EDITORIAL

Outcome Prediction and the Future of the TNM Staging System

Harry B. Burke

The prediction of patient prognosis has always been essential to the practice of medicine. By the early 20th century, Halsted *(1)* and others believed that solid tumors spread contiguously over time through a series of stages, from the primary tumor site, through the lymphatics, to distant organs, with each stage conferring an increasingly poor prognosis. A corollary of this view, supported by later research, was that, at diagnosis (clinical tumor–node–metastasis [TNM] stage) or after surgery (pathologic TNM stage), tumor size or location (T), regional lymph node involvement (N), and distant metastases (M) were indices of disease spread and could be used to predict patient outcome.

In 1953, the French surgeon Pierre Denoix proposed to the Union Internationale Centre le Cancer that these three factors be standardized and integrated into a prognostic system that could be used, with some accommodation for anatomic site, across all solid tumors *(2)*. His proposal for a common language of solid tumor prognosis was adopted as the TNM staging system, which is currently used throughout the world. The TNM system has undergone six revisions and, in the United States, these changes have been guided by the American Joint Committee on Cancer (AJCC), which was established in 1959 and which has published a succession of revisions of its *AJCC Cancer Staging Manual (3)*.

The TNM staging system is a "bin model"; the TNM prognostic factors are used to create a mutually exclusive and exhaustive partitioning of patients, so that every patient is in one and only one bin, and the bins are grouped together into larger bins called stages *(4)*. It uses the mean survival of the patients already in the bin to predict what will happen to a new patient placed in that bin. For example, if a new patient is placed in the {T1, N0, M0} bin, then that patient's 5-year disease-specific survival is predicted to be the same as the mean survival of all the patients who were placed in that bin 5 years ago. The utility of the system arises from its ability to order patients by a decreasing probability of survival. It can be used for selecting patients for therapy and for providing patients with an estimate of their prognosis.

Until the sixth edition *(3)*, the TNM stages for colon cancer had not changed substantially in 35 years. The stages in the first AJCC staging manual, published in 1976 *(5)*, were virtually identical to those in the fifth edition *(6)*, except for T2 having moved from stage II in the first edition to stage I by the fifth edition. In this issue of the Journal, O'Connell et al. *(7)* report that the sixth edition of the *AJCC Cancer Staging Manual* is substantially different from the fifth edition in terms of stages II and III. In the fifth edition, stage II was composed of the bins {T3, N0, M0} and {T4, N0, M0}. In the sixth edition, stage II is divided into two substages—IIa, which is the bin {T3, N0, M0}, and IIb, which is the bin {T4, N0, M0}. This change indicates that IIb patients will, on average, have a worse prognosis than stage IIa patients and a better prognosis than stage IIIa {T1–T2, N1, M0} patients. What O'Connell et al. found was that IIb patients had a worse prognosis than IIIa patients. This finding violates the rules of the stage model upon which the TNM system is based. The use of the colon cancer staging in the sixth edition may result in uncertain clinical trials and potentially incorrect patient therapy. Therefore, until the advent of the seventh edition, clinicians and researchers should not use the sixth edition's colon cancer staging; rather, they should rely on the fifth edition, and the AJCC should insert an erratum into the sixth edition.

Why did this problem arise? The TNM staging system is based on a temporal model. It assumes the contiguous spread of disease over time (temporal determinism), and it measures time. If a bin with a better prognosis is placed below a bin with a worse prognosis, the result will be the observed crossover in prognosis. How might this temporal error occur? The TNM staging system does not provide information regarding the natural history of the cancer. Rather, it has, since its inception, been a surgical system. The ordering of the patients assumes that all the patients will undergo surgery and no other therapy. The ordering does not take into account which patients received chemotherapy. If one places, in a separate bin, T4 colon cancer (stage IIb), and if there were an effective adjuvant therapy for lymph node–positive (stage IIIa) colon cancer, then the worse survival of the stage IIb and the improved survival of the stage IIIa patients could result in the observed survival crossover.

The future utility of the TNM staging system depends on its ability to deal with the increase in population screening for cancer, the discovery of new therapies, and the use of the new molecular (genomic and proteomic) biomarkers. First, the predictive accuracy of the TNM system depends on patients presenting across the entire temporal range of the disease. Because the system depends on the temporal progression of tumors for its predictive accuracy, anything that reduces the temporal dimension reduces its accuracy. The early detection of disease because of screening results in a shift in populating the stages to earlier disease so that the majority of patients are in stage I or at most stages I and II. This stage compression reduces the accuracy of the TNM staging system.

Adding to the stage compression problem are the trends of very small surgical specimens and to neoadjuvant therapy. Both these changes in the clinical approach to disease further reduce

Correspondence to: Harry B. Burke, MD, PhD, George Washington University School of Medicine, 2150 Pennsylvania Ave NW, Ste. 2-105, Washington, DC 20037 (e-mail: hburke@mfa.gwu.edu).

See "Notes" following "References."

DOI: 10.1093/jnci/djh293
Journal of the National Cancer Institute, Vol. 96, No. 19, © Oxford University Press 2004, all rights reserved.

the size of tumors, the detection of nodal metastases, and the accurate grading of tumors due to changes in histology. Further, the viability of the temporal model has been called into question by empirical data; for example, some patients with small tumors and no evidence of lymph node involvement die of their disease earlier than patients who present with more advanced disease *(8)*.

