

# Comparison of Chi-Squared Automatic Interaction Detection Classification Trees vs TNM Classification for Patients With Head and Neck Squamous Cell Carcinoma

F. Xavier Avilés-Jurado, MD, PhD; Ximena Terra, PhD; Enric Figuerola, MD, PhD; Miquel Quer, MD, PhD; Xavier León, MD, PhD

**Objectives:** To compare chi-squared automatic interaction detection (CHAID) classification trees vs the seventh edition of the TNM classification for patients with head and neck squamous cell carcinoma and to assess whether CHAID classification trees might improve results obtained with the TNM classification.

**Design:** Patient disease was classified according to CHAID classification trees and the TNM classification, and the results were compared.

**Setting:** Academic research.

**Patients:** A total of 3373 patients with carcinoma of the oral cavity, oropharynx, hypopharynx, or larynx.

**Main Outcome Measures:** The 2 classification methods were evaluated objectively, measuring intrastage homogeneity (hazard consistency), interstage heterogeneity (hazard discrimination), and disease stage distribution among patients (balance). In addition, to as-

sess agreement between CHAID classification trees and the TNM classification, we calculated the  $\kappa$  statistic, weighted linearly and quadratically.

**Results:** Objective evaluation of the quality of the classification methods indicated that CHAID classification trees performed better than the TNM classification in terms of hazard consistency (2.51 for CHAID and 3.01 for TNM) and hazard discrimination (70.9% for CHAID and 52.7% for TNM) but not balance (-31.7% for CHAID and -15.5% for TNM). Analysis of concordance between the classification methods showed that the quadratic  $\kappa$  statistic was 0.77 (95% CI, 0.76-0.78) and the linear  $\kappa$  statistic was 0.59 (95% CI, 0.57-0.60) ( $P < .001$  for both).

**Conclusion:** CHAID classification trees performed better than the TNM classification and offer potential inclusion of new prognostic factors.

*Arch Otolaryngol Head Neck Surg.* 2012;138(3):272-279

## Author Affiliations:

Department of Otorhinolaryngology–Head and Neck Surgery (Drs Avilés-Jurado and Figuerola) and Research Group in Applied Medicine, Departments of Medicine and Surgery (Dr Terra), Hospital Universitari de Tarragona Joan XXIII IISPV, Universitat Rovira i Virgili, Tarragona, Catalonia, and Department of Otorhinolaryngology–Head and Neck Surgery, Hospital de la Santa Creu i Sant Pau, and Networking Research Center on Bioengineering, Biomaterials, and Nanomedicine, Universitat Autònoma de Barcelona, Barcelona (Drs Quer and León), Spain.

**H**EAD AND NECK SQUAMOUS cell carcinoma (HNSCC) represents the sixth leading cancer by incidence worldwide.<sup>1</sup>

Treatment of HNSCC includes the use of radiation therapy or chemoradiotherapy, surgery, or a combination of these procedures. At present, prognostic factors are not available that can efficiently predict the response to a certain therapy.

Classification and staging of cancer enable the physician and the cancer registry to stratify patients. This leads to better treatment decisions and the development of a common language that aids in the creation of clinical trials for the future testing of cancer staging.<sup>1</sup>

The seventh edition of the TNM classification (hereafter TNM) was developed by the American Joint Committee on Cancer in cooperation with the TNM committee of the International Union

Against Cancer.<sup>1</sup> TNM has brought considerable benefits by standardizing the description and reporting of cancer worldwide. However, the system has been challenged for not including factors that have a significant bearing on tumor behavior, such as concurrent comorbidity and unique tumor characteristics. Therefore, the trend toward improving the prognostic efficacy of TNM by including markers of poor prognosis other than stage is likely to continue.<sup>2</sup>

An alternative system of classification for patients with head and neck squamous cell carcinoma is based on classification trees, which were developed more than 20 years ago. Classification trees have acquired greater importance in the past decade because of the immediate interpretation of the decision rules that they generate. Classification trees are readily accepted by professionals in clinical practice.<sup>3-6</sup>

There are many types of classification trees, including the chi-squared automatic interaction detection classification tree (hereafter CHAID), first described in 1980 by Kass<sup>7</sup> and subsequently reported by others.<sup>8,9</sup> CHAID is an alternative classification strategy that highlights the information in variable interactions and uses decision trees to maximize the probability of a correct prognosis.<sup>10</sup> A classification tree can identify variables that are not considered using the traditional multivariate method because interactions and subgroups are more easily identified and may produce better results for ordinal data. In this context, the potential advantage of CHAID is its ability to define subpopulations (groups with better or worse prognosis) using combinations of predictive factors. These combinations are analyzed to assess their significance and to select those with higher relevance (with a lower *P* value). CHAID evaluates the categories and condenses them into 1 category if they present similar survival rates.

The aim of our study was to perform an objective comparison of CHAID vs TNM in a cohort of more than 3000 patients diagnosed as having HNSCC. To address that, we determined (1) intrastage homogeneity (hazard consistency), interstage heterogeneity (hazard discrimination), and disease stage distribution among patients (balance), as proposed by Groome et al,<sup>11</sup> to quantitatively analyze the most prominent characteristics that a classification scheme should fulfill and (2)  $\kappa$  statistic between CHAID and TNM, which evaluates agreement between the 2 classification methods.

## METHODS

Patients included in this study were diagnosed and treated in the Department of Otorhinolaryngology, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, between November 1985 and November 2005. The data were obtained retrospectively from the information collected prospectively and were stored in an oncology database.<sup>12</sup> All patients gave written informed consent. The study was approved by the ethical committee of the Hospital de la Santa Creu i Sant Pau.

