EDITORIAL – THORACIC ONCOLOGY

The Devil Is in the Details

Frank C. Detterbeck, MD

Division of Thoracic Surgery, Yale University School of Medicine, New Haven, CT

Wankhede[1] has performed a systematic review and meta-analysis of studies validating the eighth edition of the tumor-node-metastasis (TNM) classification of lung cancer, which took effect in 2017. It is important to validate that stage classification actually works when applied in different independent datasets. Indeed, the International Association for the Study of Lung Cancer (IASLC) Staging and Prognostic Factors Committee (SPFC) that developed the eighth edition of lung cancer stage classification explicitly called for others to conduct such studies. Therefore, this is an important undertaking, especially because work on the development of revisions to the eighth edition are underway, with the nineth edition scheduled to replace it in 2024.

First, a few general words must be said about stage classification of lung cancer. A classification system is designed to assemble patients into relatively homogeneous groups and thus provide a foundation whereby interventions can be studied and communicated. Thus, classification is fundamentally a nomenclature that must suitable for use consistently worldwide, hence the importance of external validation. Although evolution of the classification is inevitable, the nomenclature and definitions must remain stable for a period of time, and allow translation from a previous version or to a new version as much as possible.

The study by Wankhede has several strengths.[1] It is timely. Wankhede's[1] effort to conduct a systematic review and perform a meta-analysis is noteworthy. A strength of

this effort is its recognition that the key measure of the stage classification is its ability to discriminate between groups.

On the surface, the effort by Wankhede[1] appears to be an impressive, timely, and needed analysis of lung cancer stage classification. However, scratching below the surface reveals a number of issues regarding this meta-analysis. The authors state that they analyzed only patients with non-small cell lung carcinoma (NSCLC). However, Table 2, which presents the "main characteristics" of the included studies, reports numbers that include both small cell lung carcinoma (SCLC) and NSCLC patients in the source papers. Perhaps Table 2 merely misrepresents what was included, but other inconsistencies undermine confidence in this assumption. Table 2 lists the country represented by the study, which most people would interpret as the country represented by the patients involved in the study. However, the "Egypt" study was conducted by an Egyptian investigator, but used only the SEER database, which is a U.S. database. Table 2 also lists the Chansky et al.[2] study as involving two different U.S. populations. Whereas the Chansky et al.[2] paper reports data from the U.S. National Cancer Database (NCDB), the other cohort is the from the IASLC database, which is primarily an Asian and European database with only 5% the patients from the United States.

A fundamental flaw of the Wankhede[1] study is its lack of independent datasets required for validation. The "USA-a" cohort that Wankhede[1] includes from the Chansky et al.[2] paper was a summary of the IASLC analysis that produced the tumor-node metastasis (TNM) classification (it was provided only for comparison in the paper that focused on using the NCDB as an independent external validation set). It certainly is inappropriate to use the original population used to derive a classification for an independent validation. Furthermore, approximately 90% of the patients included in the Wankhede[1] meta-analysis are from the NCDB, but this largely uses the same patients

twice. The Chansky et al.[2] USA-b cohort consisted of NCDB patients from 2000–2012, and the Yang et al.[3] study used NCDB patients from 2004–2013. Thus, about 60% of the patients in the Wankhede[1] meta-analysis are inappropriate because they are not independent patients. Despite the statement of Wankhede[1] in the Methods section that the "criteria used to define duplicate data included study period, hospital, treatment information, and any additional inclusion criteria," the study does not examine the patients included in much detail.

Another problem is the lack of clarity about what the study is validating: clinical or pathologic stage. The prognoses for these populations are fundamentally different. Wankhede[1] provides no information about what was being analyzed. In defense of Wankhede,[1] some of the source papers did not specify this. But without knowing this, it really is impossible to interpret a comparison between some mixture of clinical and pathologic stages in the seventh edition and another mixture in the eighth edition and to draw conclusions about which classification is better. The discrimination should be between adjacent groups within either the clinical stage or the pathologic stage. Wankhede[1] compares each group with stage 1A. This raises another methodologic issue that leaves me uncomfortable regarding how to interpret the findings.

