Eur J Vasc Endovasc Surg **34**, 505–513 (2007)
doi:10.1016/j.ejvs.2007.07.019, available online at http://www.sciencedirect.com on ScienceDirect

Comparison of Risk-scoring Methods in Predicting the Immediate Outcome after Elective Open Abdominal Aortic Aneurysm Surgery

T.Y. Tang,[1] S.R. Walsh,[1] T.R. Fanshawe,[2] V. Seppi,[3] U. Sadat,[1] P.D. Hayes,[1] K. Varty,[1] M.E. Gaunt[1] and J.R. Boyle[1]*

[1]*Cambridge Vascular Unit, Cambridge University Hospital NHS Foundation Trust, Cambridge, UK,* [2]*Centre for Applied Medical Statistics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK, and* [3]*James Paget University Hospitals NHS Foundation Trust, Great Yarmouth, UK*

Background & Objectives. *The aim of this study was to apply three simple risk - scoring systems to prospectively collected data on all elective open Abdominal Aortic Aneurysm (AAA) operations in the Cambridge Academic Vascular Unit over a 6 - year period (January 1998 to January 2004), to compare their predictive values and to evaluate their validity with respect to prediction of mortality and post-operative complications.*
Methods. *204 patients underwent elective open infra-renal AAA repair. Data were prospectively collected and risk assessment scores were calculated for mortality and morbidity according to the Glasgow Aneurysm Score (GAS), VBHOM (Vascular Biochemistry and Haematology Outcome Models) and Estimation of Physiologic Ability and Surgical Stress (E-PASS).*
Results. *The mortality rate was 6.3% (13/204) and 59% (121/204) experienced a post-operative complication (30-day outcome). For GAS, VBHOM and E-PASS the receiver operating characteristics (ROC) curve analysis for prediction of in-hospital mortality showed area under the curve (AUC) of 0.84 (95% confidence interval [CI], 0.76 to 0.92; $p < 0.0001$), 0.82 (95% CI, 0.68 to 0.95; $p = 0.0001$) and 0.92 (95% CI, 0.87 to 0.97; $p < 0.0001$) respectively. There were also significant correlations between post-operative complications and length of hospital stay and each of the three scores, but the correlation was substantially higher in the case of E-PASS.*
Conclusions. *All three scoring systems accurately predicted the risk of mortality and morbidity in patients undergoing elective open AAA repair. Among these, E-PASS seemed to be the most accurate predictor in this patient population.*

Keywords: Surgical audit; Complications; Aneurysm (abdominal aortic); Outcome; VBHOM; E-PASS; GAS.

Introduction

Despite advances in peri-operative care, elective open Abdominal Aortic Aneurysm (AAA) surgery mortality remains around 6–8% with associated high morbidity rates.[1–3] Subjecting patients with a high operative risk to a futile AAA repair has resource and ethical implications. Therefore, prediction of immediate post-operative outcome assumes obvious relevance because it may aid in pre-operative risk stratification and planning.

Risk assessment scoring systems built from statistical models have been used to accurately predict outcome after elective open AAA repair.[4–6] The Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (POSSUM)[7] is currently the most tested system for assessing outcomes by risk-adjusted analysis in the United Kingdom. Unfortunately POSSUM generally suffers from incomplete data collection and has been far too complex for practical use at the bedside. Furthermore, variability in the timing of data collection, inclusion of several subjective-based parameters, allocation of low (normal) scores for missing data and the inclusion of data that correct for poor surgical technique makes scoring systems based on the POSSUM less than ideal.[8]

Three simpler methods of stratifying pre-operative risk of death in patients undergoing open elective AAA repair have been proposed: the Glasgow Aneurysm Score (GAS),[9] VBHOM (Vascular Biochemistry

*Corresponding author. Mr. J. R. Boyle, MD, FRCS, Consultant Vascular Surgeon, Regional Vascular Unit, Box 201; Level 7, Cambridge University Hospital NHS Foundation Trust, Hills Road, Cambridge CB2 2QQ, UK.
E-mail address: jonboyle@doctors.org.uk

1078–5884/000505 + 09 $32.00/0

and Haematology Outcome Models)[10] and Estimation of Physiologic Ability and Surgical Stress (E-PASS),[11] which all use less number of variables and therefore have obvious advantages over POSSUM in amount of data entry needed and the complexity of the analysis.

The aim of this study was to apply the three different scoring systems to all prospectively collected data on all elective open AAA operations in the Cambridge Academic Vascular Unit over a 6 - year period to compare their predictive values and to evaluate their validity with respect to prediction of mortality and post-operative complications.

