BJA

British Journal of Anaesthesia, 115 (6): 849–60 (2015)

doi: 10.1093/bja/aev350
Advance Access Publication Date: 3 November 2015
Review Article

REVIEW ARTICLES

Risk assessment tools validated for patients undergoing emergency laparotomy: a systematic review

C. M. Oliver[1,2,3,*], E. Walker[1,2,3], S. Giannaris[3], M. P. W. Grocott[2,4,5,6] and S. R. Moonesinghe[1,2,3]

[1]UCL/UCLH Surgical Outcome Research Centre (SOuRCe), 3rd Floor, Maples Link Corridor, University College Hospital, 235 Euston Road, London NW1 2BU, UK, [2]National Institute of Academic Anaesthesia Health Services Research Centre, Royal College of Anaesthetists, London, UK, [3]Centre for Anaesthesia, University College London, London, UK, [4]University Hospital Southampton NHS Foundation Trust, Southampton, UK, [5]Integrative Physiology and Critical Illness Group, Clinical and Experimental Sciences Faculty of Medicine, University of Southampton, Southampton, UK, and [6]University Hospital Southampton NHS Foundation Trust/University of Southampton, NIHR Respiratory Biomedical Research Unit, Southampton, UK

*Corresponding author: E-mail: cmoliver@hotmail.co.uk

Abstract

Emergency laparotomies are performed commonly throughout the world, but one in six patients die within a month of surgery. Current international initiatives to reduce the considerable associated morbidity and mortality are founded upon delivering individualised perioperative care. However, while the identification of high-risk patients requires the routine assessment of individual risk, no method of doing so has been demonstrated to be practical and reliable across the commonly encountered spectrum of presentations, co-morbidities and operative procedures. A systematic review of Embase and Medline identified 20 validation studies assessing 25 risk assessment tools in patients undergoing emergency laparotomy. The most frequently studied general tools were APACHE II, ASA-PS and P-POSSUM. Comparative, quantitative analysis of tool performance was not feasible due to the heterogeneity of study design, poor reporting and infrequent within-study statistical comparison of tool performance. Reporting of calibration was notably absent in many prognostic tool validation studies. APACHE II demonstrated the most consistent discrimination of individual outcome across a variety of patient groups undergoing emergency laparotomy when used either preoperatively or postoperatively (area under the curve 0.76–0.98). While APACHE systems were designed for use in critical care, the ability of APACHE II to generate individual risk estimates from objective, exclusively preoperative data items may lead to better-informed shared decisions, triage and perioperative management of patients undergoing emergency laparotomy. Future endeavours should include the recalibration of APACHE II and P-POSSUM in contemporary cohorts, modifications to enable prediction of morbidity and assessment of the impact of adoption of these tools on clinical practice and patient outcomes.

Key words: emergency laparotomy; postoperative mortality; prognostic tool; risk adjustment; risk assessment

Editor's key points

- In this systematic review, the authors considered the effectiveness of current risk-assessment tools to predict outcome following emergency laparotomy.
- Poor study standardisation and homogeneity prevented comparison of the various tools available, but APACHE II appeared to demonstrate the most consistent discrimination of individual outcome.

Introduction

Emergency laparotomy is a commonly utilized group of intra-abdominal surgical procedures performed for a variety of acute pathologies. In excess of 30 000 emergency laparotomies are performed annually in England alone, and Emergency General Surgical (EGS) admissions are considerably more numerous.[1 2]

Internationally reported mortality rates following emergency laparotomy range from 13 to 18% at 30 days, increasing to 25% at 24 months.[3–7] This is second only to short-term mortality after emergency open repair of life-threatening ruptured abdominal aortic aneurysm (AAA).[8]

Reduction of the considerable morbidity and mortality after emergency laparotomy is the focus of several ongoing national and international audit and quality improvement programs, including the National Emergency Laparotomy Audit (NELA), the Australian and New Zealand Audit of Surgical Mortality, the American College of Surgeons National Surgical Quality Improvement Program (NSQIP), the Enhanced Peri-Operative Care for High-risk patients (EPOCH) study and the Dr Foster global comparators study.[9–12] Central to each of these programs is the identification of high-risk patients to target perioperative interventions and augmented pathways of care.

Because patients who undergo emergency laparotomy are markedly heterogeneous, the likelihood of suffering postoperative morbidity or mortality is not evenly distributed within patient populations. The delivery of individualised care and reduction of postoperative adverse events require that both the structure and delivery of perioperative care are tailored to the needs of the individual. To this end, substantial efforts have been made to characterise high-risk patient subgroups and to identify patients at the greatest risk of death and morbidity.[8 13 14]

Assessment of an individual's risk of an adverse event may be informed by clinical judgement, use of risk assessment tools, evaluation of functional capacity or plasma biomarker assay.[15] Clinical judgement may vary with experience, observations of exercise tolerance are often unfeasible in patients requiring emergency laparotomy since they are acutely unwell and evidence to support the routine use of biomarkers has yet to be established.[16–18] Risk assessment tools, which incorporate clinical variables into a score or prognostic model, currently represent the most practical means of estimating risk in patients undergoing unplanned surgery, but no tool has been widely incorporated into routine practice.

Due to prevalent co-morbidities, surgical pathologies and their systemic effects and the urgency of required intervention, patients undergoing emergency laparotomy form a population distinct from those undergoing planned general surgery,[19] evidenced by a higher incidence of adverse postoperative events.[19–21] Therefore, while there is evidence to support the routine use of selected risk assessment tools in other clinical contexts, generalisability of the performance of these tools to patients undergoing emergency laparotomy is unknown.[22–25]

The objectives of this systematic review were to identify all perioperative validation studies of risk assessment tools undertaken in adult patients undergoing emergency laparotomy and to compare the reported performance and utility of the assessed tools with the aim of identifying the best tools for routine clinical use.

Prior presentation of data

Presented at the third joint meeting of the Centre for Anaesthesia, UCL's Current Controversies in Anaesthesia and Perioperative Medicine and the Intensive Care Society of Ireland Autumn Meeting in Dingle, Ireland, September 2013.

