Optimal prediction of mortality after abdominal aortic aneurysm repair with statistical models

Vassilis Georgiou Hadjianastassiou, MRCS,[a] Leonardo Franco, PhD,[b] Jose M. Jerez, PhD,[b] Iordanis E. Evangelou, DPhil,[c] David R. Goldhill, FRCA,[d] Paris P. Tekkis, FRCS,[e] and Linda J. Hands, FRCS,[a] *Oxford, United Kingdom; Málaga, Spain; Bethesda, Md; and London, United Kingdom*

Objective: To identify the best method for the prediction of postoperative mortality in individual abdominal aortic aneurysm surgery (AAA) patients by comparing statistical modelling with artificial neural networks' (ANN) and clinicians' estimates.
Methods: An observational multicenter study was conducted of prospectively collected postoperative Acute Physiology and Chronic Health Evaluation II data for a 9-year period from 24 intensive care units (ICU) in the Thames region of the United Kingdom. The study cohort consisted of 1205 elective and 546 emergency AAA patients. Four independent physiologic variables—age, acute physiology score, emergency operation, and chronic health evaluation—were used to develop multiple regression and ANN models to predict in-hospital mortality. The models were developed on 75% of the patient population and their validity tested on the remaining 25%. The results from these two models were compared with the observed outcome and clinicians' estimates by using measures of calibration, discrimination, and subgroup analysis.
Results: Observed in-hospital mortality for elective surgery was 9.3% (95% confidence interval [CI], 7.7% to 11.1%) and for emergency surgery, 46.7% (95% CI, 42.5 to 51.0%). The ANN and the statistical models were both more accurate than the clinicians' predictions. Only the statistical model was internally valid, however, when applied to the validation set of observations, as evidenced by calibration (Hosmer-Lemeshow C statistic, 14.97; $P = .060$), discrimination properties (area under receiver operating characteristic curve, 0.869; 95% CI, 0.824 to 0.913), and subgroup analysis.
Conclusions: The prediction of in-hospital mortality in AAA patients by multiple regression is more accurate than clinicians' estimates or ANN modelling. Clinicians can use this statistical model as an objective adjunct to generate informed prognosis. (J Vasc Surg 2006;43:467-73.)

Mortality prediction models are increasingly being used in health care in various roles,[1] including:

1. in research to assess whether groups are similar in terms of underlying case mix,

2. in the audit of clinical performance of an institution or an individual, by comparing the actual with the predicted performance,[2]

3. as a quantitative surrogate measure, summarizing a patient's clinical status, to aid exchange of information between clinicians,

4. as an adjunct in the process of informed consent or to guide clinicians, patients, and their relatives in the prognosis of a patient; and,

5. the most controversial potential role[3] of influencing individual clinical management by acting as an adjunct to clinical judgement rather than replacing it.

Abdominal aortic aneurysms (AAA) form a significant part of the workload of vascular surgeons, and these patients are routinely managed postoperatively in high dependency units or intensive care unit (ICU) facilities. Predictive models have been shown to be as accurate as clinicians in predicting mortality outcome in the ICU.[4] Various risk prediction systems have been described for general ICU use, with the most extensively studied one being the Acute Physiology and Chronic Health Evaluation (APACHE-II) system.[5] The APACHE II score was significantly higher[6] in patients who died early after surgery for ruptured AAA compared with late deaths and survivors.

The Physiological and Operative Severity Score for enUmeration of Morbidity and mortality (POSSUM) methodology and its Portsmouth modification (P-POSSUM)

have been shown to be accurate risk-stratification methods for general surgical patients.[7,8] In addition, the methodology[9] successfully predicted both morbidity and mortality after arterial surgery. The P-POSSUM and vascular POSSUM (V-POSSUM) models[10] accurately predicted mortality for elective AAA surgery but did not predict outcome for both elective and emergency AAA surgery as a combined group, even when operative urgency was included in the predictor variables.

A similar attempt to model the outcome of both elective and emergency AAA again necessitated the formation of two separate models.[11] In a group of 40 patients with a ruptured AAA,[12] the APACHE II score was found to be a better predictor of outcome than the POSSUM score. The recently developed APACHE-AAA model[13] successfully predicted in-hospital mortality in a combined group of elective and emergency aneurysms by using APACHE-based physiologic variables prospectively collected in the immediate postoperative period.