Patients were categorized by an oncology committee in accord with TNM, as most recently proposed by the International Union Against Cancer.<sup>1</sup> Patients susceptible to stage change because of modifications in TNM were reviewed and restaged accordingly. In our study, changes in patient stage were limited to minor alterations in the oropharynx.<sup>1</sup>

Study inclusion was based on the diagnosis of squamous cell carcinoma confirmed by biopsy specimens from the following locations: oral cavity, oropharynx, hypopharynx, larynx, or some of these locations synchronously, with a follow-up period exceeding 2 years. Patients with a shorter follow-up period and without local recurrence or regional or distant metastasis were excluded.

Variables recorded were demographics (sex and age), tobacco use, alcohol consumption (defined as mild to moderate if <50 g/d for women or <70 g/d for men and as severe if  $\geq$ 50 g/d for women or  $\geq$ 70 g/d for men), primary tumor site, histological grade, TNM classification, and comorbidity. The functional status of patients was assessed using the index by Karnofsky et al.<sup>13</sup> Quantitative variables were described using the mean (SD). Categorical variables were described using the absolute and relative frequencies.  $\chi^2$  Test was used to analyze the relationship between categorical variables. Using the log-rank test, Kaplan-Meier actuarial survival curves were calculated to

compare groups. Data were expressed as survival (95% CI) at 5 years. Adjusted Kaplan-Meier survival curves were calculated in patients whose disease was classified according to CHAID vs TNM.

To evaluate and compare CHAID vs TNM, we used measures defined by Groome et al<sup>11</sup> for the following aspects of a staging system: (1) hazard consistency, or homogeneity of survival in the different categories included in each stage; (2) hazard discrimination, or heterogeneity between stages; predictive power, or capacity to predict the results; and (3) balance, or homogeneous distribution of all patients among stages.

An important feature of any measurement or classification device is reproducibility or reliability, which in a classification system is also referred to as concordance or agreement. Agreement between CHAID and TNM was determined using the weighted  $\kappa$  statistic, as described by Cohen.<sup>14</sup> The maximum  $\kappa$  statistic is 1.00, when agreement is perfect, and zero indicates no agreement. We assessed the weighted  $\kappa$  statistic for strength of agreement using guidelines by Fleiss et al<sup>15</sup> (poor agreement is <0.40, good agreement is 0.40 to 0.75, and excellent agreement is  $\geq$ 0.76). Linear and quadratic weights were given to the values. Linear weighting is used when the distance between categories follows a linear distribution. Quadratic weighting is used when an objective weight is unavailable for the different categories.<sup>16</sup>

All statistical analyses were performed using commercially available software (SPSS Statistics version 19.0; IBM). A specific program (Macro !KAPPA for SPSS Statistics) was used to calculate the  $\kappa$  statistic.<sup>17</sup>

## RESULTS

### PATIENT CHARACTERISTICS

**Table 1** summarizes characteristics of the patients. During the study period, 3427 patients were diagnosed as having HNSCC, but 54 patients were excluded because the follow-up period had been less than 2 years. Therefore, the study comprised 3373 patients.

Variables describing our cohort were compared using univariate analysis (Table 1). Alcohol consumption, Karnofsky index, histological grade, primary tumor site, and T and N classification had clinically relevant CIs and statistically significant *P* values for survival. In contrast, there were no significant differences in the distribution by sex, age, or tobacco use.

### CLASSIFICATION OF PATIENTS USING CHAID

**Figure 1** shows the tree created after applying CHAID. The terminal branches of the tree represent CHAID-derived homogeneous categories (terminal nodes). Terminal nodes with similar prognoses can be condensed into stages. We obtained 15 terminal nodes, which were grouped into 4 stages (I, II, III, and IV). **Table 2** gives the number of patients classified in each stage and the adjusted 5-year survival (95% CI) for each stage.

Early CHAID stages (I and II) include patients with limited tumor extension in locations associated with a good prognosis, such as the larynx, and without lymph node metastasis. Stage II includes patients with only 1 terminal node, representing advanced tumors (T3) of the larynx without lymph node metastasis. Stage III includes patients with T2N+, T3N+, and T3 tumors in locations as-

**Table 1. Baseline Patient Characteristics and Results of Univariate Analysis for Overall Survival**

Variable	No. (%) of Patients (n=3373)	5-Year Survival (95% CI), %	P Value
Sex			
Male	3119 (92.5)	67.2 (65.4-68.9)	.37
Female	254 (7.5)	65.7 (59.6-71.7)	
Age, y			
<50	531 (15.7)	65.6 (61.3-69.9)	.92
50-60	981 (29.1)	67.3 (64.1-70.4)	
>60 to 70	1106 (32.8)	67.1 (64.1-70.0)	
>70	755 (22.4)	67.9 (64.1-71.6)	
Tobacco use			
Never	271 (8.0)	69.2 (63.3-75.0)	.54
≤20 cigarettes per day	521 (15.4)	68.6 (64.2-72.9)	
>20 cigarettes per day	2581 (76.5)	66.6 (64.6-68.5)	
Alcohol consumption			
No	575 (17.0)	75.5 (71.7-79.2)	<.001
Mild to moderate	1400 (41.5)	74.3 (71.7-76.8)	
Severe	1398 (41.4)	56.3 (53.5-59.0)	
Karnofsky index			
≤70	131 (3.9)	22.9 (12.4-32.8)	<.001
80	476 (14.1)	48.1 (43.1-53.0)	
90	2766 (82.0)	72.0 (70.2-73.7)	
Histological grade			
Well	511 (15.1)	79.0 (75.2-82.7)	<.001
Moderate	2505 (74.3)	65.8 (63.8-67.7)	
Poor	357 (10.6)	58.6 (53.1-64.0)	
Primary tumor site			
Oral cavity	418 (12.4)	56.0 (51.0-61.0)	<.001
Oropharynx	651 (19.3)	44.7 (40.5-48.8)	
Hypopharynx	341 (10.1)	43.8 (37.9-49.6)	
Supraglottis	739 (21.9)	70.3 (66.7-73.8)	
Glottis	1168 (34.6)	89.1 (87.1-90.8)	
Synchronous	56 (1.7)	45.7 (31.8-59.6)	
TNM classification			
T1	926 (27.5)	89.2 (86.8-91.5)	<.001
T2	875 (25.9)	74.2 (71.0-77.3)	
T3	990 (29.4)	57.8 (54.4-61.1)	
T4	582 (17.3)	36.1 (31.9-40.2)	
N0	2143 (63.5)	80.9 (79.1-82.6)	<.001
N1	466 (13.8)	51.6 (46.7-56.5)	
N2	586 (17.4)	40.7 (36.2-45.2)	
N3	178 (5.3)	24.3 (17.3-31.3)	

