enable risk calculation using complex equations at the bedside, might increase the use of accurate risk stratification tools in day-to-day practice. Importantly, in surgical outcomes research, there is an absence of impact studies, measuring the effect of using risk stratification tools on clinician behavior, patient outcome, and resource utilization. Randomized, controlled trials to evaluate impact, further validation of existing models across healthcare systems, and establishing the infrastructure required to facilitate such work, including the routine data collection of risk and outcome data, should be of the highest priority in health services research into surgical outcome.[62]

The authors thank Judith Hulf, F.R.C.A., Past President, Royal College of Anaesthetists, London, United Kingdom.

Appendix 1. Preferred Reporting Items for Systematic reviews and Meta-analyses Checklist[12]

Section/Topic	#	Checklist Item	Reported on Page No.
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	959
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	959
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	959–60
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	960
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (*e.g.*, Web address), and, if available, provide registration information including registration number.	n/a
Eligibility criteria	6	Specify study characteristics (*e.g.*, PICOS, length of follow-up) and report characteristics (*e.g.*, years considered, language, publication status) used as criteria for eligibility, giving rationale.	960–1
Information sources	7	Describe all information sources (*e.g.*, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	960–1
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix 2

(*Continued*)

Appendix 1. (*Continued*)

Section/Topic	#	Checklist Item	Reported on Page No.
TITLE			
Study selection	9	State the process for selecting studies (*i.e.*, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	960
Data collection process	10	Describe method of data extraction from reports (*e.g.*, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	960
Data items	11	List and define all variables for which data were sought (*e.g.*, PICOS, funding sources) and any assumptions and simplifications made.	960
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	960
Summary measures	13	State the principal summary measures (*e.g.*, risk ratio, difference in means).	961
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (*e.g.*, I^2) for each meta-analysis.	n/a
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (*e.g.*, publication bias, selective reporting within studies).	960
Additional analyses	16	Describe methods of additional analyses (*e.g.*, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.	n/a
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Figure 1
Study characteristics	18	For each study, present characteristics for which data were extracted (*e.g.*, study size, PICOS, follow-up period) and provide the citations.	Tables 1–3
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 3
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and CIs, ideally with a forest plot.	n/a
Synthesis of results	21	Present results of each meta-analysis done, including CIs and measures of consistency.	n/a
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	962
Additional analysis	23	Give results of additional analyses, if done (*e.g.*, sensitivity or subgroup analyses, meta-regression [see Item 16]).	n/a
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (*e.g.*, healthcare providers, users, and policy makers).	965
Limitations	25	Discuss limitations at study and outcome level (*e.g.*, risk of bias), and at review-level (*e.g.*, incomplete retrieval of identified research, reporting bias).	970
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	970–1
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (*e.g.*, supply of data); role of funders for the systematic review.	959

n/a = not applicable.

Appendix 2. Search Strategy

MEDLINE

Risk adjustment.mp. or exp Health Care Reform/or exp Risk Adjustment/or exp "Outcome Assessment (Health Care)"/or exp Models, Statistical/or exp Risk/OR exp Risk Assessment/or risk prediction.mp. or exp Risk/or exp Risk Factors/OR predictive value of tests.mp. or exp "Predictive Value of Tests"/OR exp Prognosis/or risk stratification.mp. OR case mix adjustment.mp. or exp Risk Adjustment/OR severity of illness index.mp. or exp "Severity of Illness Index"/OR scoring system.mp.

Combined with:

Surgical Procedures, Operative/OR surgery.mp. or General Surgery/OR operation.mp. or exp Postoperative Complications/

Combined with:

mortality.mp. or exp Hospital Mortality/or exp Mortality/OR morbidity.mp. or exp Morbidity/OR outcome.mp. or exp Fatal Outcome/or exp "Outcome Assessment (Health Care)"/or exp "Outcome and Process Assessment (Health Care)"/or exp Treatment Outcome/OR postoperative complications.mp. or exp Postoperative Complications/OR intraoperative complications.mp. or exp Intraoperative Complications/OR exp Perioperative Care/or perioperative complications.mp. OR prognosis.mp. or exp Prognosis/.

