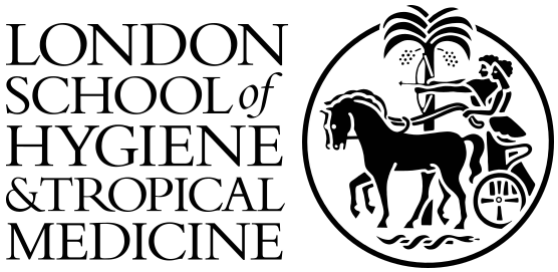

Original article

Value of risk scores in the decision to palliate patients with ruptured abdominal aortic aneurysm

M. J. Sweeting[1], P. Ulug[2], J. Roy[5], R. Hultgren[5], R. Indrakusuma[6], R. Balm[6], M. M. Thompson[4], R. J. Hinchliffe[3], S. G. Thompson[1] and J. T. Powell[2], on behalf of the Ruptured Aneurysm Collaborators, including IMPROVE, AJAX, ECAR and STAR collaborators

[1]Department of Public Health and Primary Care, University of Cambridge, Cambridge, [2]Vascular Surgery Research Group, Imperial College London, London, and [3]Bristol Centre for Surgical Research, University of Bristol, Bristol, UK, [4]Stanford School of Medicine, Stanford, California, USA, [5]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden, and [6]Department of Vascular Surgery, Academic Medical Centre, Amsterdam, The Netherlands

Correspondence to: Dr M. J. Sweeting, Department of Health Sciences, University of Leicester, George Davies Centre, University Road, Leicester LE1 7RH, UK (e-mail: michael.sweeting@le.ac.uk)

Background: The aim of this study was to develop a 48-h mortality risk score, which included morphology data, for patients with ruptured abdominal aortic aneurysm presenting to an emergency department, and to assess its predictive accuracy and clinical effectiveness in triaging patients to immediate aneurysm repair, transfer or palliative care.

Methods: Data from patients in the IMPROVE (Immediate Management of the Patient With Ruptured Aneurysm: Open *Versus* Endovascular Repair) randomized trial were used to develop the risk score. Variables considered included age, sex, haemodynamic markers and aortic morphology. Backwards selection was used to identify relevant predictors. Predictive performance was assessed using calibration plots and the C-statistic. Validation of the newly developed and other previously published scores was conducted in four external populations. The net benefit of treating patients based on a risk threshold compared with treating none was quantified.

Results: Data from 536 patients in the IMPROVE trial were included. The final variables retained were age, sex, haemoglobin level, serum creatinine level, systolic BP, aortic neck length and angle, and acute myocardial ischaemia. The discrimination of the score for 48-h mortality in the IMPROVE data was reasonable (C-statistic 0·710, 95 per cent c.i. 0·659 to 0·760), but varied in external populations (from 0·652 to 0·761). The new score outperformed other published risk scores in some, but not all, populations. An 8 (95 per cent c.i. 5 to 11) per cent improvement in the C-statistic was estimated compared with using age alone.

Conclusion: The assessed risk scores did not have sufficient accuracy to enable potentially life-saving decisions to be made regarding intervention. Focus should therefore shift to offering repair to more patients and reducing non-intervention rates, while respecting the wishes of the patient and family.

Presented to the Vascular Society, Manchester, UK, November 2017

Paper accepted 13 December 2017

Published online 6 April 2018 in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.10820

Introduction

In the UK, the mortality rate from ruptured abdominal aortic aneurysm (rAAA) is higher than in the USA, possibly because non-intervention rates are higher[1]. Patients presenting to emergency departments with suspected rAAA require rapid triaging, with a decision whether to provide immediate aneurysm repair (or immediate transfer to a specialist centre for repair), or palliative care in those for whom aneurysm repair may be futile. To help such critical decisions, guidelines for transfer have been developed[2,3] and a number of risk scores have been published to predict postoperative (usually 30-day) mortality[4–7]. The majority of deaths actually occur within the first 48 h[8]. Many of these risk scores are based on simple haemodynamic parameters, yet their predictive ability has been shown to vary when applied to other populations[9,10]. One possible reason is that a risk score performs better on the data from which it was

derived than in independent data sources; this is especially true for small studies that considered a large number of potential predictors[11].

To date, risk scores have not incorporated assessments of aortic morphology. Today these are rapidly available from emergency CT and may provide critical information regarding the complexity of any endovascular aneurysm repair (EVAR) or open repair to be undertaken, as well as predicting postoperative mortality[12,13]. The IMPROVE randomized trial (ISRCTN 48334791)[8] of patients with an in-hospital clinical diagnosis of rAAA recorded a number of preoperative morphological variables from CT images in 86 per cent of their patients with confirmed rupture. This trial, therefore, presents an opportunity to assess the value of adding morphological variables to an rAAA risk score.

The aims of this study were twofold; first, to develop a novel, point-of-care, risk score for patients presenting with an rAAA based on the IMPROVE trial data set, incorporating both physiological and imaging data that are immediately available in the emergency department; and, second, to evaluate the newly developed risk score alongside previously published scores in four independent external data sets. The external data sources considered were the AJAX (Amsterdam Acute Aneurysm Trial)[14] and ECAR (Endovasculaire ou Chirurgie dans les Anevrysmes aorto-iliaques Rompus)[15] RCTs, and the Amsterdam area[16] and Stockholm area (STAR)[17] population-based cohorts of all rAAA repairs. The previously published risk scores evaluated were the Vascular Study Group of New England (VSGNE) score[4], the Hardman index[5] and the Vancouver rAAA risk score[6], chosen because they are simple to use in an emergency department.

Attention was focused on predicting 48-h mortality rather than in-hospital or 30-day mortality because the highest death rate following repair is in the first 48 h. Established risk scores have generally been derived using either in-hospital or 30-day mortality as an outcome, usually following open repair, and are therefore sensitive to changes and improvements in critical care; they may also be sensitive to the increasing use of EVAR for emergency abdominal aortic aneurysm (AAA) repair.

Methods

Data from the IMPROVE RCT were used to develop the new score, excluding patients with a final diagnosis other than rAAA (incidental or symptomatic AAAs were excluded); patients with an aortoiliac rupture were included. All these patients who had CT were considered in the risk score, regardless of whether aneurysm repair was started.