Patients and Methods

All patients undergoing elective open infra-renal AAA surgery at the Cambridge Academic Vascular Unit between January 1998 and January 2004 were included in the study. Physiological and operative variables were collected prospectively, supplemented by case note review for observed in-hospital morbidity and 30-day mortality. This included deaths after transfer from the surgical unit to another unit within the same or another hospital. Hospital stay was also documented. Patients who presented and were turned down for elective AAA repair were not included in the study, as the Comprehensive Risk Score (CRS) component of E-PASS requires operative variables (see below). The study excluded patients with an urgent or ruptured AAA because they may already have a degree of systemic inflammation, which would have confounded the E-PASS scores. Aneurysms requiring supra-renal fixation and those who met the criteria of systemic inflammatory response syndrome (SIRS)[12] prior to surgery were also excluded.

Post-operative complications were only included when medical or interventional treatment had been carried out. Complications documented were wound infection and wound breakdown; intra-abdominal bleeding; intra-abdominal collection or abscess; any other infective complication associated with pyrexia, leucocytosis and positive cultures e.g. line infection; septic shock;[12] gastrointestinal bleeding; bowel obstruction and perforation; chest infection; urinary tract infection; pulmonary oedema; myocardial infarction; pulmonary embolus; new-onset arrhythmias; cerebrovascular accident; cardiopulmonary arrest, pleural effusion; and renal impairment. These complications have been previously defined.[13]

The morbidity score (MS) was previously defined:[14] grade 0, no complications; grade 1, mild complications that were not life-threatening; grade 2, moderate complications that were potentially life-threatening unless adequate treatment was initiated; grade 3, severe organ dysfunction that usually required mechanical support and grade 4, in hospital death as a direct result of complications. The cut-points used for CRS has been previously described,[14] for GAS by quintiles, and for VBHOM.[15]

GAS was calculated for each patient according to the following formula: risk score = (age (in years)) + (17 for shock) + (7 for myocardial disease) + (10 for cerebrovascular disease) + (14 for renal disease).[9]

VBHOM uses only seven items, which can all be obtained pre-operatively. They are: age at admission, sex of patient, haemoglobin, white cell count, urea, sodium and potassium levels. The pathology data items used were those from the first routinely collected haematology and biochemistry blood tests from admission. The following equation, which has been previously published,[10] was applied: $VBHOM = 1/(1 + \exp(-(-15.4194 + 4.4598 + (0.3290 \text{ for male sex}) + (0.1145 \times (0.1145 \times \text{age}) + (0.1110 \times \text{urea}) + (-0.0047 \times \text{sodium}) + (0.2846 \times \text{potassium}) + (-0.0383 \times \text{haemoglobin}) + (0.0048 \times \text{white cell count}))))$.

The E-PASS scoring system has been previously described[14] (Table 1) and Pre-operative Risk Score (PRS), Surgical Stress Score (SSS) and CRS scores were calculated from these equations. It is based on the results of multiple regression analysis with 6 pre-operative and 3 surgical factors identified as risk factors in gastrointestinal surgery. The PRS is calculated using the following factors: age, presence or absence of severe heart disease, severe lung disease and diabetes mellitus, American Society of Anaesthesiologists (ASA) physiological status classification[16] and performance status index (PSI) defined by the Eastern Cooperative Oncology Group.[17]

Statistical analysis

SPSS® version 13.0 (SPSS Inc., Chicago, IL, USA) and R version 2.2.1 (R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org) were used for statistical analysis of the data.

For each of the records, estimates of hospital mortality and morbidity were calculated using the predictive equations. In order to assess the incidence of mortality and morbidity for different categories of CRS, the categorisation suggested by Haga was used.[18] The categories were <0.3, 0.3−<0.5, 0.5−<1.0 and >=1.0. The two lowest categories suggested by Haga *et al.* (<0.1 and 0.1−<0.3) were grouped

Table 1. Equations for E-PASS Score

Preoperative Risk Score (PRS) is calculated with:

$$-0.0686 + 0.00345X_1 + 0.323X_2 + 0.205X_3 + 0.153X_4 + 0.148X_5 + 0.0666X_6$$

Where X_1 is age; X_2, the presence (1) or absence (0) of severe heart disease[a]; X_3, the presence (1) or absence (0) of severe pulmonary disease[b]; X_4, the presence (1) or absence (0) of diabetes mellitus[c]; X_5, the performance status index (range, 0−4); and X_6, the ASA physiologic status classification (range 1−5).

The surgical stress score (SSS) is calculated with:

$$-0.342 + 0.0139X_1 + 0.0392X_2 + 0.352X_3$$

Where X_1 is blood loss divided by body weight (ml/kg); X_2 is the operation time (hours); and X_3, the extent of the skin incision (0 = a minor incision for laparoscopic or thoracoscopic surgery including laparoscopic or thoracoscopic-assisted surgery; 1, laparotomy or thoracotomy alone; and 2, laparotomy and thoracotomy).

Comprehensive risk score (CRS) is calculated with:

$$-0.328 + 0.936(\text{PRS}) + 0.976(\text{SSS})$$

(From Haga *et al.*, 1999).[14]

[a] Severe heart disease is defined as heart failure (New York Heart Association Class III or IV) or severe arrhythmia requiring mechanical support.

