****The Clinical Implications of Integrating Additional Prognostic Factors into the TNM****

By:

Donald Earl. Henson

Division of Cancer Control and Epidemiology

The George Washington University Cancer Institute

Washington, DC 20037

[henson@gwu.edu](mailto:henson@gwu.edu)

Arnold M. Schwartz

Department of Pathology and Surgery

The George Washington University Medical Center

Washington, DC 20037

aschwartz@mfa.gwu.edu

Dechang Chen

Division of Epidemiology and Biostatistics

Uniformed Services University of the Health Sciences

Bethesda, MD 20814

[dechang.chen@usuhs.edu](mailto:dechang.chen@usuhs.edu)

Dengyuan Wu

Department of Computer Science

School of Engineering

The George Washington University

Washington, DC 20052

andrewwu@gwmail.gwu.edu

CORRESPONDING AUTHOR

Donald E. Henson, M.D.

The George Washington University Cancer Institute

2030 M Street, NW, Room 4064

Washington, DC 20036

Telephone-202-994-1966

FAX-202-994-1714

Email-henson@gwu.edu

WORD COUNTS:

Abstract= 144

Main Text= 1344

No. of References= 12

KEYWORDS: TNM, Prognostic Factors, Combinations, Staging, Survival Rates

**Abstract**

**Background and Objectives**: The survival and management of solid tumors are governed by host and tumor factors that traditionally have incorporated TNM staging with additional pathologic, biologic, and clinical host factors. Beyond the anatomic-based TNM, increasingly new prognostic and predictive factors are being discovered that have important survival and treatment implications. However, because the TNM staging system is based on a “bin” model, additional prognostic factors would rapidly overwhelm the current system. This communication demonstrates the clinical implications derived from a new algorithmic model based on clustering analysis that integrates any number and type of additional factors into the TNM.

**Methods:** A new algorithm is described that integrates additional factors into the TNM and calculates survival.

**Results:** The results indicate that additional factors can be integrated into the TNM staging system that provides additional patient stratification without changing the TNM definitions.

****Conclusions:** Integration of additional prognostic factors into the TNM by a clustering algorithm can change the stratification of patient outcome. This may guide the clinician to select a more rational management program based on the additional factors and improve cohort selection for clinical trials.**

****Introduction****

The prognosis and management of solid tumors is based on the assessment of pathologic, molecular, and epidemiologic factors that govern tumor biology. These prognostic and predictive factors include clinical presentation and functional status, tumor diagnosis and extent of disease, and biologic factors that are prognostic for survival and predictive of clinical response to therapy. Traditionally, TNM staging, based on the anatomic factors of tumor size or extent, nodal involvement, and metastatic spread, was sufficient to provide basic information for tumor assessment, treatment, and prognosis. However, given the rapid development of new prognostic and predictive factors and the limitation of the “bin” nature of the TNM, additional parameters would rapidly overwhelm the TNM. (1) Nonetheless, these newer factors are proving crucial for an expanded appreciation of tumor biology and treatment decisions and thus need to be integrated into the staging database to enhance prognostic evaluation and management decisions.

We have generated an algorithmic model based on clustering analysis that incorporates these additional factors into traditional TNM staging. (2, 3). Consequently, it provides for greater stratification of tumor stages and enhances the prognostic assessments within reference to a given anatomic stage. The algorithm enables the clinician to appreciate and quantitate the effect of additional parameters on the traditional staging system. In this communication, we demonstrate that integrating additional prognostic factors into the TNM can change the stratification of disease outcome, which has implications for the interpretation and conduct of clinical trials. For example, the algorithm predicts that higher stage cancers with favorable pathologic and molecular features may show improved survival relative to lower stage tumors with less favorable pathologic and molecular profiles.

For more than 50 years, the TNM has served as a standard classification for the anatomic extent of disease for patients with cancer.(4-9). Universally accepted, the TNM is actively promoted by professional organizations interested in cancer and results of treatment. However, cancer is no longer solely characterized by anatomic extent of disease, but by a combination of host and biologic factors. Thus, it has become increasingly apparent that the TNM needs to be expanded to accommodate additional factors. It is important to assess whether these factors add benefit by providing more accurate estimations of recurrence and death. (9-12). Any expansion, however, should preserve the TNM because of its long history in documenting disease extent in patients with cancer. (7, 9).

****Materials and Methods****

****Data Source****

The source of data was the file “Case Listings” in the SEER Program for breast cancer 1990-1998 to provide for 10 years of survival through 2008. Lung cancer data were obtained from 1992 through 2001 to allow for 5 years of survival. A total of 273,231 cases of invasive breast carcinoma were obtained and 266,646 cases of lung cancer. Cases with data missing in histological tumor type, histological grade, survival time and vital status were excluded. Inflammatory breast cancers and other T4 tumors were not considered in the analysis. Metastatic tumors to other visceral organs and sites were excluded and only M0 cancers were taken into account.