I commend Wankhede[1] for focusing on discrimination (assessed using the C-index), but it seems that the study incompletely distinguishes ability to predict prognosis from discrimination. The study refers to the C-index as a measure of prognostic value, whereas it is a measure of discrimination. I fully acknowledge that this is a confusing area. In developing the eighth edition of the TNM classification of lung cancer, the SPFC used prognosis as a tool to define how to separate tumors into groups. But because prognosis varies according to many factors (e.g., geographic region, clinical versus pathologic stage, treatment administered, type of source database), it is inappropriate to use the actual prognosis. The SPFC instead required consistent discrimination within multiple subsets (e.g., clinical, pathologic, histotype, region, N0 only, resected patients or all patients, R0 only). This sort of analysis is beyond the scope of a study analyzing external validation such as Wankhede[1] has undertaken. But an understanding of the difference between calibration (prediction of prognosis) and discrimination is important.

A classification system is inherently different from a prognostic prediction model.[4] A classification must be relatively static and universal if it is a nomenclature that allows clear communication. A prognostic model should be constantly changing, reflecting advances in real time as prognosis changes. It should be specific to predict the prognosis for patients in different locations, recognizing that the prognosis for the patients in one country or after one treatment differs from that of another. Finally, stage classification refers strictly to the anatomic extent of the tumor, whereas prognosis is determined by many factors, including anatomic extent of a cancer, biomarkers of that cancer's aggressiveness, patient-related factors (e.g., age, competing causes of death), structural factors (e.g., access to care, quality of care), and treatment-related factors.

I commend Wankhede[1] for asking a very important and timely question, and for the effort in doing an exhaustive statistical analysis. However, the reporting misses details that I think are more important than statistical output. The Wankhede[1] study does not report details of selection and data abstraction according to standards for systematic searches and meta-analyses.[5, 6] The IASLC methodology paper outlined parameters for reporting well-done external validation studies,[7] but perhaps Wankhede[1] was not aware of this. At a minimum, a study should demonstrate clarity about the number of patients analyzed, and more importantly, the number of events (deaths) analyzed. What type of validation does the study address (e.g., historic, geographic, spectrum, methodologic)? Most important, I think, is the critical thinking about underlying assumptions in the performance of an analysis. Two factors are crucial to any study: whether the authors have thought their assumptions through critically to avoid fundamental flaws and whether the study explicitly informs the reader regarding the reasons for the way the analysis was structured.

To me, this study is an example of how we often go astray through statistics. The process of systematic review and meta-analysis sounds impressive, and the 55 figures and tables of statistical output in the supplementary material are truly impressive. But I do not know how to interpret this if it is not clear what exactly the study is analyzing (i.e., type of stage), the nature and appropriateness of the patients it included, and what is needed to do an independent external validation. I recognize the intention and valiant effort Wankhede[1] has made, but I think it is unclear what the study findings actually mean.

REFERENCES

1. Wankhede D. Evaluation of eighth AJCC TNM sage for lung cancer NSCLC: a Meta-analysis. *Ann Surg Oncol.* 2020. https://doi.org/10.1245/s10434-020-09151-9.
2. Chansky K, Detterbeck FC, Nicholson AG, et al. The IASLC lung cancer staging project: external validation of the revision of the TNM stage groupings in the eighth edition of the TNM classification of lung cancer. *J Thorac Oncol.* 2017;12:1109–21.
3. Yang L, Wang S, Zhou Y, et al. Evaluation of the 7(th) and 8(th) editions of the AJCC/UICC TNM staging systems for lung cancer in a large North American cohort. *Oncotarget.* 2017;8:66784–95.
4. Detterbeck FC. Stage classification and prediction of prognosis: The difference between accountants and speculators. *J Thorac Oncol.* 2013;8:820–2.

5. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6:e1000097.
6. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1.
7. Detterbeck F, Chansky K, Groome P, et al. The IASLC lung cancer staging project: methodology and validation used in the development of proposals for revision of the stage classification of non-small cell lung cancer in the forthcoming (8th) edition of the TNM classification of lung cancer. *J Thor Oncol.* 2016;11:1433–46.