Artificial neural networks (ANNs) are computational, mathematic tools for information processing with structures inspired by the biologic nervous system[14,15] (Appendix I, online only). ANNs have been applied in clinical medicine[16] and in the pathology department[17] by using their ability to analyze nonlinear data, with large "noise" and higher order interactions between the data. It has been suggested that the complex nature of critically ill patients in the ICU, with their potential multiple interacting systemic disturbances,[14] make them an ideal scenario for outcome prediction by ANNs. Outcome in ICU patients was predicted more accurately by ANNs than both linear (conventional logistic regression) and the more complex nonlinear (correlation and regression trees) statistical models.[18] ANNs also predicted outcome in colorectal cancer patients more accurately than existing clinicopathologic staging systems and clinicians' estimates.[19] In vascular surgery, ANNs successfully predicted perioperative cardiac complications with better calibration properties than comparable logistic regression models.[20] Furthermore, ANNs successfully predicted outcome in 83% of patients from approximately 100 ruptured AAAs by using four physiologic variables selected by logistic regression analysis.[21] In contrast, other published studies[22] have suggested that traditional statistical analyses outperformed ANNs.

The purpose of this study was to compare the prognostic ability of multiple regression modelling with that of an ANN in a group of postoperative AAA patients and contrast these model predictions of in-hospital mortality with the clinicians' estimates of prognosis.

METHODS

Data sources. In the northeast Thames region of the United Kingdom (UK), a group of 24 ICUs (from 5 university/teaching and 19 community/district general hospitals) contribute information to a common database[23] for audit purposes. All ICUs admit both medical and surgical patients, and the methods and procedures of prospective data collection, validation, and maintenance of the

database have been described before.[23-26] The information collected for each patient includes demographic data, diagnostic details, and values for the components of the APACHE-II system. The study period was March 1992 to December 2000.

Inclusion and exclusion criteria. Data from all patients who underwent elective or emergency open surgical repair of AAAs and were managed postoperatively in an ICU were included in the study. Excluded from the analysis were patients transferred postoperatively from other hospitals and those with missing data for the independent predictors or the study outcome.

Study outcome and prognostic variables. The primary outcome of the study was in-hospital mortality, consistent with the APACHE-II methodology. All candidate prognostic variables (risk factors) for in-hospital mortality were chosen from the components of the APACHE-II model: (1) the Acute Physiology Score (APS) and (2) the Chronic Health (CH) status, classified as a binary variable according to whether or not the patient had any chronic health dysfunction; that is, a history of severe organ system insufficiency as defined in the original APACHE-II study[5]; (3) operative urgency classified with the help of the National Confidential Enquiry into Patient Outcome and Death (NCEPOD) classification[27] of operations as an emergency (ruptured, leaking and symptomatic classified by NCEPOD as emergency/urgent) or elective (NCEPOD scheduled/elective) surgical procedure, and (4) chronologic age. The values of these variables were the first collected as close as possible to the end of the operation, either in the operating theater or the first recorded values on admission to ICU (the latter for information on biochemical parameters). This differs from the APACHE II methodology, which is based upon the worst values in the first 24 hours of ICU admission.

Statistical selection of variables for the predictive model. Univariate logistic regression was performed to identify risk factors associated with in-hospital mortality. The variables whose univariate test had a $P < .25$ were considered as candidates for the multivariable model. Multiple regression analysis[28] with backwards, stepwise variable selection was then used to identify independent risk factors for in-hospital mortality. To avoid "over-fitting" the model, a nonparametric bootstrap resampling technique[29] with 10,000 iterations was used to calculate standard errors and to correct bias in the parameter estimation.

