

# Classification and Regression Tree Analysis of 1000 Consecutive Patients with Unknown Primary Carcinoma<sup>1</sup>

Kenneth R. Hess, Marie C. Abbruzzese,  
Renato Lenzi, Martin N. Raber, and  
James L. Abbruzzese<sup>2</sup>

Departments of Biomathematics [K. R. H.], Clinical Investigation [M. C. A., M. N. R.], and Gastrointestinal Medical Oncology and Digestive Diseases [R. L., J. L. A.], The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, Texas 77030.

## ABSTRACT

The clinical features and survival times of patients with unknown primary carcinoma (UPC) are heterogeneous. Therefore, the goals of this study were to apply a novel analytical method to UPC patients to: (a) identify novel prognostic factors; (b) explore the interactions between clinical variables and their impact on survival; and (c) illustrate explicitly how the covariates interact. The 1000 patients analyzed were referred to the University of Texas M. D. Anderson Cancer Center from January 1, 1987 through November 30, 1994. Clinical data from these patients were entered into a computerized database for storage, retrieval, and analysis. Multivariate analyses of survival were performed using recursive partitioning referred to as classification and regression tree (CART) analysis. The median survival for all 1000 consecutive UPC patients was 11 months. CART was performed with an initial split on liver involvement, and 10 terminal subgroups were formed. Median survival of the 10 subgroups ranged from 40 months (95% confidence interval, 22-66 months) for UPC patients with one or two metastatic organ sites, with nonadenocarcinoma histology, and without liver, bone, adrenal, or pleural metastases to 5 months (95% confidence interval, 4-7 months) in UPC patients with liver metastases, tumor histologies other than neuroendocrine carcinoma, age >61.5 years, and a small subgroup of patients with adrenal metastases. Two additional trees were also explored. These analyses demonstrated that important prognostic variables were

consistently applied by the CART program and effectively segregated patients into groups with similar clinical features and survival. CART also identified previously unappreciated patient subsets and is a useful method for dissecting complex clinical situations and identifying homogeneous patient populations for future clinical trials.

## INTRODUCTION

The clinical features and survival times of patients with UPC<sup>3</sup> are heterogeneous. Variables globally affecting the prognosis of UPC patients have been described (1-4), but because of the variation in presenting clinical features, it is difficult to use this information to predict an individual patient's prognosis. Past clinical research efforts have focused on identifying UPC patient subsets that are responsive to therapy. These treatment-responsive patients have been identified based on unique clinical-pathological presentations in small numbers of patients and include women with adenocarcinoma or carcinoma involving axillary lymph nodes (5, 6), selected patients with peritoneal carcinomatosis (7), and patients with squamous cell carcinoma involving cervical lymph nodes (8). Other prognostically favorable subgroups have been defined primarily on the basis of specific pathological characteristics (9). These subsets include the controversial group of patients with poorly differentiated carcinoma or poorly differentiated adenocarcinoma (3, 10) and patients with poorly differentiated neuroendocrine carcinoma (11).

The majority of UPC patients, however, fall outside of these subsets, and their prognoses are much more difficult to predict. Because of the complex presentations of patients with UPC, clinicians often experience difficulty applying standard statistical methods to assess the interactions between clinical variables, determining the cumulative effect of these variables on survival, and translating this information into appropriate management. Therefore, the goals of this study were to apply a novel statistical method to UPC patients to: (a) identify novel prognostic factors; (b) explore the interactions between clinical variables and their impact on survival; and (c) illustrate explicitly how the covariates interact.

To achieve these goals, the technique of CART analysis was explored. This method uses recursive partitioning to assess the effect of specific variables on survival, thereby ultimately generating groups of patients with similar clinical features and survival times. The partitioning of patients into groups with differing survival times using clinical variables generates a tree-structured model that can be analyzed to assess its clinical utility. A default tree generated from the unmanipulated recur-

Received 5/3/99; revised 8/18/99; accepted 8/19/99.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>1</sup> Supported in part by a grant from the University Cancer Foundation, The University of Texas M. D. Anderson Cancer Center. Presented in part at the American Society of Clinical Oncology annual meeting in Philadelphia, PA, May, 1996.