Embase

Risk Factor/or risk adjust$.mp. OR cardiovascular risk/or high risk patient/or high risk population/or risk assessment/or risk factor OR risk stratification.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR *"Scoring System"/OR "Severity of Illness Index"/OR Multivariate Logistic Regression Analysis/or Logistic Regression Analysis OR logistic models/or risk assessment/or risk factors/OR exp Scoring System OR Prediction/or possum.mp. or Scoring System/OR exp Risk Assessment/or risk stratification.mp. OR predict$.mp. OR exp Quality Indicators, Health Care/OR Risk Adjustment/.

Combined with:

exp Surgery/OR exp Surgical Procedures, Operative/OR specialties, surgical/or surgery/OR surg$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR peri-operative period.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR perioperative.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR postoperative.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR perioperative care/or intraoperative care/or postoperative care/or preoperative care.

Combined with:

complicat$.mp. [mp=title, abstract, subject headings, heading word, drug trade name, original title, device manufacturer, drug manufacturer] OR adverse outcome/or prediction/or prognosis/OR exp Postoperative Complication/co, di, ep, su, th [Complication, Diagnosis, Epidemiology, Surgery, Therapy] OR exp Perioperative Complication/or exp Perioperative Period/OR exp Mortality/or exp Surgical Mortality/OR exp Morbidity/OR outcome.mp. or "Outcome Assessment (Health Care)"/or "Outcome and Process Assessment (Health Care)" OR treatment outcome/.

Limits

1980 to August 31, 2011

Exclusions:

("all infant (birth to 23 months)" or "all child (0 to 18 years)" or "newborn infant (birth to 1 month)" or "infant (1 to 23 months)" or "preschool child (2 to 5 years)" or "child (6 to 12 years)" or "adolescent (13 to 18 years)") or (cats or cattle or chick embryo or dogs or goats or guinea pigs or hamsters or horses or mice or rabbits or rats or sheep or swine) or (communication disorders journals or dentistry journals or "history of medicine journals" or "history of medicine journals non index medicus" or "national aeronautics and space administration (nasa) journals" or reproduction journals) or Angioplasty, Balloon/or Angioplasty, Laser/or Angioplasty/or Angioplasty, Balloon, Laser-Assisted/or Angioplasty, Transluminal, Percutaneous Coronary/or ANGIOPLASTY.mp. OR Eye/or Ophthalmology/or Eye Diseases/or OPTHALMOLOGY.mp. or Hearing Loss OR CARDIAC SURGERY.mp. or HEART SURGERY.mp. or Myocardial Revascularization/or Coronary Artery Bypass/or CORONARY SURGERY.mp. or Coronary Artery Bypass, Off-Pump/.

Hand Searching of Reference Lists

The following keywords were searched separately on MEDLINE, Embase, and ISI Web of Science:

POSSUM + surgery
NSQIP
E-PASS
ACE-27
APACHE

In addition, the original development studies for all risk prediction models identified in the initial search were then snowballed by hand searching for citations on MEDLINE, Embase and ISI Web of Science.

Inclusion/Exclusion Criteria

Studies were eligible if they fulfilled the following criteria:

- Studies in adult humans undergoing noncardiac, nonneurological surgery
- Study cohorts that included at least two different surgical subspecialities
- Studies that described the predictive precision of risk models using analysis of receiver operator characteristic curves

Studies were excluded on the basis of these criteria:

- Cohorts including children (under the age of 14 yr)
- Cohorts including patients undergoing cardiac surgery
- Cohorts including patients who did not undergo surgery
- Single-speciality cohort studies (*e.g.*, vascular, orthopedic)
- Studies of ambulatory (day case) surgery
- Studies describing the development of a risk prediction model without subsequent validation in a separate cohort (either in the original study or subsequent cohorts), with the exception of studies of data from the American College of Surgeons' National Surgical Quality Improvement Programme
- Studies in which the items comprising the risk stratification tool were not disclosed in the study report or available from other sources (such as references)
- Studies using outcomes other than morbidity or mortality as their sole outcome measures (*e.g.*, discharge destination, length of stay)

Studies using only a single pathological outcome measure (*e.g.*, reoperation, cardiac morbidity, infectious complications, renal failure).