External validation of the risk score was conducted in patients from the AJAX and ECAR randomized trials. The inclusion criteria for randomization in these two RCTs, and hence the patient populations, were more selective than those in the IMPROVE trial. The IMPROVE trial randomized patients with a clinical diagnosis of rAAA, before confirmation of either rupture or suitability for EVAR, and patients did not have to be haemodynamically stable. The AJAX and ECAR trials required that patients had CT before randomization to confirm an AAA with acute haemorrhage outside the aortic wall, for the patient to be suitable for both EVAR and open repair, and for patients to be haemodynamically stable on arrival. External validation was undertaken using the wider Amsterdam cohort, which comprised patients who had rAAA repair but were unsuitable for randomization in the AJAX trial, as well as a large observational cohort of all patients admitted to hospital with rAAA in the Stockholm area (STAR cohort).

Measurements and outcomes

The outcome was death within 48 h of randomization, or within 48 h of presentation for the Amsterdam and STAR cohorts. Management included either open repair, EVAR or no repair. In secondary analyses, the 48-h risk score was also assessed for its ability to predict 30-day mortality reliably.

A set of variables was predefined for potential inclusion in the risk score, based on data available from the three rAAA trials (IMPROVE, AJAX and ECAR). These variables included: age, sex, admission systolic BP, additional admission variables for calculation of the Hardman index (haemoglobin level, serum creatinine level, acute myocardial ischaemia on ECG and loss of consciousness), and four basic morphological features of the aneurysm as measured from CT images, chosen because they are required to assess the feasibility of EVAR and are relatively easy to measure in a time-critical situation (maximum aortic diameter, aortic neck diameter, aortic neck length and proximal neck angle). None of the validation data sets provided data regarding the volume of intravenous fluids administered. Nevertheless, this variable was also assessed for its predictive ability when developing the model using the IMPROVE data. The IMPROVE data were recorded from the trial centres, which was not necessarily the hospital to which the patient presented.

Statistical analysis

Any non-linear relationship between each of the nine continuous candidate predictors (age, haemoglobin level,

serum creatinine level, admission systolic BP, volume of intravenous fluids administered, maximum aortic diameter, aortic neck diameter, aortic neck length, proximal aortic neck angle) and 48-h mortality was first assessed using fractional polynomials (FPs)[18] in a logistic model based on complete data (*Appendix S1*, supporting information). The shape of association was plotted for each predictor using the final chosen FP to check biological plausibility.

To deal with missing data for the predictors, multiple imputation was carried out (*Appendix S1*, supporting information). A multivariable logistic regression model was then chosen using a backwards selection procedure, whereby all candidate predictors were initially included using their chosen FP transformation, and variables were dropped progressively from the model if their *P* value was 0·157 or higher. The pooled coefficients of the selected variables across the multiply imputed data sets were used to define the risk score. For ease of use, a nomogram[19] was constructed based on the derived risk score.

The predicted probabilities of 48-h mortality were assessed in terms of calibration and discrimination. Calibration was assessed by plotting observed *versus* predicted risks within deciles of predicted risk, and reporting the estimated calibration slope from a logistic regression model with the risk score as the predictor[20]. Discrimination was assessed by calculating the C-statistic (the area under the receiver operating characteristic (ROC) curve).

Internal validation (within the IMPROVE data set) was conducted using tenfold cross-validation to avoid overoptimistic estimates of predictive performance caused by overfitting (*Appendix S1*, supporting information). External validation was conducted using data from the AJAX and ECAR trials, and the Amsterdam and STAR cohorts, to give an assessment of how well the derived risk score applied in other populations.

The predictive performance of other published risk scores[4–6] was assessed in each data source. Preoperative cardiac arrest was recorded only in the IMPROVE data set and so was dropped from the published risk scores in data in which it was missing. Furthermore, the VSGNE score included the use of a suprarenal clamp as a predictor, which was not specifically recorded in these data sets, and so a proxy (aortic neck length less than 10 mm) was used instead. The derivation of each score is provided in *Table S1* (supporting information).

The clinical value of any particular rAAA risk score depends on the ability to make better decisions with a model than without[21]. Following methodology developed by Vickers and Elkin[22], the net benefit of treating patients at different risk cut-offs *versus* treating none was quantified and the value of an rAAA risk score assessed. The net benefit is based on an assumed benefit to risk trade-off; for example, a surgeon treating patients with a mortality probability of 98 per cent or less quantifies the consequence of not operating when it would have been of benefit as 98 to 2, that is 49 times worse than the consequence of operating unnecessarily[20].

Results

A total of 536 patients from IMPROVE with a final diagnosis of rAAA were included, of whom 135 (25·2 per cent) died within 48 h of randomization. Some 320 patients commenced open repair, 182 started EVAR and a further 34 did not have an operation as they died before reaching theatre. A total of 327 patients (61·0 per cent) from IMPROVE arrived from another hospital; the 48-h mortality rate in these patients was 24·2 per cent, similar to that among patients who arrived directly. There were 113 and 107 patients in the AJAX and ECAR trials respectively, of whom 17 (15·0 per cent) and 15 (14·0 per cent) died within 48 h of randomization. Patients in IMPROVE were a mean of 2 years older, had larger aneurysms, and a greater proportion of women were recruited (*Table 1*). In the Amsterdam cohort, 114 of 402 patients (28·4 per cent) died within 48 h, whereas the mortality rate was much higher in the STAR cohort (107 of 284; 37·7 per cent), in part because patients were a mean of 4 years older, and one-quarter did not receive any aneurysm repair. Patients in the STAR cohort also had a lower mean systolic BP on admission, and a higher proportion had lost consciousness before arrival at the operating theatre. The mortality rate at 30 days was 40–60 per cent higher than at 48 h; the in-hospital mortality rate was similar to the 30-day mortality rate in all cohorts.