Methods

This systematic review was registered with the PROSPERO database (CRD42014009062). Methods and reporting conform to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), BMC and Cochrane guidelines.[26–28]

Definitions for the purposes of this review

Emergency

Urgent, emergent and immediately indicated surgical interventions.

Laparotomy

Open intra-abdominal surgery performed for non-aortic pathologies.

Risk assessment tool

A scoring system or prognostic model incorporating two or more variables to stratify or predict the likelihood of a specified adverse event.

Validation study

Assessment of the accuracy of one or more risk assessment tools through application to a study population. Classified as internal (application of a newly created tool to the cohort from which it was derived by practical or mathematical techniques), temporal [application of a tool to a cohort distinct in time from the derivation cohort at the institution(s) in which it was created] or external (application to patients in institutions other than that from which the tool was derived).[29 30]

Discrimination

How well a tool is able to discriminate between dichotomous outcomes (e.g. death and survival at 30 days) across a spectrum of risk profiles within a population of patients. Presentation as area under the receiver operator characteristic curve (AUC) provides a single, quantitative measure of the accuracy of a prognostic tool and also facilitates the comparison of dissimilar systems.[31] In interpreting AUC values: >0.9, good discrimination; 0.7–0.9, moderate; and <0.7, poor.[31]

Calibration

How closely a prognostic model's estimations match the observed incidence of a specified outcome across a study population. Assessed using χ^2 techniques, $P>0.05$ indicates that observed and expected outcomes are similar and $P<0.05$ differences are statistically significant.

Search methods and inclusion criteria

The literature search was undertaken with reference to methods of bias mitigation.[32] Embase (a registered trademark of Elsevier B. V.) and Medline (U.S. National Library of Medicine) were searched using database-specific search terms (a complete list of search terms used in Medline is included in the Supplementary materials (S5)). Because the term emergency laparotomy is used internationally to describe a wide and varied assortment of surgical procedures and pathologies, an inclusive search strategy was adopted to achieve the aims of this systematic review.

The search was restricted to publications relating to adult humans since 1980, but was not limited to English-language publications. The last complete search was performed on March 27, 2013. The Cochrane database of systematic reviews was accessed on November 2, 2014. Secondary searching included hand-searching of references (snow-balling) and review of citation listings in Web of Knowledge (a trademark of Thomson Reuters).

Inclusion criteria

Studies assessing the discrimination of a specified outcome, presented as the AUC, by one or more risk assessment tools in adult patients undergoing emergency laparotomy were included. Studies including both emergent and elective cases were included if discrimination was reported for patients who had undergone emergency surgery.

Exclusion criteria

In order to identify useful perioperative decision-making tools, studies were excluded if no assessment of risk was made using preoperative or intraoperative data items. Validation studies confined to cohorts undergoing emergency aortic surgery and those including extra-abdominal procedures were excluded due to overt differences in patient characteristics, operative procedures and patient outcomes.

Data extraction

Data extraction was performed by the authors CO and EW and recorded directly into purpose-built tables summarising study characteristics, design quality, patient outcomes and tool performance. Differences in extracted data were discussed and consensus reached.

Extracted study characteristics included geographical region, patient cohort size and characteristics, nature of the included surgical procedures, timing of data collection (relative to emergency laparotomy) and risk assessment tools studied. Tool applicability was then classified as either general (heterogeneous and multiple subpopulation cohorts) or subpopulation specific (applicable only to cohorts defined by patient or surgical characteristics).

Extracted indicators of quality included the number of patients in the study cohort, number of institutions collecting data (single *vs* multicentre), timing of data collection (prospective *vs* retrospective), reporting of cohort baseline characteristics, reporting of inclusion criteria and excluded patients and validation methodology.[29 33]

Extracted outcomes were as reported in the manuscript and included the incidence of mortality and morbidity at specified time points for identified pathologies and surgical indications.

Extracted tool performance characteristics included the AUC for a specified outcome and prognostic tool calibration and AUC 95% confidence interval (CI) where reported.

Data analysis

Decisions to pool data for meta-analysis of the performance of individual tools were informed by assessment of the homogeneity of extracted study characteristics, where overt heterogeneity of inclusion criteria, study design and patient characteristics would preclude statistical assessment of homogeneity.

Tool generalisability was determined by assessment of discrimination across dissimilar populations, including heterogeneous patient cohorts and subpopulations defined by patient or surgical characteristics.

Results

Overview

In total, 23 073 papers were identified in the primary electronic databases search, leaving 15 030 after restrictions. A further 802 papers were identified in the secondary search (Supplementary material, S1). After exclusions, 20 studies were eligible for data analysis and synthesis, assessing 25 risk assessment tools in >110 000 patients undergoing unplanned intra-abdominal surgery across 12 countries (Table 1). No similar systematic review was identified in the Cochrane Database of Systematic Reviews.

Study design and populations are summarised in Table 1. External validity was assessed in 13 studies and temporal validity in 2 studies. Internal validation techniques included split cohorts, crossover and bootstrapping.

Markers of methodological quality are presented in Table 1. Seven studies were conducted across multiple institutions, validation cohort size varied from 49 to >68 000 patients,[34 35] data collection was performed entirely prospectively in 15 studies and demographic data were presented in 19 papers. Reporting of inclusion criteria, exclusions and surgical procedures was universally adequate. Reporting of calibration and AUC 95% CIs was inconsistent. Statistical comparison of tool discrimination was reported in only one study.[36]

Short-term (30-day or inpatient) mortality endpoints were reported in all studies; 30-day mortality was reported in 11 studies and inpatient mortality in 9 studies. The 30-day mortality ranged from 9 to 27% and inpatient mortality from 3 to 26%, varying by operative procedure, surgical indication and patient age (Table 2). Other identified endpoints included postoperative morbidity and complications.[37 38] These were infrequently reported (Table 1).

Overt heterogeneity of study design, patient characteristics and presented outcomes precluded meta-analysis, necessitating a qualitative approach to tool comparison (Table 1).

