Using the National Cancer Database to create a scoring system that identifies patients with early-stage esophageal cancer at risk for nodal metastases

Benny Weksler, MBA, MD,[a] Kevin F. Kennedy, MS,[b] and Jennifer L. Sullivan, MD[a]

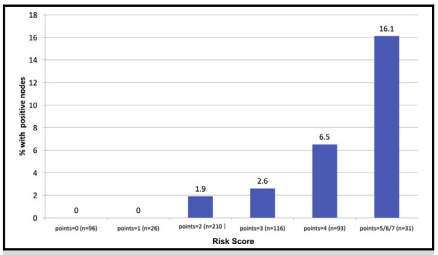
ABSTRACT

Objectives: Endoscopic resection is gaining popularity as a treatment for early-stage esophageal adenocarcinoma, particularly for T1a tumors. The goal of this study was to create a scoring system to reflect the risk of nodal metastases in early-stage esophageal adenocarcinoma to be used after endoscopic resection to better individualize treatment.

Methods: The National Cancer Database was queried for patients with T1a or T1b esophageal adenocarcinoma who underwent esophagectomy. We identified variables affecting nodal metastases using multivariable logistic regression, which we then used to create a scoring system. We stratified the model for T1a or T1b tumors, tested model discrimination, and validated the models by refitting in 1000 bootstrap samples. C-statistics greater than 0.7 were considered relevant.

Results: We identified 1283 patients with T1a or T1b tumors; 146 had nodal metastases (11.4%). Tumor category (pT1a vs pT1b), grade, and size and the presence of angiolymphatic invasion significantly affected the risk of nodal metastases. We assigned points to each variable and added them to get a risk score. In patients with T1a tumors, less than 3% of patients with a risk score of 3 or less had nodal metastases, whereas 16.1% of patients with a risk score of 5 or greater had nodal metastases. In patients with T1b tumors, less than 5% of patients with a risk score of 2 or less had nodal metastases, whereas 41% of patients with a score of 6 or greater had nodal metastases (c-statistic = 0.805).

Conclusions: The proposed scoring system seems to be useful in discriminating risk of nodal metastases in patients with T1a or T1b esophageal adenocarcinoma and may be useful in directing patients who received endoscopic resection to esophagectomy or careful follow-up. (J Thorac Cardiovasc Surg 2017;154:1787-93)



Risk of nodal metastases in patients with T1a EAC.

Central Message

A scoring system was generated to quantitate the risk of nodal metastasis in T1 esophageal tumors that may be useful in guiding patients to esophagectomy or follow-up after endoscopic resection based on this risk score.

Perspective

Although endoscopic resection is rapidly becoming the standard of care for early-stage esophageal cancer, patients with a high risk of nodal metastasis may be better treated with esophagectomy. We created a scoring system that reflects the risk of nodal metastases and may be useful in appropriately directing patients with T1 tumors to esophagectomy or follow-up after endoscopic resection.

See Editorial Commentary page 1794.

See Editorial page 1785.

The incidence of nodal metastases in patients with esophageal tumors that superficially invade the esophagus (T category pT1a and pT1b) varies from 10% to 15%.[1,2] The risk of nodal metastases in patients with T1 esophageal adenocarcinoma (EAC) increases with depth of invasion, poorly differentiated tumors, increasing size, and angiolymphatic invasion.[3-7] Nodal metastases are the most important determinant of long-term survival in patients with superficially invasive EAC.[5,8]

Scanning this QR code will take you to a supplemental video for the article.

From the [a]Division of Thoracic Surgery, Department of Surgery, University of Tennessee Health Science Center, Memphis, Tenn; and [b]Department of Biostatistics, St Lukes Health System, Mid America Heart Institute, Kansas City, Mo.

Read at the 97th Annual Meeting of The American Association for Thoracic Surgery, Boston, Massachusetts, April 29-May 3, 2017.

Received for publication April 24, 2017; revisions received July 8, 2017; accepted for publication July 16, 2017; available ahead of print Sept 1, 2017.

Address for reprints: Benny Weksler, MBA, MD, Division of Thoracic Surgery, University of Tennessee Health Science Center, 1325 Eastmoreland Ave, Suite 460, Memphis, TN 38104 (E-mail: bweksler@uthsc.edu).

0022-5223/$36.00

http://dx.doi.org/10.1016/j.jtcvs.2017.07.036

Abbreviations and Acronyms
CI = confidence interval
EAC = esophageal adenocarcinoma
LVI = lymphovascular invasion
NCDB = National Cancer Database
OR = odds ratio

EACs limited to the mucosa (T1a) are increasingly being treated with endoscopic resection, and survival after endoscopic resection of T1a tumors compares favorably with survival after esophagectomy.[9] Some practitioners have extended the indications of endoscopic resection to patients with T1b tumors with superficial invasion of the submucosa (sm1), with encouraging results.[10] Clinical staging of T1 EAC is difficult, and endoscopic ultrasound shows a concordance of only 65% in T1 esophageal carcinoma.[11] Many authors have recommended endoscopic resection as a staging tool for patients with clinical T1 EAC.[12,13] For patients with an established diagnosis of T1a or T1b adenocarcinoma after endoscopic resection, multiple authors have devised a scoring system or risk stratification for nodal metastases based on known risk factors.[6,14,15] These studies were based on a relatively small number of patients, and the largest cohort had 258 patients.

