Esophageal Cancer: Associations with pN+

Thomas W. Rice, MD, Hemant Ishwaran, PhD, Wayne L. Hofstetter, MD, Paul H. Schipper, MD, Kenneth A. Kesler, MD, Simon Law, MD, Toni E.M.R. Lerut, MD, PhD, Chadrick E. Denlinger, MD, Jarmo A. Salo, MD, Walter J. Scott, MD, Thomas J. Watson, MD, Mark S. Allen, MD, Long-Qi Chen, MD, Valerie W. Rusch, MD, Robert J. Cerfolio, MD, James D. Luketich, MD, Andre Duranceau, MD, Gail E. Darling, MD, Manuel Pera, MD, PhD, Carolyn Apperson-Hansen, MStat, Eugene H. Blackstone, MD

Conflicts of interest: None declared

NIH grant acknowledgment: This publication was made possible in part by the Clinical and Translational Science Collaborative of Cleveland, UL1TR000439 from the National Center for Advancing Translational Sciences (NCATS) component of the National Institutes of Health and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Other sources of funding: The Daniel and Karen Lee Endowed Chair in Thoracic Surgery, held by Dr. Rice, and the Kenneth Gee and Paula Shaw, PhD, Chair in Heart Research, held by Dr. Blackstone.

Address for correspondence
Thomas W. Rice, MD
Cleveland Clinic
Department of Thoracic and Cardiovascular Surgery
9500 Euclid Avenue / Desk JJ-4
Cleveland, OH 44195
USA

Phone: 216-444-6712
Fax: 216-445-0444
Email: ricet@ccf.org

Running head: Esophageal cancer: pN1 associations
AUTHOR AFFILIATIONS

Thomas W. Rice, MD
Cleveland Clinic
Cleveland, Ohio, USA

Hemant Ishwaran, PhD
University of Miami
Miami, Florida, USA

Wayne L. Hofstetter, MD
University of Texas
MD Anderson Cancer Center
Houston, Texas, USA

Paul H. Schipper, MD
Oregon Health and Science Center
Portland, Oregon, USA

Kenneth A. Kesler, MD
Indiana University
Indianapolis, Indiana, USA University

Simon Law, MD
Queen Mary Hospital
The University of Hong Kong

Toni E.M.R. Lerut, MD, PhD
University Hospital Leuven
Leuven, Belgium

Chadrick E. Denlinger, MD
Medical University of South Carolina
Charleston, South Carolina, USA

Jarmo A. Salo, MD
Helsinki University Hospital
Helsinki, Finland

Walter J. Scott, MD
Fox Chase Cancer Center
Philadelphia, Pennsylvania, USA

Thomas J. Watson, MD
University of Rochester
Rochester, New York, USA

Mark S. Allen, MD
Mayo Clinic
Rochester, Minnesota, USA

Long-Qi Chen, MD
West China Hospital of Sichuan University
Chengdu, Sichuan, China

Valerie W. Rusch, MD
Memorial Sloan-Kettering Cancer Center, USA
New York, New York, USA

Robert J. Cerfolio, MD
University of Alabama at Birmingham
Birmingham, Alabama, USA

James D. Luketich, MD
University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania, USA

Andre Duranceau, MD
University of Montreal
Montreal, Canada

Gail E. Darling, MD
Toronto General Hospital
Toronto, Canada

Manuel Pera, MD, PhD
Hospital Universitario del Mar
Institut Hospital del Mar d’Investigacions Mèdiques
Universitat Autònoma de Barcelona
Barcelona, Spain
MINI-ABSTRACT

For resected esophageal cancers, regional lymph node metastases (pN+) are associated with increasing depth of cancer invasion, increasing cancer length, decreasing cancer differentiation, and more regional lymph nodes resected. Lymphadenectomy necessary to accurately detect pN+ is 60 for short well-differentiated cancers (<2.5 cm) and 20 for longer poorly differentiated ones.

Word count = 50
ABSTRACT

Objectives: 1) To identify the association of positive lymph node metastases (pN+), number of positive nodes, and pN subclassification with cancer, treatment, patient, geographic, and institutional variables, and 2) to recommend extent of lymphadenectomy needed to accurately detect pN+ for esophageal cancer.

Summary Background Data: Limited data and traditional analytic techniques have precluded identifying intricate associations of pN+ with other cancer, treatment, and patient characteristics.