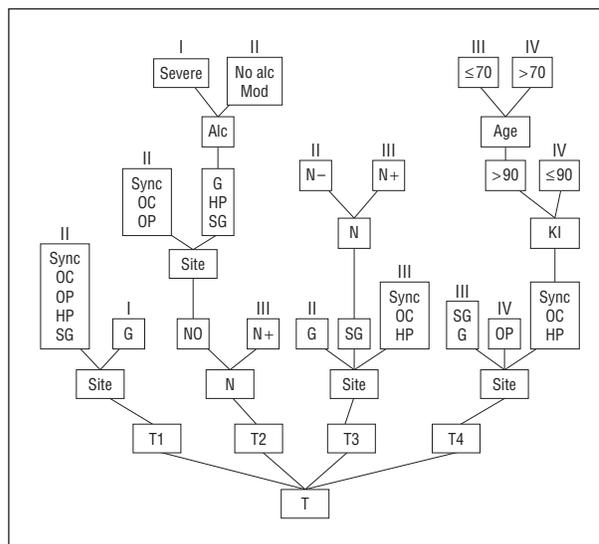
sociated with a bad prognosis, such as the hypopharynx, and patients with T4 tumors of the larynx. Stage IV includes patients with advanced tumors in unfavorable locations, such as the oropharynx, or among elderly patients or those with poor general health status.

Using CHAID, the most prognostic factor at the time of categorization was T classification. A second most prognostic factor for most T classifications (T1, T2, and T4) was primary tumor site. Multivariate analysis provided clinically relevant CIs for these variables (Table 1).

### COMPARISON OF CHAID vs TNM

#### Hazard Consistency

Different subgroups included in each cancer stage should have similar survival rates. Hazard consistency, or intrastage homogeneity, represents a weighted measure of the differences between survival in each category included



**Figure 1.** Chi-squared automatic interaction detection classification tree for overall adjusted survival. Alc indicates alcohol consumption; G, glottis; HP, hypopharynx; KI, Karnofsky index; Mod, moderate alcohol consumption; OC, oral cavity; OP, oropharynx; Severe, severe alcohol consumption; SG, supraglottis; and Sync, synchronous tumors.

**Table 2. Classification of Patients Using Chi-Squared Automatic Interaction Detection Classification Tree (CHAID)**

CHAID Stage	No. of Patients	5-Year Survival (95% CI), %
I	901	95.2 (93.6-96.8)
II	992	77.2 (74.5-79.9)
III	1174	49.1 (46.0-52.2)
IV	306	20.5 (15.6-25.4)

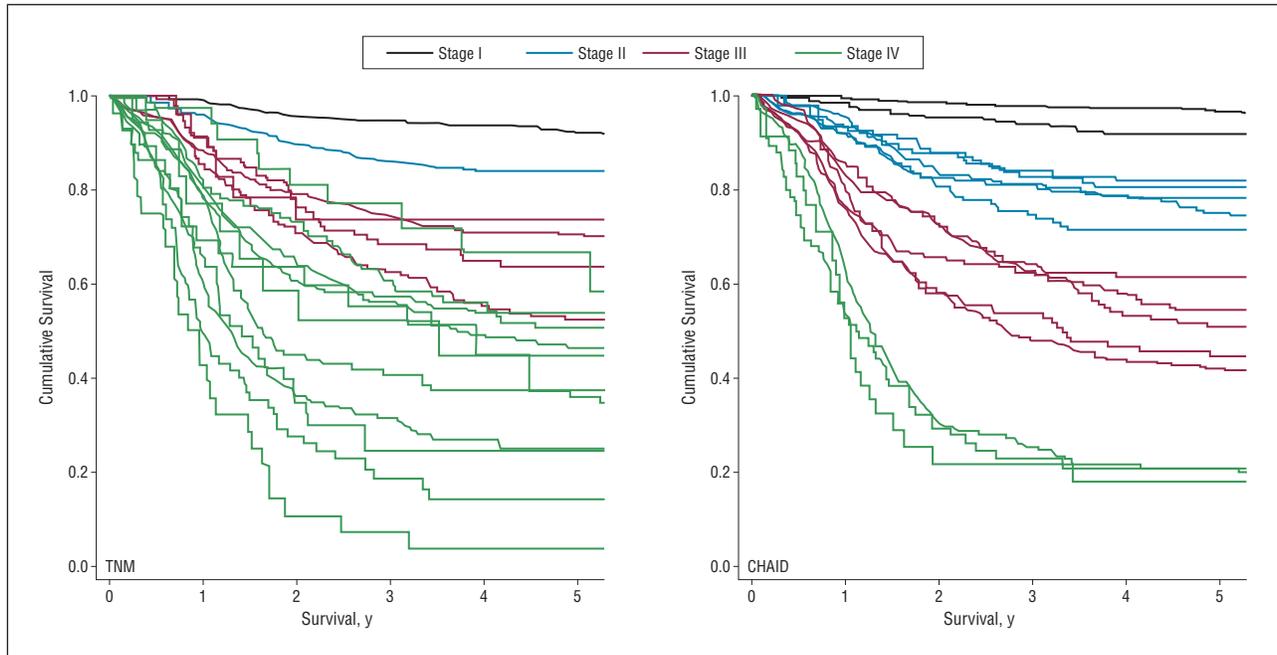
in a stage and the survival of all patients included in this stage. The lower the hazard consistency, the higher is the homogeneity of the categories included in each stage and the better the classification system.

