Supplementary Information

Statistical Analysis: Development and Evaluation of the Multivariate Logistic Model

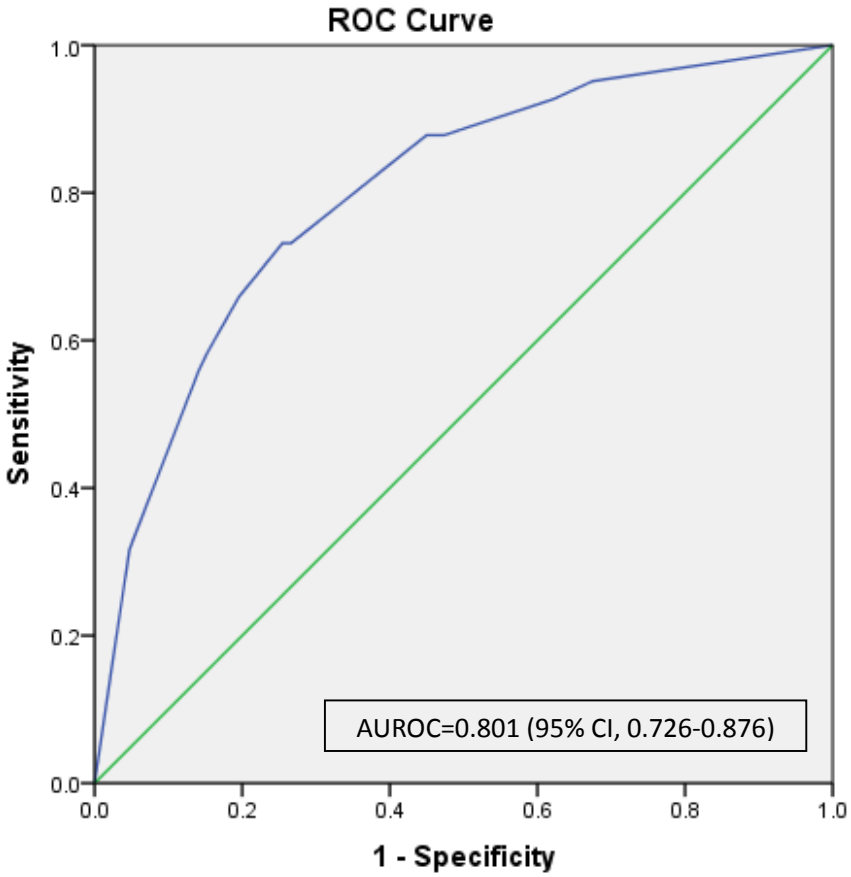
Significance of associations with the outcome of nodal metastases were first evaluated using a univariate logistic model for depth of invasion, angiolymphatic invasion, tumor grade and tumor size (Table 3). Clinically relevant pathologic variables were all included the multivariable model (Table 3). The fit of the model to the data was assessed by the Hosmer-Lemeshow statistic and standardized residuals. The multivariable model was well-calibrated to the data with no evidence for lack of fit (with a Hosmer-Lemeshow chi squared value of approximately 1.0). Exclusion of cases with standardized residuals > 3 did not significantly alter the accuracy of model predictions; hence, all cases were included. Once the final model was fit, the resulting model coefficients were applied to the cohort to calculate predicted values from the logistic equation, i.e. $\hat{y} = 1/(1 + \exp[-(X\beta)])$. Predicted probabilities were then calculated from the logistic function using the following weighted sum:

$$X\beta = -3.505 + 0.932(tumor\ size \geq 2cm) + 0.990(stage\ T1b) + 1.148(high\ grade) + 0.807(ALI)$$