Methods: Data on 5,806 esophagectomy patients from the Worldwide Esophageal Cancer Collaboration (WECC) were analyzed by Random Forest machine learning techniques.

Results: pN+, number of positive nodes, and pN subclassification were associated with increasing depth of cancer invasion (pT), increasing cancer length, decreasing cancer differentiation (G), and more regional lymph nodes resected. Lymphadenectomy necessary to accurately detect pN+ is 60 for shorter well-differentiated cancers (<2.5 cm) and 20 for longer poorly differentiated ones.

Conclusions: Not surprisingly, the finding of pN+, increasing number of positive nodes, and increasing pN classification are associated with deeper invading, longer, poorly differentiated cancers, and with extent of lymphadenectomy. In contrast to the extent of lymphadenectomy necessary to maximize survival, more deeply invasive, longer, poorly differentiated cancers are most likely to have pN+, so fewer nodes need to be removed to detect it. Conversely, superficial, shorter, well-differentiated cancers require extensive
lymphadenectomy to detect the uncommon cancer with regional lymph node metastases.

Word count: 220
Regional lymph node metastases (pN+) in esophageal cancer patients negatively affect outcome. Understanding the associations of pN+ with other cancer, treatment, and patient characteristics may aid in treatment decisions, such as extent of lymphadenectomy, and prognostication; however, these intricate associations have been difficult to identify because of limited data and traditional analytic techniques.\textsuperscript{1} To overcome prior data limitations, the Worldwide Esophageal Cancer Collaboration (WECC) uniquely provides, in a single database, cancer, treatment, patient, geographic, and institutional variables for a large cohort of esophagectomy patients.\textsuperscript{2} To overcome prior analytic limitations, Random Forest analysis,\textsuperscript{3} a modern machine-learning technique used to analyze the WECC database and to produce the 7th edition of esophageal cancer staging,\textsuperscript{4, 5} permits exploration of nonlinear, complex interrelationships.\textsuperscript{3, 6} The purpose of this study was to use this worldwide data set and machine-learning technique to identify WECC variables associated with pN+, number of positive nodes, and pN subclassification in order to develop recommendations for the number of regional lymph nodes that must be resected to accurately detect pN+.

**PATIENTS AND METHODS**

**Patients**
A total of 5,806 patients in WECC, a worldwide consortium of institutions that have contributed deidentified patient data on esophagectomy for cancer, underwent esophagectomy alone (no pre- or postoperative adjuvant therapy).\textsuperscript{2} All data sets were approved for research by each site’s Institutional Review Board, and data use agreements were executed when required. WECC variables included patient demographics (age, sex, race), region of the world (east, west), institution, cancer characteristics (location in esophagus, histopathologic cell type, histopathologic grade, cancer length, pT, pN, pM, and number of regional lymph nodes containing metastatic cancer (herein termed positive nodes), and esophagectomy variables (number of
regional lymph nodes resected, residual cancer, and year of surgery), 31 variables in all (Table 1).

**Endpoints**

The primary endpoint was pN+ disease. Secondary endpoints were number of positive nodes and pN classification: N0, no positive nodes; N1, 1 to 2 positive nodes; N2, 3 to 6 positive nodes; and N3, 7 or more positive nodes, based on the 7th edition of the American Joint Committee on Cancer (AJCC) staging manual.⁷

**Data Analysis**

**Overview**

Random Forest technology was chosen as the analytic strategy because of the known complex interactions among esophageal cancer characteristics identified in the 7th edition AJCC Cancer Staging endeavor.⁴ ⁷ The method is related to classification and decision-tree analyses, wherein the variable most related to an outcome of interest is first optimally split to improve prediction, then followed by more and more splits of it and other variables to create a tree (recursive partitioning, classification and regression trees). An individual tree “grown” by this method is inherently unstable, and this can be demonstrated by creating trees from bootstrap samples of the original data set that split much differently (the bootstrap data set is formed by random sampling with replacement until a data set of equal size is generated; there will be some duplicated patients, and an average of 37% will not be sampled). To overcome this instability, a forest of trees is grown from such bootstrap samples, permitting an ensemble average to be formulated across the individual trees.³ Because the method is completely nonparametric, with no restrictive underlying model assumptions, complex interactions among variables can be
robustly accounted for. Validity of the forest is evaluated by assessing outcome among the patients who were not selected by the bootstrap process.