**Figure 2** shows the adjusted survival curves of every category included in each stage according to CHAID and TNM among all patients. All categories with the same color represent a stage. The figure qualitatively shows the wide dispersion of TNM categories in advanced classifications (stage IV is green). In contrast, the CHAID categories included in the same stage are more homogeneous and do not overlap with stage III, as in TNM.

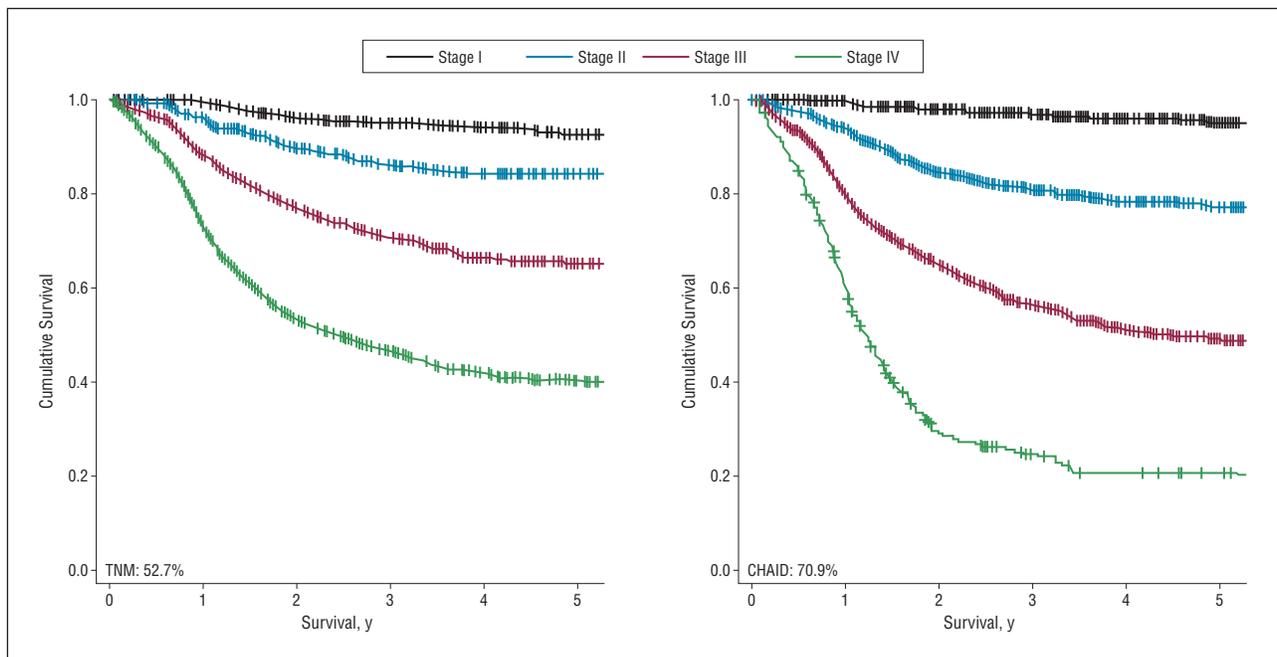
Quantitatively, intrastage homogeneity was 2.51 for CHAID and 3.01 for TNM. CHAID has a lower value and consequently has better hazard consistency than TNM in this population.

#### Hazard Discrimination

Hazard discrimination, or interstage heterogeneity, between groups addresses how evenly the stages are spaced in each classification method and how large a survival rate difference they span. The maximum hazard discrimination would be 100.0%, where the survival curves had a regular distribution with a maximum distance between stages. **Figure 3** shows the adjusted survival curves according to CHAID and TNM among all patients.



**Figure 2.** Intrastage homogeneity of TNM vs chi-squared automatic interaction detection classification tree (CHAID).



**Figure 3.** Interstage heterogeneity of TNM vs chi-squared automatic interaction detection classification tree (CHAID).

CHAID has better hazard discrimination than TNM (70.9% vs 52.7%). The 2 classification methods differ by approximately 20% in their discrimination capacity.

#### Balance

According to Groome et al,<sup>11</sup> perfect balance is achieved when each HNSCC stage represents 25% of the total population. **Table 3** gives unweighted and linearly and quadratically weighted  $\kappa$  statistics. The table summarizes the distribution of patients in each stage according to CHAID and TNM. Balance quantification indicated that TNM had

a 15.5% deviation from the expected number of patients, whereas CHAID had a 31.7% deviation, indicating that TNM was superior.

Again, a difference between classification methods is seen in stage IV. According to TNM, stage IV represented the most patients (32.6%) (Table 3). In contrast, according to CHAID, stage IV represented the fewest patients (9.1%).

