Guide to presenting clinical prediction models for use in clinical settings

Laura J Bonnett,[1] Kym I E Snell,[2] Gary S Collins,[3] Richard D Riley[2]

[1]Department of Biostatistics, University of Liverpool, Liverpool L69 3GL, UK

[2]Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University, Keele, UK

[3]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK

Correspondence to: L J Bonnett
L.J.Bonnett@liverpool.ac.uk
(ORCID 0000-0002-6981-9212)

Clinical prediction models estimate the risk of existing disease or future outcome for an individual, which is conditional on the values of multiple predictors such as age, sex, and biomarkers. In this article, Bonnett and colleagues provide a guide to presenting clinical prediction models so that they can be implemented in practice, if appropriate. They describe how to create four presentation formats and discuss the advantages and disadvantages of each format. A key message is the need for stakeholder engagement to determine the best presentation option in relation to the clinical context of use and the intended users

Introduction

Clinical prediction models estimate the risk of existing disease (diagnostic prediction model) or future outcome (prognostic prediction model) for an individual, which is conditional on the values of multiple predictors (prognostic or risk factors) such as age, sex, and biomarkers.[1] A large number of prediction models are published in the medical literature each year,[2] and most are developed using a regression framework such as logistic and Cox regression (box 1). Prediction models are also known as risk scores, prognostic indices, or prognostic scores. Examples include the Framingham risk score, which predicts 10 year risk of coronary heart disease,[3] and the APACHE (acute physiological assessment and chronic health evaluation) scores for mortality after admission to an intensive care unit.[4 5]

The transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement provides guidance on key information for authors to report when developing or validating prediction models.[6 7] Although the TRIPOD statement highlights the importance of the presentation of models, relatively little information or practical guidance exists on how to actually present a prediction model for use after development—for example, to aid implementation in clinical settings, if appropriate. A key resource is chapter 18 in the book by Steyerberg.[8]

When choosing the format to present a prediction model, researchers should carefully consider the intended users, settings, and timing of use. It is helpful to ask: "who will be accessing the model in this format, and when and in what setting will they use it?" The presentation can then be formatted accordingly. Fundamentally, the full model equation should always be presented in the journal article[6 7]; this is essential to enable independent external validation. However, additional presentation formats might be required, perhaps outside of the journal article, to enable healthcare professionals to use the model in a particular clinical setting (eg, when access to computers or mobile devices is limited). Similarly, the format might need tailoring for lay people when using the model at home (eg, patients with asthma deciding on appropriate management of their condition using Asthma UK's asthma attack risk checker[9]) to improve shared decision making.[10] User groups can help guide the best presentation choices in each situation, including healthcare professionals, patients, and the public. Patient and public involvement groups and focus groups are useful arenas for this; they highlight the need for participant and public information and engagement within health research.

Alongside the full model equation, a range of presentation formats might be required that differ according to several factors: the medium by which they are presented (paper versus electronic), the setting in which the models are to be applied (eg, clinic, bedside, or at home), the level of detail required in the predictions (eg, approximate or rounded risk estimates, or exact risk estimates), and user friendliness (simple to complex formats).[8] In this paper, we summarise four key ways of presenting clinical prediction models that could aid their use in clinical practice, if appropriate. We outline how to create each format and describe their advantages and disadvantages in relation to the clinical

SUMMARY POINTS

- Presentation format for clinical prediction models deemed suitable for use is important but receives relatively limited attention in the literature

- Clear presentation of a prediction model is fundamental to ensure other researchers can independently validate the model, and that healthcare professionals and others can implement it within healthcare

- Presentation of the full model equation is essential. There are many ways to present prediction models for end users, which range from points score systems and nomograms, to websites and mobile apps

- The best presentation is user and environment specific, and it is preferably determined through engagement of stakeholders, including patients

- If presentation requires a simplified version of the full model to be generated, then the predictive performance of this simplified model should also be validated and compared with that of the full model

Box 1: Typical format of prediction models developed using regression to enable risk predictions in new individuals (adapted from Riley and colleagues[17])

Short term clinical prediction models or diagnostic models

If the outcome can take one of only two options (eg, death or presence of the disease) and is known for all patients at a particular time, a prediction model can be developed using logistic regression. This has the following form:

$$\ln\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots$$

where p is the probability of having the outcome and $\ln(p/1-p)$ is the log odds of the outcome; α is the intercept term and the baseline log odds where "baseline" refers to those whose X values are zero. If all X predictors are centred at the mean, α is the log odds for someone with average X values. The X terms denote values of included predictors so that X_1 might represent the age of the patient in years, X_2 might represent sex and be 1 for men and 0 for women, etc. The β terms denote the change in log odds (otherwise known as the log odds ratio) for each one unit increase in the corresponding predictor. For example, β_1 might be the increase in the log odds for each one year increase in age, and β_2 might be the increase in the log odds for a man compared with a woman. Risk (outcome probability) predictions, $\hat{p}$ for a new individual can be estimated by inputting his or her predictor values into the equation and then transforming back to the probability scale:

$$\hat{p} = \frac{\exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots)}{1 + \exp(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots)}$$