Because some values were missing for some of the 31 variables, Random Forest imputation was employed to maximize use of the available data.\(^8\)

Rather than \(P\)-values, two metrics of prediction accuracy are generated based on the patients not selected (called the “out-of-bag” samples). The first ranks the importance of each variable in predicting the outcome of interest (variable importance, or VIMP).\(^3\) The second quantifies the average number of branches before a variable is split (called “minimal depth”): The closer to the trunk of the tree a variable is split, the more important that variable is to prediction accuracy.\(^9\)

In summary, predictors of outcome using Random Forest technology are identified in two steps: 1) building the forest based on cancer and other characteristics and the outcome of interest, and 2) using the resulting forest to discover the importance of variables to the prediction of the outcome and their interrelationships with respect to outcome.

Details of how this method was used for this study are given in SDC Appendix E1 and briefly summarized as follows.

**Predictors of N+**

Predictors of N+ were identified using the randomForestSRC R package.\(^10\) All 31 variables described in Table 1 plus surgical site were used to generate 1,000 random bootstrap classification trees. The average 37% of patients not included in building a given tree were used to estimate the cross-validated probability of a patient being N+. 
Predictors of Number of Positive Nodes

Predictors of number of positive nodes were identified using Random Forest nonparametric regression analysis. A forest of 1,000 trees was grown.

Predictors of pN Classification

A Random Forest strategy similar to that for identifying predictors of N+ was used for the ordinal outcome pN classification (N0, N1, N2, and N3) according to criteria in the 7th edition of the AJCC Cancer Staging Manual.

Lymphadenectomy Needed to Accurately Detect pN+

To ascertain extent of lymph node resection needed to accurately detect pN+ for each combination of pT and length of cancer (dichotomized as <2.5 or ≥2.5 cm), the number of lymph nodes resected was replaced by a fixed cutoff value and the predicted value of being pN+ calculated using the previously constructed pN+ classification forest and the out-of-bag patients. The cutoff value was varied from 0 to the maximum number of resected nodes observed within the pT and cancer length categories. The average over all such predicted values yielded the adjusted predicted probability of pN+ for the given cutoff. The point at which these curves flattened was interpreted as the lymphadenectomy needed to accurately detect pN+.

RESULTS

Predictors of pN+

More advanced cancer characteristics—longer cancer length, higher pT, and higher G—were the strongest predictors of pN+ (Figure 1, Figure 2, and SDC Figure E1). Presence of pM1 (SDC Figure E2) and squamous cell cancer were predictive, but less
so. The certainty of pN0 was improved with increasing number of negative lymph nodes resected, because the likelihood of resecting a positive node increased with increasing number of nodes resected (Figure 3 and SDC Figure E3).

The complex interplay of cancer characteristics, number of resected nodes, and pN+ is illuminated in Figure 4. The relationships were more striking for adenocarcinoma than for squamous cell carcinoma. Identifying pN+ required fewer resected lymph nodes for deeply invasive, poorly differentiated, or long cancers than for superficial invasion, well-differentiated, or short cancers. pT3, G3, or cancers longer than 4 cm were highly likely to be pN+ at all levels of nodes resected. Note that when no nodes were resected, the likelihood of pN+ became an average that increased as depth of invasion, grade, and length increased.

**Predictors of Number of Positive Nodes**

The same cancer characteristics that predicted presence of pN+ also predicted higher number of positive nodes (Figure 5). However, because number of positive nodes cannot exceed number of resected nodes, the relationship shifted to the right until about 11-20 nodes were resected. pT1 cancers that were node positive were predicted to have few positive nodes, but pT3/pT4 cancers that were node positive had a large number of positive nodes. Cancers longer than 4 cm were likely to have more than 5 positive nodes. G1 and G2 had a similar number of positive nodes, but fewer than G3.

**Predictors of pN Classification**

The association of cancer characteristics with higher classification of pN+ was more striking in adenocarcinoma than in squamous cell carcinoma (Figure 6). Superficial
cancers were likely to be pN0, and if they were not, they were likely to be pN1 rather than pN2 or, rarely, pN3. In contrast, deeply invasive cancers were more likely to be pN2 or pN3 than pN1, and if they were poorly differentiated, they were unlikely to be pN0.