Second, the TNM staging system does not take into account neoadjuvant therapy, antihormonal therapy, chemotherapy, or the new, targeted, monoclonal molecular therapies. Thus, changes in survival related to these therapies are not reflected in the TNM stages. For the TNM system to incorporate new therapies into its stages, it must stratify by therapy *(9)*. Unfortunately, adding additional bins for each therapy will cause an exponential increase in the number of bins and in the number of patients required to populate the bins. If staging does not incorporate new therapies, it will not be useful in predicting patient outcomes beyond what is predicted to happen if the patient undergoes only surgery.

Third, the TNM staging system has difficulty dealing with continuous biomarkers and adding new biomarkers. It has been shown that a computer-based statistical program that allows the T and N variables to remain continuous is more accurate statistically signficant at predicting 5- and 10-year disease-specific survival for breast and colon cancer than the TNM staging system *(10)*. Further, it is not clear how the TNM staging system would accommodate nonlinear or interactional biomarkers. In addition, just as stratifying by treatment leads to a proliferation of bins, so too does the addition of biomarkers. For example, if one adds to the TNM stages for colon cancer tumor location (eight anatomic locations) and histologic grade (four grades), one would increase the number of bins from 30 bins *(6)* ($5\ T \times 3\ N \times 2\ M = 30$) to a number that could range from a minimum of 180 bins (30×3 locations $\times$ 2 high/low grades $= 180$) to a maximum of 960 bins (30×8 locations $\times$ 4 grades $= 960$). Because the primary utility of the TNM staging system is its simplicity, it is a "look-up" table in which one looks up the stage of disease; the organization and use of a system with 180–960 bins is an impossible task. This problem is compounded by our need to add new molecular biomarkers.

We must move from a model of temporal determinism to one of biological determinism. Carcinogenesis is not defined by what stage the patient is in at detection but rather by the molecular (genomic and proteomic) characteristics of the tumor and the host. Biological determinism takes the view that the anatomic location of the disease at detection is more related to our methods of detection than to the tumor itself. Thus, all patients are at risk of metastatic disease; some are further along a biological metastatic pathway at detection than others. In this view, treatment should be driven by the molecular biology of the tumor or host and not the tumor's location at detection.

Clinical medicine desperately needs to be able to answer three questions *(11)*. 1) Does this individual patient need any therapy; i.e., is the natural history of the disease sufficiently poor to warrant therapy? 2) If this patient has a poor prognosis without therapy, which therapy, combination of therapies, or succession of therapies will provide the highest probability of survival and the best quality of life? This point requires discovering new therapies and finding therapy-specific prognostic factors that predict an individual's response to the therapy. 3) What toxic effects and side effects will this patient experience with each therapy, and are there any measures that can be taken to mitigate or eliminate these therapeutic harms? These questions cannot be answered by the current TNM staging system. Personalized medicine, to be clinically relevant, requires validated molecular biomarkers that can be combined in a computer-based system *(4)* that provides individual-patient therapy predictions.

References

(1) Halsted WS. The results of radical operations for the cure of carcinoma of the breast. Ann Surg 1907;46:1–19.

(2) Gospodarowiez M, Benedet L, Hutter RV, Fleming I, Henson DE, Sobin LH. History and international developments in cancer staging. Cancer Prevent Control 1998;2:262–8.

(3) Greene FL, Page DL, Fleming ID, Fritz AG, Balch CM, Haller DG, et al., editor. AJCC cancer staging manual. 6th ed. New York (NY): Springer, 2002.

(4) Burke HB, Henson DE. Criteria for prognostic factors and for an enhanced prognostic system. Cancer 1993:72;3131–5.

(5) Committees and Task Forces of the American Joint Committee for Cancer Staging and End Results Reporting. Classification and staging of cancer by site: a preliminary handbook. Chicago (IL): American Joint Committee on Cancer, 1976.

(6) Fleming ID, Cooper JS, Henson DE, Hutter RV, Kennedy BJ, Murphy GP, et al., editors. AJCC cancer staging manual. 5th ed. New York (NY): Springer, 1997.

(7) O'Connell JB, Maggard ME, Ko CY. Colon cancer survival rates with the new American Joint Committee on Cancer sixth edition staging. J Natl Cancer Inst 2004;1420–5.

(8) Fisher B, Gebhardt MC. The evolution of breast cancer surgery: Past, present, and future. Semin Surg Oncol 1978;5:385–94.

(9) Burke HB, Hoang A, Iglehart JD, Marks JR. Predicting response to adjuvant and radiation therapy in early stage breast cancer. Cancer 1998; 82:874–7.

(10) Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr, et al. Artificial neural networks improve the accuracy of cancer survival prediction. Cancer 1997;79:857–62.

(11) Burke HB, Henson DE. Evaluating prognostic factors. CME J Gynecol Oncol 1999;4:244–52.

Note

I thank Donald E. Henson, MD, and William E. Grizzle, MD, PhD, for their helpful and constructive comments.