Nomograms in oncology: more than meets the eye

Vinod P Balachandran , Mithat Gonen, J Joshua Smith, Ronald P DeMatteo

Nomograms are widely used as prognostic devices in oncology and medicine. With the ability to generate an individual probability of a clinical event by integrating diverse prognostic and determinant variables, nomograms meet our desire for biologically and clinically integrated models and fulfill our drive towards personalised medicine. Rapid computation through user-friendly digital interfaces, together with increased accuracy, and more easily understood prognoses compared with conventional staging, allow for seamless incorporation of nomogram-derived prognosis to aid clinical decision making. This has led to the appearance of many nomograms on the internet and in medical journals, and an increase in nomogram use by patients and physicians alike. However, the statistical foundations of nomogram construction, their precise interpretation, and evidence supporting their use are generally misunderstood. This issue is leading to an under-appreciation of the inherent uncertainties regarding nomogram use. We provide a systematic, practical approach to evaluating and comprehending nomogram-derived prognoses, with particular emphasis on clarifying common misconceptions and highlighting limitations.

Lancet Oncol 2015; 16: e173–80

Department of Surgery (V P Balachandran MD, J J Smith MD, R P DeMatteo MD) **and Department of Epidemiology and Biostatistics** (M Gonen PhD), **Memorial Sloan Kettering Cancer Center, New York, NY, USA**

Correspondence to:
Dr Vinod P Balachandran, Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA
balachav@mskcc.org

Introduction

Nomograms are a pictorial representation of a complex mathematical formula.[1] Medical nomograms use biological and clinical variables, such as tumour grade and patient age, to determine a statistical prognostic model that generates a probability of a clinical event, such as cancer recurrence or death, for a particular individual. There are two primary ways nomograms are used. One is pictorially, where each variable is listed separately, with a corresponding number of points assigned to a particular magnitude of the variable. Then, the cumulative point score for all the variables is matched to a scale of outcome (figure 1A). Alternatively, the formula is contained in a computer or smartphone based calculator, where specific variables are entered and the likelihood of an event is computed.

The gold standard for prognostication in oncology remains the tumour, node, metastasis (TNM) staging system. Proposed in 1953[3] as a common language for solid tumour prognosis, it is rooted in the Halstedian principle of temporal determinism. This proposes that solid tumours spread sequentially, first from the primary site to the lymphatic system, and then to distant organs. Patients are hence classified by both anatomical spread of disease and survival. However, the TNM system has several drawbacks. First, it is constrained by requiring a correlation between anatomical disease progression and increasing stage progression. Hence, patients with equivalent anatomical spread yet variable outcomes (recurrence or survival) are forced into the same stage, introducing heterogeneity. Second, TNM staging is unable to incorporate tumours, nodes, or metastases as continuous variables. This creates a system with a finite number of stages, complicating the determination of an individual patient's prognosis. Third, the TNM system links prognosis to descriptive, not determinant, variables—it purely states that if you are anatomically further along in the course of your disease, your prognosis will be worse, without incorporating other variables that govern prognosis, such as genetic differences, tumour mitotic rate, or histology.

In view of the limitations of TNM staging, nomograms have emerged as a simpler, yet more advanced method. One of the primary advantages of nomograms is their ability to estimate individualised risk on the basis of patient and disease characteristics. Proponents cite that nomograms can also incorporate continuous variables and relevant determinants of disease into prognosis,[4–7] are user-friendly, and are better than clinician judgment in estimating disease course.[8–10] In oncology, nomograms have the potential to affect all aspects of cancer care. Preoperative nomograms estimating the risk of positive surgical margins[11] and lymph node metastases,[12–14] could assist clinicians in identifying patients who might derive greater benefit from more extensive surgery. Postoperative nomograms estimating recurrence,[2,15,16] cancer-specific survival,[17–19] overall survival,[20–22] benefit of adjuvant therapies[23–25] and the effect of treatment on quality of life,[26,27] might assist patients and physicians alike in all aspects of decision making. Although nomograms represent a major advance in the development of prognostication methods, their proper clinical application needs a thorough understanding of the nomogram-specific question, study population, method of construction, and outcome, to clearly assess its applicability to a particular patient's clinical scenario. Additionally, the ability to interpret nomogram performance and assess specific limitations is essential to appropriately counsel patients on the meaning, accuracy, and assumptions embedded in nomogram risk estimations. Here, we consider the rationale for the use of nomograms, clarify essential components of their construction, interpretation, and application, and highlight common misconceptions.

Construction

The question, the study population, and the outcome

The most important step in nomogram construction is to identify a good question—ie, one addressable by a nomogram (panel). Nomograms are best derived to answer a focused, clinically relevant question that needs a mathematical model for the answer. Not all clinical