Risk score development

Unadjusted odds ratios for each candidate variable estimated in the IMPROVE trial after multiple imputation are shown in *Table S2* (supporting information). Most candidate predictors correlated with the outcome. FP modelling indicated a cubic effect of age, an inverse-squared effect of admission systolic BP and a log-transformed effect of neck length on the log-odds of mortality (*Fig. S1*, supporting information). Following backwards variable selection, the final variables retained in the multivariable prediction model were: age, sex, haemoglobin level on admission, serum creatinine level on admission, systolic BP on admission, aortic neck length, aortic neck angle,

Table 1 Descriptive statistics for the candidate predictors for the three RCTs and two cohorts

	IMPROVE RCT		AJAX RCT		ECAR RCT		Amsterdam cohort		STAR cohort	
	n	No. with event*	*n*	No. with event*	*n*	No. with event*	*n*	No. with event*	*n*	No. with event*
Death within 48 h	536	135 (25·2)	113	17 (15·0)	107	15 (14·0)	402	114 (28·3)	284	107 (37·7)
Death within 30 days	536	206 (38·4)	113	27 (23·9)	105	22 (21·0)	401	171 (42·6)	284	147 (51·8)
Death before primary hospital discharge	536	211 (39·4)	113	32 (28·3)	104	29 (27·9)	402	179 (44·5)	284	149 (52·5)
Operation commenced	536		113		107		402		279	
EVAR		182 (34·0)		55 (48·7)		56 (52·3)		15 (3·7)		59 (21·1)
Open aneurysm repair		320 (59·7)		58 (51·3)		50 (46·7)		329 (81·8)		147 (52·7)
No operation (palliated)		34 (6·3)		0 (0)		1 (0·9)		58 (14·4)		73 (26·2)
Age (years)†	536	76(8)	113	74(9)	107	74(11)	402	75(9)	284	79(9)
Male sex	536	424 (79·1)	113	97 (85·8)	107	97 (90·7)	402	303 (75·4)	284	215 (75·7)
Admission haemoglobin (g/dl)†	530	11·1(2·4)	113	11·5(2·3)	107	10·6(2·3)	387	11·0(2·6)	273	11·1(2·3)
Admission creatinine (μmol/l)‡	524	118 (95–153)	107	106 (91–142)	105	114 (91–136)	366	110 (89–139)	268	117 (90–140)
Admission systolic BP (mmHg)†	526	108(32)	110	120(40)	104	108(30)	328	111(36)	283	104(39)
Volume of i.v. fluids given before arrival in theatre (litres)†	391	1·06(1·13)	–	–	–	–	–	–	–	–
Maximum aneurysm diameter (mm)†	460	86(17)	92	76(16)	106	77(20)	–	–	192	81(19)
Aneurysm neck diameter (mm)†	390	25(4)	92	26(4)	106	24(5)	–	–	192	29(11)
Neck length (mm)†	435	23(17)	92	27(13)	101	25(14)	180	17(14)	192	18(18)
Proximal neck angle (°)†	432	33(20)	92	39(21)	96	34(26)	180	37(23)	192	23(16)
Acute myocardial ischaemia detected on ECG	495	38 (7·7)	65	12 (18)	107	4 (3·7)	128	29 (22·7)	139	60 (43·2)
Loss of consciousness	512	47 (9·2)	113	13 (11·5)	107	12 (11·2)	–	–	284	68 (23·9)
Preoperative cardiac arrest	536	8 (1·5)	–	–	–	–	–	–	–	–

*With percentages in parentheses unless indicated otherwise; values are †mean(s.d.) and ‡median (i.q.r.). EVAR, endovascular aneurysm repair; i.v., intravenous.

and acute myocardial ischaemia on ECG. Coefficients based on transformations of these variables are shown in *Table 2*, and these define the IMPROVE risk score. A nomogram to allow simple use of the score is presented in *Fig. 1*. A shorter neck length was strongly associated with higher 48-h mortality risk ($P < 0{\cdot}001$), whereas a smaller neck angle was weakly associated with higher risk ($P = 0{\cdot}113$). In a *post hoc* analysis, among patients who had an operation, the inverse relationship with neck angle was apparent only in those who received open repair.

Table 2 IMPROVE score coefficients from multivariable logistic regression

	β	*P*
Intercept	−1·1041 (0·7750)	
$\left(\frac{\text{Age}-50}{5}\right)^3$ (years)	0·0028 (0·0008)	0·001
Haemoglobin (g/dl)	−0·0933 (0·0524)	0·075
$\left(\frac{\text{Creatinine}}{100}\right)$ (μmol/l)	0·4391 (0·1615)	0·007
$\left(\frac{\text{Systolic BP}+1}{100}\right)^{-2}$ (mmHg)	0·4743 (0·1350)	< 0·001
$\log\left(\frac{\text{Neck length}+1}{10}\right)$ (mm)	−0·3450 (0·0912)	< 0·001
$\left(\frac{\text{Neck angle}+1}{10}\right)$ (°)	−0·0973 (0·0614)	0·113
Sex (M)	−0·4954 (0·2652)	0·062
Acute myocardial ischaemia	1·0429 (0·3684)	0·005

Values in parentheses are standard errors.

Calibration and discrimination

Calibration plots are shown in *Fig. 2* for the IMPROVE score applied to the IMPROVE data. The score was generally well calibrated, particularly for individuals at lower risk. The calibration slope was estimated as 1·08 (95 per cent c.i. 0·83 to 1·32), compatible with a value of 1, indicating good calibration.

The discriminative ability of the IMPROVE score was reasonable, but not exceptional, with a C-statistic of 0·710 when assessed in the IMPROVE trial (*Table 3*). When validated on the four external data sets, the predictive ability of