Clinical prediction models over time

If risks are predicted over time, or at a time point when some people in the dataset have previously dropped out or been lost to follow-up (that is, censored), a prediction model can be developed using a survival model; for example, a Cox model or a parametric survival model. This has the following form:

$$h(t) = h_0(t)\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots)$$

$h(t)$ is the hazard rate of the outcome at time t and $h_0(t)$ is the baseline hazard rate. The X terms denote values of included predictors, and each β denotes the change in log hazard rate (otherwise known as the log hazard ratio) for each one unit increase in the corresponding predictor. "Baseline" refers to people whose X values are zero, or if all predictors are centred at the mean, the underlying hazard rate for someone with average X values. Predictions of not having the outcome at time t ("survival" probability) for a new individual can be obtained by inputting their predictor values into the equation and then transforming back to the probability scale:

$$\hat{S}(t) = S_0(t)^{\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots)}$$

where $S_0(t)$ is the baseline survival probability at time t. Conversely, risk predictions (outcome probability) for having the outcome at time t for a new individual can be calculated as:

$$1 - \hat{S}(t) = 1 - S_0(t)^{\exp(\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots)}$$

context of use, and the intended user. Table 1 provides an overview of the different presentation formats.

Our article is not about how to develop or validate a prediction model,[8 11] or how to decide if it is fit for clinical use[1 12]; rather, we assume a model has been developed and has been deemed potentially useful for clinical practice, and so the researcher needs to consider how to present the model to aid implementation. For this purpose we use the model shown in box 2 to illustrate how to predict mortality risk over time in patients with a diagnosis of primary biliary cirrhosis. This survival model is used throughout the article, but the presentation formats described also apply to other risk prediction models developed using regression, such as those derived using logistic regression. Many of these models are relevant for prediction of a continuous outcome (eg, using linear regression).

Presentation using a points score system
Description

Within points score systems, points are assigned based on the predictor values for an individual. The total points score is then mapped to a corresponding risk of event or survival probability.[21] The intended users for points score systems are healthcare professionals and patients. These systems can be presented on a screen (eg, monitor, tablet) as part of a consultation, printed off as a take home sheet for patients, or used on the wards as a reference guide.

How to derive a points score system

To produce a points score system, a prediction model is first developed (eg, using logistic or Cox regression; see boxes 1 and 2), and then the regression coefficients of included predictors are assigned integer scores, which can be negative or positive. Unfortunately, any continuous predictors need to be categorised, and so some predictive accuracy is sacrificed. Categories do not need to be of equal size, and by having unequal categories non-linearity can be more appropriately handled. The steps to develop a points score system can be summarised:

1. Organise the continuous predictors into categories and determine the midpoint for each category.

2. Choose a reference category for each predictor (continuous, binary, and categorical).

3. For continuous variables, determine how far each category is from the reference category, and then multiply each difference by the regression coefficient for that predictor to determine the difference in "regression units." For binary and categorical variables, the "regression unit" is simply the regression coefficient for that predictor.

4. Define the number of regression units that will correspond to 1 point in the points scoring system; this definition is usually based on clinician preference.

5. Determine the points (rounded to the nearest integer) associated with each of the categories of the predictors.

6. Determine the minimum and maximum possible points totals.

7. Calculate the risk estimate for each points total across the range by using the original model with the points scores (rounded) to get the predicted probabilities. This estimate is essentially a new risk prediction model that approximates the full model.

Note that sometimes scores are derived based on hazard ratios or odds ratios rather than the corresponding regression coefficients. This approach is mathematically inappropriate as Cox (or logistic) regression models assume additivity of the log hazard (or log odds) ratios.[22]

Alongside the point score system, it is important to also present the accompanying table of probabilities (absolute risk predictions) to allow the points score to be translated to a predicted risk. Decisions such as low, intermediate, or high risk only based on a points total are uninformative unless it is clear how these decisions are defined on the predicted absolute risk scale.

A full worked example is provided in the appendix. Also, see Sullivan and colleagues for further details, including the mathematical formula associating the risk of outcome with each possible total points score for logistic and survival regression models.[21]

Advantages and disadvantages

Points score systems are easy to understand after an initial explanation or demonstration, and so instructions on their use should be given alongside these systems. Depending on the complexity of

Table 1 | Examples of formats for presenting clinical prediction models to enable outcome risks to be calculated for new individuals; these should be considered after presenting the full model equation