Artificial neural network model. The variables selected by logistic regression analysis as independent predictors were used as input values to the ANN. Continuous data were preprocessed by scaling them to a range between zero and one. Standard ANN methodology was used involving the multilayer perceptron and the back-propagation learning paradigm.[15] The patient records were randomly split into training, cross-validation, and test sets in the ratios of 50:25:25. To limit bias in the generation of the subsets, the process of randomization into these sets was repeated after first ranking the patients into 10 groups of ascending probability of death as predicted by logistic

Table I. Patient demographic characteristics and associated mortality

Variable	No. Patients (% total)	No. Deaths (% mortality)	Unadjusted OR*	95% CI*
Age (per year)			1.05	1.03-1.07
APS (per unit)			1.18	1.16-1.21
Gender				
Female	286 (16.3)	58 (20.3)	1	
Male	1465 (83.7)	309 (21.1)	1.05	0.77-1.44
Operative urgency				
Elective	1205 (68.8)	112 (9.3)	1	
Emergency	546 (31.2)	255 (46.7)	8.55	6.61-11.06
CH status				
No CH dysfunction	1331 (76.0)	251 (18.9)	1	
CH dysfunction	420 (24.0)	116 (27.6)	1.64	1.27-2.12
Total	1751	367 (21.0)		

OR, Odds ratio; *CI,* confidence interval; *APS,* Acute Physiology Score; *CH,* Chronic Health.
*ORs and 95% CIs calculated using logistic regression analysis and compared to the reference category with an OR of 1.

regression. This ensured the equal representation of patients from all levels of risk in each of the three sets.

Different numbers of processing nodes were used in one and two hidden layers to obtain the ANN architecture with the best performance. Each of the ANN architectures was tested on 10 randomly generated sets to get an estimate of the effect of randomization on the efficacy of prediction. To avoid overtraining[15] the network, the training process was stopped when the error function of the cross-validation set started increasing. The learning step size was set at 0.0325 and the momentum at 0.35 with a sigmoid transfer function and 15 processing nodes in one hidden layer.

Statistical predictive model. The ANN model with the best performance was selected for comparison with the statistical model. The input values in the training (50% of the total) and cross-validation (25% of the total) sets of this ANN model were used as the development set of the statistical model, and the ANN test set (25% of the total) was used as the statistical validation set, thus using a random 75:25 split-sample technique[30] for controlling the generalization error.

Clinicians' predictions. The resident clinicians (all trainees in their second to their sixth year of training) responsible for the postoperative admission of patients to the ICU were routinely asked to prospectively rate the patients' chances of survival, on admission to ICU, as follows: 1, expected to live; 2, likely to live; 3, even chance; 4, likely to die; and 5, expected to die; equivalent to the following approximate risks of death, respectively: 0%, 25%, 50%, 75%, and 100%.

Model evaluation. The internal validation of the predictive models was evaluated by measures of calibration and discrimination on the test (validation) set as well as subgroup analysis. Calibration or goodness-of-fit[28] refers to the ability of the model to assign the correct probabilities of outcome to individual patients. This ability was assessed using the Hosmer-Lemeshow *C* statistic[31] in which a high *P* value would indicate a good model fit. Model discrimination refers to the ability of the model to assign higher probabilities of death (outcome) to patients who actually die than those patients who live. This was evaluated by the area under the receiver operating characteristic (ROC) curve.[31]

Software. Analysis was performed using the computer software SPSS version 9 (SPSS, Chicago, Ill), Intercooled STATA 8.0 (STATA Corp, College Station, Tex), both for Windows (Microsoft, Redmond, Wash); and Neurosolutions version 4.20 (NeuroDimension, Inc, Gainesville, Fla) for Excel (Microsoft).

RESULTS

We identified 1972 patients who had undergone open AAA repair. Excluded were 40 patients who were transferred to the ICU from other hospitals, 36 patients owing to lack of recorded outcome (death), 143 patients with missing chronic health status, and 2 patients with missing records for operative urgency. Fifteen of the emergency operations were classified as "urgent" (not ruptured or leaked). Analysis of the missing CH status data did not reveal a statistically significant bias of distribution of missing values among the categories of operative urgency and their associated mortality.