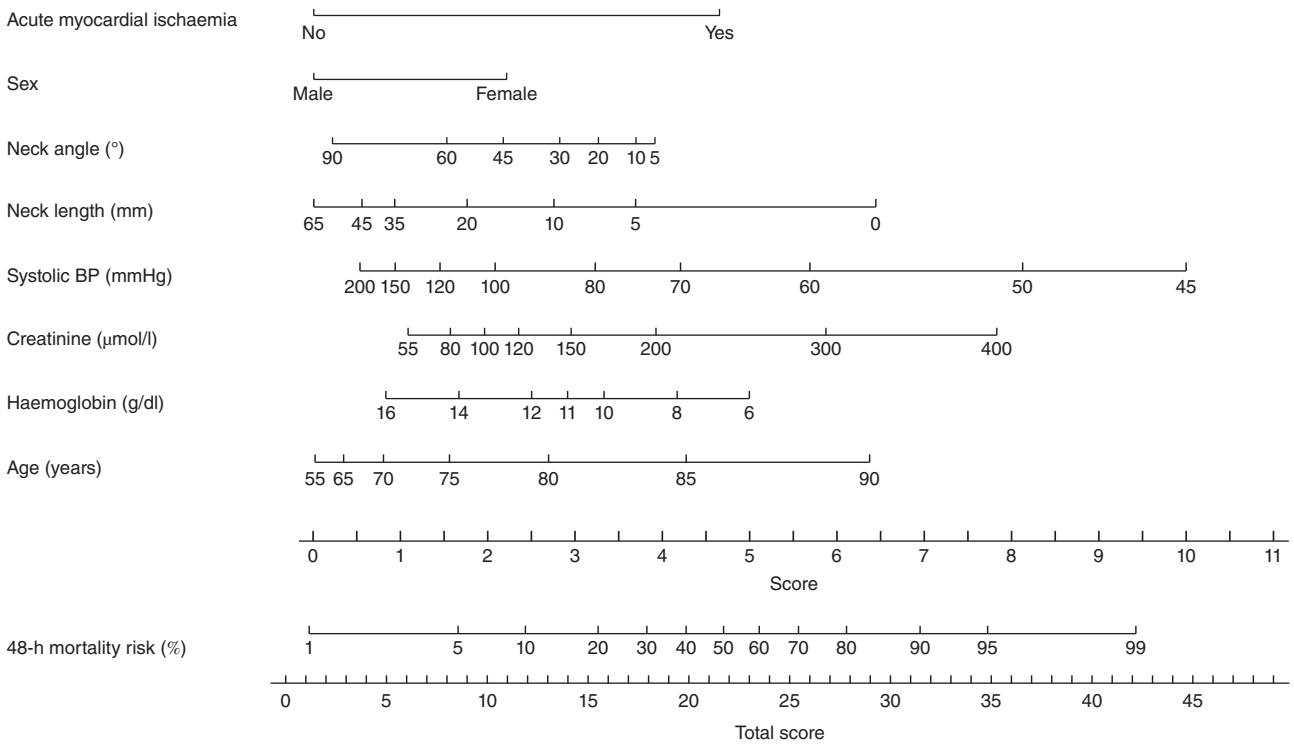

Fig. 1 Nomogram showing the IMPROVE risk score for 48-h mortality for patients with ruptured abdominal aortic aneurysm. To use this nomogram, each of the patient's characteristics is assessed and the associated score read off (upper part). The total score is obtained by summing the scores from each of the individual characteristics, and the predicted 48-h mortality risk can then be obtained (lower part)

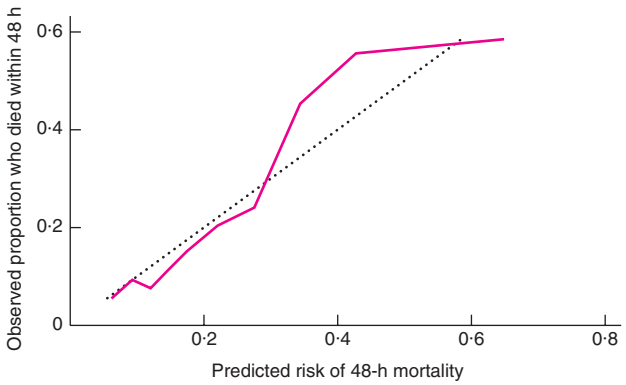

Fig. 2 Calibration plot for the IMPROVE risk score in the IMPROVE data set

the IMPROVE score was comparable to the performance seen in the IMPROVE trial, with C-statistic values ranging from 0·652 to 0·761.

There was no evidence that the IMPROVE risk score had better discriminative ability among patients who arrived directly compared with those transferred from another hospital (C-index 0·726 *versus* 0·697; $P = 0·579$).

Comparison with other published risk scores

In terms of discrimination, the IMPROVE score outperformed other published risk scores in the IMPROVE trial data (VSGNE, $P = 0·010$; Hardman, $P = 0·013$; Vancouver, $P = 0·013$) and in the Amsterdam cohort (VSGNE, $P < 0·001$; Hardman, $P < 0·001$; Vancouver, $P < 0·001$). All risk scores were comparable in the AJAX and ECAR trials, and in the STAR cohort. There was heterogeneity in the performance of each risk score when applied to different populations. Overall performance of each risk score in comparison to the predictive discrimination of using age alone is shown in *Fig. 3*, with the mean C-statistic increase pooled across cohorts and weighted by the study-specific number of events. The change in C-statistic was highest when the IMPROVE score was used, although the mean increase was still rather limited (change in C-statistic 0·08, 95 per cent c.i. 0·05 to 0·11). The performances of the risk scores in the STAR cohort were particularly modest, with only the Vancouver score showing a significant improvement in the C-statistic compared with age alone.

Table 3 Estimated C-statistics for the IMPROVE risk score and other published risk scores for predicting 48-h mortality, when applied to five different populations

	IMPROVE RCT ($n = 536$)	AJAX RCT ($n = 113$)	ECAR RCT ($n = 107$)	Amsterdam cohort ($n = 402$)	STAR cohort ($n = 284$)
IMPROVE score	0·710 (0·659, 0·760)*	0·680 (0·565, 0·796)	0·719 (0·570, 0·867)	0·761 (0·713, 0·810)	0·652 (0·586, 0·719)
VSGNE score	0·638 (0·585, 0·691)	0·634 (0·509, 0·764)†	0·674 (0·543, 0·805)†	0·640 (0·582, 0·697)‡	0·655 (0·590, 0·720)†
Hardman index	0·648 (0·597, 0·698)	0·754 (0·658, 0·850)†	0·731 (0·597, 0·865)†	0·675 (0·618, 0·732)†	0·606 (0·537, 0·674)†
Vancouver score	0·635 (0·579, 0·690)	0·609 (0·453, 0·764)†	0·725 (0·591, 0·860)†	0·654 (0·592, 0·715)‡	0·702 (0·638, 0·766)†

Values in parentheses are 95 per cent confidence intervals. *Cross-validated C-statistic (optimism-corrected); †excluding preoperative cardiac arrest, ‡excluding preoperative cardiac arrest and loss of consciousness. VSGNE, Vascular Study Group of New England.

