RESEARCH METHODS & REPORTING

Clinical prediction rules

Clinical prediction rules are mathematical tools that are intended to guide clinicians in their everyday decision making. The popularity of such rules has increased greatly over the past few years. This article outlines the concepts underlying their development and the pros and cons of their use

Simon T Adams *clinical research fellow*[1], Stephen H Leveson *professor of surgery*[2]

[1]York Hospital, York YO31 8HE, UK; [2]Hull-York Medical School, Learning and Research Centre, York Hospital

In many ways much of the art of medicine boils down to playing the percentages and predicting outcomes. For example, when clinicians take a history from a patient they ask the questions that they think are the most likely to provide them with the information they need to make a diagnosis. They might then order the tests that they think are the most likely to support or refute their various differential diagnoses. With each new piece of the puzzle some hypotheses will become more likely and others less likely. At the end of the process the clinician will decide which treatment is likely to result in the most favourable outcome for the patient, based on the information they have obtained.

Given that the above process is the underlying principle of clinical practice, and bearing in mind the ever increasing time constraints imposed on people, it is unsurprising that a great deal of work has been done to help clinicians and patients make decisions. This work is referred to by many names: prediction rules, probability assessments, prediction models, decision rules, risk scores, etc. All describe the combination of multiple predictors, such as patient characteristics and investigation results, to estimate the probability of a certain outcome or to identify which intervention is most likely to be effective.[1 2] Predictors are identified by "data mining"—the process of selecting, exploring, and modelling large amounts of data in order to discover unknown patterns or relations.[3]

Ideally, a reliable predictive factor or model would combine both a high sensitivity with a high specificity.[4 5] In other words it would correctly identify as high a proportion as possible of the patients fated to have the outcome in question (sensitivity) while excluding those who will not have the outcome (specificity).[6] In the table⇓ sensitivity can be defined as A÷(A+C) and specificity as D÷(B+D).

A good predictive factor is not the same as a strong risk factor.[4] The positive predictive value of a predictive factor or model refers to its accuracy in terms of the proportion of patients correctly predicted to have the outcome in question (A÷(A+B) in the table⇓).[7] A risk factor can be identified by calculating the relative risk (or odds ratio) of an outcome in patients with the factor in question compared with patients without it.[4] If, however, the factor identified or the outcome being used is uncommon, it is of little clinical use as a predictive factor.[4 7]

A good predictive factor or model shows a good fit between the probabilities calculated from the model and the outcomes actually observed, while also accurately discriminating between patients with and without the outcome.[4 5] For example, if all patients with a measured observation of ≥0.5 die and all patients with the measured observation <0.5 survive then the observed factor is a perfect predictor of survival.

Unfortunately, as a general rule sensitivity and specificity are mutually exclusive—as one rises the other falls. Since both are important to the development of predictive models receiver-operating characteristic (ROC) curves are used to visualise the trade-off between the two and express the overall accuracy of the model (fig 1⇓).[4 8 9] Sensitivity (true positive) is plotted on the y axis and 1−specificity (false positive) is plotted on the x axis.[4 9] The closer a point is to the top left of the graph then the higher the area under the curve and the more accurate or useful a predictive factor can be said to be.[4 8 9] Conversely a plot in the 45 degree diagonal (denoting an area under the curve of 50%) indicates a test no more accurate than chance.[4 8 9] Where the limits of acceptability are set is arbitrary and depends on several factors such as the severity of the outcome and the potential negative consequences of the test.[4 9]

Establishing a clinical prediction rule

The establishment of a prediction model in clinical practice requires four distinct phases:

Development—Identification of predictors from an observational study

Validation—Testing of the rule in a separate population to see if it remains reliable

Impact analysis—Measurement of the usefulness of the rule in the clinical setting in terms of cost-benefit, patient satisfaction, time/resource allocation, etc

Correspondence to: S Adams rpbgt@hotmail.com

Implementation—Widespread acceptance and adoption of the rule in clinical practice.

For a prediction rule to gain popularity each of the first three steps needs to be satisfactorily completed before the fourth stage.[1] Validation in a suitably powered cohort study or controlled trial is particularly important because there is no guarantee that a predictor will be accurate outside the original data set.[1 2] Indeed validation usually shows a reduction in accuracy compared to that in the original study.[1 10-12] Reliability is essentially the reproducibility of a measurement—that is, if the same test were applied under the same circumstances how similar the results would be.