**The Algorithm**

Our clustering algorithm was composed in the programming language “R”, an open source code available on the Internet. (2, 3) It offers several advantages over the current TNM in that it can integrate any number of prognostic factors, including molecular markers, into the TNM. The algorithm, however, differs from the standard TNM staging system in the following ways. First, it calculates the survival rates for every combination of prognostic factors in any order and provides a relationship between survival and the factors. If, for example, there are 4 T categories, 3 N categories, and 3 grades selected for breast cancer, then there will be 36 (4x3x3) combinations of prognostic factors and 36 survival curves. Second, by comparing and combining clusters of prognostic variable and disease outcome, the algorithm can aggregate a larger number of a survival rates into asmaller number of more comparable groups. For instance, the 36 survival rates could be aggregated into four stage groupings by differences in survival rates. Fundamentally, this represents collapsing the survival rates into a staging system instead of the individual TNM variables, which is the current practice. This clustering procedure circumvents the limitation on the number of factors that can be integrated into the TNM. Third, the effect of any factor or any combination of factors on outcome can be evaluated. Fourth, it takes into account censored survival times of cancer patients. This is in contrast to traditional algorithms for clustering data that cannot be directly applied to censored data. Finally, and most importantly, the algorithm does not change the TNM definitions, thus preserving traditional TNM classification schemes. The results of the algorithm have been validated within the SEER database by analyzing and comparing the outputs for different range of year cohorts and appreciating that the discriminatory features are maintained. The results also demonstrate an internal consistency when comparing larger and smaller cohorts of input prognostic variable on TNM survival. Although not demonstrated, the algorithm also calculates the hazard.

****Results****

The algorithm was extensively evaluated using data from the SEER Program of the National Cancer Institute because of its large size of clinical cases and outcomes. While SEER does not specify treatment protocols, treatment information can be incorporated into the algorithm if available. Presumably, nearly all patients listed in SEER have been treated.

Assessing the output of the clustering algorithm with prognostic factors on TNM designation provided additional insight for survival with clinical implications (Figures 1 & 2). Breast cancer cohorts with Stage T2N2 have a defined survival pattern (Figure 1A), but clearly represent a family of tumors with a spectrum of histologic grades, hormonal status, and other host and tumor characteristics. The survival curves separate markedly when grade and ER (estrogen receptor) status are integrated into the TNM.

In Figure 1A, all the curves represent tumors that are T2N2, yet the absolute survival difference between the subgroups exceeds 0.4 despite the identical TNM tumor stage. The result is that less favorable prognostic tumors may have a 50% decreased survival relative to better tumors with favorable prognostic indicators within the same TNM stage. One can clearly appreciate that if a similar approach were taken for T1N1, then favorable prognosis T2N2 tumors would have an improved survival relative to poor prognostic T1N1 tumors. This shift has implications for the conduct and interpretation of clinical trials. In the presence of significant nodal involvement, T1N2 tumors, the comparative stratification of tumors according to histologic grade provides meaningful separation of survival outcomes (Figure 1B). The confidence interval is inversely proportional to the number of SEER identifiable cases; however, there exists statistical relevant stratification among the tumor grades even at an advanced breast cancer stage.

The same pattern may be shown to exist for lung cancers (Figure 2). Favorable Stage I tumors of the lung (T1N0) may be stratified according to grade with a wide resultant survival pattern at 5 years. Additional factors may be considered and may demonstrate additional upward or downward shift indicative of an important prognostic variable. Favorable limited stage tumors may be treated with resection alone, whereas, unfavorable survival statistics, similar to more advanced cancers, may suggest cases for clinical trial considerations. In assessing factors, which do not produce a significant change in the survival plot, it is clear that these represent redundant, non-informative prognostic variables. Although the data are not shown, changes in survival rates were also noted with other prognostic variables such as age, racial/ethnic background, and histological tumor type. It seems therefore that survival rates are relative and depend on the selection of prognostic factors.

****Discussion****

The implications of these findings are clear when generating comparative cohorts for clinical trials. Cohort designation of tumor stage, despite randomization, may include a variety of prognostic subgroups that impact survival even in the absence of treatment. An uneven and unknown distribution of these prognostic factors among cohort groups of the same stage will either enhance or blunt true differences in outcome so that the main effect of treatment may become difficult to detect. This consideration becomes important when prognostic implications may be large and trial result absolute survival differences are relatively small. Pretrial outcome calculations for each combination of prognostic factors in comparative cohorts minimize the variation among treatment groups. Stratifying patients according to multiple and similar prognostic factors provides more homogeneous patient populations for clinical trials. Given a more heterogeneous patient population among comparative cohorts, the less likely a true treatment effect will be detected. Finally, this algorithmic approach provides a quantitative assessment, combining the results of TNM staging analysis with specific prognostic factors, which offer the opportunity for focused physician-patient discussions and more accurate outcome prediction.