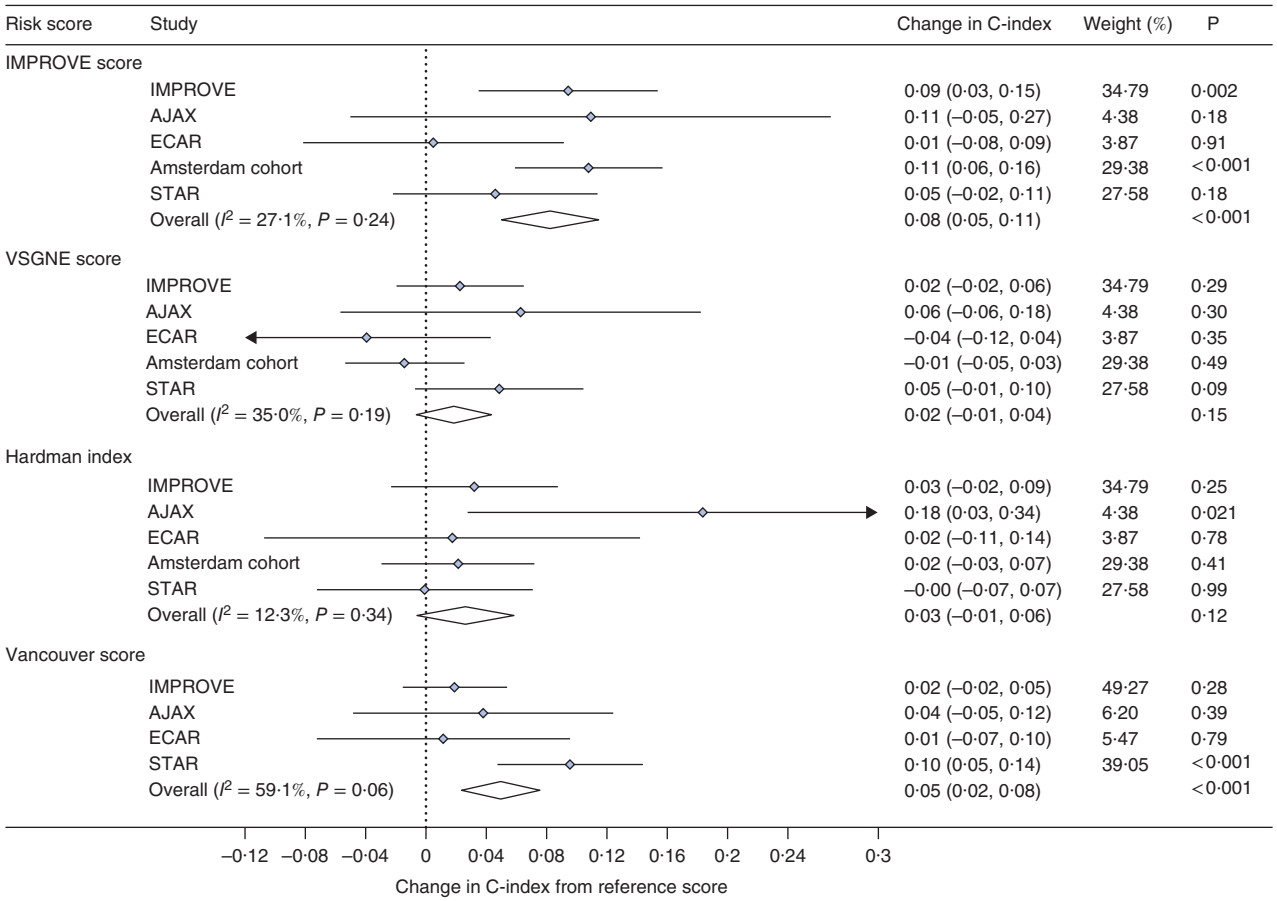

Fig. 3 Change in C-statistic for four ruptured abdominal aortic aneurysm risk scores compared with the reference score using age alone as a risk factor. The change in C-statistic was calculated in each cohort and for each risk score compared with using age alone. The changes were then pooled across cohorts. Changes in C-statistic are shown with 95 per cent confidence intervals. The Vancouver score in the Amsterdam cohort contained only the effect of age because cardiac arrest and loss of consciousness were not available; therefore, no comparisons with an age-alone model could be made

Performance of the risk scores in predicting 30-day and in-hospital mortality

The four different risk scores were also assessed for their ability to predict 30-day and (primary) in-hospital mortality (*Tables S3* and *S4*, supporting information). The predictive performance of the IMPROVE score was similar to that for prediction of 48-h mortality; in comparison with the 48-h prediction, the C-statistic increased in some studies and decreased in others for both 30-day and in-hospital mortality. For the three previously published risk scores (VSGNE, Hardman and Vancouver) the C-statistics for 30-day and in-hospital mortality were often higher than the

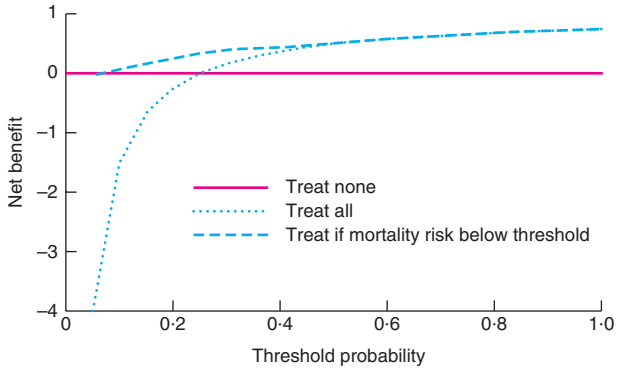

Fig. 4 Decision curve showing the benefit of the IMPROVE risk score in helping make treatment decisions. The benefit to risk trade-off inferred by a surgeon's chosen threshold probability of operating on a patient is shown on the *x*-axis. For example, a surgeon who would treat patients with a mortality probability of 98 per cent or less quantifies the consequence of not operating when it would have been of benefit as 98 to 2, that is 49 times worse than the consequence of operating unnecessarily. The dotted line shows the net benefit (relative to treating no one) of treating everyone as a function of the benefit to risk trade-off (chosen threshold). The dashed line shows the net benefit (relative to treating no one) of treating only those with a mortality risk below the chosen threshold

48-h C-statistic values in all studies except for the ECAR randomized trial, underlining the fact that these scores were developed using 30-day or in-hospital mortality as the outcome.