Despite the long running controversy concerning their usefulness and application the popularity of clinical prediction rules has been shown to be greater now than ever.[1 13 14] A Medline search by Toll and colleagues in 2008 showed that the number of papers discussing prediction rules has more than doubled in recent years (6744 papers in 1995 versus 15 662 in 2005).[1] Most publications, however, concern the development of new rules, with few articles describing validation and almost none confirming their clinical impact.[1] There are several possible reasons why validation and impact analysis are so often overlooked. Perhaps the most important are that neither validity nor reliability can be exactly quantified and that establishing validity requires investigators to consider several different aspects (face validity, content validity, construct validity, criterion validity, etc).[15 16]

Advantages and disadvantages of prediction rules

When appropriately developed and validated, prediction models have inherent advantages over human clinical decision making. Firstly, the statistical models can accommodate many more factors than the human brain is capable of taking into consideration.[17] Secondly, if given identical data a statistical model will always give the same result whereas human clinical judgment has been shown to result in both inconsistency and disparity, especially with less experienced clinicians.[17 18] Finally, and perhaps most importantly, several prediction models have been shown to be more accurate than clinical judgment alone.[14 17-21] So why are such models not used more readily in every practice?

Liao and Mark proposed in 2003 that resistance to adopting prediction models may reflect tacit acknowledgment that clinicians do not know how to take advantage of such tools.[17] They also suggested that such tools may not be thought user friendly and may not take into account the continual, dynamic way in which humans gather clinical information.[17] Their final reason for low implementation of clinical prediction rules is the sheer number of models available.[17] If multiple prediction rules exist for the same problem identifying the best one is difficult. Not only is it potentially very time consuming but differences in the methods used in the studies on which they are based may make reliable comparison impossible.[11 22] Part of the reason for the large number of prediction rules may be the wide variety of ways in which such tools can be developed.

Types of prediction model

In 2006 Grobman and Stamilio described five main methods used to develop clinical prediction models: scoring systems derived from univariate analysis, prediction models based on multivariate analysis, nomograms, artificial neural networks, and decision trees.

Scoring systems derived from univariate analysis

Factors shown to be significantly related to the outcome in observational studies are allocated a score or "weight." The cumulative final score of all the risk factors present in a patient is used as an indicator of the likelihood of the outcome occurring.[4] Well known examples of this type of prediction model include the Alvarado score for acute appendicitis and the modified Glasgow score for acute pancreatitis.[23 24] These models are simple to devise and use but their accuracy is affected by the potential inclusion of non-independent risk factors and the arbitrary manner in which factors are weighted.[4]

Prediction models based on multivariate analysis

These are developed in a similar manner to the above scoring systems except that the analysis of the results from the observational study is more refined and therefore less likely to include any non-independent factors. The models typically use logistic regression analysis, which has the added advantage of expressing the relation between the predictive factors and the outcome in the form of odds ratios (the probability of an outcome occurring versus the probability that it will not).[4] These are relatively easy to interpret and can also be used to assign weights in a less arbitrary fashion than in univariate models.[4 25] Nevertheless, multivariate analysis techniques are not completely reliable in eliminating bias from interaction of independent variables.[4] Models using logistic regression are often well suited to being represented as a nomogram (see below).[3]

Nomograms

Nomograms are graphical calculating devices that represent mathematical relations or laws and allow the user to rapidly calculate complicated formulas to a practical precision (fig 2⇓).[26] Nomograms may be as simple as the markings on a thermometer or more complex, such as the Siggaard-Andersen chart used to diagnose acid-base blood disorders.[27] The mathematics and statistics used to develop a nomogram can be equally simplistic or intricate.[4] The advantage of nomograms is that the final prediction tool created is generally comparatively simple to use and in some cases more accurate than other prediction models for the same clinical problem.[4 28] Other nomograms in common clinical use include those used to predict the likelihood of a patient having prostate cancer from their clinical examination and prostate specific antigen levels and those used to predict the peak expiratory flow rate of asthmatic patients based on their age and height.[29 30]