Appendix 3. Morbidity Definitions

Author	Model(s) Validated	Morbidity Definition
Dasgupta[35]	Detsky Index Edmonton Frail Scale	Cardiac: ischemia, congestive heart failure, new arrhythmia, or sudden death. Respiratory: pneumonia, significant bronchospasm, deep venous thrombosis or pulmonary embolism, or the excessive need for respiratory support. Delirium: required the acute onset and fluctuating course of at least one of the following symptoms as outlined in the Diagnostic and Statistical Manual of Mental Disorders, Revised third edition, occurring anytime on or after postoperative day 1. Disorganized thinking or inattention, altered level of consciousness, psychomotor agitation, disorientation or memory impairment, new perceptual disturbances, or new sleep disturbances (*e.g.*, agitation at night or excessive drowsiness during the day). If patients had a known diagnosis of dementia or were on cholinesterase inhibitors, the occurrence of delirium required more than just disorientation or memory impairment.
Davenport[28]	ASA-PS	One or more of 21 specific NSQIP defined complications: not listed
Gawande[66]	Surgical Apgar Score	According to NSQIP's established definitions: Cardiovascular: cardiac arrest requiring cardiopulmonary resuscitation, myocardial infarction Respiratory: pneumonia, unplanned intubation, pulmonary embolism, failure to wean from the ventilator 48 h after operation Renal: acute renal failure Neurological: coma for 24 h or longer, stroke Infectious: septic shock, sepsis, systemic inflammatory response syndrome Wound: wound disruption, deep- or organ-space surgical site infection Other: bleeding requiring >4 U red cell transfusion within 72 h after operation, deep venous thrombosis, and vascular graft failure
Goffi[69]	ASA-PS APACHE II	Major: cardiac failure; abdominal sepsis; hemoperitoneum; respiratory failure; intestinal obstruction; renal failure Minor: urinary infection; respiratory infection; wound infection
Haynes[72]	Surgical Apgar	NSQIP defined (see study by Gawande[66])
Hightower[70]	ASA-PS	Cardiac events: myocardial ischemia without myocardial infarction; myocardial infarction; dysrhythmias and conduction abnormalities; congestive heart failure; postoperative vasopressors; cardiac arrest with successful resuscitation Respiratory events: prolonged intubation (>24 h from end of surgery); reintubation; acute respiratory distress syndrome; hypoxemia; pneumonia; acute respiratory failure Vascular events: venous thrombus; pulmonary emboli Renal events: renal insufficiency; acute renal failure Infectious events: wound infection; sepsis Gastrointestinal events: gastrointestinal obstruction and/or paralytic lieus Reoperation Readmission

(*Continued*)

Appendix 3. (*Continued*)

Author	Model(s) Validated	Morbidity Definition
Jones[27]	POSSUM APACHE II	Cardiovascular: myocardial infarct; cardiac failure; hypotension (<90 mmHg for 2 h); respiratory failure Renal: impaired renal function (urea increase of >5 mM from preoperative level) Infection: chest; wound; urinary tract; deep; septicemia; pyrexia of unknown origin; other Wound dehiscence: superficial; deep; anastomotic leak Hemorrhage: wound; deep; other Thromboembolic: deep vein thrombosis; pulmonary embolus; cerebrovascular accident; other Other: any other complication
Kuzu[36]	Nutritional Risk Index Maastricht Index Subjective Global Assessment Mini Nutritional Assessment	Cardiovascular: myocardial infarct; cardiac failure; hypotension Respiratory: atelectasis; bronchopleural fistula; chest infection; empyema; persistent air leak; respiratory failure; pulmonary embolus Gastrointestinal/liver: gastrointestinal hemorrhage; hepatic dysfunction Renal: impaired renal function; urinary extravasation/ureterohydronephrosis; urinary infection Infectious: pyrexia of unknown origin; septicemia and bacteremia; septic shock Neurological: cerebrovascular accident Wound: abscess (intraperitoneal/extraperitoneal); anastomotic leakage; deep hemorrhage; superficial and deep surgical site infection; wound dehiscence; wound hemorrhage Thrombosis: deep venous thrombosis and/or graft thrombosis
Liebman[50]	Identification of Risk In Surgical patients	Cardiovascular: myocardial infarction Respiratory: pneumonia GI: intraabdominal abscess; anastomotic leak Renal: urinary tract infection Neurological: cerebrovascular accident Infectious: sepsis Wound: deep wound infection; rebleeding or significant wound hematoma Thrombosis and/or pulmonary emboli Pressure ulcers Other: miscellaneous; multiple organ failure
Makary[29]	ASA-PS, Lee RCRI, and Eagle Scores alone combined with Frailty Index	NSQIP defined
Pillai[75]	Otago Surgical Audit Score	Complications classed according to severity: 0: no complication; technical complications (some): *e.g.*, anesthetic complications; nonoperative complications: *e.g.*, no lesion found, pyrexia of unknown origin 1: Minor: patient discomfort *e.g.*, postoperative atelectasis; urinary retention 2: Intermediate: significant compromise: *e.g.*, prolonged ileus; deep venous thrombosis 3: Severe: major threat to life: *e.g.*, disseminate intravascular coagulation; myocardial infarction; renal failure
Regenbogen[67]	Surgical Apgar	Major complications: Acute renal failure; bleeding requiring a transfusion of 4 units or more of erythrocytes within 72 h after surgery; cardiac arrest requiring cardiopulmonary resuscitation; coma of 24 h or longer; deep venous thrombosis; myocardial infarction; unplanned intubation, ventilator use for 48 h or more; pneumonia; pulmonary embolism; stroke; wound disruption; deep- or organ-space surgical site infection; sepsis; septic shock; systemic inflammatory response syndrome; vascular graft failure.
Story[75]	Perioperative Mortality Risk Score	Unplanned ICU admission: decision made to admit to ICU, coronary care unit, or high dependency unit made during or after surgery Systemic inflammation: new finding of at least two of the following: Temperature >38.3°C or <36°C; WCC >12,000 c/ml; RR >20 beats/min; HR >90 beats/min; or a positive blood culture alone Acute renal impairment: creatinine increase >20% preoperative value or admission to ICU for RRT