[b] Severe pulmonary disease is defined as any condition with a percent vital capacity of less than 60% and/or a percentage forced expiratory volume in 1 second of less than 50%.

[c] Diabetes mellitus is defined according to the World Health Organization criteria as the presence of either fasting venous plasma glucose levels of 7.0 mmol/L (126 mg/dL) or greater, or 2-hour venous plasma glucose levels of 11.1 mmol/L (200 mg/dL) or greater after a 75 g oral glucose tolerance test.

together because only one patient fell into the <0.1 CRS category in this dataset.

The VBHOM equations were applied to the Cambridge data and tested for goodness of fit using Hosmer-Lemeshow methodology.[19,20] This involves the use of the chi-squared test to compare frequency tables obtained from prospective application of the equations. It should be noted that this is a null hypothesis test. P values less than 0.05 are indicative of a lack of fit. As this is a goodness-of-fit test, it is reasonable to conclude that a model is wrong i.e. did not predict outcome, if the null hypothesis is rejected but it is not possible to state that a different model is correct, only that it performed adequately. The statistical analysis of the overall goodness-of-fit of the model was undertaken using techniques designed to test both calibration and discrimination. Calibration is defined as the accuracy of numerical risk predictions, while discrimination is the ability of the model to appropriately rank patients in terms of risk - that is, the model's ability to ascribe high risks to high-risk patients and vice-versa. Empirical and binormal receiver operating characteristic ROC curves were plotted to assess the discriminative ability of each of the three models (CRS, GAS and VBHOM) with respect to mortality and morbidity, and the area under the ROC curve (AUC), sometimes known as the c-statistic, was used as a measure of overall diagnostic accuracy.[21] AUC values of 0.5 are given by models that is no better than chance. It is generally accepted that reasonable models produce values in the range of 0.7 to 0.8 and good models give values over 0.8.

Risk ranges were classified into meaningful categories, to give at least five predicted deaths or complications in at least 80% of the risk strata (Cochrane's rule) and to give, where possible, approximately equal predicted numbers in each risk range and to include greater than 5 percentage points.

Continuous variables were summarised by the mean and standard deviation. The statistical significance of differences in variables between groups of patients was determined using two-sided Mann-Whitney U - and Jonckheere-Terpstra tests (for continuous and ordered categorical variables, respectively). Spearman's rho was used to assess the correlation between hospital stay and CRS, GAS and VBHOM scores and Fisher's exact test to examine the association between renal risk and mortality and morbidity. A significance level of 0.05 was used for all comparisons.

Results

The unit operated on 219 patients over this six- year period. 204 (93%) patients were included in the analysis. The median age was 73 years (range 44−86) and 180 patients (88%) were male. 13 (6%) patients were excluded because their notes and details were missing at the time of analysis and 2 (1%) patients met the criteria of SIRS prior to surgery and were therefore also excluded. The decision to exclude them would not have had a major impact on the results because (i) there were not many of them (7%), and (ii) there was no reason to suspect that they would be systematically different from the included patients. No excluded patients died peri-operatively.

Patient demographics and admission data (physiological parameters) are summarised in Table 2.

There were 13 (6%) deaths during the immediate post-operative period. 121 (59%) experienced a post-operative complication. No intra-operative death occurred.

Although showing some predictive power, GAS and VBHOM appear to be weaker predictors of mortality than the E-PASS score (CRS). This is shown in Fig. 1, in which the fitted binormal ROC curves for each score is plotted. At all levels of sensitivity that would be

Table 2. Demographic data

Number of patients in analysis	204
Age (years) (median, range)	73 (44–86)
Male: Female	180: 24
Severe Heart disease	20 (10%)
Severe pulmonary disease	21 (10%)
Diabetes mellitus	20 (10%)
Shock	0 (0%)
Myocardial disease	70 (34%)
Cerebrovascular disease	25 (12%)
Renal disease	52 (26%)
[a]ASA (I:II:III:IV)	4: 86: 111: 3
[b]PSI (I:II:III:IV)	6: 91: 88: 19
Weight (kg)	80 (14)
Blood loss (ml)	864 (430)
Haemoglobin	13.8 (1.5)
White cell count (median, interquartile range)	8.0 (6.8–9.6)
Urea	7.1 (2.4)
Creatinine (median, interquartile range)	101 (85–122)
Sodium	139.8 (3.2)
Potassium	4.5 (0.6)
Operating time (hrs)	2.0 (0.4)
Observed in hospital mortality	13 (6%)
Observed in hospital morbidity	121 (59%)

Mean (standard deviation) or n (%) given unless stated.
[a] ASA = American Society of Anaesthesiologists (ASA).
[b] PSI = Physiological Status Classification.

considered useful, the E-PASS CRS score clearly outperforms both the GAS and the VBHOM scores.