Prediction using artificial neural networks

Artificial neural networks are mathematical or computational models based on the operation of biological neural networks.[31] In biology, a nerve cell (or neuron) will receive input from numerous other nerve cells. It will then process all of the input it receives and either send off an action potential or not. Because these nerve cells are all interconnected they are referred to as networks. Artificial neural networks function along similar lines: multiple sources of information (input) are fed into the software program, which interprets it and produces a dichotomous output (fig 3⇓). The main advantage of neural networks is that they can "learn" mathematical relations between a series of input variables and the corresponding output.[32-35] This is achieved by inputting a set of data containing both the input data (the predictor variables) as well as the outcomes.[32 33] With each new

data set entered the neural network is able to adjust the internal weights of the various pieces of input data and calculate the probability of a specific outcome.[32]

Neural networks require little formal statistical training to develop and can implicitly detect complex non-linear relations between independent and dependent variables as well as all possible interactions between predictor variables.[32 33] However, they have a limited ability to explicitly identify possible causal relations, they are hard to use at the bedside, and they require greater computational resources than other prediction models.[32 33] They are also prone to "overfitting"—when too many data sets are used in training the network causing it to effectively memorise the noise (irrelevant data) and reducing its accuracy.[32 33] A final drawback to neural networks is that the development model is empirical and because it is a new technique methodological problems remain.[32] In a direct comparison between neural networks and logistic regression models Tu and colleagues concluded that neural networks were better for predicting outcomes but that logistic regression was preferable when looking for possible causal relations between independent and dependent variables or when trying to understand the effect of predictor variables on an outcome.[32]

Decision trees (CART analysis)

Classification and regression tree (CART) analysis uses non-parametric tests to evaluate data and progressively divide it into subgroups based on the predictive independent variables.[4] The variables and discriminatory values used and the order in which the splitting occurs are produced by the underlying mathematical algorithm and are calculated to maximise the resulting predictive accuracy.[4] CART analysis produces "decision trees," which are generally easily understood and consequently translate well into everyday clinical practice (fig 4⇓). By following the arrows indicated by the answers to each of the questions in the boxes clinicians will be directed to the predicted outcome for the patient. Examples of CARTs used in clinical practice include those to predict large oesophageal varices in cirrhotic patients and to predict the likelihood of hospital admission in patients with asthma.[36 37] However, the CART model of prediction can be significantly less accurate than other models.[28 38] This may be because the "leaves" on the trees contain too little data to be able to predict outcomes reliably.[3]

Conclusion

Each of the five main models has advantages and disadvantages, and no single model of prediction has been clearly shown to be superior to the others in all applications. As pressure on their time increases, doctors will need to become familiar with decision making tools and the statistical principles underlying them.