#### $\kappa$ Statistic

Table 3 gives the cancer stage distribution according to CHAID and TNM among all patients. There were dis-

**Table 3. Comparison of Chi-Squared Automatic Interaction Detection Classification Tree (CHAID) vs TNM**

CHAID Stage	TNM Stage				Total No. (%)
	I	II	III	IV	
I	630	271	0	0	901 (26.7)
II	219	311	392	70	992 (29.4)
III	0	0	449	725	1174 (34.8)
IV	0	0	0	306	306 (9.1)
<b>Total No. (%)</b>	<b>849 (25.2)</b>	<b>582 (17.3)</b>	<b>841 (24.9)</b>	<b>1101 (32.6)</b>	<b>3373 (100)</b>

Agreement Assessment	$\kappa$ Statistic (95% CI)	Observed Agreement, %	P Value
Unweighted	0.35 (0.33-0.37)	50.28	<.001
Linear weighted	0.59 (0.57-0.60)	82.74	<.001
Quadratic weighted	0.77 (0.76-0.78)	93.78	<.001
Stage, TNM vs CHAID			
I	0.62 (0.59-0.65)	85.47	<.001
II	0.23 (0.19-0.26)	71.78	<.001
III	0.22 (0.19-0.25)	66.88	<.001
IV	0.34 (0.31-0.37)	76.43	<.001

crepancies with high clinical relevance between the 2 classification methods. Patients with stage IV tumors according to TNM were distributed among stages II, III, and IV according to CHAID. Specifically, 795 patients who were classified as having stage II disease (70 patients) and stage III disease (725 patients) according to CHAID were classified as having stage IV disease according to TNM. Furthermore, there were patients from all TNM stages in CHAID stage II.

Therefore, to assess the agreement or disagreement between the classification methods, we calculated weighted  $\kappa$  statistics (Table 3). Following the guidelines by Fleiss et al<sup>15</sup> for strength of agreement or concordance, our results indicated that the agreement between CHAID and TNM was moderate ( $\kappa$  statistic, 0.59; 95% CI, 0.57-0.60) when the  $\kappa$  statistic was weighted linearly. The agreement was excellent ( $\kappa$  statistic, 0.77; 95% CI, 0.76-0.78) when the  $\kappa$  statistic was weighted quadratically, but this result was at the lower limit of the CI by Fleiss et al. We also addressed concordance for each stage. Notably, stage I had the highest  $\kappa$  statistic (0.62; 95% CI, 0.59-0.65). Stage III had the lowest  $\kappa$  statistic (0.22; 95% CI, 0.19-0.25).

#### CROSS-DISTRIBUTION OF SURVIVAL CURVES USING CHAID AND TNM

We cross-distributed the survival curves using CHAID and TNM. We analyzed the resulting survival curves to evaluate the intrastage homogeneity of the classification methods.

We analyzed the survival curves resulting from applying TNM to all CHAID stages (Figure 4). All TNM stages included 2 CHAID categories, except for TNM stage IV, which included 3 CHAID categories. There were relevant differences in the adjusted survival curves for each stage ( $P < .001$ , log-rank test). The greatest difference, or highest heterogeneity, was in TNM stage III, where the adjusted survival was 78.8% for CHAID stage II vs 52.9% for CHAID stage III. In contrast, the highest homogeneity was in TNM stage IV,

where the adjusted survival was 52.3% for CHAID stage II and 47.1% for CHAID stage III.

Figure 5 shows the results after applying CHAID to all TNM stages. The homogeneity of TNM stages included in each CHAID stage was elevated. The adjusted survival values for each curve were similar in CHAID stages I and III. In CHAID stage II, there were no differences between patients included in TNM stages I, II, or III, but TNM stage IV, which represented 2.1% of the whole population, differed from the other TNM stages.