GAS

GAS was statistically significantly higher in patients who died than patients who did not (mean difference 15.3, 95% confidence interval (CI) 8.4 to 22.2, $p < 0.0001$), and significantly higher in patients who had signs of morbidity than in patients who did not (mean difference 7.2, 95% CI 3.8 to 10.7, $p < 0.0001$). The AUC was 0.84 (95% CI 0.76 to 0.92) for mortality and 0.66 (95% CI 0.58 to 0.74) for morbidity. There is no cut-off point that can be chosen in order to predict a post-operative mortality rate of 100%: even at the upper end of the GAS scale, the majority of patients do not die (see Fig. 2).

VBHOM

Similar results were obtained for VBHOM (mean difference 0.20 for mortality, $p = 0.0001$; mean difference 0.07 for morbidity, $p = 0.0004$) (Table 3). The AUC was 0.82 (95% CI 0.68 to 0.95) for mortality and 0.65 (95% CI 0.57 to 0.73) for morbidity. However, VBHOM is a scoring system designed to give risk estimates and these risk estimates were found to be ill-calibrated for this population (Table 4). Many fewer deaths were observed than would be predicted using the risk estimates from VBHOM (approximately one quarter of the total number of deaths predicted were observed in this dataset).

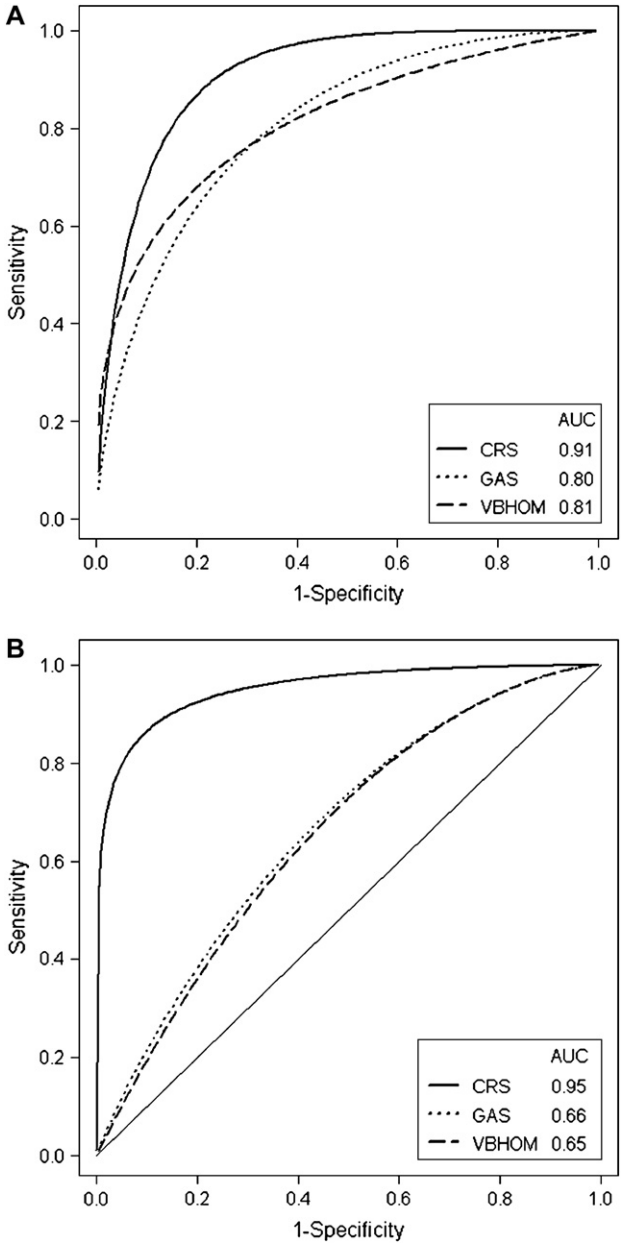

Fig. 1. Receiver Operating Characteristic (ROC) curves for mortality and morbidity for the different scoring methods. The fitted binormal ROC curve for each score is plotted. Approximate optimal cut-points for mortality: CRS 0.61 (sensitivity 87%, specificity 79%), GAS 84 (sensitivity 75%, specificity 70%), VBHOM 0.32 (sensitivity 67%, specificity 80%), Approximate optimal cut-points for morbidity: CRS 0.45 (sensitivity 86%, specificity 90%), GAS 75 (sensitivity 69%, specificity 55%), VBHOM 0.15 (sensitivity 71%, specificity 52%).

The predicted mean mortality rate for VBHOM was 0.21 (SD 0.14), which indicates that we would expect to

