SAS Journal of Surgery (SASJS) **ISSN 2454-5104**
Abbreviated Key Title: SAS J. Surg.

A Unit of Scholars Academic and Scientific Society, India

An Overview of the Process of Building and Validating Predictive Models

Mohammed Ali Kazem M.D, CABS, FRCSEd (Gen Surg), MMedSci*
Surgery and Cancer Division, Leighton Hospital, Middlewich Rd, Crewe CW1 4QJ, UK

Review Article

***Corresponding author**
Mohammed Ali Kazem

Article History
Received: 13.03.2018
Accepted: 24.03.2018
Published: 31.03.2018

DOI:
10.21276/sasjs.2018.4.3.2

Abstract: Predictive models are utilised in clinical practice and decision-making processes in cancer patients. The uptake of these models is not as common as other tools used in other clinical settings. This could be due to lack of understanding of the process of building and validating these tools. Aims: To help clinicians to understand the process of building and validating predictive models. Methods: A literature review and practical example from our published work is used to explain the process of building and externally validating a nomogram for colon cancer survival. Results: The nomogram building process should include clear steps to identify the population and outcomes of interest, the predictors, the model used to build the nomogram and the validation process. Externally validating the tool in a different population is important to ensure the tool's generalizability and its ability to predict the outcomes. There are factors to consider when externally validating any model; one important factor is the number of events in the population used to validate the model. Conclusions: Understanding the process of building and validating any predictive model is an important step for any clinician planning to utilise a predictive tool in clinical practice. This will help to ensure that only tools relevant to the target population and intended outcomes are used.
Keywords: Predictive, Models, Nomogram, Building, Validation, Colon Cancer.

INTRODUCTION

Building a Nomogram

In a previously published review article [1], we discussed the application of predictive models in cancer patients, their types, characteristics, and their limitations.

Understanding the process of building and validating predictive tools is important if we are to use these tools in our practice and decision-making process.

Nomograms are widely used tool to predict different outcomes in cancer patients [2-4]. If we are to use nomograms to predict outcomes in our patients, it is crucial to have knowledge of the process of building and validating these tools.

Although we are using nomograms to explain the process, most of the essential steps are common to the building of all predictive models, and this will help to understand the process of building other predictive tools.

The process of building a nomogram consists of five steps [5]:

- Identify the population of interest
- Identify the outcome of interest
- Identify the predictors
- Identify the model to build the nomogram
- Validation

All steps will be discussed, but we will be discussing the validation process in details as understanding this process will help us to apply the tool in our daily practice appropriately. Validating the tool is crucial to know if it is applicable to our population, and if the predications it provides are valid or not.

Identify population of interest

Identifying the data source and target population is the first step in the process of building a nomogram. Having clear and relevant inclusion and exclusion criteria is important if the tool is to be generalizable and used in a different population. The data could be from a single centre, multicentre, or population cohort. The latter two will help to produce a more generalizable tool, but problems may occur with difficulties in collecting required elements of data, and lack of consistency in the data itself [5].

Identifying the right cohort means that the right question was asked at the beginning of the process, clearly specifying the affected cohort and the problem the tool is aiming to assist with. Taking these steps will reduce the likelihood of type III error [6].