Presentation format	Explanation	Advantages	Disadvantages	Example
Points score system	Comprises two tables: one enables a total points score to be calculated based on predictor values; other provides estimate of risk based on total points score	Easy to understand	Predictions are approximate. Need predictions at each time point of interest for survival outcomes. Continuous predictors must be categorised	Renal artery stenosis[13]
Graphical score chart	Graphical representation of highly simplified points score system	Easy to understand	Can only accommodate limited number of predictors. Need one for each time point of interest for survival outcomes. Prediction of range of event probabilities and thus predictions are approximate. Continuous predictors must be categorised	Traumatic bleeding[14]
Nomogram	Graphical presentation of prediction model: points are assigned based on predictor values that are then summed and translated to estimate of risk of outcome	Can easily be applied away from computer (eg, community based medicine)	Can initially be difficult to understand. Could be inaccurate depending on size and resolution of published nomogram	Prostate cancer,[15] coronary heart disease[16]
Websites and applications	Interactive graphical user interface that provides risk estimates from underlying (often hidden) prediction model after user inputs predictor values	Visually appealing to intended users. Can automate complex modelling methods in background. Full equation is retained (albeit in background). Can quickly produce risk predictions at multiple time points for survival outcomes. Can automate extraction of predictor values (eg, from e-health records, or from available hardware such as GPS coordinates, calendar time, blood pressure)	Easy access could lead to overuse or access by those who might not be target population. Can be created by anyone so no guarantee model has been developed well. Often difficult to know how model equation has been translated to graphical tool. Often unclear whether it is of relevance or has been validated in relevant population. Possible privacy and data storage issues. Model could change over time and by location without changes being tracked	Breast cancer[10]

Box 2: Dataset and prediction model for primary biliary cirrhosis

Clinical context

These data are from an international trial on the therapeutic effect of the immunosuppressant drug azathioprine in primary biliary cirrhosis (also known as primary biliary cholangitis) of the liver and are publicly available.[18 19] Primary biliary cirrhosis is an autoimmune condition that can get worse over time and without treatment might lead to liver failure. 248 patients with primary biliary cirrhosis met eligibility criteria for the randomised placebo controlled trial, of whom 127 received the intervention and 121 the placebo. 57 (45%) deaths occurred in the treatment group and 62 (51%) in the control group. Risk prediction is important to guide patient counselling and potentially inform treatment choice.

Predictors

Predictors of interest were age (years), presence of cirrhosis (yes/no), albumin (g/dL), presence of central cholestasis (yes/no), and placebo treatment (rather than azathioprine, yes/no). Table 2 lists the characteristics of these predictors, including the ranges of the continuous predictors. Treatment was explicitly modelled.[20]

Table 2 Patient characteristics for primary biliary cirrhosis running example

Continuous predictor	Range	Mean (SD)	Binary predictor	Coding	No (%)
Age (years)	25.0-78.0	54.8 (10.6)	Cirrhosis	No: 0, yes: 1	148 (71.5), 59 (28.5)
Albumin (g/dL)	20.0-56.5	34.4 (5.9)	Central cholestasis	No: 0, yes: 1	170 (82.1), 37 (17.9)
			Treatment	Azathioprine: 0, placebo: 1	109 (52.7), 98 (47.3)

Methods to develop model

A Cox proportional hazards regression model was fit to estimate probability of death from primary biliary cirrhosis, with all predictor values centred at the mean predictor value in the study. As this was only an illustrative example, no adjustment for potential overfitting was made.

Final model equation

The estimated linear predictor (LP) for this model is the linear combination of predictors and their associated regression coefficients; that is, for individual i:

$$LP_i = (0.02 \times (age - 54.8)) + (1.06 \times (cirrhosis - 0.285)) +$$
$$(-0.06 \times (albumin - 34.4)) + (1.59 \times (cholestasis - 0.179)) +$$
$$(0.31 \times treatment)$$

The model for probability of death from primary biliary cirrhosis is shown in equation 7. Estimates of baseline survival at one year ($S_0(1)$) and three years ($S_0(3)$) are also provided, based on a Nelson-Aalen type estimator. Note that time of predictions is made relevant by clinical and patient stakeholders:

$$1 - S(t) = 1 - S_0(t)^{exp(LP_i)}$$

$$S_0(1) = 0.930; S_0(3) = 0.759$$

As predictor values are centred, the baseline ($S_0(t)$) in this equation relates to the survival probability over time for an "average" treated individual whose predictor values equal the mean predictor values in the study. As the prediction equation is derived from randomised controlled trial data the treatment effect can be assessed, which is in contrast with the more usual case when prediction equations are derived from cohort studies.

the model, and the number of included predictors, paper based point score systems can enable risks to be estimated quicker than directly inputting patient values into an online calculator or published model. However, the predictions of risk (or survival) are only approximations of the actual predicted risk from the full model; this is because information on continuous predictor values is discarded by categorisation, and regression units are rounded. Researchers must always check that the predictive performance of the simplified model based on a points score system is similar to (and has the same potential clinical impact as) the original full model.

