A guide to presenting clinical prediction models for use in the clinical setting

Laura J Bonnett, Kym IE Snell, Gary S Collins, Richard D Riley

Addresses & positions

Laura J Bonnett – University of Liverpool (Tenure Track Fellow); L.J.Bonnett@liverpool.ac.uk; Department of Biostatistics, Waterhouse Building, Block F, 1-5 Brownlow Street, University of Liverpool, Liverpool, L69 3GL; https://orcid.org/0000-0002-6981-9212

Kym IE Snell – Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University (Research Fellow in Biostatistics); k.snell@keele.ac.uk; https://orcid.org/0000-0001-9373-6591

Gary S Collins – Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford (Professor of Medical Statistics); Gary.Collins@csm.ox.ac.uk; https://orcid.org/0000-0002-2772-2316

Richard D Riley – Centre for Prognosis Research, Research Institute for Primary Care and Health Sciences, Keele University (Professor of Biostatistics); r.riley@keele.ac.uk; https://orcid.org/0000-0001-8699-0735

Corresponding Author

Correspondence to: L.J.Bonnett@liverpool.ac.uk

Standfirst

Clinical prediction models estimate the risk of existing disease or future outcome for an individual, conditional on their 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. 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 user.

Introduction

Clinical prediction models estimate the risk of existing disease (diagnostic prediction model) or future outcome (prognostic prediction model) for an individual, conditional on their 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, as outlined in 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 scores for mortality after intensive care admission.[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] However, whilst the TRIPOD Statement highlights the importance of model presentation, there is relatively little information or practical guidance 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 the excellent book chapter (chapter 18) by Steyerberg.[8]

When choosing the format to present a prediction model, researchers should carefully consider the intended user, setting and moment 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?", and to then tailor the presentation format accordingly. Fundamentally, the full model equation should always be presented in the journal publication[6 7]; this is essential to enable an independent external validation. However, additional presentation formats may be required, perhaps outside of the journal article, to enable healthcare professionals to use the model in a particular clinical setting (e.g. where their access to computers or mobile devices is limited). Similarly, the format may need tailoring for a lay person using the model at home (e.g. an asthma patient 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 general public. Patient-public involvement groups and focus groups are useful arenas for this, which aligns with the need for public and participant information and engagement (PPIE) within health research.

Therefore, alongside the full model equation, a range of presentation formats may be required that differ according to the medium by which they are presented (paper versus electronic), the setting in which the models are to be applied (e.g. clinic, bedside, or at home), the level of detail wanted in the predictions (e.g. 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 may 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 context of use, and the intended user. An overview of the different presentation formats is shown in Table 1.

We emphasise that our article is not about how to develop or validate a prediction model,[8 11] or indeed 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, which is for illustrative purposes only, predicting mortality risk over time in those diagnosed with a 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, and many are relevant for prediction of a continuous outcome (e.g. using linear regression).

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 also presenting the full model equation

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

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

Short-term clinical prediction models or diagnostic models

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

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

Here *p* is the probability of having the outcome and $ln\left(\frac{p}{1-p}\right)$ is the log odds of the outcome. α is the intercept term and the baseline log-odds where 'baseline' refers to individuals whose *X* values are all zero. If all X predictors are centred at the mean, α is the log-odds for a person 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 could represent gender and be 1 for males and 0 for females etc. The β terms denote the change in log odds (otherwise known as the log odds ratio) for each 1-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 male compared to a female. Risk (outcome probability) predictions, $\hat{p}$, for a new individual can be estimated by inputting their predictor values into the equation and then transforming back to the probability scale:

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

Clinical prediction models over time

If risks are predicted over time, or at a time-point when some individuals in the dataset have previously dropped out or been lost to follow up (i.e. censored), a prediction model can be developing using a survival model such as a Cox model or a parametric survival model. This has the following form:

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

Here $h(t)$ 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 1-unit increase in the corresponding predictor. 'Baseline' refers to individuals whose *X* values are all zero, or if all predictors are centred at the mean, the underlying hazard rate for a person 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,

