Postoperative Nomogram for Disease-Specific Survival After an R0 Resection for Gastric Carcinoma

By Michael W. Kattan, Martin S. Karpeh, Madhu Mazumdar, and Murray F. Brennan

Purpose: **Few published studies have addressed individual patient risk after R0 resection for gastric cancer. We developed and internally validated a nomogram that combines these factors to predict the probability of 5-year gastric cancer–specific survival on the basis of 1,039 patients treated at a single institution.**

Methods: **Nomogram predictor variables included age, sex, primary site (distal one-third, middle one-third, gastroesophageal junction, and proximal one-third), Lauren histotype (diffuse, intestinal, mixed), number of positive lymph nodes resected, number of negative lymph nodes resected, and depth of invasion. Death as a result of gastric cancer was the predicted end point. The concordance index was used as an accuracy measure, with bootstrapping to correct for optimistic bias. Calibration plots were constructed.**

Results: **Gastric cancer–specific survival at 5 years was 50%. A nomogram was constructed on the basis of a Cox regression model. The bootstrap-corrected concordance index was 0.80. When compared with the predictive ability of American Joint Committee on Cancer stage, the nomogram discrimination was superior ($P < .001$). Nomogram calibration appeared to be excellent.**

Conclusion: **A nomogram was developed to predict 5-year disease-specific survival after R0 resection for gastric cancer. This tool should be useful for patient counseling, follow-up scheduling, and clinical trial eligibility determination.**

J Clin Oncol 21:3647-3650.

GASTRIC CANCER continues to be a major global health problem.[1] In the United States, gastric cancer ranks 14th in cancer incidence, with approximately 22,000 new cases diagnosed each year. The prognosis of these patients is primarily related to the extent of disease at presentation, but for the majority of patients with gastric cancer, treatment is the same regardless of recurrence risk. The American Joint Committee on Cancer (AJCC) has developed stage groupings that stratify disease-specific survival after an R0 gastrectomy into six groups. The stage groupings are based on the depth of tumor invasion and the number of lymph node metastases, and stratify patients into risk groups. For the individual patient, however, risk varies substantially within stage.

Prognostic nomograms are tools designed explicitly for prediction. They do not produce risk groups. Instead, they attempt to combine all proven prognostic factors and quantify risk as precisely as possible. Previous comparisons with risk grouping approaches in prostate cancer and soft tissue sarcoma suggest improved predictive accuracy relative to the formation of risk groups.[2,3]

The purpose of this study was to develop and internally validate a prognostic nomogram for patients who have had an R0 resection for gastric cancer. We compared the predictive accuracy of this nomogram with that achieved by the AJCC stage grouping in an effort to determine whether progress has been made in estimating prognosis. Improved prediction of patient outcome would be useful for counseling patients, designing novel therapeutic strategies, and scheduling patient follow-up.

METHODS

Between July 1, 1985 and June 30, 2002, 1,136 patients had R0 resections for gastric cancer at Memorial Sloan-Kettering Cancer Center. For the end point of disease-specific survival, we believed that the following variables would be widely available and potentially prognostic: age, sex, primary site (distal one third, middle one third, proximal one third, and gastroesophageal junction), Lauren histotype (diffuse, intestinal, mixed), number of positive lymph nodes resected, number of negative lymph nodes resected, and depth of invasion as defined by the standard nomenclature.[4] Patients with one or more missing values were excluded (Lauren histotype, n = 38; size, n = 27; number of nodes negative, n = 3; depth, n = 50), leaving 1,039 complete patient records. Patients were observed until death, and cause of death was recorded.

Disease-specific survival was estimated using the Kaplan-Meier method. Cox proportional hazards regression was used for multivariable analysis. Ordinal and continuous variables were fit using restricted cubic splines[5] to relax the linearity assumptions. No variable selection was performed. This Cox model was the basis for the nomogram.

Nomogram validation comprised two activities. First, discrimination was quantified with the concordance index.[6] Similar to the area under the receiver operating characteristic curve, but appropriate for censored data, the concordance index provides the probability that, in a randomly selected pair of patients in which one patient dies before the other, the patient who died first had the worse predicted outcome from the nomogram. We used bootstrapping to obtain a relatively unbiased estimate.