Application

Tables 3a and b illustrate a points score system for the primary biliary cirrhosis example. In table 3a, scores are assigned to categories of each predictor, which have been scaled according to a 15 year increase in age (that is, the regression coefficient for age multiplied by 15); table 3b presents probabilities of the outcome that correspond to the points total. For example, an individual aged 55 (0 points), with cirrhosis (3 points), albumin of 34.4 g/dL (0 points), and central cholestasis (5 points) has a points total of 8; this corresponds to a probability of death of 0.46 at one year and 0.90 at three years. These figures are similar to the equivalent estimates of 0.44 and 0.89 for the risk of death at one and three years, respectively, directly obtained from the full model equation for this individual. These data suggest that the simplification led to only small changes in risk predictions from the points score system for this individual.

Table 3a | Points score system for probability of death for new patients with primary biliary cirrhosis based on model derived in box 2

Predictor categories	Points	Age (years):
25-34	−2	
35-44	−1	
45-54	−1	
55-64	0	
65-78	1	
Cirrhosis:		
No	0	
Yes	3	
Albumin (g/dL):		
20.0-24.9	2	
25.0-29.9	1	
30.0-34.9	0	
35.0-39.9	−1	
40.0-44.9	−2	
45.0-49.9	−3	
50.0-56.5	−4	
Central cholestasis:		
No	0	
Yes	5	
Treatment:		
Azathioprine	0	
Placebo	1	

Table 3b

Points total	Probability of death	
	At 1 year	At 3 years
−6	0.008	0.029
−5	0.011	0.040
−4	0.014	0.054
−3	0.020	0.073
−2	0.027	0.098
−1	0.036	0.131
0	0.049	0.175
1	0.067	0.231
2	0.090	0.301
3	0.121	0.387
4	0.161	0.487
5	0.213	0.598
6	0.280	0.712
7	0.361	0.818
8	0.458	0.902
9	0.566	0.958
10	0.681	0.987
11	0.790	0.997
12	0.881	1.000

Presentation using graphical score chart

Description

Graphical score charts are highly simplified, colour coded versions of points score systems. Similar to the points score system, a graphical score chart is a presentation format for a prediction model, when the intended users are healthcare professionals and patients. The chart can be used either on screen or as a print out. An example of using this approach is the SCORE (systematic coronary risk evaluation) model for predicting cardiovascular disease.[23]

How to derive a graphical score chart

The probability of the outcome must be calculated for each relevant combination of predictors as described in box 1, and is based on the average value of the category or the group of individuals in that category in the development data. Probabilities can then be tabulated and colour coded based on clinically important categories of risk. For example, those with higher risk predictions (near to 1) could be coded as red, and those with low risk predictions (near to 0) could be coded as yellow.

Advantages and disadvantages

Graphical score charts are easy to understand and the colour coding can increase ease of use compared with points score systems.[14] Additionally, decision guidelines can be coupled to the predictions; for example, dark red implies referral to intensive care. This coupling enables the rapid stratification of patients. Choosing decision thresholds needs careful thought and evaluation; in particular, arbitrary cut-off values should be avoided.[24]

This presentation usually requires some simplification of the model because it can only accommodate a limited number of predictors and requires continuous predictors to be presented as categories. Loss of information about predicted risks also occurs because the results are typically presented as ranges of predicted risks rather than specific values. Each time point of interest requires its own graphical score chart, which is a further disadvantage. Similar to the points score system, the simplified model based on a score chart should be checked for its predictive performance (at each time point of interest) compared with the full model.

Application

Figure 1 shows a graphical score chart for the primary biliary cirrhosis example. The chart was created using the point score system shown in tables 3a and b. Risks of death of widths 0.1 might be considered clinically meaningful for example, up until 0.3, and were thus chosen as the four colour categories. According to this chart, risk of death at one year is 0.46 for a patient who is aged 55, has cirrhosis, central cholestasis, albumin of 34.4 g/dL, and been treated with azathioprine. This predicted risk is similar to the value of 0.44 estimated from the full model.