**Lymphadenectomy Needed to Accurately Detect pN+**

Extent of lymphadenectomy needed to accurately detect pN+ decreased with increasing cancer length (Figure 7). Thus, for short cancers (<2.5 cm), the curves flatten at approximately 60 resected nodes. For cancers ≥2.5 cm, the curves clearly flatten at 20 nodes resected.

**DISCUSSION**

**Principal Findings**

The strongest associations with pN+ cancers reflect cancer growth, biology, and histology. Cancer growth, represented by two dimensions—depth of invasion and length—is strongly related to pN+. Adenocarcinoma invading beyond the muscularis propria (pT3) and cancers longer than 4 cm had a 60% to 80% prevalence of pN+. Cancer biology reflected by differentiation was strongly associated with pN+: G3 cancers had a 30% to 80% prevalence of pN+, depending on pT. However, we discovered a difference in prevalence of pN+ associated with histopathologic cell type; for the “same cancer,” pN+ was more likely in adenocarcinoma than squamous cell cancer. The process of metastases resulting in spread to regional lymph nodes (pN+) and distant sites (pM1) not surprisingly links these two anatomic cancer characteristics.
Rarely was a pM1 patient pN0. No cancer location was associated with a higher prevalence of pN+.

Intuitively, the more one looks, the more one finds. Thus, the only non-cancer characteristic associated with pN+ was number of regional lymph nodes resected—the extent of surgical lymphadenectomy. Proportionately, this was most important in short well-differentiated squamous cell cancers. Thus, extent of lymphadenectomy necessary to accurately detect pN+ must be greater for short, less invasive, well-differentiated cancers than for longer, more invasive, poorly differentiated ones.

We recommend that number of positive nodes and number of resected nodes be reported separately and not confounded by conversion to the continuous variable “lymph node ratio.” Its use should be discouraged because it is a puzzling blend of cancer biology and surgical technique and offers no further information than revealed by its numerator and denominator.

**The Literature**

The associations of other cancer characteristics with pN+ are rarely investigated; most publications focus instead on associations with survival. Traditional statistical analysis has demonstrated the associations of T and N. Staging of esophageal cancer in the 7th edition of the AJCC manual added three non-anatomic cancer characteristics—cancer location, histologic grade, and histopathologic cell type—providing potentially new associations to examine. It was this effort and the statistical methods behind stage groupings that led us to reexamine associations with pN+.

The overall prevalence of positive lymph nodes in esophageal cancer patients was similar for both histopathologic cell types in 1,059 esophagectomy patients at a
single institution.\textsuperscript{12} Cancer length correlated with pT/ypT, but its association with positive lymph nodes was not evaluated.\textsuperscript{13} In a study of 240 patients with esophageal squamous cell carcinoma, histologic grade was one of five variables univariably associated with positive lymph nodes; however, in multivariable analysis it did not add additional information.\textsuperscript{14} In a multi-institutional study, a larger number of positive lymph nodes predicted a higher likelihood of distant metastatic cancer\textsuperscript{15}: The probability of systemic disease exceeded 50\% when 3 or more positive nodes were present, and approached 100\% when 8 or more were present. Only one study has examined the relationship of pN+ and number of lymph nodes resected.\textsuperscript{16} Sensitivity of classifying pN+ continued to improve up to 100 nodes resected; however, maximum increase of sensitivity occurred from 0 to 6 nodes, and over 90\% sensitivity was reached at 12. Because no further associations were examined, composition of this study group could have influenced the resulting lymphadenectomy recommendations.

\textit{Strengths and Limitations}

Strengths and limitations of the WECC database have previously been outlined in detail,\textsuperscript{2,5} the most notable being that it represents advanced cancers and to a much lesser extent superficial cancers with their detailed subclassifications. This study is based on pathologic staging data rather than clinical staging data, which is necessary to establish the relationships described. However, adding length, cell type, and differentiation to depth of invasion\textsuperscript{1} improves transferability to clinical staging.

An inherent limitation of a data set such as WECC is lack of detailed information about the location of lymph nodes and variability of not only the surgical lymphadenectomy, but also the pathologic processing of the specimen. This could lead
to both over- and underestimation of cancer characteristics. Extreme values may also be influential; inherent in Random Forest methodology is resampling, which mitigates against sampling extreme values, as does averaging over many trees (in this case, over 1,000) and multivariable modeling. Institutional and region-of-the-world clustering effects were generally small, but were taken into account by incorporating sites and region of the world into risk-adjusted estimates. Nonlinearities in the data and complex interactions among cancer variables that have been identified for esophageal cancer, a challenge for traditional statistical methods, are inherently accounted for by Random Forest methods. These are the primary reasons for choosing this methodology for WECC.