APACHE II = Acute Physiology and Chronic Health Evaluation II; ASA-PS = American Society of Anesthesiologists' Physical Status Score; GI = gastrointestinal; HR = heart rate; ICU = intensive care unit; NSQIP = National Surgical Quality Improvement Program; POSSUM = Physiological and Operative Severity Score for the enUmeration of Morbidity and Mortality; RCRI = Revised Cardiac Risk Index; RR = respiratory rate; RRT = renal replacement therapy; WCC = white cell count.

Appendix 4. Risk Stratification Tools Validated in Single Studies

Author	Model	Outcome	No. of Variables	Age	Sex	Race	Smoking	Surgery Type	Surgery Urgency	ASA-PS
Dasgupta[33]	Detsky	Morbidity to hospital discharge	9	X				X		
Dasgupta[33]	Edmonton Frail Scale	Morbidity to hospital discharge								
Hadjianastassiou[70]	Surgical Mortality Score	Mortality to hospital discharge	6	X	X			X	X	
Haga[28]	E-PASS	Mortality to hospital discharge and 30 d	10	X						X
Haga[28]	mE-PASS	Mortality to hospital discharge and 30 d	7	X				X		X
Kuzu[34]	Nutritional Risk Index	Mortality and morbidity at 30 d or hospital discharge	3							
Kuzu[34]	Mini Nutritional Assessment	Mortality and morbidity at 30 d or hospital discharge	18							
Kuzu[34]	Maastricht Index	Mortality and morbidity at 30 d or hospital discharge	4							
Liebman[49]	IRIS	Mortality and morbidity to hospital discharge	4	X				X	X	
Makary[27]	Eagle Score	Morbidity to hospital discharge	5	X						
Makary[27]	Frailty Index	Morbidity to hospital discharge	5							
Nathanson[72]	MPM_0-III	Mortality to hospital discharge								
Neary[32]	BHOM	30-d and 1-yr mortality	8	X	X			X		
Neary[32]	RCRI	30-d and 1-yr mortality	6					X		
Osler[63]	ICISS	Mortality to hospital discharge								
Pillai[73]	Otago	Morbidity to hospital discharge	12	X	X	X		X	X	
Stachon[40]	SAPS II	Mortality to hospital discharge	15							

Haem	Biochem	IHD or arrhythmia	CCF	COPD	Neuro	Renal	Diabetes	Cancer	Other Preoperative Factors	Intraoperative Factors	Postoperative Factors
		X	X						General poor functional status		
									Onset time of surgery, duration of surgery		
		X*	X*	X			X		Body weight, performance status	Blood loss, duration of surgery, incision type	
		X*	X*	X			X		Performance status		
	Alb								Normal weight, current weight		
									Height, weight, BMI, nutritional history, subjective assessments of general well-being, and comorbidities		
Lymphocytes	Alb, Prealbumin								Ideal weight		
									Hospital admission status (acute *vs.* nonacute)		
		X†	X				DM				
									Shrinking, decreased grip strength, exhaustion, low physical activity, slow walking speed		
Hb WCC	Ur Na K										
		X	X		CVAR	RD	ID				
									Product of survival risk ratios of all ICD-9 classification codes		
									Admission type, number of operations, preoperative length of stay, day case *vs.* inpatient surgery	Duration of surgery, operator grade, wound category	