Presentation using a nomogram

Description

Nomograms are another graphical presentation format for a clinical prediction model (fig 2). Similar to the points score system, points are assigned based on the predictor values for a particular individual, which are then equated to a risk of event or survival probability.[25] The intended users of a nomogram are healthcare professionals. Nomograms are best used as reference guides, potentially on wards or during consultations. Similar to graphical score charts, nomograms can be colour coded to aid interpretation.[26]
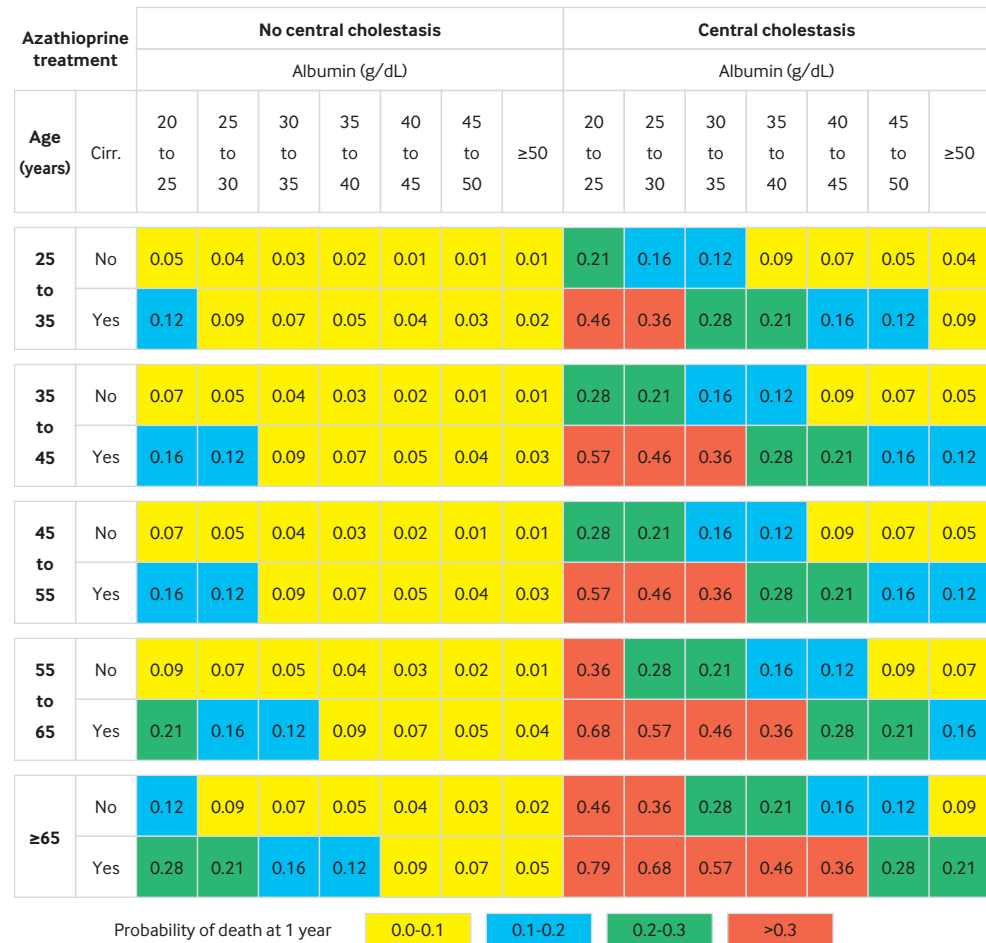
Azathioprine treatment		No central cholestasis							Central cholestasis						
		Albumin (g/dL)							Albumin (g/dL)						
Age (years)	Cirr.	20 to 25	25 to 30	30 to 35	35 to 40	40 to 45	45 to 50	≥50	20 to 25	25 to 30	30 to 35	35 to 40	40 to 45	45 to 50	≥50
25 to 35	No	0.05	0.04	0.03	0.02	0.01	0.01	0.01	0.21	0.16	0.12	0.09	0.07	0.05	0.04
	Yes	0.12	0.09	0.07	0.05	0.04	0.03	0.02	0.46	0.36	0.28	0.21	0.16	0.12	0.09
35 to 45	No	0.07	0.05	0.04	0.03	0.02	0.01	0.01	0.28	0.21	0.16	0.12	0.09	0.07	0.05
	Yes	0.16	0.12	0.09	0.07	0.05	0.04	0.03	0.57	0.46	0.36	0.28	0.21	0.16	0.12
45 to 55	No	0.07	0.05	0.04	0.03	0.02	0.01	0.01	0.28	0.21	0.16	0.12	0.09	0.07	0.05
	Yes	0.16	0.12	0.09	0.07	0.05	0.04	0.03	0.57	0.46	0.36	0.28	0.21	0.16	0.12
55 to 65	No	0.09	0.07	0.05	0.04	0.03	0.02	0.01	0.36	0.28	0.21	0.16	0.12	0.09	0.07
	Yes	0.21	0.16	0.12	0.09	0.07	0.05	0.04	0.68	0.57	0.46	0.36	0.28	0.21	0.16
≥65	No	0.12	0.09	0.07	0.05	0.04	0.03	0.02	0.46	0.36	0.28	0.21	0.16	0.12	0.09
	Yes	0.28	0.21	0.16	0.12	0.09	0.07	0.05	0.79	0.68	0.57	0.46	0.36	0.28	0.21

Probability of death at 1 year — 0.0-0.1 | 0.1-0.2 | 0.2-0.3 | >0.3

Fig 1 | Graphical score chart for probability of death for new patients with primary biliary cirrhosis (cirr) based on model derived in box 2 and restricted to patients receiving azathioprine treatment

How to derive a nomogram

The steps to build a nomogram are:

1. For each predictor, calculate the maximum change in the developed model's linear predictor by multiplying the predictor's regression coefficient by the difference between the maximum and minimum value of the predictor in the dataset. Order the predictors by their calculated maximum change.