A strength of this study is that it is a large surgery-alone series that spans the spectrum of esophageal cancer characteristics, and this will likely never be repeated in the age of neoadjuvant therapy.

**Therapeutic Implications**

**Treatment That Assumes pN Status**

Endoscopic therapies are recommended when it is reasonable to assume pN0. Absence of lymph node metastases is most likely to be true for well-differentiated short cancers with minimal invasion, and also more likely for squamous cell carcinoma than adenocarcinoma. Strictly adhering to these observations will rarely result in endoscopic mucosal therapy of a pN+ cancer.

Conversely, neoadjuvant therapy typically is recommended when it is reasonable to assume pN+. Realizing the inaccuracies of clinical staging, a deeply invasive long,
poorly differentiated cancer reported to be cN0 is likely to be pN+ and thus should receive more than surgery alone.

Clinical Implications

The aims of lymphadenectomy are twofold: 1) to achieve accurate staging and 2) to provide a possible therapeutic (survival) benefit with respect to extent of lymphadenectomy. These goals are conflicting. This study found that extent of lymphadenectomy required to accurately detect pN+ is greatest for early-stage cancers; in our previous study, extent of lymphadenectomy required to maximize survival benefit is greatest for advanced-stage cancers. Esophagectomy for deeply invasive, long, poorly differentiated (advanced-stage) cancers that are reported as pN0, but fewer than 20 nodes have been resected, are unlikely to actually be pN0 because the extent of lymphadenectomy is inadequate for both staging and a possible survival benefit. For superficial, short, well-differentiated (early-stage) cancers, one must balance the risk of extensive lymphadenectomy to accurately detect pN+ (60 nodes) and the benefit of lesser lymphadenectomy (10 nodes for pT1) to maximize 5-year survival. Although “the more one looks, the more one finds,” this must be tempered by high-risk associations with pN+.
ACKNOWLEDGMENTS

The authors thank Tess Parry for editorial assistance and Brian Kohlbacher for graphic design help. Both are Cleveland Clinic employees.
REFERENCES