Decision curve analysis

The net benefit of using the IMPROVE score at a range of possible threshold probabilities to treat patients, in contrast to treating no one, is shown in *Fig. 4*. The net benefit of treating everyone is higher than that of treating no one when the consequences of not operating are deemed much worse than the consequences of operating unnecessarily (the threshold probability is high). At a threshold probability of 1·0, surgeons would value an operation as infinitely better than no operation, and so the treat-all scenario gives the highest net benefit. If treatment decisions are based on the IMPROVE rAAA risk score, then the net benefit is almost identical to that of treating everyone for risk thresholds above 50 per cent. The benefit of using an rAAA risk score in clinical decision-making only becomes apparent for surgeons who would only treat patients with a predicted 48-h mortality risk of 40 per cent or less. A higher predicted risk threshold of 90 per cent would correctly identify for surgery all 401 survivors in IMPROVE, but would also identify for surgery 133 of 135 patients (98·5 per cent) who died within 48 h.

Discussion

The value of the rAAA risk scores assessed here when applied to different populations was poor. The IMPROVE risk score was the only one to focus on 48-h mortality, more likely to be related to bleeding than multiple organ failure, and the only score to include morphological parameters. However, like all previous risk scores, it was derived from patients admitted to vascular surgery units. Moreover, it did not perform substantially better than older scores, which were based on demographic and haemodynamic variables only and focused on 30-day mortality. As any risk score should be generalizable, the Stockholm area cohort was particularly important, because it included patients with rAAA admitted to hospitals in the Stockholm area, not just those admitted to the vascular centres or who underwent repair.

In general, all available rAAA risk scores do not have sufficient accuracy to enable surgeons to use them to make life-or-death decisions regarding surgery, as highlighted in the decision curve analyses. This conclusion is in agreement with a recent study[10] that evaluated five different scoring systems for 30-day mortality on an historical cohort of Dutch patients with rAAA presenting to two vascular centres (outside the Amsterdam area). This study evaluated the Vancouver and Hardman scores, together with the Glasgow and Edinburgh scores and the newly developed Dutch Aneurysm Score[23], with C-statistic values ranging from 0·59 to 0·72. Vos and colleagues[10] also concluded that almost perfect prediction is needed to withhold intervention, and no current scoring system is capable of that. Nevertheless, operative mortality risk is not the only important consideration, and factors such as postoperative quality of life must also play a major role in the decision-making process.

The IMPROVE risk score has shown that there are some important predictors of survival to 48 h after emergency admission, including aneurysm neck length, which complicates emergency surgical repair by influencing the feasibility of aortic grafting. Some of these predictors were found to have a non-linear relationship with mortality. An inverse relationship between aortic neck angle and risk was identified, but this was of marginal significance in the multivariable model ($P = 0{\cdot}113$), and in *post hoc* analysis was evident only in those who had open repair. This finding may be explained by easier proximal clamping in open repair when the neck is angulated. The inverse relationship between neck angle and mortality is at odds with previous

observations[24], so this finding should be treated with caution. The IMPROVE score performed reasonably among patients in the IMPROVE trial, and outperformed other published risk scores in validation on some, but not all, external patient populations. However, some of the variables used to derive the published risk scores were not available in the data sets evaluated here, which may have led to the reduced risk score that was evaluated being suboptimal.

The findings reported here suggest that further research to assess the role of risk scores in the transfer of patients, from centres without emergency vascular cover to specialist vascular centres, would not be fruitful. The inability of risk scores to predict outcomes following emergency surgical repair, with sufficient accuracy for individual clinical decision-making, indicates that any risk score should be used only for comparing different populations of patients with rAAA or risk-adjusting such data. Which risk score to use for comparing the patients in different studies is likely to be a matter for debate, but a score that includes some morphological parameters is a strong contender.

There are a number of reasons why the mortality risk of patients with rAAA cannot be predicted accurately. First, there is heterogeneity in treatment and non-intervention rates, which could result in differences in risk score performance across populations. Second, if the decision to palliate is based on external factors, the effect of the risk score could be diluted. A limitation of all derived rAAA risk scores is that the outcomes of those palliated had they undergone repair instead will never be known, and the remaining patients who do undergo repair are therefore already preselected. To provide good predictive discrimination, variation in risk factors in the population needs to be high, and preselection may influence this. The IMPROVE risk score was developed in patients deemed suitable for rAAA repair at presentation and may therefore be biased towards more physically fit patients. Inclusion and exclusion criteria were applied in all the cohorts studied, except for the STAR cohort, which was the only data set that included all patients with rAAA in the population who were admitted to hospital. None of the risk scores performed well in this cohort.

Because the mortality risk of rAAA repair cannot be predicted with sufficient accuracy, the focus should shift to offering repair to more patients and on reducing non-intervention rates for rAAA repair, which may be too high in England[1]. It is of interest that, when a Delphi consensus approach was used to formulate guidelines for the transfer of patients to specialist vascular centres, cardiac arrest in the same admission was the only condition with complete agreement that the patients should not be transferred[2]. Yet, there is some evidence that a few patients (14 per cent) may survive emergency repair of rAAA even after a preoperative cardiac arrest[25], although the quality of life in survivors has not been documented. In such life-or-death situations, perhaps it is just not ethical to consider using risk scores to withhold treatment, and the focus should remain on ensuring that the wishes of the patient and their family are respected.

Collaborators

The Ruptured Aneurysm Collaborators are as follows.

