N1 ESOPHAGEAL CARCINOMA: THE IMPORTANCE OF STAGING AND DOWNSTAGING

Thomas W. Rice, MD^a
Eugene H. Blackstone, MD^{a,b}
David J. Adelstein, MD^c
Gregory Zuccaro, Jr, MD^d
John J. Vargo, MD^d
John R. Goldblum, MD^e
Lisa A. Rybicki, MS^b
Sudish C. Murthy, MD, PhD^a
Malcolm M. DeCamp, MD^a

Objective: To evaluate the effects of clinical staging and downstaging by induction chemoradiation therapy in patients with N1 esophageal carcinoma.

Methods: Sixty-nine consecutive patients with regional lymph node metastases (cN1) according to clinical staging received induction therapy before surgery. These were compared to 75 patients both clinically and pathologically N1 (cN1/pN1) who underwent surgery without induction therapy and 79 patients clinically and pathologically not N1 (cN0/pN0) who underwent surgery without induction therapy. Analyses focused on survival and the cost and benefit of therapy.

Results: For comparison, the extremes of 5-year survival were 69% for cN0/pN0 patients who underwent surgery alone and 12% for cN1/pN1 patients who underwent surgery alone. Of 69 patients who received induction therapy, 37 were pN0 at resection (downstaged); they had an intermediate survival of 37% at 5 years. Those patients not downstaged with induction therapy had a 12% 5-year survival, similar to patients with cN1/pN1 who underwent surgery alone. After adjusting for the strongest predictors of poor outcome, pN1, and increasing N1 burden, a modest increased risk of death after induction therapy was identified. However, this cost of induction therapy was more than counterbalanced by the benefit of improved survival of downstaging to pN0.

Conclusions: (1) pN1 is the strongest determinant of poor outcome. (2) cN1 patients who are downstaged by induction chemoradiation therapy to pN0 have an intermediate outcome. (3) cN1 patients who are not downstaged by induction therapy have a poor outcome. (J Thorac Cardiovasc Surg 2001;121:454-64)

Fundamental changes in epidemiology, staging, and therapy have transformed the treatment of esophageal carcinoma.^{1,2} Previously, advanced-stage cancer and surgical therapy predominated. Now, multiple therapeutic options are available for a heterogeneous disease. These fundamental changes had several origins. First, a radical shift in the characteristics of the disease resulted in the detection of increasing numbers

of early-stage adenocarcinoma. Second, effective therapeutic alternatives were developed and include nonsurgical palliation and multimodality therapy. Third, advances in clinical staging technology allowed determination of clinical (pretreatment) stage. Along with these, the introduction of an accurate and practical staging system made it possible to use clinical stage for prognosis and treatment modification.

Patients clinically staged with cancers metastatic to regional lymph nodes (N1) have a poor prognosis if treated with surgery alone. It is known that patients who have no residual cancer after induction chemoradiation therapy have an improved survival. However, no study has clinically staged patients and quantified the benefit of downstaging from N1 to N0, regardless of primary tumor status. Therefore, the purpose of this study was to evaluate clinical staging and downstaging by induction chemoradiation therapy in patients with clinical N1 (cN1) esophageal carcinoma.

From the Departments of Thoracic and Cardiovascular Surgery,^a Biostatistics and Epidemiology,^b Hematology and Medical Oncology,^c Gastroenterology,^d and Anatomic Pathology,^e The Center for Swallowing and Esophageal Disorders, The Cleveland Clinic Foundation, Cleveland, Ohio.

Read at the Twenty-sixth Annual Meeting of The Western Thoracic Surgical Association, The Big Island, Hawaii, June 21-24, 2000. Copyright © 2001 by The American Association for Thoracic Surgery

0022-5223/2001 \$35.00 + 0 12/6/112470

doi:10.1067/mtc.2001.112470

Table I. Patient groups

Group	n	Clinical stage	Pathologic stage	Treatment	Survival
1	37	cN1	pN0	IT	Intermediate
2	32	cN1	pN1	IT	Poor
3	75	cN1	pN1	SURG	Poor
4	79	cN0	pN0	SURG	Good
5	23	cN0	pN0	IT	Intermediate
6	8	cN0	pN1	IT	Poor
7	21	cN0	pN1	SURG	Poor
8	18	cN1	pN0	SURG	Intermediate

IT, Induction therapy followed by surgery; Surg, surgery alone.

Patients and methods

Patients. Between 1983 and 1999, 577 patients with esophageal carcinoma underwent resection at The Cleveland Clinic Foundation. From this database, 293 patients were identified with known clinical (pre-induction therapy) and pathologic N status (Table I). Of these, 69 had clinical N1 (cN1) disease and received preoperative induction therapy. Thirty-seven were downstaged to pathologic N0 (cN1/pN0, group 1) and 32 were not (cN1/pN1, group 2). To provide a lower limit of comparative survival, we identified 75 patients with both clinical and pathologic N1 who underwent surgery without induction therapy (cN1/pN1, group 3). To provide an upper limit of comparative survival, we identified 79 patients with neither clinical nor pathologic N1 who underwent surgery without induction therapy (cN0/pN0, group 4).