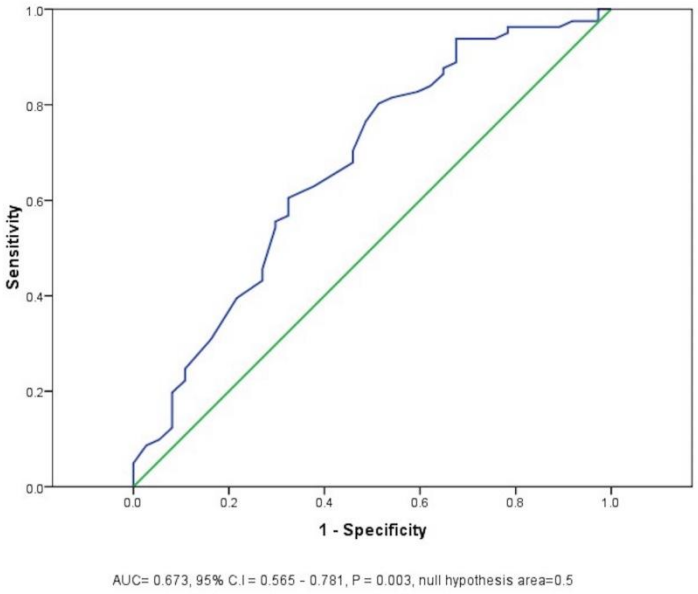

Fig-1: ROC curve for five years nomogram

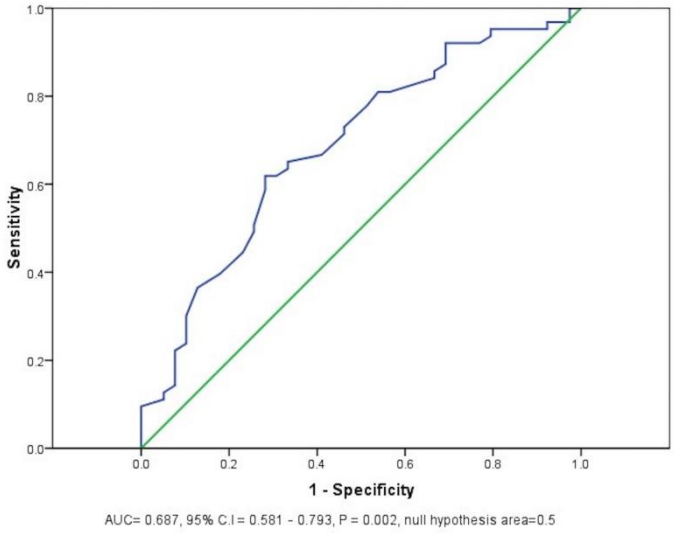

Fig-2: ROC curve for ten years nomogram

Identify the outcome of interest

Having a clear outcome to predict is crucial to building any predictive tool. The outcome is usually an event that occurs as part of the disease process, such as death, recurrence, or spread of disease. Nomograms are usually used to predict the probability of a specific outcome. Keeping this outcome clear and always linked to the predictive variables will help to reduce type III error [5, 6].

Identify the predictors

Identifying the variables that will affect the outcome and may help to predict its occurrence is an important step before building the model. These predictive variables should be easy to collect and relevant to clinical practice, as this will make the usability of the model better.

There are two ways to use the variables in building predictive models. In the full model approach all the variables that are identified will be included in the building process, which will reduce selection bias, and avoid over fitting. But this approach is not always practical or possible. Another way of using predictive variables is by choosing the ones which are most significant in affecting the outcome (significance testing), but this will lead to the risk of bias and often these models have a problem with over fitting [7].

Choosing the model to build the nomogram

Once the desired outcome has been set and the variables affecting it decided, a statistical model is then used to actually build the nomogram. Nomograms use different statistical models to express the prediction of the outcome of interest. Choosing the model depends on

what the nomogram trying to predict. If the question is death vs survival, or cure vs recurrence, in other words a binary outcome, the best model to use is logistic regression. When we wish to assess time taken for an outcome to occur or when outcome is linked to time, e.g. five-year survival, a different statistical model, the Cox proportional hazard model, is required [5,7].

Validation

Validation refers to the process of assessing the performance of a predictive model using a new set of data. It should include a comparison of observed and predicted outcome rates in patient groups (calibration), and an assessment of the ability of the model to identify the patient who will or will not develop the outcome of interest (discrimination). This process can use new data from the same cohort involved in building the model or, preferably, a new data set from different cohort of patients [8].

There are three main methods used to validate predictive tools, which include: internal validation, temporal validation and external validation.

Internal Validation

Utilising the same data set used to construct the model is a common way of validating it. If the data set is large, or from different centres, then it can be randomly split into two groups. One group will be used to build the model, and the other group will be used to validate the model. The problem with this method is that the two groups of data are likely to be very similar, as they have been derived from the same cohort, so that outcome predictions in the validation group are more likely to be favourable. Using nonrandomised splitting may help reduce this effect [8].