****Conclusions****

The clustering analysis approach provides for a greater stratification of tumor stages and enhances prognostic assessments within a given anatomic stage.The implications of these findings offer insight into the spectrum and stratification of survival outcomes within a given TNM stage and demonstrate an enhanced standardization of cohort comparisons for clinical trials. Patient stratification depends on a combination of prognostic factors each of which contributes to survival. The improved stratification utilizing this enhanced staging system, based on incorporation of prognostic variables within TNM, may guide the clinician into more rational management of patients with favorable and unfavorable biologic activity.

REFERENCES

1. Burke HB, Henson DE. Criteria for prognostic factors and for an enhanced prognostic system. Cancer 1993;72:3131-3135.
2. Chen D, Xing K, Henson D, et al. Developing Prognostic Systems of Cancer Patients by Ensemble Clustering. *Journal of Biomedicine and Biotechnology,* Volume 2009, Article ID 632786, 7 pages, doi:10.1155/2009/632786.
3. Ran Qi, Wu D, Sheng L, Henson D**,** Schwartz AM, Xu E, Xing K, Chen D. On an ensemble algorithm for clustering cancer patient data. BMC in press.
4. Harmer M. The British clinical staging of breast cancer. Brit. Med J 1958;1:767-769.
5. Copeland MM Committee on Cancer Staging and End Results Reporting – Origin, Objectives, and Program. Bull Amer. Coll. Surg. 1962;235-237.
6. Sellers AH. The clinical stage classification of malignant tumors. Canad Med Assoc J 1966;95:313-315.
7. [Gospodarowicz M](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Gospodarowicz%20M%22%5BAuthor%5D), [Benedet L](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Benedet%20L%22%5BAuthor%5D), [Hutter RV](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Hutter%20RV%22%5BAuthor%5D), et al. History and international developments in cancer staging. [Cancer Prev Control.](http://www.ncbi.nlm.nih.gov/pubmed/10470455) 1998 6:262-8.
8. Cancer Staging Manual, American Joint Committee on Cancer. Editors: Edge SB, Byrd DR, et al., 7th Edition, Springer 2010.
9. Gospodarowicz M, Mackillop W, O’Sullivan B, et al. Prognostic factors in clinical decision making. Cancer (suppl) 2001;91:1688-1695.
10. Piccirillo JF, Feinstein AR. Clinical symptoms and comorbidity: significance for the prognostic classification of cancer. Cancer 1996;77:834-842.
11. Winer EP, Carey LA, Dowsett M, Tripathy D. Beyond anatomic staging: are we ready to take the leap to molecular classification. Am Soc Clin Oncol 2005; 41st Ann. Meeting.
12. [Takes RP](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Takes%20RP%22%5BAuthor%5D), [Rinaldo A](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Rinaldo%20A%22%5BAuthor%5D), [Silver CE](http://www.ncbi.nlm.nih.gov/pubmed?term=%22Silver%20CE%22%5BAuthor%5D), et al. Future of the TNM classification and staging system in head and neck cancer. [Head Neck.](http://www.ncbi.nlm.nih.gov/pubmed/20191627) 2010;32:1693-711.

**Note:** The opinions expressed herein are those of the authors and do not necessarily represent those of the Uniformed Services University of the Health Sciences or the Department of Defense.

.

**Legends**

**Figure 1A**: Comparison of breast cancer patient survival rates based on the number of factors evaluated. One curve represents the survival rate according to the T and N (T2N2) only, and the other two according to combinations of 5 factors, which include the same T and N categories. The integration of 5 factors splits the T and N survival rates by additional “good” and “bad” factors. The differences of all curves are statistically significant at the 95% confidence limits as shown by the confidence bands. The combined 5 factors split thestage assignment at 5 years for T2N2, Grade 1, ER+, Age>50 and stage III for T2N2, Grade 3, ER-, Age>50. Grade 3 in the plot represents a combination of cases listed as grade 3 or grade 4. The T and N survival rate splits because the additional factors select patients from the population with better or less favorable outcome. The T and N survival rate alone represents a non-selected composite of patients of the selected factors.

**Figure 1B.** The effect of the histological grade on T1N2 cancers of the breast. The 3 grades split the TNM stage assignment into 3 significantly different survival rates. Note that grade remains prognostic even in the presence of category N2 nodes. In T1N2 cases, grade 3 significantly reduces the survival compared to grade 1.

**Figure 2:** Five-year survival rates of male lung cancer patients after histological grades 1, 2, and 3 have been integrated into the TNM. Note how the grade splits the T and N survival curve. T1N0 indicates stage I, which implies a survival of greater than 75% at 5 years according to TNM staging. Cases labeled T1N0, grade 1 and T1N0, grade 3 are significantly different from T1N0, grade 2 cases and from overall T1N0 cases according to 95% confidence limits. Cases labeled T1N0, grade 2 are not significantly different from cases assigned T1N0 only. T1N0, grade 3 is comparable to stage II, which has a survival of 50%-74%. Thus, integrating the histological grade into the TNM not only splits the T1N0 survival rate for lung cancer, but also may change the AJCC stage assignment. Data were SEER lung cancer cases for 1992-2001 for men with adenocarcinomas only. Survival rates were calculated by the Kaplan-Meier procedure.