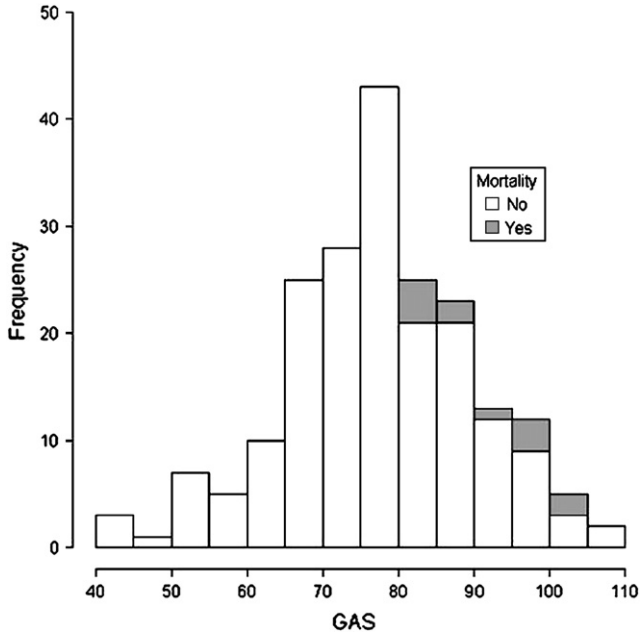

Fig. 2. Mortality rate with respect to Glasgow Aneurysm Score.

see approximately 43 deaths in this sample. Only 13 were observed (Table 4). This provides strong evidence that VBHOM is not well-calibrated, and that mortality risk is being overestimated, for this population. VBHOM, however, under-predicts morbidity. The Hosmer-Lemeshow test based on quartiles (to increase predicted number of events per category) formalises the clear lack of fit of VBHOM for this dataset (Table 5).

E-PASS

Prospective application of the E-PASS equations on this group of patients have been previously reported on[11] but in summary:

Table 3. Mean scores for E-PASS, GAS and VBHOM

Score	Group		n	Mean (standard deviation)
CRS	All patients		204	0.52 (0.27)
	Mortality	Yes	13	0.98 (0.26)
		No	191	0.49 (0.24)
	Morbidity	Yes	121	0.66 (0.25)
		No	83	0.30 (0.10)
GAS	All patients		204	78.4 (12.8)
	Mortality	Yes	13	92.7 (10.0)
		No	191	77.4 (12.4)
	Morbidity	Yes	121	81.3 (11.8)
		No	83	74.1 (13.0)
VBHOM	All patients		204	0.21 (0.14)
	Mortality	Yes	13	0.40 (0.17)
		No	191	0.20 (0.13)
	Morbidity	Yes	121	0.24 (0.15)
		No	83	0.17 (0.14)

Table 4. VBHOM

Risk band	Number of patients	Predicted deaths	Observed deaths	Observed morbidity
0 to 0.05	25	0.68	0	7
0.05 to 0.1	25	1.93	1	13
0.1 to 0.15	24	2.97	0	16
0.15 to 0.25	62	12.08	2	37
0.25 to 0.5	57	19.60	5	38
0.5 to 1	11	6.22	5	10

As the CRS increased, the incidence of post-operative morbidity and mortality significantly increased ($p < 0.0001$) (Table 3, Fig. 3). CRS had extremely good predictive power for both mortality and morbidity as demonstrated by high areas under the ROC curve in both cases. The AUC was 0.92 (95% CI 0.87 to 0.97) for mortality and 0.95 (95% CI 0.93 to 0.98) for morbidity (Fig. 3). E-PASS had the largest area under the curve and its best approximate cut-off value (0.61) had a better accuracy than the other scoring systems. Individually, PRS and SSS also demonstrated a strong relationship with mortality and development of complications, which have been previously reported.[11] (PRS: AUC 0.91 (95% CI 0.85 to 0.96) for mortality, 0.93 (95% CI 0.90 to 0.97) for morbidity; SSS: AUC 0.80 (95% CI 0.65 to 0.95) for mortality, 0.80 (95% CI 0.74 to 0.87) for morbidity). In particular, the AUC values for PRS were similar to those of the combined score, CRS.

Hospital stay

There was a statistically significant correlation between length of hospital stay and each of the three scores, but the correlation was substantially higher in the case of CRS than it was for either GAS or VBHOM:

CRS: 0.53 (95% CI 0.42 to 0.62)
GAS: 0.31 (95% CI 0.18 to 0.43)
VBHOM: 0.34 (95% CI 0.20 to 0.46)

Table 5. VBHOM

Risk band	Number of patients	Predicted deaths	Observed deaths	Observed morbidity
0 to 0.1011	51	2.71	1	21
0.1011 to 0.1846	51	7.46	0	30
0.1846 to 0.2881	52	12.36	3	35
0.2881 to 1	50	20.96	9	35

$X^2 = 30.9$ for mortality, $p < 0.0001$: evidence of lack of fit.
$X^2 = 280.6$ for morbidity, $p < 0.0001$: evidence of lack of fit.
(Chi-squared distribution with df = 2; critical value for test of size 0.05 is 6.0).