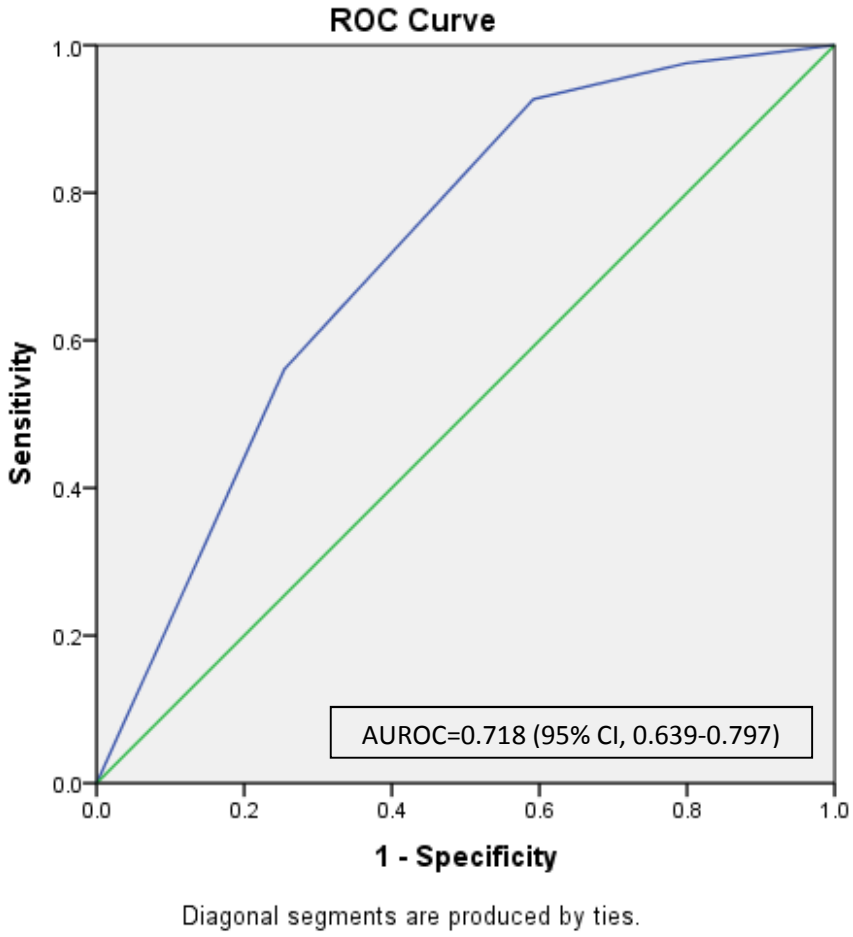
where presence of a given binary risk factor (in italics) is denoted by a value of 1 (versus 0 for absence of that variable). The model coefficients and classification accuracy rate was validated using a 10-fold cross validation procedure, finding a pattern of coefficients similar to the overall model in each iteration (data not shown). Also, the average correct classification rate for nodal metastasis in the 10-fold cross validation (83.0±8.1%) was similar to the correct classification rate of the overall model (82.9%).

The area under the receiver operating characteristic (ROC) curve (AUROC) was calculated for the data to produce an estimate of the model's discriminatory ability. Predicted risk groups were defined *a priori* based on clinically motivated cut-off values at 5%, 10% and 20% predicted probability of nodal metastasis.

Supplementary Figure 1: Receiver Operator Curve (ROC) Analysis of Model Estimated Probabilities of Nodal Metastasis for Predicting Nodal Metastasis at Esophagectomy



Supplementary Figure 2: Receiver Operator Curve (ROC) Analysis of a Logistic Model based only on Depth of Invasion (classified as T1a-superficial, T1a-deep, T1b-superficial and T1b-deep) as a Predictor of Nodal Metastasis at Esophagectomy



Supplementary Table 1: Validation Cohort Cases Diagnosed by Endoscopic Resection with Subsequent Esophagectomy