If the data set is limited, data re-sampling methods need to be applied to validate the model. In general, data re-sampling works by choosing part of the data set to build the model and the remaining samples will be used to assess its performance. This process is repeated multiple times and the results collected and summarised [6]. Different methods for re-sampling exist, but the main difference between them is how the samples are chosen. The two main re-sampling techniques used are cross-validation and bootstrapping.

- Cross validation: Both k fold cross validation and leave-one-out cross validation are ways to assess a model's performance. In k fold validation the sample is randomly split into k number of groups (usually five or ten) with almost the same size, one group is left out and the rest of the groups are used to fit the model. The process is then repeated multiple times (at least 200), taking it in turns to leave each group out [5,6].

  In leave-one-out cross validation the k is the number of observations. In this technique as one sample is left out each time, the final assessment of performance is calculated using the k observations held out [6].
- Bootstrap validation: in this technique, the bootstrap sample is the same size as the original data set, but with replacement. This means some observations may be present in that sample multiple times, and others none at all. The observations that are not selected are used later to assess the performance of the model built using the bootstrap data set. This process is repeated multiple times, ensuring at least two thirds of the observations are represented each time [5,6].

Temporal validation:

In temporal validation the data set used for assessing performance of the model is collected prospectively from the same centre as the original data set. The new cohort of patients may share common characteristics with the original cohort, and same clinical practices will have been used in their management. However, it is still a new data set independent from the old one, and it could be argued that it is external in relation to time of collection [8].

External validation:

The only way to assess generalizability of any model, including a nomogram, is by using a different data set collected from a different cohort of patients. In external validation the data set could be retrospective as long as the variables required for the nomogram are available [5,8]. This is useful especially when long term follow up is required, such as when five- or ten-year survival is the desired outcome, but can be complicated by changes in practice and available treatments, which may affect patient outcomes.

Validating a Nomogram, a practical example:

Having an understanding of the validating process is important for any clinician planning to utilise a predictive model in their practice. Explaining this process using a practical example will help to simplify the steps and enable clinicians to be able to assess if the tool has been validated appropriately to use in their cohort of patients.

Our team externally validated the MSKCC colon cancer disease free survival nomogram [9]. We will go through the steps required to validate the nomogram and use our published work [10] as an example.

Understanding the Nomogram:

Having a clear understanding of the nomogram we are planning to use or validate is important; this should include what the nomogram is predicting, the cohort used to build it, the variables required and their relevance, and its validation process.

The MSKCC nomogram predicts the probability of five and ten year recurrence-free survival

after undergoing a curative resection for colon cancer. The cohort used to build the nomogram was identified from patients who had curative resection of colon cancer (TNM stage I to III), with no evidence of distant metastases, in Memorial Sloan Kettering Cancer Centre (MSKCC) between January 1990 and December 2000 (total of 1320 patients). The tool was internally validated using bootstrapping, and performance was assessed by performing a concordance index and calibration curve [9].

The variables included in the tool are age, sex, tumour location, pre-operative CEA level, tumour differentiation, number of positive lymph nodes, number of negative lymph nodes, lymphovascular invasion, perineural invasion, depth of tumour penetration into the colon wall, and whether the patient received chemotherapy or not [9,11].

Choosing the cohort:

Identifying the appropriate cohort to use in the validation process is essential to ensure we assess the generalizability and the accuracy of the tool.

Our cohort was patients who underwent an elective curative colonic cancer resection, and have been followed up for up to 10 years. Making sure that this cohort is as close to the original nomogram cohort is essential, as the results have to reflect how the nomogram will predict the outcomes in another cohort from the one used to build it (assessing generalizability).