Of the 8 possible combinations of cN/pN and induction therapy, 4 additional groups were identified: 23 patients receiving induction therapy because of advanced clinical depth of tumor invasion (cT), with neither clinical nor pathologic N1 (cN0/pN0, group 5); 8 patients receiving induction therapy who were either understaged or had disease progression (cN0/pN1, group 6); 21 patients undergoing surgery without induction therapy who were understaged (cN0/pN1, group 7); and 18 patients undergoing surgery without induction therapy who were overstaged (cN1/pN0, group 8).

The addition of groups 5 through 8 was the strategy used to quantify and adjust for clinical staging inaccuracies.

Of the 293 patients, 240 (82%) were men, 276 (94%) were white, and age ranged from 26 to 84 years (mean $\pm$ SD, 62 $\pm$ 11). A total of 232 (79%) patients had adenocarcinoma, 57 (19%) patients had squamous cell carcinoma, and 4 (1%) had adenosquamous carcinoma.

Staging. All patients underwent endoscopic esophageal ultrasound (EUS) before treatment to determine clinical stage. Clinical depth of tumor invasion (cT) and clinical regional lymph node status (cN) were recorded by means of the 1997 criteria of the American Joint Committee on Cancer Staging.³ Ultrasound characteristics were primary determinants of cN and included nodal size, shape, border, internal echo characteristics, and proximity of the regional node to the primary tumor. Distant metastatic disease (cM1b) was identified by computed tomographic scanning.

Pathologic stage (pT and pN) was determined by examination of the resection specimen and regional lymph nodes that were either sampled or included in the lymphadenectomy specimen.

Induction therapy. Since 1991, patients have been offered induction therapy for cN1 and cT3 or cT4 cancers. Induction therapy included two courses of chemotherapy, either cisplatin and 5-fluorouracil⁴ or cisplatin and taxol.⁵ Concurrent with chemotherapy, accelerated fractionation radiation therapy (1.5 Gy bid to a total dose of 45 Gy) was administered in two courses over 2 months. Surgery was accomplished 4 to 6 weeks later. These studies were approved and reviewed annually by The Cleveland Clinic Foundation Institutional Review Board, and written informed consent was obtained from all patients before beginning treatment.

Surgery. Of 293 patients, 234 (80%) had esophagectomy with thoracotomy and 2-field lymphadenectomy, and 59 (20%) had transhiatal esophagectomy and lymph node sampling. Patients with cN1 and cT3 or cT4 cancers not receiving induction therapy underwent surgery as the primary treatment. Six patients had pT4 tumors, all infracardinal. Ninety-four percent had R0 resections.

Study end point. The end point of the study was all-cause mortality. Patients were followed up by periodic clinic visits; however, a cross-sectional systematic follow-up was made in December 1999. Mean follow-up was 26 $\pm$ 15 months (43 $\pm$ 35 months in patients alive at last follow-up).

Methods—data analysis

Descriptive. Descriptive statistics are summarized as the mean and standard deviation for continuous variables and as frequencies and percentages for categorical variables.

Survival. Nonparametric estimates of survival are obtained by the method of Kaplan and Meier.⁶ A parametric method was used to resolve the number of phases of instantaneous risk of death (hazard function) and to estimate their shaping parameters.⁷ Estimates of survival were considered reliable to a maximum of 10 years.

So that the prognostic relevance of clinical staging could be established, a multivariable time-related analysis of survival was performed of clinical staging, demographic, and surgical experience variables (Appendix I), in the following manner. Initial screening of variables possibly related to survival used

the log-rank test and the univariable Cox proportional hazards model. Continuous and ordinal variables were assessed univariably by decile risk analysis to suggest transformations of scale required to assure that the relationships of these measured variables were well calibrated to model of outcome.

Multivariable analysis was performed simultaneously for the two hazard phases identified from the data. We used a directed technique of entry of variables into the multivariable models.⁸ A $P = .1$ criterion for retention of variables in the final models was used. Additionally, bootstrap resampling was used to validate the models.⁹

Who will respond to induction therapy? Among the 69 cN1 patients who received induction therapy, 37 were found to have pN0 at operation and are classified as "responders." Multivariable logistic regression analysis was performed of all pre-induction therapy variables (Appendix I) to identify factors that would predict pN0.

Cost and benefit of therapy. Two avenues of inquiry were taken to investigate the question, "Do patients receiving induction therapy survive as long as their counterparts who did not receive induction therapy?" If they do not, this would suggest a toxic effect of therapy, advancement of the disease during therapy, the biologic process of downstaging, or other therapy-related differences.

Matched analysis. Even in a randomized study, finding counterparts for such a comparison is impossible because pathologic stage is unknown at the time of randomization. Therefore, patients were multivariably propensity-matched for their demography, pathologic stage, operation, and date of operation (Appendix I). Matching factors associated with induction therapy were identified by means of multivariable logistic regression analyses and strategies for model building similar to the survival analysis described above. After these factors had been identified, statistically nonsignificant factors, including some components of pathologic stage, sex, race, cell type, degree of differentiation of the tumor, surgical approach, and date of operation were added to the analyses. The resulting logistic equation yielded a propensity score for each patient. Patients were then sorted according to their score. This revealed that about one third (tercile I) of the patients were unmatched (less than 10% chance of receiving induction therapy). In addition, the 19 patients downstaged by induction therapy to pT0/pN0 were unmatched. The remaining two terciles were well matched (Appendix Table I). The effect of induction therapy on survival was determined within each tercile as a whole and separately for the pN1 and pN0 cohorts.