(*Continued*)

Appendix 4. (*Continued*)

Author	Model	Outcome	No. of Variables	Age	Sex	Race	Smoking	Surgery Type	Surgery Urgency	ASA-PS
Stachon[40]	APACHEN	Mortality to hospital discharge	15							
Stachon[40]	SAPSN	Mortality to hospital discharge	16							
Stachon[74]	DELAWARE	Mortality to hospital discharge	9	X						
Story[75]	Perioperative risk score	30-d mortality	6	X						X

* Cardiac comorbidity classed as single variable. † Eagle criteria: score separately for history of angina *vs*. history of myocardial infarction.

Alb = serum albumin; ALT = alanine transaminase; ASA-PS = American Society of Anesthesiologists' Physical Status score; APACHE II = Acute Physiology and Chronic Health Evaluation II; APACHEN = Acute Physiology and Chronic Health Evaluation-Nucleated; BHOM = Biochemistry and Hematology Outcome Model; BMI = body mass index; CCF = congestive cardiac failure; Chol = cholesterol; CK = creatine kinase; COPD = chronic obstructive pulmonary disease; CRP = C-reactive protein; CVAR = cerebrovascular accident with residual deficit; DELAWARE = Dense Laboratory Whole Blood Applied Risk Estimation; DM = Any definition of diabetes mellitus; Hb = hemoglobin; ICD = International Classification of Diseases; ICISS = International Classification of Disease Illness Severity Score; ICU = intensive care unit; ID = insulin dependent; IHD = ischemic heart disease; IRIS = Identification of Risk In Surgical Patients; K = potassium; (m)E-PASS = (modified) Estimation of Physiologic Ability and Surgical Stress; MPM_0 = Mortality Prediction Model; Na = serum sodium; Plt = platelet count; RCRI = Revised Cardiac Risk Index; RD = Other definition of renal dysfunction; SAPS = Simplified Acute Physiology Score; SAPSN = Simplified Acute Physiology Score-Nucleated; TGC = serum triglycerides; Ur = serum urea; WCC = white cell count.

References

1. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R: European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999; 16:9–13
2. Adams ST, Leveson SH: Clinical prediction rules. BMJ 2012; 344:d8312
3. Grobman WA, Stamilio DM: Methods of clinical prediction. Am J Obstet Gynecol 2006; 194:888–94
4. Saklad M: Grading of patients for surgical procedures. Anesthesiology 1941; 2:281–4
5. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, Sugarbaker DJ, Donaldson MC, Poss R, Ho KK, Ludwig LE, Pedan A, Goldman L: Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. Circulation 1999; 100:1043–9
6. Hennis PJ, Meale PM, Grocott MP: Cardiopulmonary exercise testing for the evaluation of perioperative risk in non-cardiopulmonary surgery. Postgrad Med J 2011; 87:550–7
7. Liao L, Mark DB: Clinical prediction models: Are we building better mousetraps? J Am Coll Cardiol 2003; 42:851–3
8. Noble D, Dent T, Greenhalgh T: Re: Comparisons of established risk prediction models for cardiovascular disease: Systematic review. (Rapid response). BMJ 2012; 345:e4357
9. Mallen C, Peat G, Croft P: Quality assessment of observational studies is not commonplace in systematic reviews. J Clin Epidemiol 2006; 59:765–9
10. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009; 6:e1000097
11. Altman DG: Systematic reviews of evaluations of prognostic variables. BMJ 2001; 323:224–8
12. Altman DG: Systematic reviews of evaluations of prognostic variables, Systematic Reviews in Health Care. Meta-analysis in Context, 2nd edition. Edited by Egger M, Davey Smith G, Altman DG. London, BMJ Books, 2001, pp 228–47
13. Altman DG, Vergouwe Y, Royston P, Moons KG: Prognosis and prognostic research: Validating a prognostic model. BMJ 2009; 338:b605
14. Moons KG, Altman DG, Vergouwe Y, Royston P: Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. BMJ 2009; 338:b606
15. Swets JA: Measuring the accuracy of diagnostic systems. Science 1988; 240:1285–93
16. Arvidsson S, Ouchterlony J, Sjöstedt L, Svärdsudd K: Predicting postoperative adverse events. Clinical efficiency of four general classification systems. The project perioperative risk. Acta Anaesthesiol Scand 1996; 40:783–91