Management of the cohort in relation to inclusion and exclusion criteria has to be clear and reflects what the nomogram is assessing. In our case we made sure to exclude patients who did not fit with the model (emergency resections), and any patients who did not have all the predictor variables available. When it came to validating the model we excluded any patients who did not complete the follow up period (five or ten years) either as result of being lost to follow up or if they died from non-cancer causes.

Establishing strict inclusion criteria helped to ensure that our cohort was as close to the original cohort as practically possible. This meant that the tool would function in our cohort similarly to how it functioned in the original cohort, improving the efficacy of the validation process.

Validating the model

The model (a nomogram), predicts five and ten years disease free survival following curative colon cancer resection. The main aim of our work was to assess the ability of the MSKCC nomogram to predict disease free survival and identify patients at high risk of developing recurrence. To do this we used the receiver operator characteristic (ROC) curve, and calculated the area under the curve (AUC).

The ROC curve was developed by the British engineers during world war two to assess the ability of radar receivers to discriminate between German planes and other false signals like flocks of birds or friendly planes. Since then the ROC curve has been used to assess the ability of a diagnostic test to discriminate between patients who do or do not have a disease [12].

A ROC curve is a graph representing sensitivity (true predictive rate) on the y-axis and 1-specificity (false predictive rate) on the x-axis for different cut off points of test value [13]. In our case the test value was the nomogram predictions for five and ten year disease free survival. A ROC curve shows the inverted relationship between sensitivity and specificity, so as the test sensitivity increases its specificity decreases and vice versa [14]. The purpose of a ROC curve is to help identify the best cut off point for a test at which it shows the most discriminatory result [13].

The AUC represents the performance of a test and its ability to distinguish between patients who have or do not have a disease [13]. In our case it showed the ability of the nomogram to predict correctly five and ten year recurrence free survival. The larger the AUC the better the performance of the test; if AUC equals one this will mean the test has the ability to always identify the outcome. As the AUC gets smaller the performance of the test reduces [12].

In our published study [10], to validate the model we used a ROC curve and calculated the AUC for the five years recurrence free survival predictions of the nomogram against the actual recurrence free survival of our cohort. A ROC curve was plotted (Fig 1) and AUC was calculated (AUC= 0.673, 95% C.I = 0.565 – 0.781, P = 0.003, null hypothesis area=0.5).

Similarly, the ROC curve was plotted (Fig 2) and AUC was calculated (AUC= 0.687, 95% C.I = 0.581 – 0.793, P = 0.002, null hypothesis area=0.5), for the ten years recurrence free survival predictions of the nomogram against the actual recurrence free survival of our cohort.

The above process demonstrated the ability of the nomogram to predict the outcome it has been built for. The closer the AUC to one the more sensitive the test is, as AUC gets lower the poorer the performance of the tool is. In our case the AUC for five and ten year disease free survival were 0.673 and 0.687 respectively, which is considered acceptable.

Limitations of validation process

We need to remember that validating any model in a different cohort always has its limitations. Differences in the validating cohort compared with the original cohort will affect the outcomes and predictions of the tool. However, validating the model in a new

cohort is the best way to assess the generalizability of the predictive tool [15].

In our case, a significant difference between the original cohort and the validating cohort was that the original cohort was from the United States and our validating cohort from the UK. This mean that the cohorts were treated in two different health systems, which may have affected clinical decision-making and available treatment options. Despite these differences the nomogram performance was acceptable and we think it was valid in our cohort.

In our case, another limitation to consider is the size of our cohort. Having strict inclusion and exclusion criteria helped to make the cohort very similar to the cohort used to build the nomogram, but also made the cohort size smaller which may have affected the assessment of its performance [15].

Validation study versus impact study

The main aim of a validation study is to assess the performance of the predictive model, and examine whether the model could be used with a new cohort. On the other hand, an impact study aims to assess the effect that using the model has on daily clinical practice, whether it changes doctors' behaviour, and if it improves patient outcome [16].

Impact studies differ from validation studies in their design. Validation studies can be either prospective or retrospective, they do not need a control group, and they report on predicted outcomes and model performance. In contrast, impact studies are before and after studies with randomisation, they will require a control group of doctors not using the model, and long term follow up is not required if the study is only looking to assess change in clinicians' behaviour. Impact studies report on change in behaviour, patient outcomes, and cost-effectiveness [16-18].