Second, calibration was assessed. This was done by grouping patients with respect to their nomogram-predicted probabilities and then comparing the mean of the group with the observed Kaplan-Meier estimate of disease-specific survival. Again, bootstrapping correction was used for this activity. All analyses were performed using S-plus 2000 Professional software (Statistical Sciences, Seattle, WA) with the Design and Hmisc libraries added.[7]

RESULTS

Descriptive statistics for this cohort appear in Table 1. At last follow-up, 427 patients had died as a result of their disease, and

From the Departments of Urology, Epidemiology and Biostatistics, and Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY.

Submitted January 31, 2003; accepted July 23, 2003.

Supported by a grant from the Gelb Foundation.

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Murray F. Brennan, MD, Memorial Sloan-Kettering Cancer Center, Department of Surgery, 1275 York Ave, C1275 New York, NY 10021; e-mail: brennanm@mskcc.org.

0732-183X/03/2119-3647/$20.00

Table 1. Descriptive Statistics for Gastric Cohort

Patient Characteristic	No.	%
Sex		
Male	415	36.5
Female	721	63.5
Primary site		
Antrum or pyloric	325	28.6
Body or middle one third	248	21.8
Gas esophageal function	376	33
Proximal or upper one third	187	16.5
Lauren histotype		
Diffuse	359	31.6
Intestinal	641	56.4
Mixed	98	8.6
Not available	38	3.3
Stage 97		
0	17	1.5
IA	211	18.6
IB	181	15.9
II	244	21.5
IIIA	255	22.5
IIIB	135	11.9
IV	93	8.2
Depth		
Mucosa	94	8.3
Submucosa	156	13.7
Propria muscularis	138	12.1
Subserosa	245	21.6
Suspected serosal invasion	24	2.1
Definite serosal invasion	389	34.2
Adjacent organ involvement	40	3.5
No. of positive nodes		
Minimum	0	
First quartile	0	
Median	1	
Mean	4	
Third quartile	6	
Maximum	60	
No. of negative nodes		
Minimum	0	
First quartile	10	
Median	17	
Mean	20	
Third quartile	27	
Maximum	84	
Not available	3	
Size		
Minimum	0	
First quartile	2.5	
Median	4.3	
Mean	4.8	
Third quartile	6.5	
Maximum	21	
Not available	27	

57% of the living patients were assessed within 1 year of analysis. Disease-specific survival by AJCC stage grouping is shown in Figure 1. There appears to be a sufficient number of patients at risk at 9 years to make a nomogram for this time-point prediction in addition to the 5-year prediction.

From the Cox model, age at diagnosis ($P < .001$), primary site ($P < .001$), number of positive nodes ($P < .001$), and depth ($P < .001$) were associated with disease-specific survival, whereas sex

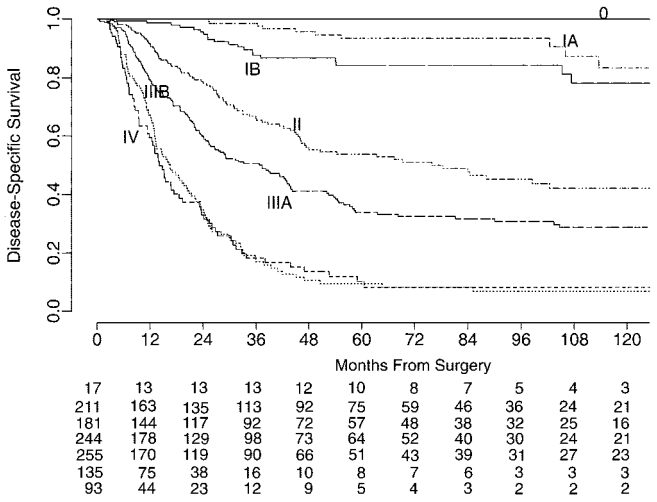


Fig 1. Disease-specific survival by American Joint Committee on Cancer stage grouping. Numbers beneath *x*-axis indicate patients at risk.

($P = .06$), Lauren histotype ($P = .26$), and size ($P = .58$) were not. A nomogram on the basis of this Cox model appears in Figure 2. The concordance index for the model was 0.80.