The patient demographic characteristics and associated, unadjusted, in-hospital mortality risk for the various patient related risk factors are summarized in Table I. In-hospital operative mortality for this specific ICU subgroup of patients was 9.3% for elective surgery (95% confidence interval [CI], 7.7% to 11.1%) and 46.7% for emergency surgery (95% CI, 42.5% to 51.0%). Gender was not found to be a significant predictor of outcome (*P* = .758) using univariate logistic regression analysis and was therefore excluded from the multiple regression model.

There were no significant changes over the 9 years of the study in the ratio of emergency to elective operations (χ^2 test for trend, *P* = .818) or in the mortality rate (χ^2 test for trend, *P* = .505). The year of operation was not a significant predictor of outcome in univariate logistic regression analysis or after adjustment for other predictors (*P* = .828).

Table II. Calibration and discrimination properties of the predictive models

	Area Under ROC* (95% CI)	P†	*H-L C Statistic**	P†
Development set (n = 1313)		.116 (LR vs ANN)		
LR	0.828 (0.800-0.856)	0.116	4.35	.824
ANN	0.831 (0.803-0.859)	.038 (vs ANN)	32.80	.000
Clinicians	0.800 (0.769-0.832)	.055 (vs LR)	7.87	.020
Validation set (n = 438)		.490 (LR vs ANN)		
LR[4]	0.869 (0.824-0.913)	0.49	14.97	.060
ANN[5]	0.870 (0.826-0.914)	.008 (vs ANN)	25.69	.001
Clinicians	0.816 (0.786-0.845)	.011 (vs LR)	1677.69	.000

CI, confidence interval; *LR,* logistic regression statistical model; *ANN,* artificial neural network model.

*χ^2 test.

†Receiver Operating Characteristic curve *(ROC)* is a measure of the discrimination, and the Hosmer-Lemeshow *(H-L) C* statistic is a measure of calibration.

The results of the multiple regression analysis are shown in Appendix III (online only), after adjusting for case-mix. The four risk factors of age, Acute Physiology Score, Chronic Health status, and operative urgency were found to be independent predictors of outcome.

The 10 random 75:25 split-samples of the study population were used to develop the ANN with the different architectures. The mean ROC area (0.840, SD 0.001) of the ANN with a single hidden layer of processing nodes was the same as the result of the ANN with two hidden layers. Ranking the patients according to predicted risk before randomization into the split samples did not yield a significant difference to the models compared with simple randomization alone (without prior ranking). The best ANN architecture (single hidden layer of 15 processing nodes) was selected for comparison with the other methods of analysis. The 10 randomly generated 75:25 split-sample sets for this architecture produced a mean ROC area of 0.842 (SD 0.014).

The relative internal validity of the models can be found in Table II, which lists the summary calibration and discrimination statistics of the predictive models compared with the clinicians' predictions for the single random 75:25 split used in this study. The calibration properties of the multiple regression model were better than those of the ANN or the clinicians. In addition, it is also apparent that the discrimination properties of the clinicians' model fared worse than the other two models. The bootstrap technique used in the study to minimize the generalization error did not produce a significantly different statistical model compared with the split-sample technique.

The calibration properties of the predictive models in comparison with the clinicians' predictions, as applied to the validation set (438 observations, 25% of the total), are shown in Fig 1. The corresponding calibration graph for the development set (1313 observations, 75% of the total) is depicted in Appendix IV (online only).

As a further measure of the internal validity of the models, subgroup analysis was performed (Fig 2) that contrasts the predictive models with the clinicians' predictions with respect to the Chronic Health status and the urgency of the operation. It is evident from the graphs that the clinicians tended to underestimate the risk of high-risk