1   Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *J Clin Epidemiol* 2008;61:1085-94.
2   Cook CE. Potential pitfalls of clinical prediction rules. *J Man Manip Ther* 2008;16:69-71.
3   Bellazzi R, Zupan B. Predictive data mining in clinical medicine: current issues and guidelines. *Int J Med Inform* 2008;77:81-97.
4   Grobman WA, Stamilio DM. Methods of clinical prediction. *Am J Obstet Gynecol* 2006;194:888-94.
5   Braitman LE, Davidoff F. Predicting clinical states in individual patients. *Ann Intern Med* 1996;125:406-12.
6   Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. *BMJ* 1994;308:1552.
7   Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994;309:102.
8   Collins JA. Associate editor's commentary: mathematical modelling and clinical prediction. *Hum Reprod* 2005;20:2932-4.
9   Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. *BMJ* 1994;309:188.
10 Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;144:201-9.
11 Yealy DM, Auble TE. Choosing between clinical prediction rules. *N Engl J Med* 2003;349:2553-5.
12 Verma S, Hamilton K, Hawkins HH, Kothari R, Singal B, Buncher R, et al. Clinical application of the Ottawa ankle rules for the use of radiography in acute ankle injuries: an independent site assessment. *AJR Am J Roentgenol* 1997;169:825-7.
13 Kleinmuntz B. Why we still use our heads instead of formulas: toward an integrative approach. *Psychol Bull* 1990;107:296-310.
14 Marchese MC. Clinical versus actuarial prediction: a review of the literature. *Percept Mot Skills* 1992;75:583-94.
15 Bland JM, Altman DG. Statistics notes: validating scales and indexes. *BMJ* 2002;324:606-7.
16 Cookson J. A critique of the specialty certificate examinations of the Federation of Royal Colleges of Physicians of the UK. *Clin Med* 2010;10:141-4.
17 Liao L, Mark DB. Clinical prediction models: are we building better mousetraps? *J Am Coll Cardiol* 2003;42:851-3.
18 Gandara E, Wells PS. Diagnosis: use of clinical probability algorithms. *Clin Chest Med* 2010;31:629-39.
19 Bandiera G, Stiell IG, Wells GA, Clement C, De Maio V, Vandemheen KL, et al. The Canadian C-spine rule performs better than unstructured physician judgment. *Ann Emerg Med* 2003;42:395-402.
20 Gardner W, Lidz CW, Mulvey EP, Shaw EC. Clinical versus actuarial predictions of violence of patients with mental illnesses. *J Consult Clin Psychol* 1996;64:602-9.
21 Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C. Clinical versus mechanical prediction: a meta-analysis. *Psychol Assess* 2000;12:19-30.
22 Stiell IG, Clement CM, McKnight RD, Brison R, Schull MJ, Rowe BH, et al. The Canadian C-spine rule versus the NEXUS low-risk criteria in patients with trauma. *N Engl J Med* 2003;349:2510-8.
23 Alvarado A. A practical score for the early diagnosis of acute appendicitis. *Ann Emerg Med* 1986;15:557-64.
24 Taylor SL, Morgan DL, Denson KD, Lane MM, Pennington LR. A comparison of the Ranson, Glasgow, and APACHE II scoring systems to a multiple organ system score in predicting patient outcome in pancreatitis. *Am J Surg* 2005;189:219-22.
25 Bland JM, Altman DG. Statistics notes. The odds ratio. *BMJ* 2000;320:1468.
26 Doerfler R. On jargon: the lost art of nomography. *UMAP* 2009;30:457-93.
27 Siggaard-Andersen O. The acid-base status of the blood. *Scand J Clin Lab Invest* 1963;15(suppl 70):1-134.
28 Chun FK, Karakiewicz PI, Briganti A, Walz J, Kattan MW, Huland H, et al. A critical appraisal of logistic regression-based nomograms, artificial neural networks, classification and regression-tree models, look-up tables and risk-group stratification models for prostate cancer. *BJU Int* 2007;99:794-800.
29 Eastham JA, May R, Robertson JL, Sartor O, Kattan MW. Development of a nomogram that predicts the probability of a positive prostate biopsy in men with an abnormal digital rectal examination and a prostate-specific antigen between 0 and 4 ng/mL. *Urology* 1999;54:709-13.
30 Lam KK, Pang SC, Allan WG, Hill LE, Snell NJ, Fayers PM, et al. Predictive nomograms for forced expiratory volume, forced vital capacity, and peak expiratory flow rate, in Chinese adults and children. *Br J Dis Chest* 1983;77:390-6.
31 Neuro AI. Intelligent systems and neural networks. 2007. www.learnartificialneuralnetworks.com/.
32 Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49:1225-31.
33 Ayer T, Chhatwal J, Alagoz O, Kahn CE Jr, Woods RW, Burnside ES. Informatics in radiology: comparison of logistic regression and artificial neural network models in breast cancer risk estimation. *Radiographics* 2009;30:13-22.
34 Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J Clin Epidemiol* 2010;63:826-33.
35 Hermundstad AM, Brown KS, Bassett DS, Carlson JM. Learning, memory, and the role of neural network architecture. *PLoS Comput Biol* 2010;7:e1002063.
36 Hong WD, Dong LM, Jiang ZC, Zhu QH, Jin SQ. Prediction of large esophageal varices in cirrhotic patients using classification and regression tree analysis. *Clinics (Sao Paulo)* 2011;66:119-24.
37 Tsai CL, Clark S, Camargo CA Jr. Risk stratification for hospitalization in acute asthma: the CHOP classification tree. *Am J Emerg Med* 2010;28:803-8.
38 Austin PC, Tu JV, Lee DS. Logistic regression had superior performance compared with regression trees for predicting in-hospital mortality in patients hospitalized with heart failure. *J Clin Epidemiol* 2010;63:1145-55.