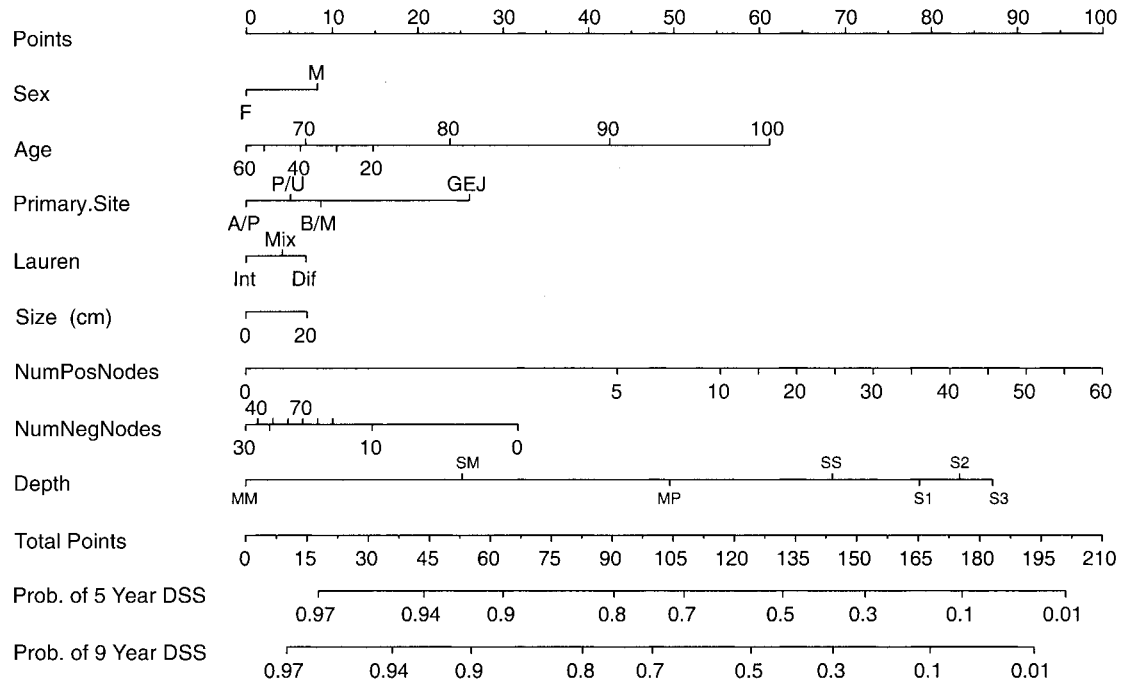
Figure 3 illustrates the calibration of the nomogram. Calibration appears to be accurate for both the 5- and 9-year predictions.

Finally, we compared predictions from the nomogram with those obtained by using the AJCC stage groupings. Individual AJCC stage groups and nomogram predictions were compared for their ability to rank the patients (eg, concordance index). To correct for overfit, nomogram predictions were calculated on a leave-one-out basis so that each patient was not included in the model that produced his or her prediction. Nomogram discrimination was superior to that of AJCC stage grouping (concordance index 0.80 *v* 0.77; $P < .001$). This difference is difficult to appreciate clinically, and therefore, Figure 4 illustrates the discrepancies between the two prediction methods. Within each AJCC stage grouping is a histogram of nomogram-predicted probabilities. Note the heterogeneity in several of the AJCC stage groups, particularly groups II and IIIA. Unlike the nomogram, the AJCC system would provide the same prognostic estimate for all group II or IIIA patients.

DISCUSSION

On the basis of our patient series, we have developed and internally validated a tool for predicting disease-specific survival after surgery for gastric cancer, assuming that all gross disease was resected with negative resection margins. The tool predicts with a concordance index of 0.80, and thus is not perfect, but it is more predictive than AJCC stage grouping alone. The clinical magnitude of these discrepancies is greatest in stages II and IIIA disease (Fig 4). For example, the average 5-year probability of death as a result of gastric cancer for a stage II patient in this study was 0.57 but ranged from 0.09 to 0.90. Under the current standard of care, all of these patients are given the same treatment and are expected to have the same prognosis solely on the basis of their AJCC stage group.

Interestingly, the number of negative nodes removed and age appear to have nonmonotonic relationships with disease-specific



Instructions for Physician: Locate the patient's sex on the **Sex** axis. Draw a line straight upwards to the **Points** axis to determine how many points towards gastric cancer-specific death the patient receives for his or her sex. Repeat this process for the other axes, each time drawing straight upward to the **Points** axis. Sum the points achieved for each predictor and locate this sum on the **Total Points** axis. Draw a line straight down to the disease-specific survival axes to find the patient's probability of surviving gastric cancer assuming he or she does not die of another cause first.