Table 1. Patient and Esophageal Cancer Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. (%) or Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>5,664</td>
<td>63 ± 11</td>
</tr>
<tr>
<td>Male</td>
<td>5,672</td>
<td>4,392 (77)</td>
</tr>
<tr>
<td>Race</td>
<td>4,063</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2,834 (70)</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1,171 (28)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>58 (2.0)</td>
<td></td>
</tr>
<tr>
<td>East (part of world)</td>
<td>5,673</td>
<td>1,171 (21)</td>
</tr>
<tr>
<td>Location of cancer</td>
<td>4,602</td>
<td></td>
</tr>
<tr>
<td>Upper third</td>
<td>209 (5.0)</td>
<td></td>
</tr>
<tr>
<td>Middle third</td>
<td>1,209 (26)</td>
<td></td>
</tr>
<tr>
<td>Lower third</td>
<td>3,184 (69)</td>
<td></td>
</tr>
<tr>
<td>Cancer length (cm)</td>
<td>3,150</td>
<td>3.5 ± 2.6</td>
</tr>
<tr>
<td>pT</td>
<td>5,673</td>
<td></td>
</tr>
<tr>
<td>is</td>
<td>370 (7.0)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1,393 (25)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>932 (16)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2,787 (49)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>191 (3.0)</td>
<td></td>
</tr>
<tr>
<td>pN</td>
<td>5,639</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>3,199 (57)</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>2,440 (43)</td>
<td></td>
</tr>
<tr>
<td>Number of regional lymph nodes</td>
<td>5,238</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>3,199 (61)</td>
<td></td>
</tr>
<tr>
<td>N1</td>
<td>952 (18)</td>
<td></td>
</tr>
<tr>
<td>N2</td>
<td>634 (12)</td>
<td></td>
</tr>
<tr>
<td>N3</td>
<td>453 (9.0)</td>
<td></td>
</tr>
<tr>
<td>Number of regional lymph nodes resected</td>
<td>4,250</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>438 (10)</td>
<td></td>
</tr>
<tr>
<td>1-5</td>
<td>1,019 (24)</td>
<td></td>
</tr>
<tr>
<td>6-10</td>
<td>625 (15)</td>
<td></td>
</tr>
<tr>
<td>11-15</td>
<td>492 (12)</td>
<td></td>
</tr>
<tr>
<td>16-25</td>
<td>774 (18)</td>
<td></td>
</tr>
<tr>
<td>≥26</td>
<td>902 (21)</td>
<td></td>
</tr>
<tr>
<td>pM</td>
<td>5,673</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>5,258 (93)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>415 (7.0)</td>
<td></td>
</tr>
<tr>
<td>G (histologic grade)</td>
<td>4,202</td>
<td></td>
</tr>
<tr>
<td>G1 (well differentiated)</td>
<td>1,422 (34)</td>
<td></td>
</tr>
<tr>
<td>G2 (moderately differentiated)</td>
<td>1,349 (32)</td>
<td></td>
</tr>
<tr>
<td>G3 (poorly differentiated)</td>
<td>1,425 (34)</td>
<td></td>
</tr>
<tr>
<td>G4 (undifferentiated)</td>
<td>6 (&lt;1)</td>
<td></td>
</tr>
<tr>
<td>Histopathologic cell type</td>
<td>5,673</td>
<td></td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>2,154 (38)</td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>3,519 (62)</td>
<td></td>
</tr>
<tr>
<td>Resection margins</td>
<td>5,194</td>
<td></td>
</tr>
<tr>
<td>R0</td>
<td>4,605 (89)</td>
<td></td>
</tr>
<tr>
<td>R1</td>
<td>464 (9.0)</td>
<td></td>
</tr>
<tr>
<td>Year of esophagectomy</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>1970s</td>
<td>78 (1.3)</td>
<td></td>
</tr>
<tr>
<td>1980s</td>
<td>1,265 (22)</td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>2,303 (40)</td>
<td></td>
</tr>
<tr>
<td>2000s</td>
<td>2,160 (37)</td>
<td></td>
</tr>
</tbody>
</table>

R2

5,806

1,256 (2.0)

a. Patients with data available.

Key: SD, standard deviation.
FIGURE LEGENDS

Figure 1: Variable importance for each of the 31 variables from Random Forest analysis of pN status.

Figure 2: Frequency of pN+ according to G, pT, and histopathologic cell type (red=G1, green=G2, blue=G3). Key: adeno, adenocarcinoma; squam, squamous cell cancer.

Figure 3: Out-of-bootstrap predicted probability of pN+ cancer as a function of number of nodes resected, stratified by G, pT, and histopathologic cell type (red=G1, green=G2, blue=G3). Individual dots represent predicted probabilities and solid lines are LOESS (locally weighted scatterplot smoothing) values of predicted probabilities. Key: adeno, adenocarcinoma; squam, squamous cell cancer.

Figure 4: Probability of pN+ according to number of resected nodes and various esophageal cancer characteristics.

Figure 5: Predicted number of positive nodes according to number of resected nodes and various esophageal cancer characteristics.

Figure 6: Out-of-bootstrap predicted probability of pN0, pN1 (1-2 positive nodes), pN2 (3-6 positive nodes), and pN3 (7 or more positive nodes) cancers, stratified by G, pT, and histopathologic cell type (red=pN0, green=pN1,
blue=pN2, turquoise=pN3). Key: *adeno*, adenocarcinoma; *squam*, squamous cell cancer.