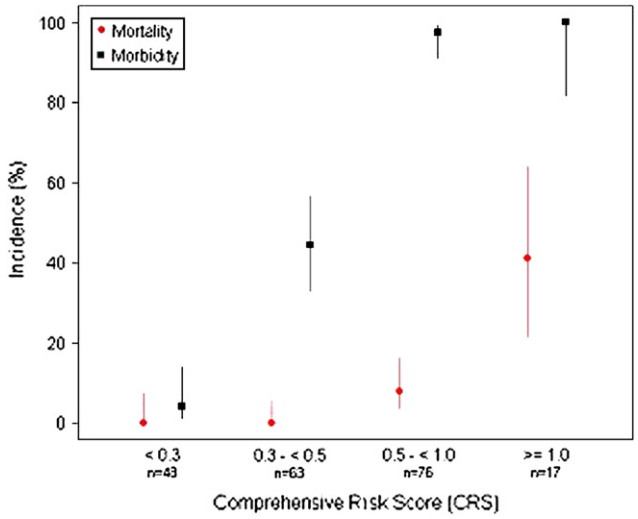

Fig. 3. Incidence of mortality and morbidity according to CRS. The graph appears to demonstrate that patients in the $>=1.0$ category are at particularly high risk of mortality, and in the 0.5−<1.0 and $>=1.0$ categories at particularly high risk of morbidity. Bars show 95% confidence intervals.

Renal disease

Of the 52 patients with renal disease, 4 died (8%) and 35 (67%) showed signs of morbidity; there was no significant difference in either mortality ($p=0.74$) or morbidity ($p=0.19$) compared to patients without renal disease ($9/152=6\%$ mortality, $86/152=57\%$ morbidity).

Morbidity and morbidity score

This relationship can be seen via the ROC curve for morbidity (Fig. 1) and the tables of morbidity score by CRS, GAS and VBHOM category (Table 6). For each of the three scores, there is a strongly statistically significant trend of increasing score with increasing morbidity score ($p<0.0001$ in each case). However the ROC curves indicate that CRS would be the best of the three for making individual predictions of morbidity.

Discussion

This study involved patients undergoing elective open infra-renal AAA surgery at one tertiary referral vascular centre in the United Kingdom. The overall mortality rate after operation of 6% is similar to those of other reported series.[22,23] All three scoring systems, using a variety of different mathematical regression equations, significantly predicted outcome. This

Table 6. Morbidity score

	Morbidity Score				
CRS	0	1	2	3	4
<0.3	46	1	1	0	0
0.3−< 0.5	35	3	20	5	0
0.5−<1.0	2	20	39	9	6
>1.0	0	1	7	2	7
GAS	0	1	2	3	4
>=68	26	4	10	1	0
69 to 75	14	8	14	2	0
76 to 80	20	5	15	3	0
81 to 88	13	5	13	3	6
>=89	10	3	15	7	7
VBHOM	0	1	2	3	4
0 to 0.05	18	2	4	1	0
0.05 to 0.1	12	3	8	1	1
0.1 to 0.15	8	4	8	4	0
0.15 to 0.25	25	8	24	3	2
0.25 to 0.5	19	7	20	6	5
0.5 to 1	1	1	3	1	5

externally validates previous studies of using the GAS in successfully predicting outcome after elective open AAA repair[23,24] but has the added advantage that data were all collected prospectively. This is also the first study to properly validate the use of VBHOM in predicting individual patient risk after this type of surgery. A recent article by Hadjianastassiou *et al.*[25] also compared the accuracy of several contemporary mortality prediction models including VBHOM after open abdominal aortic aneurysm (AAA) surgery. However, post-operative data seem to have been used to feed the POSSUM and VBHOM models, which all require pre-operative data. It is therefore not surprising that their APACHE-AAA model performed the best.

A proper comparison of models would have required appropriate use of pre- and post-operative data in the respective equations. Interestingly, reasonable discrimination values were still achieved suggesting that these models or at the very least the parameters they use are determinants of adverse outcome.

The E-PASS scoring system, which was originally generated based on the quantification of pre-operative risk and surgical stress applied in elective gastrointestinal surgery[14] was previously shown to be a useful decision making tool in patients undergoing elective open AAA surgery.[11]

The question is which system is more valuable to the clinician?

This comparison study found that E-PASS outperformed the other two scores in terms of accurately predicting both mortality and morbidity. Furthermore the correlation with length of hospital stay was substantially higher with E-PASS. It was previously reported that E-PASS was useful in estimating surgical

costs in gastrointestinal surgery.[26] E-PASS had a significant positive correlation to the duration and costs of hospital stay. The authors showed an equation for estimating surgical costs and compared a real to estimated costs among hospitals, proposing a risk-based payment system because hospitals that treat more high-risk patients would not only show higher mortality and morbidity rates but also surgical costs of hospital stay. This may be a useful costing exercise to do in AAA surgery in view of the fact that there was a strong positive correlation with hospital stay demonstrated. Although E-PASS uses far fewer variables and therefore has obvious advantages over POSSUM in amount of data entry needed and the complexity of the analysis, one of the disadvantages of E-PASS compared to GAS and VBHOM is that operative data is still required. However, we have found that the CRS can be quickly calculated immediately after the operation and the different parameters to calculate the pre-operative component of the score (PRS) were relatively easy to collect as demonstrated by the low number of cases excluded.[11] The authors also found a strong correlation between PRS and outcome ($p < 0.0001$ for mortality and morbidity), which may allow the vascular surgeon to predict risk in an individual patient before surgery. Furthermore this risk can be discussed confidently with both patient and relatives whilst gaining informed consent. If the risk predicted by PRS is too high for a patient a less invasive procedure such as endovascular stenting or conservative management may be considered.

