EDITORIAL COMMENT

Filipe Leite [1]
[1] Portuguese Oncology Institute of Porto – Thoracic Surgery Department

Concerning the TNM 9th edition - Go forth and validate

It has been almost ten years since the Eight Edition of the TNM Classification of Lung Cancer was published. This period has seen an enormous change in every area of diagnosis and treatment of lung cancer.

As of January 2025, the new, Ninth Edition of the TNM Classification of Lung Cancer will take effect.

It is indeed a notable occasion.

Still, we must consider one especially important key point.

The TNM classification is a risk model.

The first variables identified were Tumour, mediastinal lymph Nodes, and Metastasis.

Over time, with every subsequent edition and the enlargement of the patient database, more and more variables were found, and smaller and smaller differences became clear.

Nowadays, these span from gene mutations to cell histology, tumour location, lymph node stations and volume, pattern of spread and of metastatic dissemination.[1]

The process of creating a risk model is not random and involves several steps, with Steyerberg et al. considering seven fundamental ones: [2]

Step 1: Problem definition and data inspection
Step 2: Coding of predictors
Step 3: Model specification
Step 4: Model estimation
Step 5: Model performance
Step 6: Model validity
Step 7: Model presentation

As seen above, one of the steps is determining the model validity. This is usually composed of two parts: internal and external validation.

Internal validation refers to the validity of the statements made within the same population from which the sample originated, in other words, it concerns the reproducibility, or the capacity to replicate with accuracy the

results within the same population. [3]

Nevertheless, after the model presentation a curious phenomenon tends to happen. We assume almost without questioning that the model is applicable to all populations, even if they are completely different from the one used to create it.

This is where external validation enters. It addresses the potential to extend conclusions drawn from a sample of one original population to another population that is plausibly related.

Once again, Steyerberg et al. offers us a four-point blueprint on how to perform a correct external validation: [2]

Point A: Alpha: calibration-in-the-large.

Calibration refers to the agreement between observed endpoints and predictions. In other words, if we predict a 2% risk that a patient submitted to a pneumectomy will die within 30 days, the observed proportion should be approximately 2 deaths per 100.

with such a prediction.

Point B: Beta: calibration slope.

This term has evolved over time. It began as a measure of "spread". Over time it was correlated with "discrimination", before finally acquiring this interpretation as a measure of calibration.

The calibration slope evaluates the spread of the estimated risks and has a target value of 1. A slope < 1 suggests that estimated risks are too extreme, i.e., too high for patients who are at high risk and too low for patients who are at low risk. A slope > 1 suggests the opposite, i.e., that risk estimates are too moderate. The calibration intercept, which is an assessment of calibration-in-the-large, has a target value of 0; negative values suggest overestimation, whereas positive values suggest underestimation.

Still the slope does not by itself measure calibration and should not be presented independently of the previous "calibration-in-the-large'". [4-5]

Point C: Concordance statistic: discrimination.

This refers to the ability of a model to distinguish a patient with the endpoint, for example death, from a patient without it.

A caveat here: a model that predicts the same risk for all patients, equal to the actual incidence, has perfect calibration, but zero discrimination, as it cannot tell patients apart.

For binary endpoints, this is equal to the concordance (c), or the area under the receiver operating characteristic (ROC) curve, which plots the sensitivity (true-positive rate) against 1 – specificity (false-positive rate) for consecutive cut-offs for the predicted risk.

Point D: Decision-curve analysis.

The previous points, whilst important, do not assess the clinical usefulness of the model.

Decision curve analysis is a statistical method that evaluates models and tests in terms of their clinical consequences. This is unlike traditional accuracy measures - such as the area-under-the-curve or Brier score - which do not consider events such as, for instance, it being worse to miss a cancer (false negative) than do an unnecessary biopsy (false positive). Decision curve analysis evaluates the net benefit of a model or test in comparison to the two default strategies of treat all patients and treat no patients.[6]

We are ill-prepared and equipped to go through all these steps, but others have done it, and we should follow suit.[7-8]

Nevertheless, we do not have to master the finer points.

But, to interpret a study, or our own results, we must know of these steps' existence, and if we want to apply a model to our own population, we need to know how to use them.

Let us finish as we started.

It has been almost ten years since the Eight Edition of the TNM Classification of Lung Cancer was published. This period has seen an enormous change in every area of diagnosis and treatment of lung cancer.

As of January 2025, the new, Ninth Edition of the TNM Classification of Lung Cancer will take effect.

It is indeed a notable occasion.

But this is no justification to mindlessly start using this powerful new weapon.

We need and should assess it. Not due to hubris or simple disagreement. But because it gives it strength.

By doing so we assess its reproducibility and transportability, leading to generalizability, which when paired with accuracy are the foundations for the TNM Classification of Lung Cancer and the scientific discussion for the upcoming decade.

So, use the Ninth Edition of the TNM Classification of Lung Cancer and see if the predicted results match your observed data.

Go forth and validate.

REFERENCES

1. Detterbeck FC, Asamura H, Rami-Porta R, Rusch VW. The Only Constant Is Change: Introducing the International Association for the Study of Lung Cancer Proposals for the Ninth Edition of TNM Stage Classification of Thoracic Tumors. J Thorac Oncol. 2023;18(10):1258-60.
2. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J. 2014;35(29):1925-31.
3. Justice AC CK, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med. 1999;130(6):515-24.
4. Stevens RJ, Poppe KK. Validation of clinical prediction models: what does the "calibration slope" really measure? J Clin Epidemiol. 2020;118:93-9.
5. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic t, et al. Calibration: the Achilles heel of predictive analytics. BMC Med. 2019;17(1):230.
6. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making. 2006;26(6):565-74.
7. Cai J, Li Y, Yang F. Caucasian validation of downstaging from IIB to IIA in T1N1M0 patients within the 9th edition of the non-small cell lung cancer tumor-node-metastasis staging. Cancer Med. 2024;13(14):e70018.
8. Wang RR, Li MJ, Peng Q, Huang ZY, Wu LL, Xie D. Validation of the 9th edition of the TNM staging system for non-small cell lung cancer with lobectomy in stage IA-IIIA. Eur J Cardiothorac Surg. 2024;65(3).