Haem	Biochem	IHD or arrhythmia	CCF	COPD	Neuro	Renal	Diabetes	Cancer	Other Preoperative Factors	Intraoperative Factors	Postoperative Factors
									Nucleated red cell assay added to APACHE II score as an independent variable		
									Nucleated red cell assay added to SAPS II score as an independent variable		
Plts WCC	ALT CK Chol K TGC CRP Alb										
											Acute renal impairment, unplanned ICU admission, inflammation

17. Copeland GP, Jones D, Walters M: POSSUM: A scoring system for surgical audit. Br J Surg 1991; 78:355–60
18. Ding LA, Sun LQ, Chen SX, Qu LL, Xie DF: Modified physiological and operative score for the enumeration of mortality and morbidity risk assessment model in general surgery. World J Gastroenterol 2007; 13:5090–5
19. Carneiro AV, Leitão MP, Lopes MG, De Pádua F: [Risk stratification and prognosis in critical surgical patients using the Acute Physiology, Age and Chronic Health III System (APACHE III)]. Acta Med Port 1997; 10:751–60
20. Zhang H, Zhu D-M, Xue Z-G, Luo J-F, Jiang H: Performance of APACHE II models in surgical intensive care unit. Fudan Univ J Med Sci 2004; 31:417–20
21. Saba V, Goffi L, Jassem W, Ghiselli R, Necozione S, Mattei A, Carle F: Prognostic value of the Apache II scoring system daily preoperative use in major general surgery. Chirurgia 1997; 10:187–94
22. Martin Graczyk AI, Molina Hernandez MJ, Vazquez PC, Mora FJ, Hierro VM, Gomez PJ, Ribera Casado JM: Preoperative geriatric assessment in major surgery in the aged. Anales de Medicina Interna 1995; 12:270–4
23. Kuo HS, Chuang JH, Tang GJ, Hou CC, Chou SS, Lui WY, P'eng FK: Development of a new prognostic system and validation of APACHE II for surgical ICU mortality: A multicenter study in Taiwan. Chung Hua i Hsueh Tsa Chih - Chin Med J 1999; 62:673–81
24. Krenzien J, Roding H, Mummelthey R: Surgical risk in old age: Prospective evaluation of a prognosis index. Zentralblatt fur Chirurgie 1990; 115:717–27
25. Jones DR, Copeland GP, de Cossart L: Comparison of POSSUM with APACHE II for prediction of outcome from a surgical high-dependency unit. Br J Surg 1992; 79:1293–6
26. Davenport DL, Bowe EA, Henderson WG, Khuri SF, Mentzer RM Jr: National Surgical Quality Improvement Program (NSQIP) risk factors can be used to validate American Society of Anesthesiologists Physical Status Classification (ASA PS) levels. Ann Surg 2006; 243:636–41; discussion 641–4
27. Makary MA, Segev DL, Pronovost PJ, Syin D, Bandeen-Roche K, Patel P, Takenaga R, Devgan L, Holzmueller CG, Tian J, Fried LP: Frailty as a predictor of surgical outcomes in older patients. J Am Coll Surg 2010; 210:901–8
28. Haga Y, Ikejiri K, Wada Y, Takahashi T, Ikenaga M, Akiyama N, Koike S, Koseki M, Saitoh T: A multicenter prospective study of surgical audit systems. Ann Surg 2011; 253:194–201
29. Donati A, Ruzzi M, Adrario E, Pelaia P, Coluzzi F, Gabbanelli V, Pietropaoli P: A new and feasible model for predicting operative risk. Br J Anaesth 2004; 93:393–9
30. Brooks MJ, Sutton R, Sarin S: Comparison of Surgical Risk Score, POSSUM and p-POSSUM in higher-risk surgical patients. Br J Surg 2005; 92:1288–92