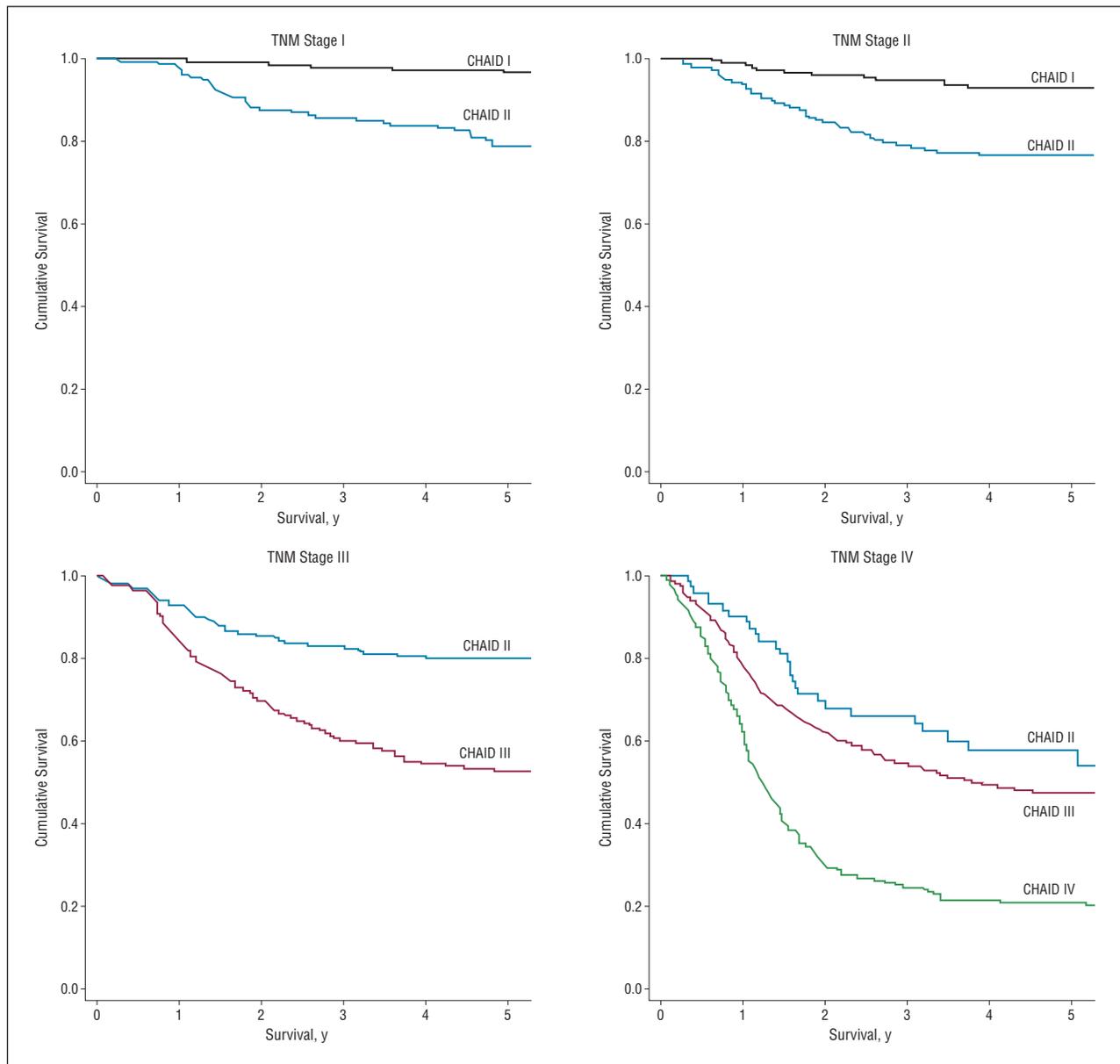
#### COMMENT

To date, this is the largest study to perform an objective comparison of CHAID vs TNM in patients with HNSCC. It is unique because it is the first such study among more than 3000 patients.

The most outstanding finding was that CHAID performed better than TNM in hazard consistency and hazard discrimination. We suggest that CHAID could enhance the current classification scheme for HNSCC.

Staging systems in cancer express the relative severity, or extent, of the disease. The finest detail is preferred, looking for the best discrimination among patients. Such description is meant to facilitate prognosis and to provide useful information for treatment decisions. The assumptions underlying the use of TNM in HNSCC are that it groups together patients with similar disease severity and that the difference in severity between groups is meaningful. It has been shown for HNSCC that these assumptions do not hold.<sup>11</sup>

In clinical practice, TNM is the most widely used classification system for prognosis. It is based on the simple concept that tumors grow following an ordered progression of facts, using local extension, regional extension, and the presence of distant metastases as classification variables. However, HNSCC does not always follow this pattern. In these types of tumors, other variables influence the outcome. CHAID, but not TNM, includes these



**Figure 4.** Cross-distribution of TNM by chi-squared automatic interaction detection classification tree (CHAID) Kaplan-Meier survival curves for overall adjusted survival.  $P < .001$  for all.

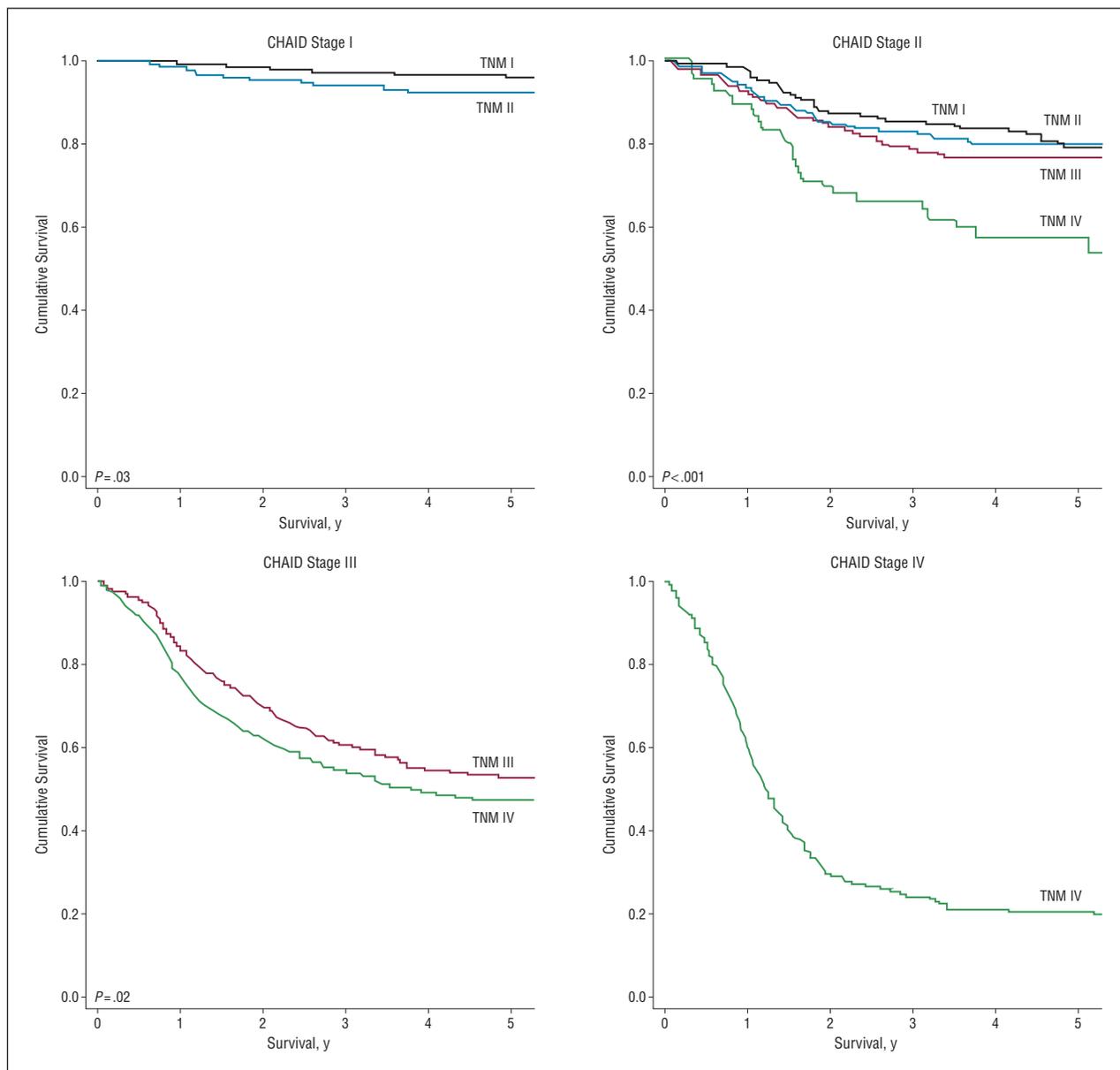
variables in the evaluation. This is the main reason why classification trees may improve the current classification method. Furthermore, additional research on cancer prognosis could provide new molecular markers of outcome, and classification trees such as CHAID would allow the inclusion of these variables.