Multivariable analysis. The second avenue of inquiry was multivariable risk factor analysis. The analysis was directed toward (1) identifying an incremental risk for death of induction therapy in both early and late hazard phases after accounting for the strongest predictors of outcome (pathologic stage variables) and (2) quantifying the cost if induction therapy indeed downstaged the patient to pN0. Because we anticipated a wide variability of risk, preliminary analyses were performed on 5 subgroups of patients: (1) cN1/pN0 or cN1/pN1, who received induction therapy; (2) the addition to

1 of cN1/pN1 patients undergoing surgery without induction therapy; (3) all pN1 patients; (4) all pN0 patients; and (5) pN0 patients, who underwent surgery without induction therapy. We then performed a combined analysis with interaction terms representing differences identified by the preliminary subgroup analyses.

So that risk-adjusted estimates of the effect of being downstaged by induction therapy to pN0 could be obtained, the final multivariable model was solved for two 62-year-old patients, both pT2 N0 M0 at thoracotomy, one downstaged to that pathology by induction therapy. These estimates are compared with a third 62-year-old patient at thoracotomy with pN1 disease and three positive nodes. The cost of induction therapy in the patient downstaged to pN0 was estimated as the area between the survival curves for the two pN0 patients. The benefit of induction therapy was estimated as the area between the survival curve of the patient downstaged to pN0 and the patient with pN1 disease.

Differences of areas represent the reduced and added years of life within the 10 years after surgery contributed by downstaging. For reasons previously given, these differences in lifetime are presented with 90% confidence limits (CL).¹⁰

Presentation. Confidence limits of survival estimates are equivalent to 1 standard error (68% CL). Tables of multivariable analyses are presented with their regression coefficients, standard errors, and P values, rather than an expression of relative risk. This is because the transformation of scale of continuous and ordinal variables makes conversion of regression coefficients to either odds ratios or hazard ratios uninterpretable. In addition, the hazard function models are not of proportional hazards, further reducing the interpretability of hazard ratios based solely on regression coefficients.

To obtain risk-adjusted estimates of the effect of N1 burden on 1- and 5-year survival, we solved the final multivariable equation for number of nodes from 1 to 10 with all other factors held constant: age 62 years, pN1 disease, no induction therapy, use of thoracotomy, and no metastatic disease.

Results

Staging and downstaging

Surgery groups. A total of 193 patients were clinically staged, received no induction therapy, and were treated with surgery (groups 3, 4, 7, and 8). Assessment of clinical staging of regional lymph node status is provided by the comparison of clinical to pathologic stage in these patients (Appendix Table II). The clinical determination of N1 had 80% accuracy, 78% sensitivity, 81% specificity, 81% positive predictive value, and 79% negative predictive value.

Induction therapy groups. One hundred patients were clinically staged and received induction therapy followed by surgery (groups 1, 2, 5, and 6). Clinical stage before induction and pathologic stage at resection are listed in Appendix Table III. Efficacy of induction therapy can be estimated by comparing the concordance

Fig 1. Survival in surgery group (no induction therapy). Survival, confidence limits, and number of patients still alive (in parentheses) at 1 and 5 years were as follows: group 3, 49% $\pm$ 6% (30) and 9% $\pm$ 4% (4); group 4, 89% $\pm$ 4% (58) and 68% $\pm$ 7% (21); group 7, 52% $\pm$ 12% (8) and 13% $\pm$ 8% (2); group 8, 76% $\pm$ 10% (13) and 43% $\pm$ 13% (4).

Table II. Pathologic depth of tumor invasion for patient groups

Group	Treatment	n	cT3 and cT4		pT3 and pT4	
			No.	%	No.	%
1*	IT	37	30	81	6	16
2	IT	32	25	78	26	81
3	SURG	75	69	92	70	93
4	SURG	79	15	19	11	14
5*	IT	23	15	65	5	22
6	IT	8	4	50	4	50
7	SURG	21	15	71	15	71
8	SURG	18	15	83	11	61

IT, Induction therapy followed by surgery; Surg, surgery alone.

*Downstaging of T and N by induction therapy.

and discordance of cN and pN between the surgery and induction therapy groups. For surgery groups, concordance of cN and pN was 80% (accuracy) and discordance, 20%; for induction therapy groups, concordance was 55% and discordance, 45%.

Survival

Surgery groups. Survival for the four surgical groups (groups 3, 4, 7, and 8) is displayed in Fig 1. The extremes of survival are seen in groups 3 (cN1/pN1) and 4 (cN0/pN0). Patients who were understaged and who were actually pN1 (group 7) had poor survival similar to group 3. Patients who were overstaged and who were actually pN0 (group 8) had intermediate survival. Compared with group 4, group 8 patients had a similar cT, but a higher pT status (61% vs 14%), despite both groups being pN0 (Table II).

Risk factors for death after esophagectomy without induction therapy, restricted to those factors known before induction therapy (Appendix I), included cN1 and increasing cT in cN0 patients (Table III).