31. Sutton R, Bann S, Brooks M, Sarin S: The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. Br J Surg 2002; 89:763–8
32. Neary WD, Prytherch D, Foy C, Heather BP, Earnshaw JJ: Comparison of different methods of risk stratification in urgent and emergency surgery. Br J Surg 2007; 94:1300–5
33. Dasgupta M, Rolfson DB, Stolee P, Borrie MJ, Speechley M: Frailty is associated with postoperative complications in older adults with medical problems. Arch Gerontol Geriatr 2009; 48:78–83
34. Kuzu MA, Terzioğlu H, Genç V, Erkek AB, Ozban M, Sonyürek P, Elhan AH, Torun N: Preoperative nutritional risk assessment in predicting postoperative outcome in patients undergoing major surgery. World J Surg 2006; 30:378–90
35. Copeland GP, Sagar P, Brennan J, Roberts G, Ward J, Cornford P, Millar A, Harris C: Risk-adjusted analysis of surgeon performance: A 1-year study. Br J Surg 1995; 82:408–11
36. Whiteley MS, Prytherch DR, Higgins B, Weaver PC, Prout WG: An evaluation of the POSSUM surgical scoring system. Br J Surg 1996; 83:812–5
37. Organ N, Morgan T, Venkatesh B, Purdie D: Evaluation of the P-POSSUM mortality prediction algorithm in Australian surgical intensive care unit patients. ANZ J Surg 2002; 72:735–8
38. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: APACHE II: A severity of disease classification system. Crit Care Med 1985; 13:818–29
39. Charlson ME, Pompei P, Ales KL, MacKenzie CR: A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987; 40:373–83
40. Stachon A, Becker A, Kempf R, Holland-Letz T, Friese J, Krieg M: Re-evaluation of established risk scores by measurement of nucleated red blood cells in blood of surgical intensive care patients. J Trauma 2008; 65:666–73
41. Charlson M, Szatrowski TP, Peterson J, Gold J: Validation of a combined comorbidity index. J Clin Epidemiol 1994; 47:1245–51
42. Atherly A, Fink AS, Campbell DC, Mentzer RM Jr, Henderson W, Khuri S, Culler SD: Evaluating alternative risk-adjustment strategies for surgery. Am J Surg 2004; 188:566–70
43. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA: New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. J Clin Epidemiol 2004; 57:1288–94
44. Haynes SR, Lawler PG: An assessment of the consistency of ASA physical status classification allocation. Anaesthesia 1995; 50:195–9
45. Grocott MP, Levett DZ, Matejowsky C, Emberton M, Mythen MG: ASA scores in the preoperative patient: Feedback to clinicians can improve data quality. J Eval Clin Pract 2007; 13:318–9
46. Aronson WL, McAuliffe MS, Miller K: Variability in the American Society of Anesthesiologists Physical Status Classification Scale. AANA J 2003; 71:265–74
47. Mak PHK, Campbell RCH, Irwin MG: The ASA physical status classification: Inter-observer consistency. Anaesth Intensive Care 2002; 30:633–40
48. Snowden CP, Prentis JM, Anderson HL, Roberts DR, Randles D, Renton M, Manas DM: Submaximal cardiopulmonary exercise testing predicts complications and hospital length of stay in patients undergoing major elective surgery. Ann Surg 2010; 251:535–41
49. Liebman B, Strating RP, van Wieringen W, Mulder W, Oomen JL, Engel AF: Risk modelling of outcome after general and trauma surgery (the IRIS score). Br J Surg 2010; 97:128–33
50. Robinson PJ, Wilson D, Coral A, Murphy A, Verow P: Variation between experienced observers in the interpretation of accident and emergency radiographs. Br J Radiol 1999; 72:323–30
51. Trzeciak S, Erickson T, Bunney EB, Sloan EP: Variation in patient management based on ECG interpretation by emergency medicine and internal medicine residents. Am J Emerg Med 2002; 20:188–95
52. Dindo D, Hahnloser D, Clavien PA: Quality assessment in surgery: Riding a lame horse. Ann Surg 2010; 251:766–71
53. Mohammed MA, Deeks JJ, Girling A, Rudge G, Carmalt M, Stevens AJ, Lilford RJ: Evidence of methodological bias in hospital standardised mortality ratios: Retrospective database study of English hospitals. BMJ 2009; 338:b780
54. Hall BL, Hirbe M, Waterman B, Boslaugh S, Dunagan WC: Comparison of mortality risk adjustment using a clinical data algorithm (American College of Surgeons National Surgical Quality Improvement Program) and an administrative data algorithm (Solucient) at the case level within a single institution. J Am Coll Surg 2007; 205:767–77
55. Copeland GP: The POSSUM system of surgical audit. Arch Surg 2002; 137:15–9
56. Tilford JM, Roberson PK, Lensing S, Fiser DH: Differences in pediatric ICU mortality risk over time. Crit Care Med 1998; 26:1737–43
57. Kramer AA, Zimmerman JE: Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Crit Care Med 2007; 35:2052–6
58. Parikh P, Shiloach M, Cohen ME, Bilimoria KY, Ko CY, Hall BL, Pitt HA: Pancreatectomy risk calculator: An ACS-NSQIP resource. HPB (Oxford) 2010; 12:488–97
59. Gupta PK, Franck C, Miller WJ, Gupta H, Forse RA: Development and validation of a bariatric surgery morbidity risk calculator using the prospective, multicenter NSQIP dataset. J Am Coll Surg 2011; 212:301–9
60. Cohen ME, Bilimoria KY, Ko CY, Hall BL: Development of an American College of Surgeons National Surgery Quality Improvement Program: Morbidity and mortality risk calculator for colorectal surgery. J Am Coll Surg 2009; 208:1009–16
61. Gupta PK, Gupta H, Sundaram A, Kaushik M, Fang X, Miller WJ, Esterbrooks DJ, Hunter CB, Pipinos II, Johanning JM, Lynch TG, Forse RA, Mohiuddin SM, Mooss AN: Development and validation of a risk calculator for prediction of cardiac risk after surgery/clinical perspective. Circulation 2011; 124:381–7
62. Grocott MP: Improving outcomes after surgery. BMJ 2009; 339:b5173
63. Osler TM, Rogers FB, Glance LG, Cohen M, Rutledge R, Shackford SR: Predicting survival, length of stay, and cost in the surgical intensive care unit: APACHE II *versus* ICISS. J Trauma 1998; 45:234–7; discussion 237–8
64. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ: POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. Br J Surg 1998; 85:1217–20
65. Gawande AA, Kwaan MR, Regenbogen SE, Lipsitz SA, Zinner MJ: An Apgar score for surgery. J Am Coll Surg 2007; 204:201–8
66. Regenbogen SE, Ehrenfeld JM, Lipsitz SR, Greenberg CC, Hutter MM, Gawande AA: Utility of the surgical apgar score: Validation in 4119 patients. Arch Surg 2009; 144:30–6; discussion 37
67. Haynes AB, Regenbogen SE, Weiser TG, Lipsitz SR, Dziekan G, Berry WR, Gawande AA: Surgical outcome measurement for a global patient population: Validation of the Surgical Apgar Score in 8 countries. Surgery 2011; 149:519–24
68. Goffi L, Saba V, Ghiselli R, Necozione S, Mattei A, Carle F: Preoperative APACHE II and ASA scores in patients having major general surgical operations: Prognostic value and potential clinical applications. Eur J Surg 1999; 165:730–5
69. Hightower CE, Riedel BJ, Feig BW, Morris GS, Ensor JE Jr, Woodruff VD, Daley-Norman MD, Sun XG: A pilot study