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

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:

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

Box 2: Primary Biliary Cirrhosis – example dataset & prediction model

Clinical Context

This data is 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 is publically available.[18 19] Primary biliary cirrhosis is an autoimmune condition which can get worse over time and without treatment may lead to liver failure. A total of 248 primary biliary cirrhosis patients met eligibility criteria for the randomised placebo controlled trial of whom 127 received the intervention and 121 the placebo. There were 57 (45%) deaths 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). Characteristics of these predictors, including the ranges of the continuous predictors, can be seen in Table 2. Treatment was explicitly modelled.[20]

Table 2: Patient characteristics for the primary biliary cirrhosis running example

Continuous Predictor	Range	Mean (SD)	Binary Predictor	Coding	N (%)
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 the 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. I.e. for individual i,

$$LP_i = \big(0.02 \times (Age - 54.8)\big) + \big(1.06 \times (Cirrhosis - 0.285)\big) + \big(-0.06 \times (Albumin - 34.4)\big) + \big(1.59 \times (Cholestasis - 0.179)\big) + (0.31 \times Treatment).$$

The model for probability of death from primary biliary cirrhosis is shown below. Estimates of baseline survival at one ($S_0(1)$) 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 centered, 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. This is in contrast to the more usual case where prediction equations are derived from cohort studies.

Presentation using a points score system

Description

Within points score systems, points are assigned based on the predictor values for a particular 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. They can be presented on a screen (e.g. monitor, iPad) 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 (e.g., using logistic or Cox regression; see Boxes 1 and 2), and the regression coefficients of included predictors are then assigned integer scores which can be negative or positive. Unfortunately, any continuous predictors need to be categorised to facilitate this, sacrificing some predictive accuracy. Categories do not need to be of equal size and by having unequal categories nonlinearity can be more appropriately handled. A summary of the steps to develop a points score system are as follows:

1. Organise the continuous predictors into categories and determine the mid-point 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 just the regression coefficient for that predictor.
4. Define the number of regression units that will correspond to one point in the points scoring system. This 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 is essentially a new risk prediction model that is an approximation of 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), in order 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 are defined on the predicted absolute risk scale.

A full worked example is provided in the supplementary material. For further details, including the mathematical formula associating the risk of outcome with each possible total point score for logistic and survival regression models, see Sullivan et al.[21]

Advantages and limitations

Points score systems are easy to understand following an initial explanation or demonstration, and therefore instructions on how to use it should be described alongside the points score system. Depending on the complexity of the model, and the number of included predictors, paper-based point

score systems can enable risks to be estimated quicker than inputting patient values directly 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 being rounded. Researchers must always check that the predictive performance of the simplified model based on a points score system is similar (and has same potential clinical impact) as the original full model.

Application

Table 3 illustrates a points score system for the primary biliary cirrhosis example. The table on the left assigns scores to categories of each predictor which have been scaled according to a 15-year increase in age (i.e. the regression coefficient for age multiplied by 15), while the table on the right provides probabilities of the outcome that correspond to the points total. For example, an individual aged 55 years (0 points), with cirrhosis (3 points), albumin of 34.4g/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 of 0.90 at three years. These are very 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, suggesting that the simplification led to only small changes in risk predictions from the point scoring system for this individual.

Table 3: Points score system for probability of death for new patients with primary biliary cirrhosis based on the model derived in Box 2.