Induction therapy groups. Survival for the four groups (groups 1, 2, 5, and 6) receiving induction therapy is displayed in Fig 2. Patients downstaged to pN0 (group 1) or cN0 patients who received IT (group 5) had intermediate survival and a similar pT (16% and 22%, Table II). Patients not downstaged (group 2) or patients understaged or progressing during induction therapy (group 6) had poor survival similar to groups 3 and 7.

Who will respond to induction therapy? Cell type was found to be associated with the “response” of cN1 disease. Fourteen of 17 patients with squamous cell

Fig 2. Survival in induction therapy groups. Survival, confidence limits, and number of patients still alive (in parentheses) at 1 and 5 years were as follows: group 1, 68% $\pm$ 8% (25) and 37% $\pm$ 9% (5); group 2, 52% $\pm$ 9% (16) and 12% $\pm$ 6% (3); group 5, 52% $\pm$ 10% (12), and 35% $\pm$ 10% (5); group 6, 50% $\pm$ 18% (4) and 0% (0).

Table III. Clinical factors associated with mortality after esophagectomy without preceding induction therapy (groups 3, 4, 7, and 8)

Incremental risk factor	Coefficient $\pm$ SD	P value
cN1 disease	4.2 $\pm$ 0.71	<.0001
cT in cN0 patients*	1.07 $\pm$ 0.20	<.0001

SD, Standard deviation.

*Clinical depth of tumor invasion has been expressed as an ordinal variable: 1 = cT1; 2 = cT2; 3 = cT3; and 4 = cT3 or cT4.

carcinoma converted to N0 (82%), in contrast to only 23 of 51 patients with adenocarcinoma (45%, $P = .01$). The one patient with adenosquamous carcinoma underwent induction therapy and did not respond.

Cost and benefit of therapy

Matched analysis. In general, patients receiving induction therapy were younger, had N1 disease, and greater T. This was true for both clinical stage (Table IV) and pathologic stage (Table V). In well-matched patients, there was a small, consistent, nonsignificant difference in survival for similar pathologic stages in patients receiving induction therapy (Appendix Table VI). Survival of patients in well-matched terciles II and III (Appendix Table I) had similar pattern of survival ($P = .4$). In both terciles II and III, patients receiving induction therapy had a slightly poorer survival at any

given point. In both terciles II and III, this pattern was seen in pN0 patients ($P = .4$ and $.6$) and pN1 patients ($P = .8$ and $.2$).

Multivariable analysis. Risk factor analysis based on pN at esophagectomy (Table VI) revealed that induction therapy, after accounting for the substantial risk of pN1, modestly increased early risk (Fig 3). In the typical patient, as shown by the areas beneath the survival curves in Fig 3, induction therapy reduced expected lifetime for a patient downstaged to pN0 by about one third compared with a patient with pN0 disease not receiving induction therapy. Nevertheless, the cost is offset by a large, 17-fold increase in lifetime for such a downstaged patient compared with pN1 disease.

Also, in the early hazard phase, there was an increased risk of death for pN0 patients who underwent thoracotomy. Late survival was affected by greater pT in pN0 patients and by a greater number of N1 nodes in pN1 patients (Fig 4). Pathologically identified metastatic disease further reduced survival.

Discussion

Staging and downstaging. EUS is a powerful clinical staging tool. In this study, accuracy in determination of regional lymph node status (N1) was 80%. This may be improved with future modifications of the equipment and the addition of EUS-directed fine-needle aspiration of regional lymph nodes. This technologic advance in clini-

Table IV. Clinical factors associated with induction therapy

Factor	Logistic coefficient $\pm$ SD	P value
Younger age*	-4.9 ± 1.76	.005
Older age	0.29 ± 0.116	.01
cN1 disease	0.62 ± 0.28	.03
Depth of tumor invasion† (cT)	-1.76 ± 0.81	.03
Intercept	-1.14	

SD, Standard deviation.

*exp[age/50] exponential transformation.

†[1/cT] inverse transformation, where clinical depth of tumor invasion has been expressed as an ordinal variable: 1 = cT < 1; 2 = cT1; 3 = cT2; and 4 = cT3 or cT4.

Table V. Factors associated with induction therapy, considering only pathologic information regarding the cancer (19 patients with pT0 are excluded)

Factor	Logistic coefficient $\pm$ SD	P value
Younger age*	-3.89 ± 1.12	.0005
Older age†	12.5 ± 4.3	.003
Higher pathologic stage‡	-3.3 ± 1.00	.0008
N1 burden§	-0.83 ± 0.21	.0001
Intercept	-36.9	

SD, Standard deviation.

*exp[age(years)/50] exponential transformation.

†ln[age(years)] natural logarithmic transformation.

‡1/[pStage] inverse transformation, where pathologic stage has been expressed as an ordinal variable: 0 = stage 0; 1 = stage I; 2 = stage IIA; 3 = stage IIB; 4 = stage III; 5 = stage IVA; 6 = stage IVB.

§ln[number of positive nodes + 1] natural logarithmic transformation.

cal staging has been aided by an accurate staging system that can be used to group patients with similar prognoses.