evaluating predictors of postoperative outcomes after major abdominal surgery: Physiological capacity compared with the ASA physical status classification system. Br J Anaesth 2010; 104:465–71

70. Hadjianastassiou VG, Tekkis PP, Poloniecki JD, Gavalas MC, Goldhill DR: Surgical mortality score: Risk management tool for auditing surgical performance. World J Surg 2004; 28:193–200
71. Hobson SA, Sutton CD, Garcea G, Thomas WM: Prospective comparison of POSSUM and P-POSSUM with clinical assessment of mortality following emergency surgery. Acta Anaesthesiol Scand 2007; 51:94–100
72. Nathanson BH, Higgins TL, Kramer AA, Copes WS, Stark M, Teres D: Subgroup mortality probability models: Are they necessary for specialized intensive care units? Crit Care Med 2009; 37:2375–86
73. Pillai SB, van Rij AM, Williams S, Thomson IA, Putterill MJ, Greig S: Complexity- and risk-adjusted model for measuring surgical outcome. Br J Surg 1999; 86:1567–72
74. Stachon A, Becker A, Holland-Letz T, Friese J, Kempf R, Krieg M: Estimation of the mortality risk of surgical intensive care patients based on routine laboratory parameters. Eur Surg Res 2008; 40:263–72
75. Story DA, Fink M, Leslie K, Myles PS, Yap SJ, Beavis V, Kerridge RK, McNicol PL: Perioperative mortality risk score using pre- and postoperative risk factors in older patients. Anaesth Intensive Care 2009; 37:392–8