AJAX Trial investigators: R. Balm, M. J. W. Koelemay, M. M. Idu, C. Kox, D. A. Legemate, L. C. Huisman, M. C. M. Willems, J. A. Reekers, O. M. van Delden, K. P. van Lienden (Academic Medical Centre, Amsterdam, The Netherlands; 39 patients); L. L. Hoornweg, J. J. Reimerink, S. C. van Beek (Academic Medical Centre, Amsterdam, The Netherlands; trial coordinators); A. C. Vahl, V. J. Leijdekkers, J. Bosma, A. D. Montauban van Swijndregt, C. de Vries, V. P. M. van der Hulst, J. Peringa, J. G. A. M. Blomjous, M. J. T. Visser, F. H. W. M. van der Heijden (Onze Lieve Vrouwe Gasthuis, Amsterdam, The Netherlands; 46 patients); W. Wisselink, A. W. J. Hoksbergen, J. D. Blankensteijn, M. T. J. Visser, H. M. E. Coveliers, J. H. Nederhoed, F. G. van den Berg, B. B. van der Meijs, M. L. P. van den Oever, R. J. Lely, M. R. Meijerink, I. Westra (VU University Medical Centre, Amsterdam, The Netherlands; 31 patients); A. Voorwinde, J. M. Ultee, R. C. van Nieuwenhuizen (Sint Lucas Andreas Ziekenhuis, Amsterdam, The Netherlands; referring centre); B. J. Dwars, T. O. M. Nagy (Slotervaartziekenhuis, Amsterdam, The Netherlands; referring centre); P. Tolenaar, A. M. Wiersema (Boven IJ Ziekenhuis, Amsterdam, The Netherlands; referring centre); J. A. Lawson, P. J. van Aken, A. A. Stigter (Ziekenhuis Amstelland, Amsterdam, The Netherlands; referring centre); T. A. A. van den Broek, G. A. Vos (Waterlandziekenhuis, Purmerend, The Netherlands; referring centre); W. Mulder, R. P. Strating (Zaans Medisch Centrum, Zaandam, The Netherlands; referring centre); D. Nio, G. J. M. Akkersdijk, A. van der Elst (Spaarne Ziekenhuis, Hoofddorp, The Netherlands); P. van Exter (regional ambulance services).

ECAR Trial investigators: P. Desgranges, J.-P. Becquemin, E. Allaire, F. Cochennec, J. Marzelle, N. Louis, J. Schneider, M. Majewski (Centre Hospitalier Universitaire (CHU) Henri Mondor, Créteil, France); Y. Castier, G. Leseche, F. Francis (CHU Bichat, Paris, France); E. Steinmetz, J.-P. Berne, C. Favier (CHU Dijon, Dijon, France); S. Haulon, M. Koussa, R. Azzaoui, D. Piervito (Centre Hospitalier Régional Universitaire (CHRU) Lille, Lille, France); Y. Alimi, M. Boufi, O. Hartung, P. Cerquetta (Hôpital Nord Marseille, Marseille, France); P. Amabile, P. Piquet, J. Penard, M. Demasi (CHU Marseille, Marseille, France); P. Alric, L. Cannaud, J.-P. Berthet (CHU Montpellier, Montpellier, France); P. Julia, J.-N. Fabiani, J. M. Alsac (CHU Hôpital Européen Georges-Pompidou, Paris, France); P. Gouny, A. Badra, J. Braesco (CHU Brest, Brest, France); J.-P. Favre, J.-N. Albertini (CHU Saint Etienne, Saint Etienne, France); R. Martinez (CHRU Tours, Tours, France); R. Hassen-Khodja, M. Batt, E. Jean, M. Sosa, S. Declemy (CHU Nice, Nice, France); L. Destrieux-Garnier (Centre Hospitalier Régional Annecy, Annecy, France); P. Lermusiaux, P. Feugier (CHU Lyon, Lyon, France).

IMPROVE Trial investigators: management committee – J. T. Powell (Chair; Imperial College London, London, UK); R. Ashleigh (University Hospital of South Manchester, Manchester, UK), M. Gomes (London School of Hygiene and Tropical Medicine, London, UK), R. M. Greenhalgh (Imperial College London, London, UK), R. Grieve (London School of Hygiene and Tropical Medicine, London, UK), R. Hinchliffe (St George's Hospital, London, UK), M. Sweeting (University of Cambridge,

Cambridge, UK), M. M. Thompson (St George's Hospital, London, UK), S. G. Thompson (University of Cambridge, Cambridge, UK), P. Ulug (Imperial College London, London, UK). N. J. Cheshire (Imperial College Healthcare NHS Trust, London, UK; 20 patients); J. R. Boyle (Addenbrooke's Hospital, Cambridge, UK; 40 patients); F. Serracino-Inglott, J. V. Smyth (Manchester Royal Infirmary, Manchester, UK; 69 patients); M. M. Thompson, R. J. Hinchliffe (St George's Hospital, London, UK; 75 patients); R. Bell (Guy's and St Thomas' Hospital, London, UK; 81 patients); N. Wilson (Kent and Canterbury Hospital, Canterbury, UK; 23 patients); M. Bown, M. Dennis (Leicester Royal Infirmary, Leicester, UK; 18 patients); M. Davis (Royal Free Hospital, London, UK; 1 patient); R. Ashleigh (University Hospital of South Manchester, Manchester, UK; 21 patients); S. Howell (Leeds General Infirmary, Leeds, UK; 23 patients); M. G. Wyatt (Freeman Hospital, Newcastle upon Tyne, UK; 23 patients); D. Valenti (King's College Hospital, London, UK; 2 patients); P. Bachoo (Aberdeen Royal Infirmary, Aberdeen, UK; 4 patients); P. Walker (James Cook University Hospital, Middlesbrough, UK; 5 patients); S. MacSweeney (Queen's Medical Centre, Nottingham, UK; 34 patients); J. N. Davies (Royal Cornwall Hospital, Truro, UK; 5 patients); D. Rittoo, S. D. Parvin (Royal Bournemouth Hospital, Bournemouth, UK; 22 patients); W. Yusuf (Royal Sussex County Hospital, Brighton, UK; 5 patients); C. Nice (Queen Elizabeth Hospital, Gateshead, UK; 5 patients); I. Chetter (Hull Royal Infirmary, Hull, UK; 32 patients); A. Howard (Colchester General Hospital, Colchester, UK; 24 patients); P. Chong (Frimley Park Hospital, Frimley, UK; 14 patients); R. Bhat (Ninewells Hospital, Dundee, UK; 8 patients); A. Gordon, I. Lane (University Hospital of Wales, Cardiff, UK; 4 patients); S. Hobbs (New Cross Hospital, Wolverhampton, UK; 3 patients); W. Pillay (Doncaster Royal Infirmary, Doncaster, UK; 8 patients); T. Rowlands, A. El-Tahir (Royal Derby Hospital, Derby, UK; 13 patients); J. Asquith (University Hospital of North Staffordshire, Stoke-on-Trent, UK; 15 patients); S. Cavanagh (York Hospital, York, UK; 3 patients); L. Dubois, T. L. Forbes (London Health Sciences Centre, University of Western Ontario, London, Ontario, Canada; 13 patients).