Risk assessment tools

Thirteen general tools and 12 subpopulation-specific tools were assessed in the identified studies (Table 1). In addition to the Physiological and Operative Severity Score for the enumeration of Mortality and morbidity (POSSUM), several previously derived POSSUM systems and novel coefficients were assessed.[39 40]

Many tools were not created for the purpose of perioperative risk assessment, including tools for critical care [Acute Physiology and Chronic Health Evaluation II (APACHE II), APACHE III, the Simplified Acute Physiology Score II (SAPS II) and the Mortality Probability Model II (MPM II)],[41 42] the Early Warning Score and 'sepsis score'[43] and the POSSUM systems, which were created for comparative audit.

Two tools specific to emergency laparotomy were identified: the preoperative and perioperative NSQIP Emergency Laparotomy models.[4] These were assessed in a single, large internal validation study.

Tool characteristics are summarised in Supplementary material S2. There was notable variation in the required number of data items [ranging from the one composite measure of the

Table 1 Study characteristics and quality. EL, emergency laparotomy

First Author	Year	Region	Cohort size	Single/multi centre	Data acquisition	Baseline data reported	Inclusion criteria	Exclusions	Models assessed	Outcome	Validation methodology
Al-Temimi	2012	USA	37 553	Multi	Prospective	Yes	>16 years, EL for general surgical indications or mesenteric insufficiency	Missing data, urgent or vascular surgery, laparoscopic procedure converted to open	1) EL perioperative model, 2) EL preoperative model	30-day mortality	Internal: crossover by year
Biondo	2000	Spain	55	Single	Prospective	Yes	Consecutive patients, emergency surgery for distal colonic peritonitis	'Complicated non-specific inflammatory disease of the colon'	1) Colonic peritonitis severity score	Inpatient mortality	Internal: split
Biondo	2006	Spain	156	Single	Prospective	Yes	Clinical diagnosis of peritonitis, emergency surgery	None declared	1) Colonic Peritonitis Severity Score, 2) Mannheim Peritonitis Index	Inpatient mortality	Temporal
Buck	2012	Denmark	117	Multi	Prospective	Yes	Surgical treatment for perforated peptic ulcer	Pregnant or breastfeeding, malignant ulcers, perforation of another organ	1) Boey score, 2) ASA-PS, 3) APACHE II, 4) sepsis score	30-day mortality, septic shock	External
Ertan	2008	Turkey	102	Single	Retrospective	Yes	Emergency colorectal surgery for complications of colorectal carcinoma	Uncertain diagnosis, insufficient data	1) APACHE III, 2) MPM II, 3) CR-POSSUM	30-day mortality	External
Ferjani	2007	UK	158	Single	Prospective	Yes	Consecutive patients, histologically confirmed colorectal cancer, abdominal surgery to remove primary tumour	Laparoscopic surgery	1) ACPGBI, 2) POSSUM, 3) P-POSSUM, 4) CR-POSSUM	30-day mortality	External
Garcea	2010	UK	280	Single	Retrospective	Yes	EL for suspected perforation of a viscus/obstruction/fulminant colitis/upper gastrointestinal bleeding/surgery for strangulated ventral or groin hernia	None declared	1) Early Warning Score, 2) ASA-PS, 3) POSSUM, 4) APACHE II	ICU and total LOS, inpatient mortality,	External
Goffi	1999	Italy	49	Single	Prospective	Yes	Major emergency operations, including trauma	None declared	1) APACHE II	30-day mortality	External

Kermani	2013	USA	68 344	Multi	Retrospective	Yes	ICD-9 coding: >18 years, admitted non-electively, colectomy	Missing data	1) Practical mortality risk score for emergent colectomy	Inpatient mortality	Internal: crossover (10-fold k-partitions)
Koc	2007	Turkey	75	Single	Prospective	Yes	Emergency surgery for perforated peptic ulcer	Surgery for perforated ulcer at the site of a previous anastomosis, malignancy	1) APACHE II, 2) APACHE III, 3) SAPS II, 4) MPM II	30-day mortality	External
Kologlu	2001	Turkey	473	Single	Retrospective	No	Operation for intra-abdominal infection without continuous postoperative peritoneal lavage	Missing data, percutaneous drainage of intra-abdominal abscess, uncomplicated appendicitis, uncomplicated cholecystitis, planned repeat laparotomy	1) Mannheim Peritonitis Index, 2) Peritonitis Index of Altona, 3) Combined peritonitis score	Inpatient mortality	External
Kulkarni	2007	India	50	Single	Prospective	Yes	Peritonitis due to perforation of hollow viscus	Blunt abdominal trauma with associated solid organ/vascular/ neurological injury or fracture	1) APACHE II	Inpatient mortality	External
Kwok	2011	USA	372	Multi	Prospective	Yes	>80 years, CPT code: emergency colectomy	Laparotomy resulting in 'diversion only', i.e. without colonic resection, missing data	1) Targeted risk prediction score, 2) ASA-PS, 3) Surgical Risk Scale, 4) ACS colorectal surgery risk calculator	30-day mortality	Temporal
Lohsiriwat	2008	Thailand	152	Single	Prospective	Yes	Emergency surgery for perforated peptic ulcer	Perforated gastric cancer	1) Boey score, 2) ASA-PS, 3) Mannheim Peritonitis Index	30-day mortality, complications	External
Merad	2012	France	575	Multi	Prospective	Yes	> 16 years, major digestive surgery	>1 missing P-POSSUM value	1) P-POSSUM	Inpatient mortality	External
Moller	2012	Denmark	2668	Multi	Prospective	Yes	Surgical treatment of benign gastric or duodenal perforated peptic ulcer	Malignant peptic ulcer	1) Peptic ulcer prediction score, 2) ASA-PS, 3) Boey score	30-day mortality	Internal: bootstrapping for PULP