Predictor	Categories	Points
Age	25-35	-2
	35-45	-1
	45-55	-1
	55-65	0
	65-78	1
Cirrhosis	No	0
	Yes	3
Albumin	20.0-25.0	2
	25.0-30.0	1
	30.0-35.0	0
	35.0-40.0	-1
	40.0-45.0	-2
	45.0-50.0	-3
	50.0-56.5	-4
Central cholestasis	No	0
	Yes	5
Treatment	Azathioprine	0
	Placebo	1

Points total	Probability of death at 1 year	Probability of death 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 a graphical score chart

Description

Graphical score charts are highly-simplified, colour-coded versions of points score systems. As for the points score system, a graphical score chart is a presentation format for a prediction model with intended users of healthcare professionals and patients. It can be used either on screen, or as a print-out. An example adopting this approach is the SCORE model for predicting cardiovascular disease.[23]

How to derive a graphical score chart

First, the probability of the outcome must be calculated for each relevant combination of predictors using the formula in Box 1 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 limitations

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

This presentation will usually require some simplification of the model as it can only accommodate a limited number of predictors and requires continuous predictors to be presented as categories. There is also a loss of information regarding predicted risks as 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 too which is a further disadvantage. As for 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 to the full model.

Application

Table 4 demonstrates a graphical score chart for the primary biliary cirrhosis example. It was created using the point score system shown in Table 3. 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, risk of death at one year is 0.46 for a patient who is aged 55 years, with cirrhosis, central cholestasis, albumin of 34.4g/dl, and treated with azathioprine. This predicted risk is similar to the value of 0.44 estimated from the full model.

Table 4: Graphical score chart for probability of death for new patients with primary biliary cirrhosis based on the model derived in Box 2 – restricted to patients receiving azathioprine treatment

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

Azathioprine treatment		No Central Cholestasis							Central Cholestasis						
		Albumin (g/dl)							Albumin (g/dl)						
Age (yrs)	**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
Over 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

Presentation using a nomogram

Description

Nomograms are another graphical presentation format for a clinical prediction model. As for 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 user of a nomogram is healthcare professionals. They are best used as reference guides potentially on the wards or during a consultation. Like graphical score charts, nomograms can be coloured to aid interpretation.[26]

How to derive a nomogram

The steps to build a nomogram are as follows:

1. For each predictor, calculate the maximum change in the developed model's linear predictor by multiplying the predictors' 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 x (maximum change for the predictor / maximum change for predictor A).
3. Calculate the minimum and maximum possible total points based on all possible combinations of predictors and 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 (via Harrell's 'rms' package) and Stata (via 'nomolog' for logistic regression and 'nomocox' for Cox regression).[27 28]

Advantages and limitations

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 including 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. Also, interaction and nonlinear 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 also easily be applied away from a computer, especially when a model includes only a small number of predictors. However, nomograms can appear relatively complex at first sight and they require an explanation as to how they should be used (as also highlighted in the TRIPOD Statement[6]). Additionally, they can be inaccurate depending on the size and resolution of the published figure, and the larger the number of predictors included in the model, the more challenging the nomogram is to interpret. Rounding of coefficients may also be required.

Application

Figure 1 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 1, we find that someone aged 55 years (24 points), with cirrhosis (42 points), albumin of 34.4g/dl (65 points), central cholestasis (62 points), and treated with azathioprine (0 points) has a total point score of 193.

This equates to a one year probability of death of 0.40 and a three year probability of 0.85, which are again very similar to the estimates of 0.44 and 0.89 obtained directly from the full regression formula. If the individual was not treated (32 points), but all other characteristics were unchanged, their total point score would be 171 which equates to a one and three year probability of death of 0.25 and 0.68 respectively.

Presentation within websites and applications (apps)

Description

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

How to develop a website

Websites are built using a building platform or content management system, and also require a domain name and web host. There are a variety of website building platforms, 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 both healthcare professionals and the general public. They can be used by interested individuals from anywhere in the world, or they can be designed for use in specific circumstances such as requiring log-in details from registered users, or via a National Health Service server to ensure that the information is delivered to the patient via 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/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.

Advantages and limitations

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 could be automated – for example, in general practice the age and gender of the patient is likely to already be recorded in the medical centre's computer system. Prediction models could be implemented within electronic health records to provide real-time feedback to clinicians although missing data and implausible values can be problematic. Digital applications easily enable switching between units for both laboratory results and anthropometrics such as height recorded in either metres or feet.