Once tumors have metastasized to regional lymph nodes, survival is poor. Surgery in these patients is little more than palliation. The identification of N1 disease and downstaging with induction therapy to pN0 are possible. This is evident by the greater discordance of cN and pN in the induction therapy group compared with the surgery groups. The discordance (inaccuracy of staging) seen in the surgery group was 20%. This discordance rose to 45% in the induction therapy groups. However, this is a combination of staging inaccuracy and downstaging. Thus, about 25% to 30% of patients will be downstaged to N0 by induction therapy. Unfortunately, accurate identification of responders to induction therapy is not possible at present.

Survival. Regardless of treatment, patients with pN1 disease (groups 2, 3, 6, and 7) have poor survival. There is slight modulation of survival in these patients by N1

Table VI. Risk factors for death after esophagectomy

Incremental risk factor	Coefficient $\pm$ SD	P value
Early hazard phase (~3 y)		
Older age*	-2.6 ± 0.96	.007
pN1 disease	2.2 ± 0.66	.0008
Induction therapy	0.71 ± 0.26	.006
Thoracotomy in pN0 disease	1.75 ± 0.66	.008
Late hazard phase (>~3 y)		
Greater depth of invasion (pT) in pN0 disease†	1.08 ± 0.24	< .0001
N1 burden‡	-7.9 ± 1.39	< .0001
Distant metastatic disease (M1a or M1b)	1.02 ± 0.41	0.01

SD, Standard deviation.

*[50/age(years)] inverse transformation.

†Numerically expressed as an ordinal variable: 1 = T0; 2 = THGD; 3 = T1; 4 = T2; 5 = T3; 6 = T4.

‡[1/(number of positive nodes + 1)] inverse transformation.

Table VII. Use of thoracotomy

Group	N	Thoracotomy	
		No.	%
1	37	36	97
2	32	32	100
3	75	65	87
4	79	37	47
5	23	22	96
6	8	8	100
7	21	19	90
8	18	15	83
Total	293	234	80

burden. In pN0 disease (groups 1, 4, 5, and 8), survival is intermediate to good and is substantially modulated by induction therapy, pT, and type of operation.

Cost and benefit of therapy. Induction therapy in pathologically matched patient groups produces a small difference in survival and in multivariable analysis a statistically significant difference. This difference is the result of multiple factors, including induction toxicity, the biologic process of downstaging that produces a lower stage that may not be equivalent to that same stage in a patient not receiving induction therapy (hysteresis effect of downstaging), possible progression of disease during therapy, and additional unknown factors.

The cost of induction therapy in responders is considerably outweighed by the survival benefit of conversion to pN0. In the future, if nonresponders can be identified, they can be spared the risks of induction therapy.

Fig 3. Cost and benefit of induction therapy. Estimates are risk-adjusted using the multivariable equation from Table VI. **A**, Time-related survival after esophagectomy in two 62-year-old patients, both pT2 N0 M0 at thoracotomy, one of whom had been downstaged to pN0 by induction therapy. **B**, As in **A**, except that for comparison, a similar pN1 patient with three positive lymph nodes is shown. The areas between the survival curves represent the loss of lifetime (cost) or gain of lifetime (benefit) of induction therapy. The cost to the patient downstaged by induction therapy is a loss of 1.7 years (90% CL, 0.4-2.9) of lifetime over a 10-year period, from 5.3 to 3.6 years (*uppermost shaded area*). The benefit is a gain of 3.4 years (90% CL, 2.5-4.4) of lifetime over the same 10-year period, from 0.2 to 3.6 years (*lower shaded area*).

Implications. Previous studies of induction therapy have ignored clinical stage and randomized patients into surgery-only and multimodality arms, assuming that randomization will be equal across stages.¹¹⁻¹⁴ This study illustrates several problems with such trials. For patients with early-stage tumors and excellent survival, small studies would be underpowered to detect a therapeutic effect in this growing

segment of patients. In addition, including such patients without accounting for stage may dilute the therapeutic effect of induction therapy. Another problem of randomized trials in this setting is that they may still be susceptible to the very selection bias they are designed to protect against.¹⁵ Patients who are deemed to have either good or bad therapy risk may or may not be included at all in the trial. Further, as

Fig 4. Survival decrement by N1 burden. The figure is a nomogram from the multivariable equation (Table VI) illustrating risk-adjusted 1- and 5-year survival according to number of positive regional lymph nodes. See “Methods” for details.

shown by this study, there is inaccuracy in clinical staging of about 20%. This is similar in magnitude to the proportion of patients who may respond to induction therapy. Both inaccuracy and lack of response will tend to dilute and hide the beneficial effect of induction therapy (Appendix II).

Presently we depend on clinical stage to direct treatment. The finding of cN1 predicts poor survival, and those patients who are suitable candidates are offered induction chemoradiation therapy and surgery in a protocol setting. Those patients who are understaged and found to have N1 disease at resection are offered post-operative adjuvant chemoradiation therapy. Although this treatment protocol is not perfect, it has allowed stratification of patients for both prognosis and evaluation of therapy. We strongly recommend that all patients with esophageal carcinoma be clinically staged and have their treatment modified accordingly.

Limitations

Limitations of this study include the inaccuracy of clinical staging and the inability to restage patients accurately after induction therapy. This prompted the addition of groups 5 through 8 to highlight these inaccuracies and the realities of treatment of esophageal cancer. Another limitation is that this is a modest-sized nonrandomized study with treatment biases over time. We have attempted to balance these problems with propensity matching and multivariable analysis.