Table

Table 1| Tabular representation of predicted versus actual outcomes of a predictive model

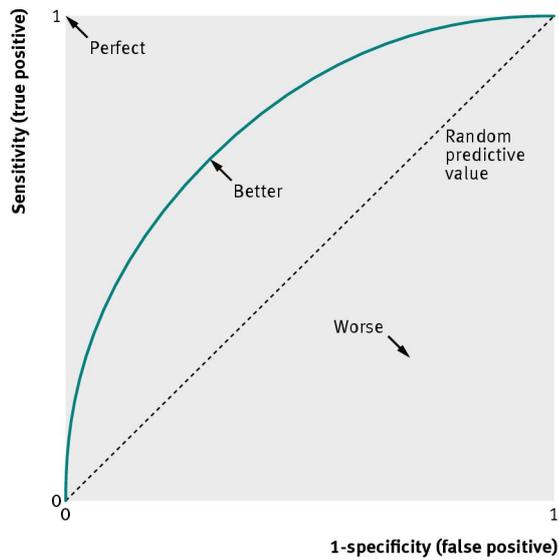|	Actual outcome	
Predicted outcome	**Positive**	**Negative**
Positive	A (true positive)	B (false positive)
Negative	C (false negative)	D (true negative)

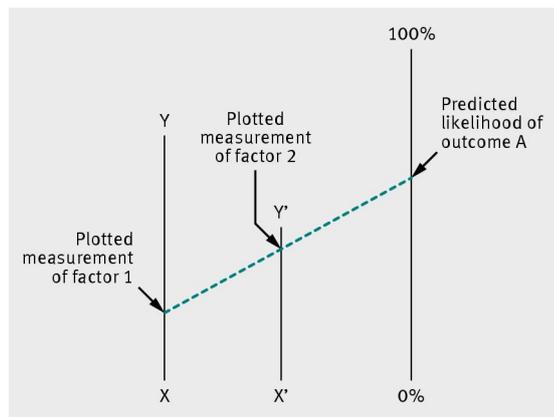# Figures

Fig 1 Receiver-operating characteristic (ROC) curve

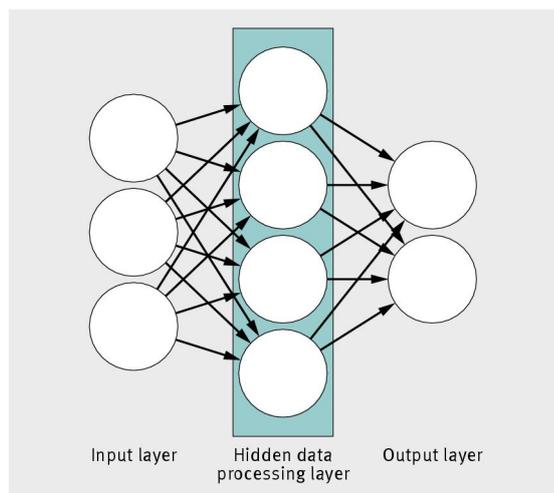**Fig 2** Simplified representation of a basic nomogram

Fig 3 Schematic representation of an artificial neural network. The first column (input layer) represents a piece of data that can be put in to the neural network programme. The circles in the second column (hidden layer) represent the neural network programme assigned weight or numerical significance of each piece of data entered in the input layer. The final column (output layer) represents the dichotomous predicted outcome for the information entered
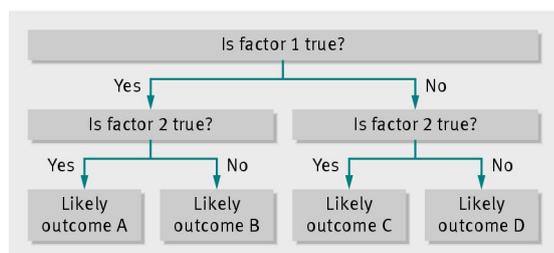

Fig 4 Simplified representation of a basic decision tree