STAR Cohort investigators: R. Hultgren, J. Roy, S. Zommorodi, M. Lindquist Liljeqvist, A. Siika, O. Nilsson, A. Garcia Värild (Karolinska Institute, Stockholm, Sweden).

Acknowledgements

This project was funded by the UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) programme (project number 07/37/64). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, National Health Service or the Department of Health.

Disclosure: The authors declare no conflict of interest.

References

1 Karthikesalingam A, Holt PJ, Vidal-Diez A, Ozdemir BA, Poloniecki JD, Hinchliffe RJ *et al.* Mortality from ruptured abdominal aortic aneurysms: clinical lessons from a comparison of outcomes in England and the USA. *Lancet* 2014; **383**: 963–969.

2 Hinchliffe RJ, Ribbons T, Ulug P, Powell JT. Transfer of patients with ruptured abdominal aortic aneurysm from general hospitals to specialist vascular centres: results of a Delphi consensus study. *Emerg Med J* 2013; **30**: 483–486.

3 Mell MW, Schneider PA, Starnes BW. Variability in transfer criteria for patients with ruptured abdominal aortic aneurysm in the western United States. *J Vasc Surg* 2015; **62**: 326–330.

4 Robinson WP, Schanzer A, Li Y, Goodney PP, Nolan BW, Eslami MH *et al.* Derivation and validation of a practical risk score for prediction of mortality after open repair of ruptured abdominal aortic aneurysms in a US regional cohort and comparison to existing scoring systems. *J Vasc Surg* 2013; **57**: 354–361.

5 Hardman DT, Fisher CM, Patel MI, Neale M, Chambers J, Lane R *et al.* Ruptured abdominal aortic aneurysms: who should be offered surgery? *J Vasc Surg* 1996; **23**: 123–129.

6 Chen JC, Hildebrand HD, Salvian AJ, Taylor DC, Strandberg S, Myckatyn TM *et al.* Predictors of death in nonruptured and ruptured abdominal aortic aneurysms. *J Vasc Surg* 1996; **24**: 614–620.

7 Srinivasan A, Ambler GK, Hayes PD, Chowdhury MM, Ashcroft S, Boyle JR *et al.* Premorbid function, comorbidity, and frailty predict outcomes after ruptured abdominal aortic aneurysm repair. *J Vasc Surg* 2016; **63**: 603–609.

8 IMPROVE Trial Investigators. Endovascular or open repair strategy for ruptured abdominal aortic aneurysm: 30 day outcomes from IMPROVE randomised trial. *BMJ* 2014; **348**: f7661.

9 van Beek SC, Reimerink JJ, Vahl AC, Wisselink W, Peters RJ, Legemate DA *et al.* Editor's choice – external validation of models predicting survival after ruptured abdominal aortic aneurysm repair. *Eur J Vasc Endovasc Surg* 2015; **49**: 10–16.

10 Vos CG, de Vries JP, Werson DA, van Dongen EP, Schreve MA, Unlu C. Evaluation of five different aneurysm scoring systems to predict mortality in ruptured abdominal aortic aneurysm patients. *J Vasc Surg* 2016; **64**: 1609–1616.

11 Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009; **338**: b605.

12 Dick F, Diehm N, Opfermann P, von Allmen R, Tevaearai H, Schmidli J. Endovascular suitability and outcome after open surgery for ruptured abdominal aortic aneurysm. *Br J Surg* 2012; **99**: 940–947.

13 IMPROVE Trial Investigators. The effect of aortic morphology on peri-operative mortality of ruptured abdominal aortic aneurysm. *Eur Heart J* 2015; **36**: 1328–1334.

14 Reimerink JJ, Hoornweg LL, Vahl AC, Wisselink W, van den Broek TA, Legemate DA *et al.* Endovascular repair *versus* open repair of ruptured abdominal aortic aneurysms: a multicenter randomized controlled trial. *Ann Surg* 2013; **258**: 248–256.

15 Desgranges P, Kobeiter H, Katsahian S, Bouffi M, Gouny P, Favre JP *et al.* Editor's choice – ECAR (Endovasculaire ou Chirurgie dans les Anevrysmes aorto-iliaques Rompus):

A French randomized controlled trial of endovascular *versus* open surgical repair of ruptured aorto-iliac aneurysms. *Eur J Vasc Endovasc Surg* 2015; **50**: 303–310.
16 van Beek SC, Reimerink JJ, Vahl AC, Wisselink W, Reekers JA, van Geloven N *et al.* Effect of regional cooperation on outcomes from ruptured abdominal aortic aneurysm. *Br J Surg* 2014; **101**: 794–801.
17 Hultgren R, Zommorodi S, Gambe M, Roy J. A majority of admitted patients with ruptured abdominal aortic aneurysm undergo and survive corrective treatment: a population-based retrospective cohort study. *World J Surg* 2016; **40**: 3080–3087.
18 Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999; **28**: 964–974.
19 Zlotnik A, Abraira V. A general-purpose nomogram generator for predictive logistic regression models. *Stata Journal* 2015; **15**: 537–546.
20 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35**: 1925–1931.
21 Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010; **21**: 128–138.
22 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006; **26**: 565–574.
23 von Meijenfeldt GC, van Beek SC, Bastos Goncalves F, Verhagen HJ, Zeebregts CJ, Vahl AC *et al.* Development and external validation of a model predicting death after surgery in patients with a ruptured abdominal aortic aneurysm: the Dutch Aneurysm Score. *Eur J Vasc Endovasc Surg* 2017; **53**: 168–174.
24 Sternbergh WC III, Carter G, York JW, Yoselevitz M, Money SR. Aortic neck angulation predicts adverse outcome with endovascular abdominal aortic aneurysm repair. *J Vasc Surg* 2002; **35**: 482–486.
25 Harris DG, Garrido D, Oates CP, Kalsi R, Huffner ME, Toursavadkohi S *et al.* Repair of ruptured abdominal aortic aneurysm after cardiac arrest. *J Vasc Surg* 2016; **64**: 1497–1502.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.

Editor's comments

Is this the definitive word on preoperative risk scores ruptured AAA? In my opinion, yes. This elegant and comprehensive research shows convincingly that in this dire circumstance the computer cannot assist and decisions must be made by surgeons, taking into account their patient's best interests. Perhaps reassuringly, artificial intelligence may not conquer all aspects of medicine and surgery.

J. J. Earnshaw
Editor-in-Chief, BJS