We thank Diane Baisden for data collection and follow-up and Lucinda Mitchin for manuscript preparation.

Received for publication June 15, 2000; revisions requested Sept 21, 2000; revisions received Oct 13, 2000; accepted for publication Oct 24, 2000.

Address for reprints: Thomas W. Rice, MD, The Cleveland Clinic Foundation, 9500 Euclid Ave, Desk F25, Cleveland, OH 44195 (E-mail: ricet@ccf.org).

REFERENCES

1. Devesa SS, Blot WJ, Fraumeni JF Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer* 1998;83:2049-53.
2. Rice TW, Adelstein DJ. Precise clinical staging allows treatment modification of patients with esophageal carcinoma. *Oncology* 1997;11:58-62.
3. American Joint Committee on Cancer. *AJCC Cancer Staging Manual*, 5th ed. Philadelphia: Lippincott-Raven; 1997. p. 65-8.
4. Adelstein DJ, Rice TW, Becker M, Larto MA, Kirby TJ, Koka A, et al. Use of concurrent chemotherapy, accelerated fractionation radiation, and surgery for patients with esophageal carcinoma. *Cancer* 1997;80:1011-20.
5. Adelstein DJ, Rice TW, Rybicki LA, Larto MA, Ciezki J, Saxton J, et al. Does paclitaxel improve the chemoradiotherapy of locoregionally advanced esophageal cancer? A nonrandomized comparison with fluorouracil-based therapy. *J Clin Oncol* 2000;18:2032-9.
6. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;53:457-81.
7. Blackstone EH, Naftel DC, Turner ME Jr. The decomposition of time-varying hazard into phases, each incorporating a separate

- stream of concomitant information. *J Am Stat Assoc* 1986;81:615-24.
8. Baskerville JC, Toogood JH. Guided regression modeling for prediction and exploration of structure with many explanatory variables. *Technometrics* 1982;24:9-17.
 9. Chen CH, George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Stat Med* 1985;4:39-46.
 10. American College of Cardiology/American Heart Association Task Force on Assessment of Diagnostic and Therapeutic Cardiovascular Procedures (Subcommittee on Coronary Artery Bypass Graft Surgery. Kirklin JW, Akins CW, Blackstone EH, Booth DC, Califf RM, Cohen LS, Hall RJ, Harrell FE Jr, Kouchoukos NT, McCallister BD, Naftel DC, Parker JO, Sheldon WC, Smith HC, Wechsler AS, Williams JF Jr). Guidelines and indications for coronary artery bypass graft surgery. *J Am Coll Cardiol* 1991;17:543-89 and *Circulation* 1991;83:1125-73.
 11. Walsh TN, Noonan N, Hollywood D, Kelly A, Keeling N, Hennessy TP. A comparison of multimodal therapy and surgery for esophageal adenocarcinoma. *N Engl J Med* 1996;335:462-7.
 12. Bosset JF, Gignoux M, Triboulet JP, Tiet E, Manton G, Elias D, et al. Chemoradiotherapy followed by surgery compared with surgery alone in squamous cell cancer of the esophagus. *N Engl J Med* 1997;337:161-7.
 13. Law S, Fok M, Chow S, Chu KM, Wong J. Preoperative chemoradiotherapy versus surgical therapy alone for squamous cell carcinoma of the esophagus: a prospective randomized trial. *J Thorac Cardiovasc Surg* 1997;114:210-7.
 14. Kelsen DP, Ginsberg R, Pajak TF, Sheahan DG, Gunderson L, Mortimer J, et al. Chemotherapy followed by surgery compared with surgery alone for localized esophageal cancer. *N Engl J Med* 1998;339:1979-84.
 15. Piantadosi S. Bias and random error. In: Piantadosi S. *Clinical trials: a methodological perspective*. New York: John Wiley; 1997. p. 106-25.

Appendix I. Variables examined in multivariable analyses

Demography

Sex, race, age (years)

Pre-induction therapy (or preoperative) clinical staging

Each clinical depth of tumor invasion (cT) stage (0, Tis, 1-4), an ordinal variable representing these (1 = cT0, 2 = cTis, 3 = cT1, 4 = cT2, 5 = cT3, 6 = cT4), and a dichotomous variable for cT > 2; clinical N1 disease.

Pre-esophagectomy therapies

Previous esophagogastric operations

Preoperative induction therapy

Surgical approach

Thoracotomy versus transhiatal

Tumor characteristics

Cell type (adenocarcinoma, squamous cell, adenosquamous); each histologic grade of tumor differentiation and an ordinal variable representing differentiation (0 = well, 1 = moderately well, 2 = moderate, 3 = moderately poor, and 4 = poor); R0 resection; presence of distant metastases, either M1a or M1b.

Pathologic staging

Pathologic depth of tumor invasion (pT) expressed similar to clinical staging; pathologic N1 disease (pN1), number of nodes sampled, number of nodes positive, and percent of nodes sampled found to be positive (both are expressions of N1 burden); pathologic stage (pStage) (ordinal variable, where 0 = stage 0, 1 = stage I, 2 = stage IIA, 3 = stage IIB, 4 = stage III, 5 = stage IVA, 6 = stage IVB).