cases, such as the emergencies and the patients with significant comorbidity. The ANN predictions tended to overestimate the risk of low-risk cases and underestimate the risk of the high-risk patients. In contrast, the multiple regression model had the best internal validity, as evidenced by both the uniformly good fit of the model in the calibration graphs and in subgroup analysis the model's predictions rested within the 95% CI of the observed mortality across both the categories of operative urgency and Chronic Health status.
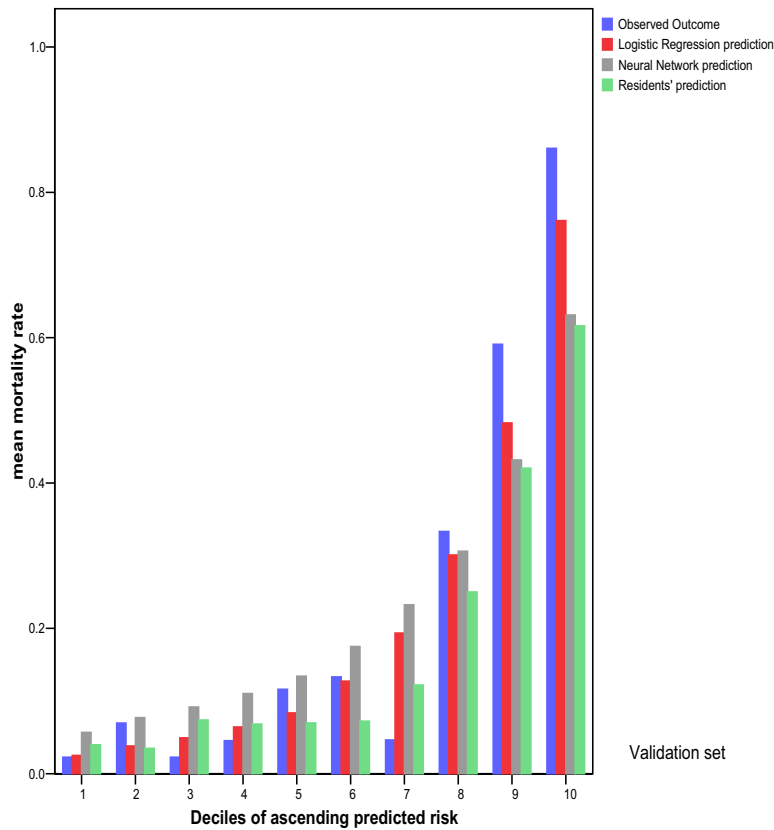
DISCUSSION

The study compares and contrasts two prognostic models, based on conventional logistic regression and ANN methodologies, in relation to clinicians' estimates of the mortality risk following AAA surgery and ICU admission.

Data validity. The mortality rates for both elective and emergency cases are consistent with the literature.[32] The finding that gender was not a significant predictor of outcome is also consistent with previously published UK work.[33,34] The study included data collected during a 9-year period with the risk of temporal changes in patterns of surgical practice, but the year of operation was not found to be a significant predictor of outcome.

Validation of the database[25] used and the proportion of missing data in the study (9%) compares well with rates in similar studies, such as the original APACHE II study[5] (13% missing data), the UK APACHE II study[35] (20% missing data), and the recent Vascular Biochemistry and Haematology Outcome Model for vascular surgery from the National Vascular Database[11] (36% missing data). In addition, validation of the database[26] has shown that inaccuracies recorded in physiologic values did not confer a statistically significant difference to the mortality ratio.

The data were collected from a particular region of the UK, not subject to random selection, so inferences about the applicability of the models elsewhere cannot be made until the models are externally validated in another region.