Fig 2. Nomogram for disease-specific survival (DSS). NumPosNodes, number of positive nodes; NumNegNodes, number of negative nodes; Prob.; probability; A/P, antrum or pyloric, B/M, body or middle one third, GEJ, gastroesophageal junction; P/U, proximal or upper one third; int, intestinal; mix, mixed; Dif, diffuse; MM, mucosa; MP, propria muscularis; S1, suspected serosal invasion; S2, definite serosal invasion; S3, adjacent organ involvement; SM, submucosa; and SS, subserosa.

survival. The predictive ability of the nomogram decreased when these relationships were forced to be linear (results not shown). We assume that because the differences over the nonmonotonic range for negative lymph nodes (< 10) are small, this selects for a small number of patients that are inadequately staged, and emphasizes the importance of examining more than 15 nodes for accurate staging. The observation that young patients have a worse prognosis than those at the common age of incidence has been seen in a number of prior studies.

Improving the accuracy of our prognostic estimates is exceedingly important for making treatment recommendations. Patients deserve accurate predictions when they request them, and we should therefore use the most accurate methods currently available when formulating these predictions. For example, the elderly patient with stage II gastric cancer may be quite relieved to hear that, on the basis of a more accurate prognostic tool, his or her probability for 5-year disease-specific survival exceeds 50%. On the other hand, a young patient with stage IB disease may accept the recommendation to receive postoperative adjuvant therapy more easily with a nomogram that predicts a probability of 30% 5-year disease-specific survival.

Illustrations such as this call into question how postoperative treatment recommendations are made to patients. In principle, the patients should be chosen for adjuvant treatment whose risk after conventional therapy exceeds some defined threshold where evidence of potential benefit exists.

Our results suggest that choosing patients solely on the AJCC stage is, at a minimum, inexact. Instead, we now have the ability to calculate risk of treatment failure more accurately, and offer the experimental treatment if risk exceeds a threshold. For example, depending on the threshold chosen, use of the nomogram (Fig 2) rather than the AJCC stage grouping may have dramatic effects on who is offered postoperative adjuvant therapy and who is not. This can be seen by identifying a single threshold, or a range of probabilities (Fig 4), that transcends the AJCC stage. It is important to emphasize the need for adequate (> 15) lymph node sampling to improve the accuracy of risk projection.[8]

There are important limitations to our study. In particular, our nomogram is not perfectly accurate. It merely improves on the existing ability to predict patient outcome (ie, AJCC stage groupings). Furthermore, external validation needs to be assessed. Our methods of nomogram construction, when applied in other diseases, have been validated,[9,10] despite the fact that these nomograms are also models of data from single institutions. Our nomogram predicts to 9 years, and gastric cancer–specific death beyond that time point is possible, but uncommon.

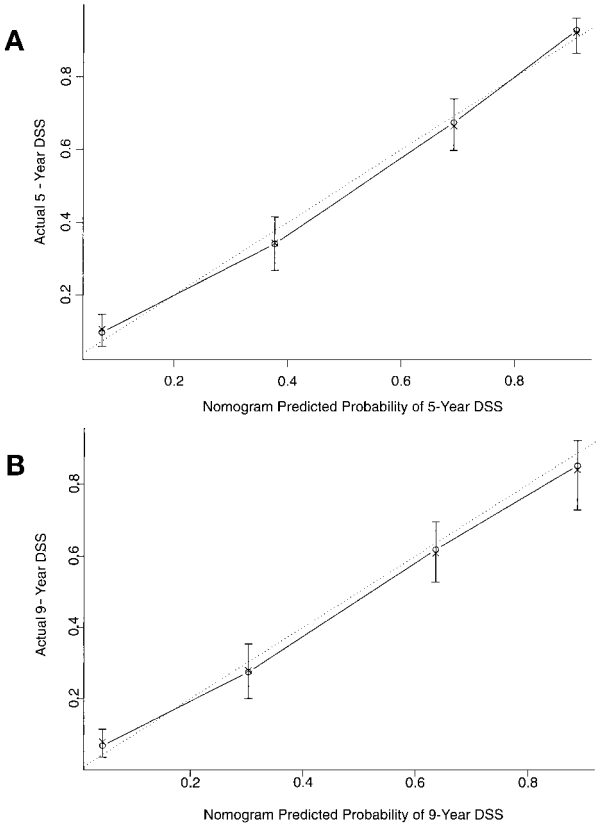


Fig 3. Calibration curves for 5- and 9-year nomogram predictions. DSS, disease-specific survival; (- - - -) an ideal nomogram; (——) the current nomogram; vertical bars, 95% CIs.