Downstaging

Difference in T, difference in N.

Postoperative

Use of postoperative chemotherapy and/or radiotherapy

Surgical experience

Date of operation, expressed continuously as years between January 1, 1985, and the date of esophagectomy

Note: All dichotomous variables were coded as 0 for absence of the factor and 1 for presence of the factor. Thus, the factor pN1 disease (Table VI) is 0 for pN0 disease and 1 for pN1 disease.

Appendix II: Simulation of randomized trial of induction therapy

We simulated a randomized clinical trial with these nonrandomized patients by a propensity-matched analysis based on preinduction therapy clinical staging (Table IV) and other non-significant factors, as described in the "Methods" section. As in the propensity matching using pathologic stage, only two terciles of patients were well matched. In tercile II, survival was 51% and 16% at 1 and 5 years in patients not receiving induction therapy, and 45% and 14% at these same times in patients receiving induction therapy ($P = .7$). In tercile III, survival was 54% and 36% at 1 and 5 years in patients not receiving induction therapy, and 65% and 28% at these same times in patients receiving induction therapy ($P = .9$). However, when these same patients were included in a multivariable analysis of survival, adjusting for propensity and pathologic stage at esophagectomy, a clear benefit of downstaging became evident ($P = .0002$). Thus, the unadjusted comparison hid the benefit of induction therapy for those who responded by downstaging.

Discussion

Dr Richard Whyte (*Stanford, Calif*). The objective was to determine the relevance of clinical staging and the importance of downstaging. The results, not surprisingly, showed that the earlier stage patients did the best, the most advanced stage patients did the worst, and the patients who received induction therapy were somewhere in between. Furthermore, of those that received induction therapy, patients with N1 disease did

Appendix Table I. Propensity-matched terciles, based on pathologic information

Variable	Tercile I*		Tercile II			Tercile III		
	No IT	IT	No IT	IT	P	No IT	IT	P
No.	88	3	67	24		38	54	
Male (%)	82	100	82	88	.5	84	80	.6
Age (y)	66 ± 12	70 ± 8	64 ± 10	61 ± 10	.2	58 ± 6.8	59 ± 7.5	.6
White (%)	100	100	96	92	.5	89	92	.6
Thoracotomy (%)	35	67	100	100	1.0	100	100	1.0
pT3/pT4 (%)	48	0	67	71	.7	53	44	.4
pN1 (%)	41	33	63	67	.7	47	43	.7
Positive nodes†	7	—	4	3.5	.8	2	2	.12
Adenocarcinoma (%)	92	33	81	88	.4	71	74	.8
Distant metastatic disease (M1a or M1b) (%)	0	0	15	17	.8	5	7	.7
pStage III or IV (%)	41	0	55	71	.18	39	32	.4
Poorly differentiated (%)	36	67	40	54	.2	55	52	.7
Date of operation‡	10 ± 3.2	8.5 ± 3.6	9.1 ± 3.4	9.0 ± 3.1	.9	9.9 ± 3.9	9.9 ± 2.3	.9

IT, Induction therapy.

*No valid comparisons (essentially unmatched for induction therapy). Also note that no comparison could be made for the 19 patients downstaged by induction therapy to pT0/pN0.

†Median number in pN1 patients.

‡Number of years between January 1, 1985, and date of esophagectomy.

Appendix Table II. Clinical (cN) and pathologic (pN) stage in surgery groups

Pathologic stage	Clinical stage		Total
	cN1	cN0	
pN1			
Group	Group 3	Group 7	
No. (%)	75 (39%)	21 (11%)	96
pN0			
Group	Group 8	Group 4	
No. (%)	18 (9%)	79 (41%)	97
Total	93	100	193

The percentages are of n = 193. Groups 3 and 4 represent true “positives” and “negatives,” respectively, and groups 8 and 7 false “positives” and “negatives,” respectively.

Appendix Table III. Preinduction therapy and pathologic stage in induction therapy groups

Pathologic stage	Pre-IT stage		Total
	cN1	cN0	
pN1			
Group	Group 2	Group 6	
No. (%)	32 (32%)	8 (8%)	40
pN0			
Group	Group 1	Group 5	
No. (%)	37 (37%)	23 (23%)	60
Total	69	31	100

Clinical (cN) and pathologic (pN) stage in preinduction therapy (pre-IT) groups. Groups 2 and 5 represent true “positives” and “negatives,” respectively, and groups 1 and 6 false “positives” and “negatives,” respectively.

worse than those with N0 disease. This, I thought, was fairly straightforward. The manuscript, on the other hand, was anything but straightforward. Somehow the 1-page abstract was transformed into 44 pages, including 9 tables, 6 figures, 2 appendixes, and enough statistics, including *P* values, adjusted coefficients, confidence limits, terciles, propensity matching, and mathematical modeling to make my head spin. To be fair, however, much of the text and several of the figures are contained in two appendixes, which detail the complex statistical methods used to analyze the data. Nonetheless, I did indeed find the statistics a challenge. I would like to thank both Craig Miller, my colleague, for convincing me that the statistics are both appropriate and, as he terms it, “the wave of the future,” and Tom Rice for a quick tutorial on these statistics on