Continued

Table 1 *Continued*

First Author	Year	Region	Cohort size	Single/multi centre	Data acquisition	Baseline data reported	Inclusion criteria	Exclusions	Models assessed	Outcome	Validation methodology
Notash	2005	Iran	80	Single	Prospective	Yes	Consecutive patients, EL for secondary peritonitis	Primary peritonitis (no identified cause)	1) Mannheim Peritonitis Index	Inpatient mortality	External
Ohmann	1993	Europe	271	Multi	Prospective	Yes	Diffuse or localised intra-abdominal infection, confirmed at laparotomy	None declared	1) APACHE II, 2) Mannheim Peritonitis Index, 3) Peritonitis Index of Altona	30-day mortality	External
Poon	2005	Hong Kong	160	Single	Prospective	Yes	Consecutive patients, urgent surgery for malignant colorectal obstruction	None declared	1) P-POSSUM	30-day mortality	External
Ren	2009	China	90	Single	Retrospective	Yes	Colorectal cancer	None declared	1) P-POSSUM, 2) modified P-POSSUM, 3) modified CR-POSSUM	Inpatient mortality	Internal: split cohorts

Table 2 Summary mortality rates after emergency laparotomy

Surgical indication or procedure		Postoperative mortality (%)	
	Subgroup	30 days	Inpatient
Emergency laparotomy		8–20	9–21
Peritonitis		21	16–26
Colorectal cancer		11–20	3
	>80 years	26	
Perforated peptic ulcer		9–27	

ASA physical status classification (ASA-PS) to the 41 variables comprising the NSQIP perioperative Emergency Laparotomy model],[4] the preoperative availability of data items, complexity of calculation and the requirement for subjective interpretation of clinical data items.

General tools

Of the 13 identified general tools (Table 3), the most frequently assessed were APACHE II (7 studies), ASA-PS (5 studies) and P-POSSUM (4 studies). Each tool was assessed in both heterogeneous patient cohorts and subpopulations defined by demographic characteristics, surgical indication or operative procedure.

The ability of APACHE II to discriminate between short-term death and survival was moderate to good (AUC 0.76–0.98) when applied to heterogeneous cohorts undergoing unplanned intra-abdominal surgery for a variety of indications including peritonitis, colorectal malignancy and perforated peptic ulcer.[35 37 44–48] Notably, APACHE II was scored using exclusively preoperative data in four studies: at hospital admission in three studies and on booking for a surgical theatre in one study.[35 37 45 46]

Discrimination between short-term outcomes by the ASA-PS was moderate or good in patient cohorts undergoing emergency laparotomy or repair of perforated peptic ulcer (AUC 0.81 and AUC 0.73–0.91, respectively), but poor in an elderly cohort (AUC 0.66).[37 38 47 49 50]

Reported AUC values for the discrimination of short-term outcomes by P-POSSUM in cohorts undergoing unplanned intra-abdominal surgery ranged from 0.65 to 0.82. Discrimination was moderate or poor in patients with colorectal cancer (AUC 0.65, 0.66, 0.75).[39 51 52]

Discrimination of 30-day survival by NSQIP emergency laparotomy models was moderately good (AUC 0.87–0.88) in a single internal validation study.[4]

Reporting of prognostic tool calibration was poor and, where reported, performance was inconsistent (Table 3).

Subpopulation-specific tools

A variety of indication-specific and co-morbidity-specific tools were identified (Supplementary materials (S3)).

Colorectal surgery. Of six tools specific to colorectal surgery, CR-POSSUM alone was assessed in multiple studies. Discrimination of 30-day outcome after unplanned surgery for colorectal cancer by CR-POSSUM was moderate or poor (AUC 0.65, 0.72).[48 52]

Peritonitis. Of four identified studies of tools specific to peritonitis, the Mannheim Peritonitis Index (MPI) and Peritonitis Index of Altona II (PIA II) underwent multiple assessments of external validity. Discrimination of short-term outcome by the MPI and PIA II

Table 3 Discrimination and calibration of general tools. GOF, goodness of fit; PPU, perforated peptic ulcer; H-L, Hosmer-Lemeshow statistic

Tool	Total patients	Region	First author (year)	Cohort size	Surgery subtype	Inclusion criteria	Primary endpoint	Incidence of primary endpoint	AUC (95% CI)	Calibration P-value	H-L value (unless stated)
APACHE II	944	Turkey	Ertan (2008)	102	General Surgery	Colorectal cancer	30-day mortality	17%	0.78	0.49	4.448
		Italy	Goffi (1999)	49	General Surgery	Mixed	30-day mortality	20%	0.87	0.63	χ^2 quintiles
		Europe	Ohmann (1993)	271	General Surgery	Peritonitis	30-day mortality	21%	0.87	>0.05	
		Turkey	Koc (2007)	75	PPU Repair	PPU	30-day mortality	11%	0.87	0.007	17.58
		Denmark	Buck (2012)	117	PPU Repair	PPU	30-day mortality	17%	0.76		
		UK	Garcea (2010)	280	General Surgery	Mixed	inpatient mortality	15%	0.76		
		India	Kulkarni (2007)	50	General Surgery	Perforative peritonitis	Inpatient mortality	16%	0.98		
APACHE III	177	Turkey	Ertan (2008)	102	General Surgery	Colorectal cancer	30-day mortality	17%	0.77	0.9	2.208
		Turkey	Koc (2007)	75	PPU Repair	PPU	30-day mortality	11%	0.84	0.01	15.08
ASA-APS	3589	USA	Kwok (2011)	372	General Surgery	>80 years: colectomy	30-day mortality	26%	0.66	0.14	Residual GOF
		Denmark	Buck (2012)	117	PPU Repair	PPU	30-day mortality	17%	0.73		
		Denmark	Moller (2012)	2668	PPU Repair	PPU	30-day mortality	27%	0.78 (0.76–0.80)		
		UK	Garcea (2010)	280	General Surgery	Mixed	Inpatient mortality	15%	0.81		
		Thailand	Lohsiriwat (2008)	152	PPU Repair	PPU	30-day mortality	9%	0.91 (0.85–0.95)		
Early Warning Score	280	UK	Garcea (2010)	280	General Surgery	Mixed	Inpatient mortality	15%	0.71		
Emergency Laparotomy perioperative model	37 553	USA	Al-Temimi (2012)	37 553	General Surgery	Mixed	30-day mortality	14%	0.88	<0.001	
Emergency Laparotomy preoperative model	37 553	USA	Al-Temimi (2012)	37 553	General Surgery	Mixed	30-day mortality	14%	0.87	<0.001	
Mortality Probability Model (MPM) II	177	Turkey	Ertan (2008)	102	General Surgery	Colorectal cancer	30-day mortality	17%	0.71	0.46	7.736
		Turkey	Koc (2007)	75	PPU Repair	PPU	30-day mortality	11%	0.98	0.99	1.36
POSSUM	438	UK	Ferjani (2008)	158	General Surgery	Colorectal cancer	30-day mortality	20%	0.63 (0.55–0.70)	0.037	
		UK	Garcea (2010)	280	General Surgery	Mixed	Inpatient mortality	15%	0.81		