Statistical methodology. The statistical model was developed using multiple logistic regression analysis. Recent work[13] on this database has not demonstrated a significant variation in outcome between ICU units, which
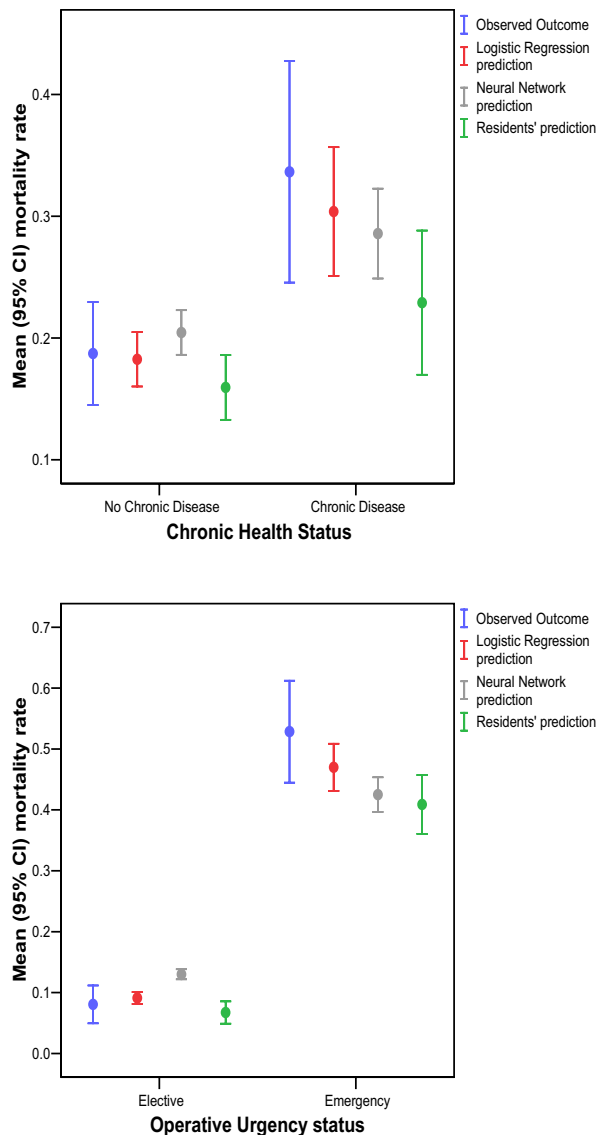
Fig 1. Calibration chart for the predictive models as applied to the validation set of observations. Bars represent the observed and predicted in-hospital percent mortality risk.

has negated the use of a hierarchical model[36] as there was no clustering of patients among the different ICU units. The bootstrap resampling methodology for minimizing the generalization error did not significantly improve the model's performance compared with the split-sample technique, confirming that the latter technique did not introduce a sampling bias.

Artificial neural network methodology. The model was developed using a relatively simple architecture with connections between the processing nodes in series and conventional back-propagation learning. The inclusion of a second hidden layer of nodes did not improve the model's performance. Provision was made to avoid over-fitting the network to the training data by using a cross-validation set to minimize the generalization error when the model was applied to a test set. Furthermore, ranking the patients in 10 groups of progressively increasing mortality risk before randomization into the training, cross-validation, and test sets, ensured that the randomization process would not introduce a sampling bias in the sets. In practice, this has not proven to be better than simply randomizing the nonranked group of patients into the three sets, consistent with the large number of outcome events in the database. Finally, repeating the randomization process 10 times has again confirmed that the sampling error was minimal.

Application of the model and future studies. The models presented have kept the basic physiologic principles of the APACHE methodology in risk-stratifying patients, reflecting the patients' physiologic reserve and comorbid state, the severity of illness, and the surgical status of the patients. The values of these variables were collected immediately after the AAA operation to represent the case-mix of the patients at the end of the operation before ICU care has had any influence. Although the POSSUM methodology has been applied in general surgery,[7,8] including arterial surgery,[9,10] the APACHE-based methodology presented in this study is more specific to the critical care setting and, hence, may be best suited as a tool to quantify prognosis in postoperative AAA patients.

This study has shown that prognostic modelling techniques can be more accurate than trainee clinicians in quantifying prognosis. These models can thus be invaluable tools for ICU trainee clinicians, supplementing their clinical judgement and empowering them to provide an "informed prognosis." Although the two models had similar discrimination properties, the ANN was not as accurate in predicting outcome in the widely differing underlying risk separating elective from emergency patients, as shown in the calibration and subgroup analyses.

Fig 2. Subgroup analysis by Chronic Health and operative urgency status for the predictive models (validation set). Error bars represent the 95% confidence interval of the in-hospital percent mortality risk.

In the group of patients under study, statistical modelling was the most accurate predictor of outcome at all levels of risk, making it the model of choice for predicting outcome for the individual patient. However, an error rate of approximately 15% in ICU prognostic models[37] precludes models from being used as the sole determinants of clinical decision-making. There is evidence[38] to support the use of predictive models to decrease uncertainty and promote communication among clinicians, patients, and patients' families. Because prognostic models are derived from data larger than any one clinician's experience, they can be more credible and reproducible than an individual clinician's estimates.[39]

CONCLUSION

The postoperative model presented here would be an ideal tool for clinicians who wish to quantify the prognosis of a patient in response to the question, "How did the operation go, Doctor?" In contrast, a prognostic model using the worst values in the first 24 hours after the operation would not be in a position to answer this question until after the first 24 hours have passed.