CONCLUSIONS

This review provides clinicians with a degree of knowledge of the building and validating process of predictive tools. This knowledge is crucial if the clinician is to use any predictive tool.

It is for clinicians to decide if any predictive tool they are going to utilise in their clinical practice is appropriate for use and fits with their cohort of patients. To ensure this is the case one needs to review the original building process, including the original cohort and its characteristics, predictor variables and how relevant are they for the predicated outcomes, and the validation process. External validation of a tool in a different cohort will give the clinician more confidence in its use, however, it is important to understand the limitations of the validation process, namely the appropriateness of the cohort used and the number of events included.

REFERENCES

1. Kazem MA. Predictive Models in Cancer Management: A Guide for Clinicians http://dx.doi.org/10.1016/j.surge.2016.06.002.
2. Bochner BH, Kattan MW, Vora KC: Postoperative nomogram predicting risk of recurrence after radical cystectomy for bladder cancer. J Clin Oncol 2006; 24:3967-3972
3. Kattan MW, Karpeh MS, Mazumdar M, Brennan MF. Postoperative nomogram for disease-specific survival after an R0 resection for gastric carcinoma. J Clin Oncol. 2003;21:3647–50
4. Marrelli D, De Stefano A, de Manzoni G, Morgagni P, Di Leo A, Roviello F. Prediction of recurrence after radical surgery for gastric cancer: a scoring system obtained from a prospective multicenter study. Ann Surg. 2005;241:247–55
5. Lasonos A, Schrag D, Raj G, Panagea K. How to Build and Interpret a Nomogram for Cancer Prognosis. J Clin Oncol 26:1364-1370
6. Kuhn M, Johnson K. Applied Predictive Modeling. NewYork; Springer Science; 2013
7. Royston P, Moons KG, Altman DG, et al: Prognosis and prognostic research: Developing a prognostic model. BMJ 2009, 338:b604
8. Kattan M. Nomograms are superior to staging and risk grouping systems for identifying high-risk patients: Preoperative application in prostate cancer. Curr Opin Urol 2003;13:111– 6
9. Weiser MR, Landmann RG, Kattan MW. Individualized prediction of colon Cancer recurrence using a nomogram. J Clin Oncol 2008; 26:380-5
10. Kazem M A, Khan A U, Selvasekar C R. Validation of nomogram for disease free survival for colon cancer in UK population: A prospective cohort study, https://doi.org/10.1016/j.ijsu.2015.12.069
11. Colorectal cancer nomogram: post-surgery. http://nomograms.mskcc.org/Colorectal/PostSurgery.aspx. Accessed February 2018
12. Receiver Operating Characteristic Curves, http://ebp.uga.edu/courses/Chapter%204%20-%20Diagnosis%20I/8%20-%20ROC%20curves.html, accessed February 2018
13. Kumar R, Indrayan A. Receiver Operating Characteristic (ROC) Curve for Medical Researchers. Indian Paediatrics;2011;48; 277-287. http://medind.nic.in/ibv/t11/i4/ibvt11i4p277.pdf
14. Tape TG, Interpreting diagnostic tests. http://gim.unmc.edu/dxtests/Default.htm, accessed February 2018
15. Collins G. S, Ogundimu E. O, Altman D. G. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. DOI: 10.1002/sim.678
16. Moons KG, Altman DG, Vergouwe Y, et al: Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009, 338:b606

17. Meyer G, Kopke S, Bender R, Muhlhauser I. Predicting the risk of falling—efficacy of a risk assessment tool compared to nurses' judgement: a cluster-randomised controlled trial. BMC Geriatr 2005;5:14
18. Marrie TJ, Lau CY, Wheeler SL, Wong CJ, Vandervoort MK, Feagan BG. A controlled trial of a critical pathway for treatment of community-acquired pneumonia. JAMA 2000;283:749-55.