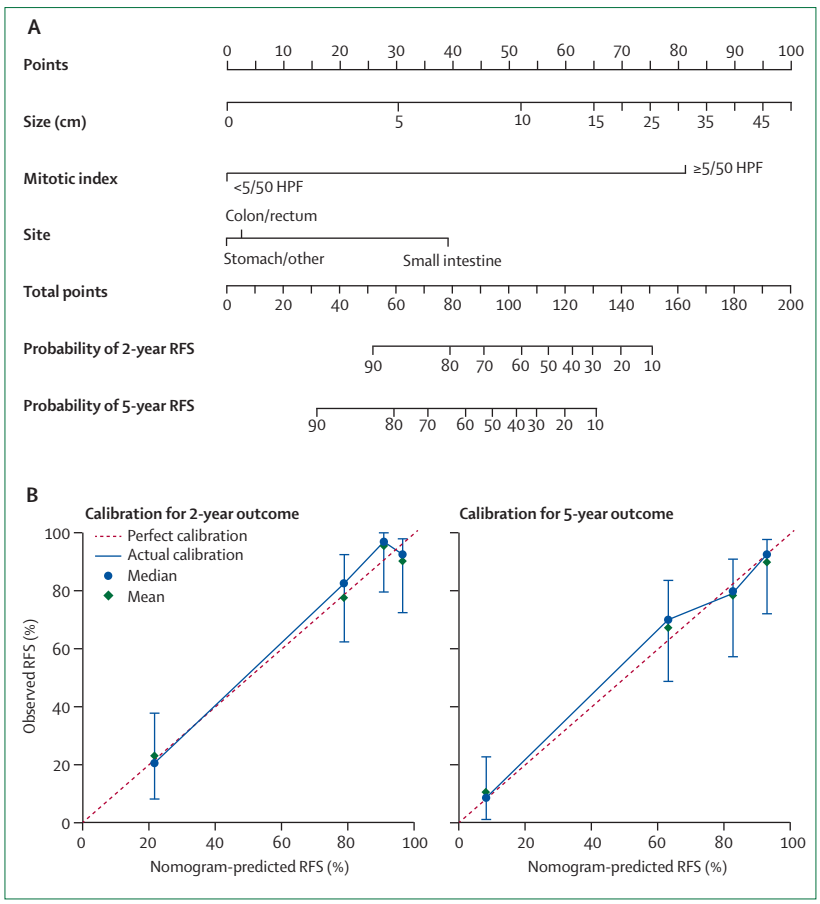

Figure 1: **Use and interpretation of a nomogram**
(A) An example of a nomogram—estimating RFS in resected primary gastrointestinal stromal tumour. Draw an upward vertical line from the covariate to the points bar to calculate points. Based on the sum of the covariate points, draw a downward vertical line from the total points line to calculate RFS.[2] (B) Calibration curves of a nomogram estimating RFS in resected primary gastrointestinal stromal tumour showing the 2-year and 5-year outcome. Error bars=95% CIs.[2] RFS=recurrence-free survival. HPF=high power field.

questions need a nomogram—for instance, in view of the scarcity of benefit showed with routine nasogastric tube decompression[28] and their decreasing routine use in clinical practice, a nomogram to estimate nasogastric tube insertion distance would have little value.[29] Next, the patient cohort that will be used to derive the nomogram is selected. It should be representative of the general population with the disease, and its definition transparent so readers can assess its applicability to their patients. Single institution cohorts might have more complete datasets, yet they might be biased by institutional practice patterns, which can be overcome by the use of multi-institutional or national databases. Next, the outcome defining the question is chosen—typically various types of recurrence (local, distant, or both) or survival. Attention should be paid to disease-specific survival that reflects the natural history of a patient's disease versus overall survival, which reflects the cumulative effect of competing diseases and age on a patient's survival. The primary outcome should have a clear, well accepted definition, and be easily and reproducibly measured.

Method

The next step involves the selection of variables (covariates) that could determine the outcome based on a priori clinical hypotheses. This approach avoids exclusion of covariates based on incomplete data and selection purely based on statistical significance. Covariates could be tumour specific, such as tumour size, depth of penetration, lymphovascular invasion, and patient specific, such as age and sex. Treatment per se should be avoided as a covariate unless there are validated data from a randomised clinical trial.

After variable selection, one must choose a statistical model. The most common model for fitting Kaplan-Meier survival curves is the Cox proportional hazards model. The Cox model generates a hazard function, h(t) (failure rate at time t for patients surviving to time t), as a function of the covariates. It estimates the number of new events in unit time among the population at risk, by contrast with a logistic regression model that evaluates the proportion of new events per unit time in the entire population. A logistic regression can be used when a single timepoint (eg, 5-year survival) is of interest and all the patients who are alive have a follow-up beyond that timepoint. After a statistical model is selected, multivariate analyses are done to measure the association between the covariates and the outcome, adjusting for all the other variables in the model. Covariate inclusion in a multivariate analysis should follow Harrell's guideline (the number of events should exceed the number of covariates by at least ten-times).[30] However, inclusion of more covariates does not necessarily lead to higher accuracy, but instead to overfitting, and should be avoided. The model is then derived using the formula: probability of event at time,

$$t{=}S_0(t)^{\exp(\beta 1x1+\beta 2x2\ldots)}$$

where β are the regression coefficients and x are the reported values of the covariates. $S_0(t)$ is called the baseline survival function and is also estimated from the data. Regression coefficients are used to construct the variable axes in the nomogram and S_0 is used in the translation from total points to predicted probability (figure 1).[31]

Performance

Validation

Validation is the process of testing the model in different populations to obtain unbiased estimates of model performance (discrimination, calibration, and clinical usefulness) and judging its applicability to these populations. External validation, preferably in many disparate datasets, is the gold standard and should be obtained whenever possible. Unfortunately, most nomograms (including those at our own