2. Assign up to 100 points for each predictor. First assign 100 points to the predictor with the largest maximum change as identified from step 1. Call this predictor A. Then provide a points score for the other predictors equal to 100×(maximum change for predictor/maximum change for predictor A).

3. Calculate the minimum and maximum possible total points based on all possible combinations of predictors; project the points onto the probability scale by fitting a prediction model as in box 1 with total points as the only predictor.

Nomograms can be drawn using statistical software programmes such as R (through Harrell's "rms" package) and Stata (by using "nomolog" for logistic regression and "nomocox" for Cox regression).[27][28]

Advantages and disadvantages

The main advantages of nomograms over the other presentation formats is that continuous predictors do not need to be categorised, and multiple time points can be included in a single nomogram by incorporating multiple probability scales based on the possible total points. Additionally, the relative importance of predictors can be judged by the length of the lines within the nomogram. Furthermore, interaction and non-linear terms can be well handled.[8] Complex models, for example those with time dependent predictors, can also be presented in this way.[29] Nomograms can easily be applied away from a computer, especially when a model includes only a small number of predictors.

Nomograms can, however, appear relatively complex at first sight and they require an explanation as to how they should be used (highlighted in the TRIPOD statement[6]). Additionally, they can be inaccurate depending on the size and resolution of the published figure; the larger the number of predictors included in the model, the more challenging the nomogram is to interpret. Rounding of coefficients might also be required.

Application

Figure 2 shows the nomogram associated with the primary biliary cirrhosis example from box 2. To determine the survival probability at a specified time point, the user identifies the points score associated with each predictor value by reading up from that predictor value to the points scale at the top. Once a score has been assigned to each predictor value, a total points score is calculated. Translation from total points to the probability of the outcome is then made by reading down to the associated probability of the outcome from the total points scale. Therefore, using figure 2, it can be seen that an individual aged 55 (24 points), with cirrhosis (42 points), albumin of 34.4 g/dL (65 points), and central cholestasis (62 points), and who has been treated with azathioprine (0 points) has a total points score of 193. This equates to a one year and three year probability of death of 0.40 and 0.85, respectively, which are again similar to the estimates of 0.44 and 0.89 obtained directly from the full regression model. If the individual was not treated (32 points), but all other characteristics were unchanged, the total points score would be 171, which equates to a one year and three year probability of death of 0.25 and 0.68, respectively.

Presentation within websites and mobile apps
Description

Increasingly, prediction models are being made available through a website calculator or within an app for a tablet or smartphone device. These calculators and apps are generally interactive graphical user interfaces that provide individualised risk estimates from an underlying prediction model, which are conditional on the user's inputted predictor values. Often access is free, but sometimes a fee is charged.

How to develop a website

Websites are developed using a building platform or content management system; they also require a domain name and web host. A variety of website building platforms exist, including specific tools that enable statistical software packages to run web apps—for example, Shiny for R and SWire for Stata.[30][31] Websites and apps are available to healthcare professionals and the public, and they can be used by interested individuals from anywhere in the world. These websites and apps can also be designed for use in specific circumstances, such as requiring log in details from registered users, or through a National Health Service server to ensure that the information is delivered to the patient through a healthcare professional.

Websites need to be explicit about the target user and target population, and any website or app should clearly state how to use the model. References to manuscripts describing the model development and subsequent validation (and potentially clinical impact evaluation) should also be provided. The website or app calculator should be checked to ensure that the predicted probability agrees with the predictions from the underlying regression model. For models with continuous predictors, entering values outside the range (of the development dataset) should also be restricted to avoid extrapolation, or at least provide a warning to the user.
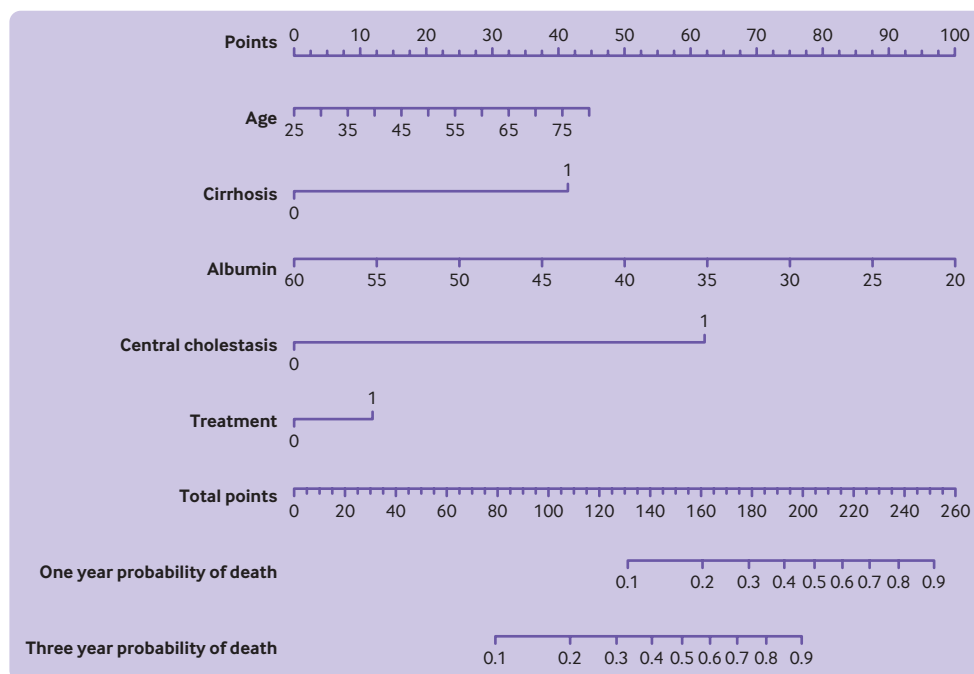