Continued

Table 3 *Continued*

Tool	Total patients	Region	First author (year)	Cohort size	Surgery subtype	Inclusion criteria	Primary endpoint	Incidence of primary endpoint	AUC (95% CI)	Calibration P-value	H-L value (unless stated)
P-POSSUM	983	China	Ren (2009)	90	General Surgery	Colorectal cancer	Inpatient mortality	3%	0.66	0.25	2.81
		Hong Kong	Poon (2005)	160	General Surgery	Colorectal cancer - obstructing	30-day mortality	11%	0.75	0.11	5.98
		France	Merad (2012)	575	General Surgery	Mixed	Inpatient mortality	21%	0.82 (0.78–0.86)	<0.001	68.7
		UK	Ferjani (2008)	158	General Surgery	Colorectal cancer	30-day mortality	20%	0.65 (0·57–0·73)	<0.001	
mP-POSSUM (local modification)	90	China	Ren (2009)	90	General Surgery	Colorectal cancer	Inpatient mortality	3%	0.66	0.8	4.99
SAPS II	177	Turkey	Ertan (2008)	102	General Surgery	Colorectal cancer	30-day mortality	17%	0.83	0.98	1.079
		Turkey	Koc (2007)	75	PPU Repair	PPU	30-day mortality	11%	0.86	0.085	8.2
Sepsis score	117	Denmark	Buck (2012)	117	PPU Repair	PPU	30-day mortality	17%	0.69		
Surgical Risk Scale	2721	USA	Kwok (2011)	372	General Surgery	>80 years: colectomy	30-day mortality	26%	0.66	0.14	Residual GOF

was inconsistent for patients undergoing general surgery for peritonitis (AUC 0.73, 0.97 and 0.69, 0.95, respectively).[36 44 53 54]

Perforated peptic ulcer (PPU). Of two tools specific to PPU, the Boey score alone was assessed in multiple studies. Discrimination of outcome at 30 days by the Boey score was moderate or poor (AUC 0.63–0.86), but in discrimination of development of post-operative septic shock and complications, performance was moderate (AUC 0.72 and 0.80, respectively) (Supplementary material S4).[37 38 49]

Discussion

Risk assessment tools can support clinical judgement in determining appropriate treatment plans, informing consent and tailoring perioperative care and may also be used in risk adjustment, thus supporting quality improvement.

The objectives of this systematic review were to identify all validation studies of risk assessment tools in cohorts of adults undergoing emergency laparotomy and to compare the performance and utility of these tools in order to determine which are best suited to perioperative clinical practice.

Despite the publication of guidance, study design and quality was variable, with some studies not meeting basic criteria.[55] Calibration was infrequently reported.

Short-term mortality after emergency laparotomy varied considerably and was greatest in patient cohorts defined by increased age (Table 2).

The most frequently assessed tools were not created for the perioperative prediction of individual risk. APACHE II, which was created for assessment of the severity of critical illness, demonstrated the most consistent accuracy across a variety of patient cohorts undergoing emergency laparotomies when used either preoperatively or postoperatively. In contrast with many other tools, APACHE II also has the advantage that it may be used to generate individual percentage risk estimates using objective clinical data items routinely available in this setting (Supplementary material S2).

Performance

Because patients who require emergency laparotomy are markedly heterogeneous, the capacity to tolerate a cascade of acute surgical pathology, massive surgical insult and resulting post-operative organ dysfunction varies between individuals. Furthermore, underlying surgical pathologies may not be apparent prior to surgery. For a tool to be useful for emergency laparotomy it must therefore be both applicable to and accurate across the spectrum of patient characteristics, surgical pathologies and operative procedures encountered in clinical practice. One general tool is therefore preferable to multiple subpopulation-specific tools.

When determining the best tool for emergency laparotomy, demonstration of satisfactory and consistent performance is essential and must precede these other considerations.[29–31]

Comparisons of tool performance

Assessment of external validity provides the best measure of a tool's generalisability and the performance of tools that are validated only internally may not be replicable in external patient populations due to factors including overfitting.[30 55] Of the identified general risk assessment tools, only seven were assessed in multiple patient cohorts (Table 3). The two identified tools that were developed specifically for emergency laparotomy were assessed in a single large database cohort and validated only internally.

For those tools that were assessed in multiple cohorts, pooling of data for head-to-head comparisons of discriminatory performance was not feasible due to significant heterogeneity of both study design and patient cohorts and the infrequent reporting of AUC CIs and within-study statistical comparisons of performance.

Demonstration of prognostic tool calibration is essential both for clinicians to have confidence in estimates of risk and for the process of risk adjustment. It is therefore notable that calibration was infrequently reported in the identified validation studies of prognostic tools.

Disparities were evident in the discriminatory performance (APACHE II, 0.76–0.98; ASA-PS, 0.66–0.91; P-POSSUM, 0.65–0.82) and calibration of the most frequently assessed tools. This variation is most likely to reflect differences in study cohorts and methodologies, poor generalisability of performance to some patient subgroups and poor reliability of subjectively scored tools. However, no identified tool incorporates measures of organisational structure and processes of care or of geographical variations in patient-level risk, which may account for some of the variation observed.[56]

The reduced accuracy of the ASA-PS and P-POSSUM in patients with colorectal malignancy and older people suggests that the performance of these tools is not generalisable to these subgroups of patients.