This study, as discussed, validates previous work based on the VBHOM concept that the risk of in-hospital mortality can be modelled in patients undergoing index arterial operations such as AAA repair using a small number of commonly used laboratory and administrative items.[15] Although there was a significant correlation between VBHOM and outcome and the score had reasonable discriminative ability, the risk estimates were poorly calibrated for this subset of patients: VBHOM over-predicted mortality and under-predicted peri-operative complications. Perhaps the results are not surprising as this may be a consequence of the fact that the original VBHOM predictor equations (as was the case of GAS) were developed using a dataset that included patients undergoing emergency operations leading necessarily to a high death rate amongst patients with high risk estimates. However, it is important to note that VBHOM is the only score than attempts to predict mortality rates for individuals, and so is the only candidate which can be assessed by methods for determining calibration. Furthermore, a recent study focussed on the development of a new VBHOM model and its validation.[27] It was built from data collected prospectively from around the United Kingdom and seems to provide a single unified model that allows good prediction of surgical mortality after both open elective and ruptured AAA repair.

This study also showed that GAS had a predictive value for prognosis after elective open AAA repair. This successfully validates previous European studies, which demonstrated that GAS was highly predictive of post-operative outcome in different geographical settings[23,24,28,29] but with the added advantage that data were collected prospectively. However, previous work found that GAS's accuracy in predicting post-operative complications was less but this could be the way the data were interpreted and its retrospective collection.[28]

Limitations

Although all data were collected prospectively and are likely to be accurate, drawbacks of the study include the relative small sample size used and low mortality rate. A larger cohort of patients with more definite endpoints (mortality and morbidity) would have been more conducive to risk estimation. E-PASS currently does not give individual predicted percentage mortality or morbidity rather a range of mortality, whereas other scoring systems such as VBHOM do.[11] E-PASS also only targets the elective setting. Patients who have emergency surgery will already have a degree of inflammation, which will affect the E-PASS score pre-operatively. Like the POSSUM predictor equations it suffers from the weakness, which is by definition that they exclude patients who were either not offered or refused surgery. This highlights the importance of good patient selection.

GAS, though for its simplicity and its potential use pre-operatively, has limitations for clinical decision - making. The different groups describe low positive predictive value and positive likelihood ratio of this scoring system for post-operative mortality and morbidity, which implies that the system was not helpful for the individual high - risk patient.[24,28] However, it is important to appreciate that these values have a strong relationship with disease prevalence, that depends very much on the setting in which the score is used.

Other limitations of GAS are that it does not account for the protective effects of coronary artery bypass grafting or percutaneous transluminal coronary angioplasty for cardiac disease, carotid endarterectomy or stenting for cerebrovascular disease and the use of medications in general.

VBHOM is essentially a minimalist approach and does not use any data collected at operation. Some may therefore question its validity but it needs to be appreciated that all risk models can only predict risk within the ''dimensions'' of the data items used within the model. It is certain that there are numerous other factors, many of which would only be found at operation, which influence the risk of adverse outcome for individual patients.

Conclusions

Although these models performed well prospectively in this subset of patients, they should not be taken as definitive predictor equations and should only be used to aid the individual surgeon who ultimately must make the decision to operate on each patient with an AAA. The reader must remember that a particular group of patients in one geographical setting are likely to be different to the population that was used to build the original model. Statistical fit and model performance are therefore likely to be variable. All three scoring systems require further validation in vascular surgery at different geographical locations.

Competing interests

None declared.

Funding

SRW receives funding from the Moulton Charitable Foundation.