Fig 2 | Nomogram enabling calculation of predicted probabilities of death at one year and three years for new patients with primary biliary cirrhosis. The underlying statistical model is described in box 2

Advantages and disadvantages

A major advantage is that the full model equation can be embedded behind the scenes, and thus no approximation is required, and any complexity is "hidden" from the end user. Websites and apps can provide a user friendly interface in front of complex statistical models, which include large numbers of predictors, non-linear terms, and interactions. Additionally, much of the data input can be automated; for example, in general practice the age and sex of patients are probably already recorded in the medical centre's computer system. Prediction models can be implemented within electronic health records to provide real time feedback to clinicians, although missing data and implausible values might be problematic. Digital applications easily enable switching between units for laboratory results and anthropometrics, such as height recorded in either metres or feet.

Because anyone can create a web calculator, there is currently no assurance that the underlying model is appropriate for use, has been developed adequately, or is validated for the relevant populations accessing the website.[32] Additionally, it is often difficult to know how the model has been translated into the graphical tool. Furthermore, the target user (eg, healthcare professionals or patients) might not be clear, and public access websites could lead to overuse or access by people for whom the model is not intended.

Clearly, access to the internet is required. Data privacy and storage can be a concern, particularly if a website is designed to collect and present data; this concern should be clearly signposted on the website or app. Furthermore, accompanying graphical presentations such as colour coded stick men, smiley faces, or similar, to show the proportion of people predicted to have the outcome depend on the target user and need to be carefully considered.[33 34] Beside graphical presentations, many other metrics can be used in risk communication derived from prediction models. An example is the heart age metric recommended by the European Society of Cardiology.[35 36] Finally, the model might be updated over time to reflect changes in the underlying population characteristics. The web address could also change. Sometimes these changes might not have been tracked. Version control is therefore vital and the reasons for any model update recorded. These changes should be clearly signposted on the website.

Application

An example of a website is Your Heart Forecast tool from New Zealand.[37] This tool provides a graphic design that compares a patient's predicted cardiovascular disease risk with that of the healthy population of the same age. Because risk can be hard for patients to understand, the model provides a graphical depiction of heart age and a future projection that depends on whether the individual does or does not modify their predictors. Other examples include GRACE (for acute coronary events),[38] ASCVS Plus (for atherosclerotic cardiovascular disease),[39] and Predict (for breast cancer).[10] Two examples of websites for primary biliary cirrhosis, but with different predictors than those in this article's running example, are UK-PBC[40] and GLOBE.[41]

Summary

The format of presentation is an important consideration when a clinical prediction model is deemed suitable for use in clinical practice. In addition to providing the full equation (which is essential), there are many ways to present models to aid clinical use, ranging from points score systems and nomograms, to websites and mobile apps. If a model is to be presented in a reduced format (eg, predictors based on categorised values even though they were originally continuous in the full model), this reduced model should undergo the same validation process as the full model before it can be deemed suitable for clinical use.

The best format is user and environment specific, with bedside tools for healthcare professionals requiring different options than patients at home on a computer or tablet. For this reason, means of presentation are best determined through stakeholder engagement, including healthcare professionals and patients. Empirical evidence is now required to determine whether certain formats promote better uptake, use, or understanding. In due course, a similar guide will be required for models developed using advanced or alternative modelling techniques, such as landmarking and machine learning.