The reliability of the output of risk assessment tools may be reduced by subjective interpretation of data items and inconsistent application, which may account for some of the variation observed in the performance of the ASA-PS and P-POSSUM. Both tools require the interpretation and scoring of clinical data items (including chest X-ray and ECG),[40] the POSSUM system 'multiple procedures' item is variably interpreted by clinicians and linear and exponential analyses are variably used in the calculation of percentage predicted risk with POSSUM systems.[57]

APACHE II was created for the assessment of risk in critical care admissions,[58] incorporating physiological parameters, markers of chronic disease and age. The comparatively good performance of APACHE II in these patients undergoing emergency laparotomy may therefore reflect the associations of age, magnitude of systemic insult and relevant co-morbidities with adverse postoperative outcomes in such cohorts of patients.

Tool performance in core subgroups

Complications of intra-abdominal malignancy are a common indication for emergency laparotomy,[59] the incidence of colorectal cancer in older people is increasing[60] and outcomes after emergency laparotomy vary with age and the timing of presentation with malignancy.[61] Because elderly patients and individuals with colorectal malignancy thus represent core subgroups of patients undergoing emergency laparotomy, it is essential that the performance of tools for emergency laparotomy is generalisable to these patients.

While the significance of the observation is uncertain on the basis of the data analysed, it is therefore notable that discrimination of outcome by APACHE II in a cohort of patients with colorectal cancer was not evidently reduced (AUC 0.78), whereas P-POSSUM and CR-POSSUM discriminated less well in patients with colorectal cancer; similarly the accuracy of the ASA-PS was reduced in a cohort of elderly patients.[39 50–52]

Adding weight to reports that existing scoring systems are inaccurate and unreliable in elderly cohorts undergoing emergency surgery, it is notable that in a cohort of elderly patients undergoing emergency colectomy, neither age-specific nor general tools demonstrated adequate discrimination.[50 62 63]

Tool characteristics

Even if consistent and generalisable performance can be demonstrated, a tool is unlikely to be adopted into routine clinical practice if it is unwieldy. This may be true of tools requiring numerous data points, exacting a high burden of data collection or due to the complexity of required calculations.

Utility

The ASA-PS requires only one composite variable, APACHE II requires 12 data points that are routinely available for many patients undergoing emergency laparotomy and P-POSSUM requires 18 data points, whereas the NSQIP emergency laparotomy models require a minimum of 39 variables (Supplementary material S2). Tools that require numerous data items might be expected to better capture all relevant risk factors, but performance data did not suggest better discrimination by complex tools (Table 3).

Percentage estimates of individual risk may be used to inform a variety of shared and clinical decisions, including consent for emergency laparotomy, clinician seniority and the location of postoperative care.[9–12] The ASA-PS lacks this capability, but such estimates may be generated with prognostic tools including P-POSSUM and APACHE II. And while logarithmic equations are required, the widespread availability of online and 'app'-based calculators means that bedside estimation of individual risk is feasible and could be incorporated into routine clinical practice.

Timing of risk assessment

The accuracy of tools that require intra-operative and postoperative data items may not be superior to tools that require only preoperative data.[64] Indeed, accuracy increased only negligibly with the inclusion of intra-operative data (AUC 0.87 and 0.88, respectively) in the NSQIP emergency laparotomy tools derivation study.[4]

In clinical practice, the preoperative availability of individualised estimates of risk will better inform shared decisions than estimates that are available only postoperatively. It is therefore notable that while the accuracy of APACHE II estimates calculated using only preoperative data was comparable with studies using perioperatively or postoperatively collected data (AUC 0.76–0.98 and 0.78–0.87 respectively), P-POSSUM-predicted risk was not calculated preoperatively in any of the studies identified.[35 37 44–48] Preoperative performance of P-POSSUM may therefore have been overestimated by the findings of this review.

Limitations

Few of the identified tools were assessed in multiple validation studies and many cohorts were small and drawn from single centres, limiting both statistical power and assessment of the generalisability of tool performance.

Only studies reporting tool discrimination as AUC were included in this systematic review. While the benefits of doing so are widely accepted, a small proportion of excluded studies used alternative methods of assessing and reporting tool performance in patient cohorts undergoing emergency laparotomy. Quantitative comparative analysis of tool performance (discrimination and calibration) was not feasible due to the significant heterogeneity of study design and cohort characteristics, poor reporting and infrequent within-study statistical comparisons of tool performance.[33]

Organisational factors in the delivery of health care have undergone considerable change over recent years. Because this review includes studies from the 1990s, the reported accuracy and calibration of identified tools may not describe performance in contemporary patient cohorts.[65]

Bias identification and mitigation was limited by variable reporting of study methodology, potential for preferential publication of studies with positive findings and the predominance of English-language papers in the accessed electronic databases.[66]

Conclusions and direction of future research

The identification of high-risk patients for targeted perioperative interventions and augmented pathways of care is central to current initiatives to improve patient outcomes after emergency laparotomy. Parallel initiatives include ensuring that the care of all patients meets agreed standards, that novel perioperative interventions continue to be developed[67] and that, where possible, individuals at risk of emergency presentation are identified, counselled and managed before the acute episode.

Comparison of the performance of risk assessment tools for emergency laparotomy was not possible with existing data, highlighting the need for consensus in the reporting of perioperative outcomes.[68] However, due to consistent performance across patient subgroups and its capacity to generate individual preoperative risk estimates using routinely available objective data items, APACHE II is promoted as a practical means of estimating individual risk in patients undergoing emergency laparotomy.

Unfocussed efforts to derive novel tools are likely to achieve only marginal improvements in performance, increase uncertainty and struggle to gain the limited acceptance currently achieved by the best known systems. Future efforts should therefore seek to update the performance (primarily the calibration) of APACHE II and P-POSSUM in large, well-conducted contemporary studies in heterogeneous emergency general surgical cohorts, assess the implications of dynamic risk scores (e.g. preoperative and postoperative scores), develop the ability to accurately predict postoperative morbidity due to associations with prolonged excess mortality and assess the impact of adoption of these tools on patient outcomes and clinical practice.[29 30 69]

Finally, it should be recognised that in addition to the numerous potential benefits to patients of individual risk prediction, the implementation of nationally coordinated methods for the assessment of individual risk, as is true of the National Emergency Laparotomy Audit (NELA), the Intensive Care National Audit and Research Centre (ICNARC) and the American College of Surgeons (ACS) NSQIP augments and promotes the collection of risk data, health services research and quality improvement initiatives.