CaseID	Reason for Esophagectomy	EMR Tstage	Grade	Size	ALI	Estimated Risk Group	Final Tstage	pN stage	pM stage	# LN Pos	# LN Eval	Induction?
EMR1	Positive margins on ER	T1a	LG	<2cm	-	<5%	T1a	N0	M0	0	29	No
EMR2	ALI+	T1a	LG	<2cm	+	5-10%	T1a	N1	M0	1	11	No
EMR3	ALI+	T1a	LG	<2cm	+	5-10%	T1a	N0	M0	0	22	No
EMR4	Elective	T1a	LG	<2cm	-	<5%	T1a	N0	M0	0	16	No
EMR5	Positive margins on ER	T1a	LG	<2cm	-	<5%	T1a	N0	M0	0	26	No
EMR6	Positive margins on ER; ALI+	T1a	LG	<2cm	-	<5%	T1b	N0	M0	0	37	No
EMR7	Long seg BE with multifocal T1a and extensive HGD	T1a	LG	≥2cm	-	5-10%	T1a	N0	M0	0	28	No
EMR8	Long seg BE with T1a and extensive HGD	T1a	LG	≥2cm	-	5-10%	T1a	N0	M0	0	16	No
EMR9	Positive margins on ER; ALI+	T1a	LG	<2cm	+	5-10%	T1b	N0	M0	0	21	No
EMR10	Elective	T1a	LG	<2cm	-	<5%	T1a	N0	M0	0	13	No
EMR11	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N0	M0	0	8	No
EMR12	T1b on ER	T1b	LG	<2cm	-	5-10%	T1b	N0	M0	0	14	No
EMR13	T1b on ER	T1b	HG	<2cm	-	20%	T1b	N0	M0	0	21	No
EMR14	T1b on ER	T1b	HG	≥2cm	+	30-60%	T1b	N1	M0	2	24	No
EMR15	T1b on ER	T1b	LG	<2cm	+	5-10%	T1b	N0	M0	0	23	No
EMR16	T1b on ER	T1b	HG	<2cm	-	20%	T1b	N0	M0	0	23	No
EMR17	T1b on ER	T1b	LG	≥2cm	-	5-10%	T1b	N0	M0	0	13	No
EMR18	T1b on ER	T1b	HG	≥2cm	+	30-60%	T1b	N1	M0	2	23	Yes
EMR19	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N0	M0	0	26	No
EMR20	T1b on ER	T1b	LG	<2cm	-	5-10%	T1b	N0	M0	0	29	No
EMR21	T1b on ER	T1b	LG	≥2cm	+	30-60%	T1b	N0	M0	0	16	No
EMR22	T1b on ER	T1b	LG	<2cm	-	5-10%	T1b	N0	M0	0	25	No
EMR23	T1b on ER	T1b	LG	≥2cm	+	30-60%	T1b	N0	M0	0	26	No
EMR24	T1b on ER	T1b	LG	<2cm	+	15-20%	T1b	N0	M0	0	5	No
EMR25	T1b on ER	T1b	HG	<2cm	-	20%	T1b	N0	M0	0	23	No
EMR26	T1b on ER	T1b	LG	≥2cm	+	30-60%	T1b	N0	M0	0	19	No
EMR27	T1b on ER	T1b	LG	≥2cm	+	30-60%	T1b	N2	M0	3	33	No
EMR28	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N1	M0	2	23	No
EMR29	T1b on ER	T1b	LG	<2cm	-	5-10%	T1b	N0	M0	0	39	No
EMR30	T1b on ER	T1b	LG	≥2cm	-	15-20%	T2	N0	M0	0	11	No
EMR31	T1b on ER	T1b	LG	<2cm	-	5-10%	T1b	N0	M0	0	17	No
EMR32	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N1	M0	2	47	No

EMR33	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N0	M0	0	25	No
EMR34	T1b on ER	T1b	LG	≥2cm	-	15-20%	T3	N2	M0	3	38	No
EMR35	T1b on ER	T1b	HG	<2cm	-	20%	T1b	N0	M0	0	23	No
EMR36	T1b on ER	T1b	HG	<2cm	+	30-60%	T1b	N0	M0	0	18	No
EMR37	T1b on ER	T1b	HG	<2cm	+	30-60%	T1b	N2	M0	6	22	No
EMR38	T1b on ER	T1b	LG	≥2cm	-	15-20%	T1b	N0	M0	0	31	No
EMR39	T1b on ER	T1b	HG	<2cm	+	30-60%	T1b	N0	M0	0	31	No