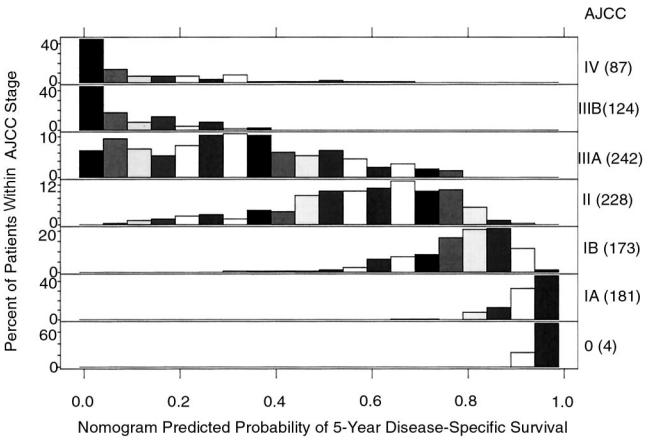


Fig 4. Distribution of nomogram predictions within each American Joint Committee on Cancer (AJCC) stage grouping. Values in parentheses indicate number of patients in each AJCC stage.

Another limitation of the nomogram is that drawing lines and adding points on Figure 2 can be cumbersome. For this reason, we plan to provide free Palm, Pocket PC, and Web software for its implementation, as we currently do for other nomograms (http://www.nomograms.org).

The decision to model gastric cancer–specific survival rather than overall survival is debatable. The addition of death as a result of other causes would warrant the inclusion of several other predictors (eg, comorbidity and socioeconomic status) to avoid the assumption that patients with the same disease-specific covariate values have the same risk of death as a result of other causes. However, our motivation was to provide a tool that predicted the end point potentially modifiable by additional gastric cancer–specific treatment. Another limitation of the nomogram is that it relies on postoperative variables, making it an inadequate preoperative patient-counseling tool, and it is important to emphasize that it is a nomogram for use after R0 resection.

In conclusion, we have developed an internally validated nomogram for predicting gastric cancer–specific survival at 5 and 9 years. The tool is accurate and appears to provide an improvement over existing prognostic methods for this disease. This tool should be useful for patient counseling, clinical trial design, and follow- up scheduling. It is hoped that additional validation will lead to a greater willingness to use risk assessment as the basis for postoperative treatment recommendations.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following authors or their immediate family members have indicated a financial interest as follows. No conflict exists for drugs or devices used in a study if they are not being evaluated as part of the investigation. Served as an officer or member of the board of a company: Murray F. Brennan, Antigenics Inc. Received more than $2,000 a year from a company for either of the past 2 years: Murray F. Brennan, Antigenics Inc.

REFERENCES

1. Karpeh MSJ, Brennan MF: Gastric carcinoma. Ann Surg Oncol 5:650-656, 1998

2. Kattan MW, Zelefsky MJ, Kupelian PA, et al: Pretreatment nomogram for predicting the outcome of three-dimensional conformal radiotherapy in prostate cancer. J Clin Oncol 18:3352-3359, 2000

3. Kattan MW, Leung DHY, Brennan MF: A postoperative nomogram for 12-year sarcoma-specific death. J Clin Oncol 20:791-796, 2002

4. Karpeh MSJ, Kelsen D, Tepper J: Cancer of the stomach, in DeVita VT Jr, Hellman S, Rosenberg SA (eds): Cancer: Principles and Practice of Oncology. Philadelphia, PA, Lippincott Williams & Wilkins, 2001, pp 1092-1125

5. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-387, 1996

6. Harrell FE Jr, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. JAMA 247:2543-2546, 1982

7. Harrell FE Jr: Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. New York, NY, Springer-Verlag, 2001

8. Kim H, Karpeh MSJ, Brennan MF: Standardization of the extent of lymphadenectomy for gastric cancer: Impact on survival. Adv Surg 35:203-223, 2001

9. Graefen M, Karakiewicz P, Cagiannos I, et al: A validation study of the accuracy of a postoperative nomogram for recurrence after radical prostatectomy for localized prostate cancer. J Clin Oncol 20:951-956, 2002

10. Graefen M, Karakiewicz PI, Cagiannos I, et al: International validation of a preoperative nomogram for prostate cancer recurrence following radical prostatectomy. J Clin Oncol 20:3206-3212, 2002