A limitation of the study is the fact that the models were compared with predictions generated by trainee clinicians, and evidence suggests that more experienced physicians may be better able to assess risk,[4] although the difference did not reach statistical significance. This difference may be a reason why these models may be of more use to resident clinicians than more experienced physicians who may be more accurate in their predictions. A recent report[40] from the UK revealed that there is low involvement of experienced clinicians in the first 24 hours of a patient's admission to ICU. This supports the idea that the resident clinicians are the frontline staff with the most contact with patients and their relatives in the first 24 hours of ICU admission and, therefore, would be the most likely physician group to find these models useful. A Bayesian approach[4,41] of combining the risk estimates of physicians (either trainees or specialists) with the model predictions may form the basis of a future study to improve the prognostic accuracy for the individual patient. It would be important to note that the population under study did not include any AAAs treated by endovascular means, although by 1999, only 5.8% of all AAA repairs in the UK were performed in that way.[42]

In this era of accountability and evolving professional responsibility for trust in the doctor-patient relationship, the concept of "informed prognosis" should become an integral component of healthcare delivery by grasping the opportunities offered by predictive modelling techniques.

AUTHOR CONTRIBUTIONS

Conception and design: V.G.H, LF, I.E.E, D.R.G, P.P.T, L.J.H
Analysis and interpretation: V.G.H, LF, J.M.J
Data collection: V.G.H, I.E.E, D.R.G
Writing the article: V.G.H
Critical revision of the article: V.G.H, LF, J.M.J, I.E.E, D.R.G, P.P.T, L.J.H
Final approval of the article: V.G.H, LF, J.M.J, I.E.E, D.R.G, P.P.T, L.J.H
Statistical analysis: V.G.H, LF, J.M.J, I.E.E, P.P.T
Overall responsibility: V.G.H

REFERENCES

1. Gunning K, Rowan K. Outcome data and scoring systems in ABC of intensive care. BMJ 1999;319:241-4.
2. Hadjianastassiou VG, Tekkis PP, Poloniecki JD, Gavalas M, Goldhill DR. The surgical mortality score. A risk management tool for auditing surgical performance. World J Surg 2004;28:193-200.

3. Schuster DP. Predicting outcome after ICU admission. The art and science of assessing risk. Chest 1992;102:1861-70.

4. Marcin JP, Pollack MM, Patel KM, Ruttimann UE. Combining physicians' subjective and physiology-based objective mortality risk predictions. Crit Care Med 2000;28:2984-90.

5. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13:818-29.

6. Maziak DE, Lindsay TF, Marshall JC, Walker PM. The impact of multiple organ dysfunction on mortality following ruptured abdominal aortic aneurysm repair. Ann Vasc Surg 1998;12:93-100.

7. Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Br J Surg 1998;85:1217-20.

8. Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg 1991;78:356-60.

9. Prytherch DR, Ridler BF, Beard JD, Earnshaw JJ. Audit and Research Committee; The Vascular Surgical Society of Great Britain and Ireland. A model for national outcome audit in vascular surgery. Eur J Vasc Endovasc Surg 2001;21:477-83.

10. Prytherch DR, Sutton GL, Boyle JR. Portsmouth POSSUM models for abdominal aortic aneurysm surgery. Br J Surg 2001;88:958-63.

11. 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;92:714-8.

12. Lazarides MK, Arvanitis DP, Drista H, Staramos DN, Dayantas JN. POSSUM and APACHE II scores do not predict the outcome of ruptured infrarenal aortic aneurysms. Ann Vasc Surg 1997;11:155-8.

13. Hadjianastassiou VG, Tekkis PP, Hands LJ, Goldhill DR. Quantification of mortality risk after abdominal aortic aneurysm repair. Br J Surg 2005;92:1092-8.

14. Drew PJ, Monson JRT. Artificial neural networks. Surgery 2000;127:3-11.

15. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet 1995;346:1075-9.