Using CHAID, our study showed that the most decisive variable at the time of classification was T classification. The variable having the closest relationship to prognosis for most T classifications (T1, T2, and T4) was the primary tumor site. The 15 resulting terminal nodes were grouped into 4 stages with similar prognoses. One of the drawbacks of the comparison of the prognosis systems is the need to perform an objective evaluation of the quality of the classifications considered. There are several methods that compare staging systems.<sup>18</sup> The method proposed by Groome et al<sup>11</sup> ana-

lyzes important aspects in the classification of oncology patients.

Comparison of the survival curves generated by the 2 classification methods indicated that TNM presents higher intrastage dispersion than CHAID, mainly in stage IV. In addition, the objective evaluation of the quality of the classification methods demonstrated that CHAID performed better than TNM in terms of hazard consistency and hazard discrimination but not balance. The latter was included in the assessment of the schemes because maximizing the number of patients in each group helps to maximize statistical power, but it has no direct relevance to clinicians.<sup>11</sup> Groome et al<sup>11</sup> considered balance the least relevant criterion, and it is given a smaller weight when comparing systems.

In addition, we analyzed intrastage homogeneity of the methods by cross-distributing the survival curves of one



**Figure 5.** Cross-distribution of chi-squared automatic interaction detection classification tree (CHAID) by TNM Kaplan-Meier survival curves for overall adjusted survival.

classification scheme present in the stages of the other method. CHAID showed a higher discrimination capacity than TNM in each stage analyzed.

The fundamental objective of a classification method for prognosis is to classify each patient into the correct stage and to discriminate adequately the survival among different groups of patients. The first conclusion drawn from our results is that classification trees, which include more prognostic factors than TNM, might improve the classification scheme of HNSCC.

The inclusion of different variables as potential prognostic factors is a matter of debate because some (eg, treatment) could introduce bias in the results. Because patients may be recruited over a long period, treatment strategies could have changed for patients with similar primary tumor site and stage. Also, therapeutic strategies might differ across centers; therefore, treatment as

a variable cannot be standardized in a generalizable prognostic system. Moreover, according to characteristics proposed by the American Joint Committee on Cancer,<sup>19</sup> the CHAID model fulfills the most important criteria for a prognostic system.

When a new classification scheme is proposed, researchers become aware of its reliability. In this study, CHAID and TNM had excellent strength of agreement when the  $\kappa$  statistic was weighted quadratically. This coefficient indicates that the 2 methods are similar. However, they are far from reaching perfect agreement ( $\kappa$  statistic of 1.00). The independence of the methods is summarized in Table 3. A striking finding from  $\kappa$  statistic analysis is that stage III had the lowest agreement. In addition, discrepancies were shown in TNM stage IV, where 725 patients were classified as CHAID stage III and 70 patients as CHAID stage II. Although the group of 70

patients represented only 2.1% of the whole population, this discrepancy implies different treatments and highlights the need for stronger classification systems to make correct treatment decisions. This has an additional relevance in head and neck oncology because of the aggressive treatments applied. Taken together, the discrepancies between CHAID and TNM indicate that the classification methods evaluate different prognostic factors and that one does not subrogate the other.

This study has some potential limitations. First, the weighted  $\kappa$  statistic gives rise to the problem of weight selection,<sup>20</sup> although it is possible to determine the weight pattern arbitrarily. In general, it is more serious to wrongly classify as a stage II instead of a stage IV than to miss a stage III instead of a stage IV because different therapeutic decisions must be taken. By weighting the distances between stages, we can optimize the rules for making decisions according to the clinical importance associated with each missed classification. However, this makes it difficult to compare results obtained by different studies or researchers. Furthermore, in the quadratic weighting, the increase in the  $\kappa$  statistic is directly related to the number of categories included.