Author's contributions

Study design, search strategy development, first draft and revisions: C.M.O., S.R.M.
Data extraction and analysis: C.M.O., E.W., S.G.
Critical revision of the manuscript: C.M.O., E.W., S.G., M.G., S.R.M.

Supplementary material

Supplementary material is available at *British Journal of Anaesthesia* online.

References

1. Shapter SL, Paul MJ, White SM. Incidence and estimated annual cost of emergency laparotomy in England: is there a major funding shortfall? *Anaesthesia* 2012; **67**: 474–8
2. Symons NRA, Moorthy K, Almoudaris AM, *et al*. Mortality in high-risk emergency general surgical admissions. *Br J Surg* 2013; **100**: 1318–25
3. Saunders DI, Murray D, Pichel AC, Varley S, Peden CJ. Variations in mortality after emergency laparotomy: the first report of the UK emergency laparotomy network. *Br J Anaesth* 2012; **109**: 368–75
4. Al-Temimi MH, Griffee M, Enniss TM, *et al*. When is death inevitable after emergency laparotomy? Analysis of the American College of Surgeons National Surgical Quality Improvement Program database. *J Am Coll Surg* 2012; **215**: 503–11
5. Vester-Andersen M, Lundstrom LH, Moller MH, *et al*. Mortality and postoperative care pathways after emergency gastrointestinal surgery in 2904 patients: a population-based cohort study. *Br J Anaesth* 2014; **112**: 860–70
6. Abbas SM, Kahokehr A, Mahmoud M, Hill AG. The Simple Prognostic Index (SPI)—a pathophysiologic prognostic scoring tool for emergency laparotomy. *J Surg Res* 2010; **163**: e59–65
7. Awad S, Herrod PJ, Palmer R, *et al*. One- and two-year outcomes and predictors of mortality following emergency laparotomy: a consecutive series from a United Kingdom teaching hospital. *World J Surg* 2012; **36**: 2060–7
8. Pearse RM, Harrison DA, James P, *et al*. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006; **10**: R81
9. NELA Project Team. *First patient report of the National Emergency Laparotomy Audit*. London: RCoA, 2015
10. The Australian and New Zealand Audit of Surgical Mortality. Available from http://www.surgeons.org/for-health-professionals/audits-and-surgical-research/anzasm/#national_report
11. American College of Surgeons National Surgical Quality Improvement Program. Available from http://site.acsnsqip.org/
12. The Dr Foster Global Comparators Study. Available from http://drfosterintelligence.co.uk/global-comparators/
13. Moonesinghe SR, Mythen MG, Grocott MP. High-risk surgery: epidemiology and outcomes. *Anesth Analg* 2011; **112**: 891–901
14. Ackland GL, Edwards M. Defining higher-risk surgery. *Curr Opin Crit Care* 2010; **16**: 339–46
15. Moonesinghe SR, Mythen MG, Grocott MP. Patient-related risk factors for postoperative adverse events. *Curr Opin Crit Care* 2009; **15**: 320–7
16. Pettigrew RA, Hill GL. Indicators of surgical risk and clinical judgment. *Br J Surg* 1986; **73**: 47–51
17. Hartley MN, Sagar PM. The surgeon's 'gut feeling' as a predictor of post-operative outcome. *Ann R Coll Surg Engl* 1994; **76**: 277–8
18. Markus PM, Martell J, Leister I, Horstmann O, Brinker J, Becker H. Predicting postoperative morbidity by clinical assessment. *Brit J Surg* 2005; **92**: 101–6
19. Ingraham AM, Cohen ME, Bilimoria KY, *et al*. Comparison of hospital performance in nonemergency versus emergency colorectal operations at 142 hospitals. *J Am Coll Surg* 2010; **210**: 155–65
20. Pearse RM, Moreno RP, Bauer P, *et al*. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012; **380**: 1059–65
21. Sorensen LT, Malaki A, Wille-Jorgensen P, *et al*. Risk factors for mortality and postoperative complications after gastrointestinal surgery. *J Gastrointest Surg* 2007; **11**: 903–10

22. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006; **144**: 201–9
23. Wyatt JC, Altman DG. Prognostic models: clinically useful or quickly forgotten? *BMJ* 1995; **311**: 1539–41
24. Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? *J Am Coll Cardiol* 2003; **42**: 851–3
25. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MPW. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery. Qualitative systematic review. *Anesthesiology* 2013; **119**: 959–81
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Open Med* 2009; **3**: e123–30
27. Deville WL, Buntinx F, Bouter LM, *et al*. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002; **2**: 2–14
28. Higgins JPT Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org (accessed March 2011)
29. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**: 1432–5
30. Toll DB, Janssen KJM, Vergouwe Y, Moons KGM. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008; **61**: 1085–94
31. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988; **240**: 1285–93
32. Egger M, Smith GD. Meta-analysis: bias in location and selection of studies. *BMJ* 1998; **316**: 61–6
33. Altman DG. Systematic reviews in health care: systematic reviews of evaluations of prognostic variables. *BMJ* 2001; **323**: 224–8
34. Kermani R, Coury JJ Jr, Dao H, *et al*. A practical mortality risk score for emergent colectomy. *Dis Colon Rectum* 2013; **56**: 467–74
35. 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
36. Kologlu M, Elker D, Altun H, Sayek I. Validation of MPI and PIA II in two different groups of patients with secondary peritonitis. *Hepatogastroenterology* 2001; **48**: 147–51
37. Buck DL, Vester-Andersen M, Moller MH. Accuracy of clinical prediction rules in peptic ulcer perforation: an observational study. *Scand J Gastroenterol* 2012; **47**: 28–35
38. Lohsiriwat V, Prapasrivorakul S, Lohsiriwat D. Perforated peptic ulcer: clinical presentation, surgical outcomes, and the accuracy of the Boey scoring system in predicting postoperative morbidity and mortality. *World J Surg* 2009; **33**: 80–5
39. Ren L, Upadhyay AM, Wang L, Li L, Lu J, Fu W. Mortality rate prediction by Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (POSSUM), Portsmouth POSSUM and Colorectal POSSUM and the development of new scoring systems in Chinese colorectal cancer patients. *Am J Surg* 2009; **198**: 31–8
40. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; **78**: 355–60
41. Legall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS-II) based on a European/North American multicenter study. *JAMA* 1993; **270**: 2957–63
42. Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM-II) based on an international cohort of intensive care unit patients. *JAMA* 1993; **270**: 2478–86
43. Levy MM, Fink MP, Marshall JC, *et al*. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003; **31**: 1250–6
44. Ohmann C, Wittmann DH, Wacha H. Prospective evaluation of prognostic scoring systems in peritonitis. *Eur J Surg* 1993; **159**: 267–74
45. Kulkarni SV, Naik AS, Subramanian N Jr. APACHE-II scoring system in perforative peritonitis. *Am J Surg* 2007; **194**: 549–52
46. Koc M, Yoldas O, Kilic YA, *et al*. Comparison and validation of scoring systems in a cohort of patients treated for perforated peptic ulcer. *Langenbecks Arch Surg* 2007; **392**: 581–5
47. Garcea G, Ganga R, Neal CP, Ong SL, Dennison AR, Berry DP. Preoperative early warning scores can predict in-hospital mortality and critical care admission following emergency surgery. *J Surg Res* 2010; **159**: 729–34
48. Ertan T, Yoldas O, Kilic YA, *et al*. External validation of prognostic models among cancer patients undergoing emergency colorectal surgery. *Am J Surg* 2008; **195**: 439–41
49. Moller MH, Engebjerg MC, Adamsen S, Bendix J, Thomsen RW. The Peptic Ulcer Perforation (PULP) score: a predictor of mortality following peptic ulcer perforation. A cohort study. *Acta Anaesthesiol Scand* 2012; **56**: 655–62
50. Kwok AC, Lipsitz SR, Bader AM, Gawande AA. Are targeted preoperative risk prediction tools more powerful? A test of models for emergency colon surgery in the very elderly. *J Am Coll Surg* 2011; **213**: 220–5
51. Poon JT, Chan B, Law WL. Evaluation of P-POSSUM in surgery for obstructing colorectal cancer and correlation of the predicted mortality with different surgical options. *Dis Colon Rectum* 2005; **48**: 493–8
52. Ferjani AM, Griffin D, Stallard N, Wong LS. A newly devised scoring system for prediction of mortality in patients with colorectal cancer: a prospective study. *Lancet Oncol* 2007; **8**: 317–22
53. Biondo S, Ramos E, Fraccalvieri D, Kreisler E, Marti Rague J, Jaurrieta E. Comparative study of left colonic Peritonitis Severity Score and Mannheim Peritonitis Index. *Br J Surg* 2006; **93**: 616–22
54. Notash AY, Salimi J, Rahimian H, Fesharaki Ms, Abbasi A. Evaluation of Mannheim peritonitis index and multiple organ failure score in patients with peritonitis. *Indian J Gastroenterol* 2005; **24**: 197–200
55. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000; **19**: 453–73
56. Ghaferi AA, Birkmeyer JD, Dimick JB. Variation in hospital mortality associated with inpatient surgery. *N Engl J Med* 2009; **361**: 1368–75
57. Wijesinghe LD, Mahmood T, Scott DJA, Berridge DC, Kent PJ, Kester RC. Comparison of POSSUM and the Portsmouth predictor equation for predicting death following vascular surgery. *Br J Surg* 1998; **85**: 209–12
58. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification-system. *Crit Care Med* 1985; **13**: 818–29
59. Faiz O, Warusavitarne J, Bottle A, *et al*. Nonelective excisional colorectal surgery in English National Health Service Trusts: a study of outcomes from Hospital Episode Statistics data between 1996 and 2007. *J Am Coll Surg* 2010; **210**: 390–401
60. Kesisoglou I, Pliakos I, Sapalidis K, Deligiannidis N, Papavramidis S. Emergency treatment of complicated colorectal cancer in the elderly. Should the surgical procedure be influenced by the factor 'age'? *Eur J Cancer Care* 2010; **19**: 820–6

61. Ingraham AM, Cohen ME, Raval MV, Ko CY, Nathens AB. Variation in quality of care after emergency general surgery procedures in the elderly. *J Am Coll Surg* 2011; **212**: 1039–48
62. Rix TE, Bates T. Pre-operative risk scores for the prediction of outcome in elderly people who require emergency surgery. *World J Emerg Surg* 2007; **2**: 16
63. Pelavski AD, Lacasta A, De Miguel M, Rochera MI, Roca M. Mortality and surgical risk assessment among the extreme old undergoing emergency surgery. *Am J Surg* 2013; **205**: 58–63
64. Ragg JL, Watters DA, Guest GD. Preoperative risk stratification for mortality and major morbidity in major colorectal surgery. *Dis Colon Rectum* 2009; **52**: 1296–303
65. Kramer AA. Predictive mortality models are not like fine wine. *Crit Care* 2005; **9**: 636–7
66. Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Reporting studies of diagnostic accuracy according to a standard method; the Standards for Reporting of Diagnostic Accuracy (STARD). *Ned Tijdschr Geneeskd* 2003; **147**: 336–40
67. Cruz DN, Antonelli M, Fumagalli R, *et al*. Early use of polymyxin B hemoperfusion in abdominal septic shock. The EUPHAS randomized controlled trial. *JAMA* 2009; **301**: 2445–52
68. Gargon E, Gurung B, Medley N, *et al*. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One* 2014; **6**: 1–12
69. Moonesinghe SR, Harris S, Mythen MG, *et al*. Prolonged postoperative morbidity is an independent risk factor for reduced long-term survival. *Br J Anaesth* 2013; **110**: 868

Handling editor: J. G. Hardman