16. Baxt WG. Application of artificial neural networks to clinical medicine. Lancet 1995;346:1135-8.

17. Dybowski R, Gant V. Artificial neural networks in pathology and medical laboratories. Lancet 1995;346:1203-7.

18. Dybowski R, Weller P, Chang R, Gant V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. Lancet 1996;347:1146-50.

19. Bottaci L, Drew PJ, Hartley JE, Hadfield MB, Farouk R, Lee PWR, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet 1997;350:469-72.

20. Lapuerta P, L'Italien GJ, Paul S, Hendel RC, Leppo JA, Fleisher LA, et al. Neural network assessment of perioperative cardiac risk in vascular surgery patients. Med Decis Making 1998;18:70-5.

21. Turton EP, Scott DJ, Delbridge M, Snowden S, Kester RC. Ruptured abdominal aortic aneurysm: a novel method of outcome prediction using neural network technology. Eur J Vasc Endovasc Surg 2000;19:184-9.

22. Wyatt J. Nervous about artificial neural networks. Lancet 1995;346:1175-77.

23. Goldhill DR, Withington PS, Birch NJ. The North East Thames regional intensive care audit system. Theor Surg 1993;8:1-5.

24. Goldhill DR, Withington PS. The effect of casemix adjustment on mortality as predicted by APACHE II. Intensive Care Med 1996;22:415-9.

25. Goldhill DR, Withington PS. Mortality predicted by APACHE II: the effect of changes in physiological values and post-ICU hospital mortality. Anaesthesia 1996;51:719-23.

26. Goldhill DR, Sumner A. APACHE II, data accuracy and outcome prediction. Anaesthesia 1998;53:937-43.

27. National Confidential Enquiry into Perioperative Deaths. Classification of Operations (NCEPOD definitions). http://www.ncepod.org.uk/2004report/appendices.a.htm. (accessed Jul 1 2005).

28. Hosmer JA, Lemeshow S, editors. Applied logistic regression. New York: Wiley & Sons, Inc; 1989;140-5.

29. Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman and Hall; 1993.

30. Picard RR, Berk KN. Data splitting. Am Stat 1990;44:140-7.

31. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361-87.

32. Ashley S, Riddler B on behalf of the Audit and Research Committee of the Vascular Surgical Society of Great Britain and Ireland. National Vascular Database Report 2002. Dendrite Clinical Systems Ltd, April 2003.

33. Evans SM, Adam DJ, Bradbury AW. The influence of gender on outcome after ruptured abdominal aortic aneurysm. J Vasc Surg 2000;32:258-62.

34. Earnshaw J, Riddler B on behalf of the Audit and Research Committee of the Vascular Surgical Society of Great Britain and Ireland. National Vascular Database Report 2001. Dendrite Clinical Systems Ltd, May 2001.

35. Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's APACHE II study in Britain and Ireland-II: outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. BMJ 1993;307:977-81.

36. Goldstein H, Spiegelhalter D. League tables and their limitations: statistical issues in comparison of institutional performance. J R Stat Soc (A) 1996;159:1-25.

37. Schafer JH, Maurer A, Jochimsen F, Emde C, Wegscheider K, Arntz HR, et al. Outcome prediction models on admission in a medical intensive care unit: do they predict individual outcome? Crit Care Med 1990;18:1111-8.

38. Knaus WA, Harrell FE Jr, Lynn J, Goldman L, Phillips RS, Connors AF Jr, et al. The SUPPORT prognostic model. Objective estimates of survival for seriously ill hospitalized adults. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med 1995;122:191-203.

39. Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991;100:1619-36.

40. National Confidential Enquiry into Patient Outcome and Death (2005). An Acute Problem? ICU admission process. http://www.ncepod.org.uk/2005report/summary.pdf. (accessed Oct 1 2005).

41. Gill CJ, Sabin L, Schmid CH. Why clinicians are natural bayesians. BMJ 2005;330:1080-3.

42. Earnshaw J, Ridler BMF. The Vascular Surgical Society of Great Britain and Ireland. Registry 1999. http://www.vascularsociety.org.uk/docs/registry1999.pdf. (accessed Oct 1 2005).