1 Steyerberg EW, Moons KGM, van der Windt DA, et al, PROGRESS Group. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLoS Med* 2013;10:e1001381. doi:10.1371/journal.pmed.1001381

2 Bouwmeester W, Zuithoff NP, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med* 2012;9:1-12. doi:10.1371/journal.pmed.1001221

3 Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837-47. doi:10.1161/01.CIR.97.18.1837

4 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;13:818-29. doi:10.1097/00003246-198510000-00009

5 Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36. doi:10.1378/chest.100.6.1619

6 Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63. doi:10.7326/M14-0697

7 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1-73. doi:10.7326/M14-0698

8 Steyerberg EW. Clinical Prediction Models. New York: Springer Science+Business Media, LLC 2009.

9 Asthma UK. [18th August 2017]. www.asthma.org.uk/advice/manage-your-asthma/risk/

10 Public Health England. [18th August 2017]. Available from: http://www.predict.nhs.uk/.

11 Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how?*BMJ* 2009;338:b375. doi:10.1136/bmj.b375

12 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565-74. doi:10.1177/0272989X06295361

13 Krijnen P, van Jaarsveld BC, Steyerberg EW, Man in 't Veld AJ, Schalekamp MA, Habbema JD. A clinical prediction rule for renal artery stenosis. *Ann Intern Med* 1998;129:705-11. doi:10.7326/0003-4819-129-9-199811010-00005

14 Perel P, Prieto-Merino D, Shakur H, et al. Predicting early death in patients with traumatic bleeding: development and validation of prognostic model. *BMJ* 2012;345:e5166. doi:10.1136/bmj.e5166

15 Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90:766-71. doi:10.1093/jnci/90.10.766

16 Van Belle V, Van Calster B. Visualizing risk prediction models. *PLoS One* 2015;10:e0132614. doi:10.1371/journal.pone.0132614

17 Riley RD, Ensor J, Snell KIE, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ* 2016;353:i3140. doi:10.1136/bmj.i3140

18 Christensen E, Neuberger J, Crowe J, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis. Final results of an international trial. *Gastroenterology* 1985;89:1084-91. doi:10.1016/0016-5085(85)90213-6

19 Therneau TM, Grambsch PM. *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag, 2000. doi:10.1007/978-1-4757-3294-8.

20 Groenwold RHH, Moons KGM, Pajouheshnia R, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *J Clin Epidemiol* 2016;78:90-100. doi:10.1016/j.jclinepi.2016.03.017

21 Sullivan LM, Massaro JM, D'Agostino RBSr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med* 2004;23:1631-60. doi:10.1002/sim.1742

22 Moons KGM, Harrell FE, Steyerberg EW. Should scoring rules be based on odds ratios or regression coefficients?*J Clin Epidemiol* 2002;55:1054-5. doi:10.1016/S0895-4356(02)00453-5

23 Conroy RM, Pyörälä K, Fitzgerald AP, et al, SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24:987-1003. doi:10.1016/S0195-668X(03)00114-3

24 Pepe MS. *The statistical evaluation of medical tests for classification and prediction.* 2003. Oxford Statistical Science Series. Oxford University Press.

25 Harrell FE. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer-Verlag, 2001.

26 Van Belle V, Van Calster B, Van Huffel S, Suykens JA, Lisboa P. Explaining Support Vector Machines: A Color Based Nomogram. *PLoS One* 2016;11:e0164568. doi:10.1371/journal.pone.0164568

27 rms: Regression Modeling Strategies [program]. R package version 5.1-2 version, 2018.

28 Zlotnik A, Abraira V. A general-purpose nomogram generator for predictive logistic regression models. *Stata J* 2015;15:537-46. doi:10.1177/1536867X1501500212

29 Kattan MW, Heller G, Brennan MF. A competing-risks nomogram for sarcoma-specific death following local recurrence. *Stat Med* 2003;22:3515-25. doi:10.1002/sim.1574

30 shiny: Web Application Framework for R [program], 2017.

31 SWire Web Apps [program], 2017.

32 Gulland A. Doctors developing apps to try to modernise the NHS. *BMJ* 2018;360:k685. doi:10.1136/bmj.k685

33 Fagerlin A, Wang C, Ubel PA. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics?*Med Decis Making* 2005;25:398-405. doi:10.1177/0272989X05278931

34 Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *Science* 2011;333:1393-400. doi:10.1126/science.1191181

35 Piepoli MF, Hoes AW, Agewall S, et al, ESC Scientific Document Group. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur Heart J* 2016;37:2315-81. doi:10.1093/eurheartj/ehw106

36 Leening MJG, Cook NR, Ridker PM. Should we reconsider the role of age in treatment allocation for primary prevention of cardiovascular disease?*Eur Heart J* 2017;38:1542-7.

37 The University of Auckland. Your Heart Forecast. 2014. https://heart.bmj.com/content/96/9/708

38 University of Michigan Medical School. GRACE ACS Risk Model 2018www.outcomes-umassmed.org/grace/acs_risk/acs_risk_content.html accessed 7th October 2018.

39 American College of Cardiology. ASCVD Risk Estimator Plus 2013 www.acc.org/tools-and-practice-support/mobile-resources/features/2013-prevention-guidelines-ascvd-risk-estimator.

40 UK-PBC. www.uk-pbc.com/resources/tools/riskcalculator/ accessed 18th January 2019.

41 The Global PBC Study Group. www.globalpbc.com/globe accessed 19th January 2019.

Supplementary information: Points score system—worked example