The National Cancer Database (NCDB) collects data from more than 1500 facilities and includes 70% of all new cancer diagnosis in the United States.[16] The NCDB has data points for all factors previously associated with a high risk of nodal metastases in patients with EAC. The aim of this study was to create a simple scoring system that can assess the risk of nodal metastases in patients with T1 EAC using this large national database.

MATERIALS AND METHODS

The NCDB was queried for all patients from 2010 to 2013 who underwent esophagectomy with pathologically confirmed T1a or T1b EAC. Patients were excluded if they received preoperative chemotherapy or radiation therapy, if they had metastases at diagnosis, and if data on T status or N status were incomplete. From the participant user files provided by the NCDB, we extracted basic patient demographics, pathologic information, including tumor category, tumor grade, presence of angiolymphatic invasion, number of lymph nodes examined, number of lymph nodes positive for metastases, and any neoadjuvant therapy received by the patient. The NCDB categorizes tumor grade as well as differentiated, moderately differentiated, poorly differentiated, and anaplastic. We combined poorly differentiated and anaplastic into 1 category of poorly differentiated tumors. The NCDB data are completely deidentified; therefore, this study was deemed exempt from approval and informed consent by the Institutional Review Board of the University of Tennessee Health Science Center.

Statistical Analysis

Data are represented as mean ± standard deviation or median and interquartile range for continuous variables and as n (%) for categoric variables. Unadjusted differences between positive and negative nodes were tested with Student *t* test and chi-square analysis, where appropriate. Missing data for candidate variables were substituted using multiple imputation methods with sequential regression using IVware software.[17] The association between tumor size and positive node status also was evaluated as a continuous variable by means of spline regression, using methods as described by Desquilbet and Mariotti.[18] For this analysis, restricted cubic spline functions were used in the adjusted logistic regression model relating positive node status to natural tumor size to identify the shape of the curve and to test the hypothesis of nonlinearity of this relationship. Using this functional relationship and the Youden index,[19] we created 3 categories for tumor size: less than 15 mm, 15 to 25 mm, and greater than 25 mm.

To identify independent predictors of positive nodes, we developed a multivariable logistic regression model with the preidentified variables: age, sex, T status, tumor differentiation, tumor size, Charlson comorbidity score, academic institution, and lymphovascular invasion (LVI). Model discrimination was tested with the c-statistic.

Next, we created a point-scoring system using the most prognostic variables for positive nodal status as determined by the beta weights of the variables.[20] We intentionally did not give a weight to tumor status (T1a/T1b) because we used this variable for stratification. We then graphically assessed the ability of the point system to discriminate between positive and negative nodal status, and computed the c-statistic using only the point system.

Finally, after fitting the model in the entire dataset, we conducted internal validation by refitting the model in 1000 bootstrap samples with replacement. This method of model validation has been found to have lower variability and lower bias potential compared with traditional split-sample validation and k-fold cross-validation.[21] All measures of model performance were corrected for optimism and a calculated "shrinkage" factor derived from the calibration slope. This analysis led to minimal adjustments to the full model (data not shown).

RESULTS

We identified 1283 patients with T1a or T1b tumors; 146 had nodal metastases (11.4%). There were 572 patients with T1a tumors (44.6%) and 711 patients with T1b tumors (55.4%) (Table 1). The most frequently missing variables in the dataset were tumor size (missing in 18%), tumor grade (missing in 13%), and angio-LVI (missing in 16%). Patients with T1a tumors (18/572) had a 3.1% incidence of nodal metastases, and patients with T1b tumors (128/711) had an 18% incidence of nodal metastases. The median age for the full cohort was 65 (interquartile range, 59-71), and 1095 were male (85.3%). In multivariable analysis, tumor category (pT1a vs pT1b; odds ratio [OR], 3.45; 95% confidence interval [CI], 2.47-4.81; $P < .001$), tumor differentiation (moderately differentiated vs well differentiated; OR, 4.39; 95% CI, 1.55-12.42; $P = .006$; poorly differentiated vs well differentiated; OR, 6.69; 95% CI, 2.33-19.18; $P < .001$), tumor size (15-25 mm vs < 15 mm; OR, 2.07; 95% CI, 1.21-3.53; $P = .008$; >25 mm vs <15 mm; OR, 2.98; 95% CI, 1.82-4.90; $P < .001$), and the presence of LVI ($P < .001$) were identified as significantly affecting the risk of nodal metastases (Figure 1). On the basis of our multivariable analysis, we assigned points to each variable to create a scoring system (Table 2) and then added the points to get the patients' risk scores (Figures 2 and 3).

In patients with T1a tumors, the majority (448/572, 78.3%) had a risk score of 3 or less and had a risk of nodal metastases less than 3%. There were 93 patients (93/572,