**Figure 7:** Adjusted predicted probability of pN+ for various cutoff values of number of resected nodes, according to pT and cancer length. Where curves plateau is interpreted as the lymphadenectomy necessary to accurately detect pN+.
**Figure 1.** Variable importance for each of the 31 variables from Random Forest analysis of pN status.
Figure 2. Frequency of pN+ according to G, pT, and histopathologic cell type (red=G1, green=G2, blue=G3). Key: adeno, adenocarcinoma; squam, squamous cell cancer.
Figure 3. Out-of-bootstrap predicted probability of pN+ cancer as a function of number of nodes resected, stratified by G, pT, and histopathologic cell type (red=G1, green=G2, blue=G3). Individual dots represent predicted probabilities and solid lines are LOESS (locally weighted scatterplot smoothing) values of predicted probabilities.
Figure 4. Probability of pN+ according to number of resected nodes and various esophageal cancer characteristics.
Figure 5. Predicted number of positive nodes according to number of resected nodes and various esophageal cancer characteristics.
Figure 6. Out-of-bootstrap predicted probability of pN0, pN1 (1-2 positive nodes), pN2 (3-6 positive nodes), and pN3 (7 or more positive nodes) cancers, stratified by G, pT, and histopathologic cell type (green=pN0, yellow=pN1, orange=pN2, red=pN3). Key: adeno, adenocarcinoma; squam, squamous cell cancer.
Figure 7. Adjusted predicted probability of pN+ for various cutoff values of number of resected nodes, according to pT and cancer length. Where curves plateau is interpreted as the lymphadenectomy necessary to accurately detect pN+.
**Figure E1.** Relative frequency that cancer is pN+ as a function of number of resected lymph nodes, stratified by histopathologic cell type, pT, and G (red=G1, green=G2, blue=G3). Dots represent raw frequency data. Key: adeno=adenocarcinoma; squam=squamous cell carcinoma.
Figure E2. Out-of-bag (OOB) predicted probability of pN+ cancer stratified by G, pT, pM, and histopathologic cell type (red=G1, green=G2, blue=G3). Key: adeno=adenocarcinoma; squam=squamous cell carcinoma.
**Figure E3.** Relative frequency of number of resected lymph nodes, presented as histograms of 5-node groups. These groups are stratified by histopathologic cell type, pT, and G (red=G1, green=G2, blue=G3). Key: adeno=adenocarcinoma; squam=squamous cell carcinoma.
Appendix E1: Random Forest Methodology Details

Predictors of \( pN^+ \)

Predictors of \( pN^+ \) were identified in 2 steps using Random Forest technology: 1) building a forest by regressing \( pN \) status (\( pN^0, pN^+ \)) on cancer and other characteristics, and 2) using the forest for discovering the importance of variables and their relationships. Random forest classification and regression was implemented using the randomForestSRC R package.\(^{10}\)

The first step was building a forest of 1,000 random bootstrap classification trees, grown under Gini index splitting with random input selection. Each random classification tree was constructed by sampling the data with replacement to build a new data set of size equal to the original. This bootstrap sampling procedure on average included 63% of the patients (some patient data are duplicates); the remaining 37%, referred to as out-of-bag (OOB) data, were used to construct OOB (cross-validated) estimates of a patient's probability of being \( pN^+ \). Trees were grown as deeply as possible under the restriction that terminal nodes contained no fewer than 2 patients (nodesize). Nodesize was determined by fitting Random Forest systematically under different nodesize specifications and choosing the nodesize value minimizing OOB misclassification error rate, where OOB error was calculated using OOB predicted \( pN^+ \). All other Random Forest parameters were set to default settings of the software. Missing data were imputed using the Random Forest method described by Ishwaran and colleagues.\(^{8}\)

The second step was to quantify the predictive importance of each variable, estimated using Breiman permutation variable importance (VIMP).\(^{3}\) On a scale from
-100% to 100%, VIMP estimates the expected change in misclassification error of predicted pN+ classification if the variable were removed from the multivariable forest analysis. To determine whether a variable’s predictiveness was statistically significant, minimal depth variable selection was used.\(^9\) Minimal depth equals the shortest distance from the root of a tree to the first node where a given variable produces a split (branch). Shorter distances indicate more predictiveness. Forest-averaged minimal depth for each variable was compared with a threshold value determined from a null minimal depth distribution to test whether the variable was predictive.\(^9\)

**Predictors of Number of Positive Nodes**
Random Forest nonparametric regression, implemented by the randomForestSRC R package,\(^{10}\) was used to regress number of regional lymph nodes, as for pN+. A forest of 1,000 Random Forest regression trees was grown using weighted mean-squared error splitting with random input selection to obtain OOB estimates for the predicted number of nodes. The OOB optimized nodesize value of 6 was used; all other Random Forest parameters were set to default values of the software.

**Predictors of pN Classification**
A Random Forest strategy similar to that for pN+ was used to regress pN classification (N0, N1, N2, and N3) on the 31 independent variables to obtain OOB estimates for pN classification and their probabilities. Identical Random Forest tuning parameters were used.