As anyone can create a web calculator, there is currently no assurance that the underlying model is appropriate for use, has been developed adequately, or 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. The target user (e.g. healthcare professionals or patients) may also not be clear, and public access websites may lead to over-use or access by those for whom the model is not intended.

Clearly, access to the internet is also required. Data privacy and storage can additionally be a concern, particularly if a website is designed to collect as well as present data – this should be clearly signposted on the website/app. Also, accompanying graphical presentations such as colour-coded stick men, smiley faces or similar to demonstrate proportion of people predicted to have the outcome need to be carefully considered depending on the target user.[33 34] Beside graphical presentations there are a large number of other metrics that can be used in risk communication derived from prediction models. Examples include the heart age metric recommended by the European Society of Cardiology.[35 36] Finally, the model may be updated over time to reflect changes in the underlying population characteristics. The web address may also change. In both cases the changes made may not have been tracked. Version control is therefore vital, as is reason for model update, both of which should be clearly sign-posted on the website.

Application

An example of a website is Your Heart Forecast tool from New Zealand.[37] It provides a graphic design that compares a patient's predicted cardiovascular disease risk against that of the healthy population of the same age. As risk can be hard to understand for patients, the model provides a graphical depiction of heart age and provides a future projection conditional 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, although with different predictors to those in this manuscript'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 (e.g. predictors based on categorised values even though they were originally continuous in the full model) then this reduced model should be subject to 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 to 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.

Contributors and sources

GSC had the idea for the article. His research interests are focused on methodological aspects of prediction model development and validation and he led an international collaboration to produce the TRIPOD consensus guidance. The article was written by LJB with feedback and revisions undertaken by all authors. LJB has extensive experience of developing and validating prediction models for people with chronic conditions, and her models have informed the Driver and Vehicle Licensing Agency's guidelines for time off driving for people with epilepsy. LJB is the guarantor for this work. RDR leads the PROGnosis RESearch Strategy that seeks to improve the standards of prognosis

research. KIES works in the field of prognosis research and leads a training course on statistical methods for risk prediction and prognostic models.

License for publication

Funding

This article is independent research arising from a Post-Doctoral Fellowship (Dr Laura Bonnett - PDF-2015-08-044) supported by the National Institute for Health Research (https://www.nihr.ac.uk/). GSC was supported by the NIHR Biomedical Research Centre, Oxford. RDR and KIES are supported by funding from the Evidence Synthesis Working Group, which is funded by the National Institute for Health Research School for Primary Care Research (NIHR SPCR) [Project Number 390]. KIES is also supported by an NIHR SPCR launching fellowship. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, or the Department of Health.

Competing interests declaration

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: LJB had financial support from the National Institute for Health Research for the submitted work; all authors declare no other financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Public and Participant Involvement

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy.

Key Messages

- Presentation format is an important issue for clinical prediction models deemed suitable for use, but receives relatively limited attention within the literature.
- Clear presentation of a prediction model is fundamental to ensure other researchers can independently validate it, and that healthcare professionals and individuals can implement it within healthcare.
- Presentation of the full model equation is essential. In addition, there are many ways to present prediction models for end-users ranging from points score systems and nomograms, to websites and mobile apps.

- The best presentation is user- and environment-specific and is best determined through stakeholder engagement, including patients
- If presentation requires the generation of a simplified version of the full model, then the predictive performance of this simplified model should also be validated and compared to that of the full model.

Acknowledgements

With thanks to the BMJ Manuscript Committee and peer reviewers for their assistance with the development of this manuscript.

Figure Legends

Figure 1: Nomogram enabling predicted probabilities of death to be calculated at 1 and 3 years for new individuals with primary biliary cirrhosis. The underlying statistical model is that described in Box 2.