Second, the design of prognosis models commonly leads to an overestimation of results because the definitions are created from a specific cohort of patients and the results are thus fitted to this particular population. When prognosis indexes or predictive instruments are tested on a second population, they often perform less well.<sup>21</sup> It is important to validate new scales to assess whether prognosis information can be extrapolated.<sup>22</sup>

In conclusion, after an objective comparison, we found that the cancer stage classification based on CHAID performed better than that based on TNM. Although CHAID agreed well with TNM, discrepancies exist that must be considered.

CHAID highlights the relative importance and potential interactions of variables, which should allow more accurate stratification of patients. Our results suggest that this model is worth adopting, although it may be more complex. To evaluate extrapolation, CHAID should be externally validated.

**Submitted for Publication:** September 27, 2011; final revision received October 26, 2011; accepted December 13, 2011.

**Correspondence:** F. Xavier Avilés-Jurado, MD, PhD, Department of Otorhinolaryngology–Head and Neck Surgery, Hospital Universitari de Tarragona Joan XXIII IISPV, Universitat Rovira i Virgili, 4 Mallafré Guasch, 43007 Tarragona, Spain (fxavilesj@gmail.com).

**Author Contributions:** Drs Avilés-Jurado, Terra, Quer, and León had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Avilés-Jurado, Terra, Quer, and León. *Acquisition of data:* Avilés-Jurado, Quer, and León. *Analysis and interpreta-*

*tion of data:* Avilés-Jurado, Terra, Figuerola, Quer, and León. *Drafting of the manuscript:* Avilés-Jurado and Quer. *Critical revision of the manuscript for important intellectual content:* Avilés-Jurado, Terra, Figuerola, Quer, and León. *Statistical analysis:* Avilés-Jurado and León. *Study supervision:* Terra, Figuerola, Quer, and León. **Financial Disclosure:** None reported.

## REFERENCES

1. International Union Against Cancer. *TNM Classification of Malignant Tumors*. 7th ed. Chichester, England: Wiley-Blackwell; 2009.
2. Paleri V, Mehanna H, Wight RG. *TNM Classification of Malignant Tumours*, 7th edition: what's new for head and neck? *Clin Otolaryngol*. 2010;35(4):270-272.
3. Jones GW, Browman G, Goodyear M, Marcellus D, Hodson DI. Comparison of the addition of T and N integer scores with TNM stage groups in head and neck cancer. *Head Neck*. 1993;15(6):497-503.
4. Snyderman CH, Wagner RL. Superiority of the T and N integer score (TANIS) staging system for squamous cell carcinoma of the oral cavity. *Otolaryngol Head Neck Surg*. 1995;112(6):691-694.
5. Hart AA, Mak-Kregar S, Hilgers FJ, et al. The importance of correct stage grouping in oncology: results of a nationwide study of oropharyngeal carcinoma in the Netherlands. *Cancer*. 1995;75(11):2656-2662.
6. Hall SF, Groome PA, Rothwell D, Dixon PF. Using TNM staging to predict survival in patients with squamous cell carcinoma of head & neck. *Head Neck*. 1999; 21(1):30-38.
7. Kass GV. An exploratory technique for investigating large quantities of categorical data. *Appl Stat*. 1980;29(2):119-127.
8. Chan F, Cheing G, Chan JY, Rosenthal DA, Chronister J. Predicting employment outcomes of rehabilitation clients with orthopedic disabilities: a CHAID analysis. *Disabil Rehabil*. 2006;28(5):257-270.
9. Trujillano J, Badia M, Serviá L, March J, Rodríguez-Pozo A. Stratification of the severity of critically ill patients with classification trees. *BMC Med Res Methodol*. 2009;9(1):e83. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2797013/?tool=pubmed>. Accessed January 6, 2012.
10. Biggs D, De Ville B, Suen E. A method of choosing multiway partitions for classification and decision trees. *J Appl Stat*. 1991;18(1):49-62.
11. Groome PA, Schulze K, Boysen M, Hall SF, Mackillop WJ. A comparison of published head and neck stage groupings in carcinomas of the oral cavity. *Head Neck*. 2001;23(8):613-624.
12. León X, Orús C, Quer M. Design, maintenance, and exploitation of an oncologic database for patients with malignant tumors of the head and neck. *Acta Otorrinolaringol Esp*. 2002;53(3):185-190.
13. Karnofsky DA, Abelmann WH, Graver LF. The use of nitrogen mustards in the palliative treatment of carcinoma. *Cancer*. 1948;1:634-656.
14. Cohen J. Weighted  $\kappa$ : nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213-220.
15. Fleiss JL, Levin B, Paik MC. *Statistical Methods for Rates and Proportions*. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2003.
16. Doménech JM. *Fundamentos de Diseño y Estadística: Medida del Cambio: Análisis de Diseños con Medidas Intrajeto* [in Spanish]. 12th ed. Barcelona, Spain: Signo; 2011.
17. Doménech JM, Granero R. *Macro! KAPPA for SPSS Statistics: Weighted  $\kappa$*  [computer program]. Bellaterra, Spain: Universitat Autònoma de Barcelona; 2000.
18. Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18(17-18): 2529-2545.
19. Burke HB, Henson DE. The American Joint Committee on Cancer: criteria for prognostic factors and for an enhanced prognostic system. *Cancer*. 1993;72(10): 3131-3135.
20. Maclure M, Willett WC. Misinterpretation and misuse of the  $\kappa$  statistic. *Am J Epidemiol*. 1987;126(2):161-169.
21. Charlson ME, Ales KL, Simon R, MacKenzie CR. Why predictive indexes perform less well in validation studies: is it magic or methods? *Arch Intern Med*. 1987;147(12):2155-2161.
22. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130(6):515-524.