References

1 Department of Health. Hospital episode statistics 1989–1996. London: Department of Health; 1997.
2 Brady AR, Fowkes FG, Greenhalgh RM, Powell JT, Ruckley CV, Thompson SG. Risk factors for postoperative death following elective surgical repair of abdominal aortic aneurysm: results from the UK Small Aneurysm Trial. On behalf of the UK Small Aneurysm Trial participants. *Br J Surg* 2000; **87**(6):742–749.
3 Dardik A, Lin JW, Gordon TA, Williams GM, Perler BA. Results of elective abdominal aortic aneurysm repair in the 1990s: a population-based analysis of 2335 cases. *J Vasc Surg* 1999; **30**(6):985–995.
4 Hadjianastassiou VG, Tekkis PP, Goldhill DR, Hands LJ. Quantification of mortality risk after abdominal aortic aneurysm repair. *Br J Surg* 2005;**92**(9):1092–1098.
5 Hadjianastassiou VG, Franco L, Jerez JM, Evangelou IE, Goldhill DR, Tekkis PP *et al*. Informed prognosis [corrected] after abdominal aortic aneurysm repair using predictive modeling techniques [corrected]. *J Vasc Surg* 2006;**43**(3):467–473.
6 Prytherch DR, Sutton GL, Boyle JR. Portsmouth POSSUM models for abdominal aortic aneurysm surgery. *Br J Surg* 2001; **88**(7):958–963.
7 Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991;**78**(3):355–360.
8 Bann D, Sarin S. Comparative audit: the trouble with POSSUM. *J R Soc Med* 2001;**94**:632–634.
9 Samy AK, Murray G, MacBain G. Glasgow Aneurysm Score. *Cardiovasc Surg* 1994;**2**(1):41–44.
10 Prytherch DR, Ridler BM, Ashley S. Risk-adjusted predictive models of mortality after index arterial operations using a minimal data set. *Br J Surg* 2005;**92**(6):714–718.
11 Tang T, Walsh SR, Fanshawe TR, Gillard JH, Sadat U, Varty K *et al*. Estimation of physiologic ability and surgical stress (E-PASS) as a predictor of immediate outcome after elective abdominal aortic aneurysm surgery. *Am J Surg* 2007;**194**(2): 176–182.
12 ACCP/SCCM. Consensus Conference Committee/Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. *Chest* 1992;**101**:1644–1655.
13 Cook EJ, Walsh SR, Farooq N, Alberts JC, Justin TA, Keeling NJ. Post-operative neutrophil-lymphocyte ratio predicts complications following colorectal surgery. *Int J Surg* 2007;**5**(1):27–30.
14 Haga Y, Ikei S, Ogawa M. Estimation of Physiologic Ability and Surgical Stress (E-PASS) as a new prediction scoring system for post-operative morbidity and mortality following gastrointestinal surgery. *Surg Today* 1999;**29**:219–225.
15 Prytherch DR, Sirl JS, Weaver PC, Schmidt P, Higgins B, Sutton GL. Towards a national clinical minimum data set for general surgery. *Br J Surg* 2003;**90**(10):1300–1305.
16 Owens WD, Felts JA, Spitznagel Jr EL. ASA physical status classifications: a study of consistency of ratings. *Anesthesiology* 1978; **49**(4):239–243.
17 Oken M, Creech RH, Tormey DC, Horton J, Davies TE, McFadden ET *et al*. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *Am J Clin Oncol* 1982;**5**: 649–655.
18 Haga Y, Ikei S, Wada Y, Takeuchi H, Sameshima H, Kimura O *et al*. Evaluation of an Estimation of Physiologic Ability and Surgical Stress (E-PASS) scoring system to predict postoperative risk: a multicenter prospective study. *Surg Today* 2001;**31**(7): 569–574.
19 Hosmer DJ, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. *Commun Stat Theor Meth* 1980;**A9**: 1043–1069.
20 Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;**115**:92–106.
21 Zhou X-H, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. Wiley, UK; 2002.
22 Galland RB. Mortality following elective infrarenal aortic reconstruction: a Joint Vascular Research Group study. *Br J Surg* 1998; **85**(5):633–636.
23 Biancari F, Leo E, Ylonen K, Vaarala MH, Rainio P, Juvonen T. Value of the Glasgow Aneurysm Score in predicting the immediate and long-term outcome after elective open repair of infrarenal abdominal aortic aneurysm. *Br J Surg* 2003;**90**(7):838–844.
24 Hirzalla O, Emous M, Ubbink DT, Legemate D. External validation of the Glasgow Aneurysm Score to predict outcome in elective open abdominal aortic aneurysm repair. *J Vasc Surg* 2006; **44**(4):712–716.
25 Hadjianastassiou VG, Tekkis PP, Athanasiou T, Muktadir A, Young JD, Hands LJ. Comparison of mortality prediction models after open abdominal aortic aneurysm repair. *Eur J Vasc Endovasc Surg* 2007;**33**(5):536–543.
26 Haga Y, Wada Y, Takeuchi H, Sameshima H, Kimura O, Furuya T. Estimation of surgical costs using a prediction scoring system: estimation of physiologic ability and surgical stress. *Arch Surg* 2002;**137**(4):481–485.
27 Tang T, Walsh SR, Prytherch DR, Lees T, Varty K, Boyle JR. VBHOM, a data economic model for predicting the outcome

after open abdominal aortic aneurysm surgery. *Br J Surg* 2007; **94**(6):717–721.
28 Nesi F, Leo E, Biancari F, Bartolucci R, Rainio P, Satta J *et al*. Preoperative risk stratification in patients undergoing elective infrarenal aortic aneurysm surgery: evaluation of five risk scoring methods. *Eur J Vasc Endovasc Surg* 2004; **28**(1):52–58.
29 Biancari F, Heikkinen M, Lepantalo M, Salenius JP. Glasgow Aneurysm Score in patients undergoing elective open repair of abdominal aortic aneurysm: a Finnvasc study. *Eur J Vasc Endovasc Surg* 2003;**26**(6):612–617.

Accepted 22 July 2007
Available online 14 September 2007