Appendix Table IV. Comparison of 1- and 5-year survival after surgery without and with preceding induction therapy in well-matched terciles

Tercile	Survival (%)			
	Surgery		Induction therapy	
	1 y	5 y	1 y	5 y
II	58	28	48	16
pN0	79	49	50	38
pN1	44	16	48	14
III	65	23	61	22
pN0	73	33	68	35
pN1	57	11	52	7

Thursday morning. Fortunately, Tom has chosen not to present all of the statistics today but instead has chosen two key features: the hazard functions and a simulated survival curve of a hypothetical patient based on his mathematical model. This model allows one to perform a simulated randomized trial of two treatment modes. While the model takes into account many of the quantitative aspects of patient selection, it fails to account for some of the more qualitative issues such as patient bias, physician bias, and patient performance status. To borrow from Dr Matloff's presentation, if one were to compare clinical trials to government, I would submit that a randomized prospective study is like democracy: it is not perfect but it is the best system we have.

I have 3 questions for Dr Rice. First, numerous other trials of neoadjuvant therapy for esophageal cancer have shown that complete pathologic response including both tumor and nodal response is a predictor of better outcome. Why did this study instead ignore tumor response and use an assumption of nodal response? I say assumption since there is generally no pretreatment proof of N1 disease. What was the complete pathologic response rate in these patients with post-treatment N0 disease?

Dr Rice. The problem with clinical staging is that the pathologic stage is not known until all the treatment is completed. If induction therapy is successful, then the pathologic stage is changed. I agree that patients who have a complete response, no residual tumor, are the best group of patients. Twenty percent of our patients had complete sterilization of the tumor and the regional lymph nodes with induction therapy. However, those patients who were T0 but still N1 at resection had a poor survival. So sterilization of the tumor is not the key. Regional lymph node downstaging is crucial. We did not include the data on T downstaging because it made the presentation and paper even more complex. I plan to write a companion article about the implications of T staging and downstaging. Twenty percent of our patients had sterilization of the tumor and regional nodes, and approximately 30%, a third of the patients, had sterilization of regional nodes.

Dr Whyte. My second question concerns the clinical staging. You had an error rate of 20% based on clinical staging—including both false positives and false negatives. Would this justify more aggressive surgical staging such as more liberal use of needle biopsies, EUS-guided needle biopsies, or thoracoscopic or laparoscopic staging?

Dr Rice. Staging is in its infancy and is certainly evolving. No doubt clinical staging will improve with improved instruments and technology. The addition of EUS-guided fine needle aspiration of regional nodes will improve the determination of regional lymph node status. Thoracoscopic and laparoscopic staging, which is not the equivalent of mediastinoscopy in lung cancer, is hard to do. It is a difficult procedure that is essentially another operation. It should be reserved for those questions that are not answered by clinical staging. There is a role for invasive staging, but not in every patient.

Dr White. Should we now consider a simulated or, to use the computer parlance of the day, a "virtual" randomized study to settle these types of questions, particularly when there are true randomized studies or randomized clinical trials that address the same question yet fail to find significant differences in the two treatment arms?

Dr Rice. Simulated randomization would never replace a truly randomized phase III study. However, the present studies as they were conceived do not randomize patients by stage. This is a major problem, because the small effect of induction therapy on a few patients is hidden by the inclusion of very good risk patients. Not until we have sophisticated studies that randomize patients by stage, are well conducted, and are adequately analyzed are we ever going to have a phase III study that will answer the questions. If we continue to conduct studies of esophageal cancer as we have, every study will be a negative study. The answers will be hidden because of the inadequacies of the preparation and conduct of the study.

Dr Donald Low (Seattle, Wash). Dr Rice, I enjoyed your paper as well. Previous publications, however, have tended to take the nodal status issue and look on it not only positive-negative but number of nodes positive and number of stations positive. Did you stratify the patients who were pN1 positive to the number of nodes and to the number of stations positive?

Dr Rice. We did analyze by number of positive nodes. There is an exponential decreasing survival. The break point is at three positive nodes. After, survival plateaus at a very low level. With one node positive, 5-year survival is about 40%, with two nodes positive it is a bit less than 20%, and once there are three nodes or more positive, 5-year survival is less than 5%. The percentage of lymph nodes positive in resection specimens is also important, and the break point is around 20%. We did not look at the number of perihilar stations. There is no doubt that lymph node burden is crucial in further defining N1 disease.

James Mark (Stanford, Calif). I was struck with the staging question also. You had some cN1 pN0 patients who were operated on alone. In other words, nothing had been given them and they were overstaged before the operation. I would suggest that if you are going to use patients like that as a control or use noninvasive staging to get those answers, you subtract that number from the ones that are downstaged by induction therapy, because some are downstaged with no therapy.

Dr Rice. Understaging/overstaging was included in the analysis. The analysis in the paper includes 8 groups: the 4 additional groups were patients overstaged, patients understaged, N0 patients who received induction therapy because of advanced T, and patients who were cN0 at induction but were pN1 at resection; this represents either misstaging or progression on treatment. The analysis of all 8 groups gets pretty complex. For simplicity, I presented 4 groups that encompass the two main concepts, staging and downstaging. The detailed analysis of all 8 groups is included in the paper.