References

1. Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLoS medicine* 2013;10(2):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 medicine* 2012;9(5):e1001221.
3. Wilson PWF, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97(18):1837-47.
4. Knaus WA, Draper EA, Wagner DP, et al. APACHE II: a severity of disease classification system. *Critical care medicine* 1985;13(10):818-29.
5. Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. *Chest* 1991;100(6):1619-36.
6. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): The tripod statement. *Annals of Internal Medicine* 2015;162(1):55-63. doi: 10.7326/M14-0697
7. Moons KM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): Explanation and elaboration. *Annals of Internal Medicine* 2015;162(1):W1-W73. 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]. Available from: http://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, et al. Prognosis and prognostic research: what, why, and how? *BMJ* 2009;338:b375. doi: 10.1136/bmj.b375 [published Online First: 2009/02/25]
12. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Medical decision making : an international journal of the Society for Medical Decision Making* 2006;26(6):565-74. doi: 10.1177/0272989X06295361
13. Krijnen P, van Jaarsveld BC, Steyerberg EW, et al. A clinical prediction rule for renal artery stenosis. *Annals of Internal Medicine* 1998;129(9):705-11.
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 : British Medical Journal* 2012;345 doi: 10.1136/bmj.e5166

15. Kattan MW, Eastham JA, Stapleton AM, et al. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *JNCI: Journal of the National Cancer Institute* 1998;90(10):766-71.
16. Van Belle V, Van Calster B. Visualizing risk prediction models. *PloS one* 2015;10(7):e0132614.
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 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(5):1084-91.
19. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. New York: Springer-Verlag 2000.
20. Groenwold RHH, Moons KGM, Pajouheshnia R, et al. Explicit inclusion of treatment in prognostic modeling was recommended in observational and randomized settings. *Journal of Clinical Epidemiology* 2016;78:90-100. doi: https://doi.org/10.1016/j.jclinepi.2016.03.017
21. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: the Framingham Study risk score functions. *Statistics in medicine* 2004;23(10):1631-60.
22. Moons KGM, Harrell FE, Steyerberg EW. Should scoring rules be based on odds ratios or regression coefficients? *Journal of Clinical Epidemiology* 2002;55(10):1054-55. doi: https://doi.org/10.1016/S0895-4356(02)00453-5
23. Conroy RM, on behalf of the Spg, Pyörälä K, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal* 2003;24(11):987-1003. doi: 10.1016/S0195-668X(03)00114-3
24. Pepe MS. The statistical evaluation of medical tests for classification and prediction: Medicine 2003.
25. Harrell FE. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer-Verlag 2001.
26. Van Belle V, Van Calster B, Van Huffel S, et al. Explaining Support Vector Machines: A Color Based Nomogram. *PloS one* 2016;11(10):e0164568.
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(2):537-46.
29. Kattan MW, Heller G, Brennan MF. A competing-risks nomogram for sarcoma-specific death following local recurrence. *Statistics in medicine* 2003;22(22):3515-25.
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 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? *Medical decision making* 2005;25(4):398-405.
34. Spiegelhalter D, Pearson M, Short I. Visualizing uncertainty about the future. *science* 2011;333(6048):1393-400.
35. Piepoli MF, Hoes AW, Agewall S, et al. 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). *European heart journal* 2016;37(29):2315-81.
36. Leening MJ, Cook NR, Ridker PM. Should we reconsider the role of age in treatment allocation for primary prevention of cardiovascular disease? *European heart journal* 2017;38(20):1542-47.

37. Auckland TUo. Your Heart Forecast. 2014
38. University of Michigan Medical School. GRACE ACS Risk Model 2018 [Available from: https://www.outcomes-umassmed.org/grace/acs_risk/acs_risk_content.html accessed 7th October 2018.
39. American College of Cardiology. ASCVD Risk Estimator Plus 2013 [Available from: https://www.acc.org/tools-and-practice-support/mobile-resources/features/2013-prevention-guidelines-ascvd-risk-estimator.
40. UK-PBC. [Available from: http://www.uk-pbc.com/resources/tools/riskcalculator/ accessed 18th January 2019.
41. The Global PBC Study Group. [Available from: https://www.globalpbc.com/globe accessed 19th January 2019.