<sup>2</sup> To whom requests for reprints should be addressed, at Department of Gastrointestinal Medical Oncology and Digestive Diseases, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Box 78, Houston, Texas 77030. Phone: (713) 792 2828; Fax: (713) 745-1163; E-mail: jabbruzz@notes.mdacc.tmc.edu.

<sup>3</sup> The abbreviations used are: UPC, unknown primary carcinoma; CART, classification and regression tree; UPT, unknown primary tumor; CI, confidence interval.

Table 1 Twenty-six clinical variables included in CART analysis

Demographic variables
Age
Ethnicity
Sex
Pathologic variables
Histology
Differentiation
Tumor burden and specific metastatic sites
No. of involved organ sites
Any lymph node involvement
Involvement of specific nodal sites
Cervical
Axillary
Supraclavicular
Mediastinal
Thoracic
Retroperitoneal
Abdominal
Inguinal
Bone
Brain
Bone marrow
Adrenal
Liver
Lung
Abdomen
Pelvis
Pleura
Peritoneum
Skin

sive partitioning algorithm and two alternative trees exploring the effects of alternative partitioning schemes were generated and analyzed.

**PATIENTS AND METHODS**

**Patients.** The patient population analyzed in this study was derived from 1609 patients initially referred from community-based physicians to the UPT clinic at The University of Texas M. D. Anderson Cancer Center between January 1, 1987, and November 30, 1994. The medical records of these patients were reviewed for results of radiological studies and pathological diagnosis before referral, and all referred patients were entered into the UPT database at the time of their initial registration after excluding from further analysis those who (a) received an inadequate work-up before referral; (b) were inappropriately referred to the UPT clinic; or (c) had an obvious primary tumor identified at the time of their initial visit.

Of 1609 patients referred with a diagnosis of UPT, 148 were excluded from further analysis based on the criteria outlined above. Thus, from this initial group, 1461 patients were identified with suspected UPTs. From this group of 1461 patients, 81 patients had no evidence of cancer, leaving 1380 patients with suspected UPTs. Of the 1380 patients referred with a suspected diagnosis of UPT, a primary tumor was identified in 380 patients using the diagnostic evaluation previously defined by our group (12), leaving a total of 1000 UPC patients for evaluation.

Twenty-six clinical variables were analyzed within the following general categories: demographic variables, pathologic

Table 2 Demographic characteristics of 1000 consecutive UPC patients

Parameter	No. of patients	%
Sex		
Male	519	52
Female	481	48
Age (median, 59 yr; range, 17–89)		
0–39	97	10
40–49	161	16
50–59	252	25
60–69	318	32
70–79	147	15
80+	25	2
Ethnic origin		
White	866	87
Hispanic	72	7
Black	43	4
Other	19	2

Table 3 Tumor-related characteristics of 1000 consecutive UPC patients

Parameter	No. of patients	%
Histology		
Adenocarcinoma	603	60
Carcinoma	292	29
Squamous carcinoma	62	6
Neuroendocrine carcinoma	43	4
Principal metastatic sites		
Lymph nodes	418	42
Liver	331	33
Bone	289	29
Lung	263	26
Pleura	112	11
Peritoneal	90	9
Brain	64	6
Adrenal	60	6
Skin	38	4
Bone marrow	34	3
No. of involved organ sites		
1	408	41
2	295	29
3	177	18
4	120	12
Therapy <sup>a</sup>		
Surgery	153	15
Radiotherapy	271	27
Chemotherapy	523	52

<sup>a</sup> Some patients had treatment with more than one modality or supportive care only.

ical variables, tumor burden, and involvement of specific metastatic sites (Table 1).

**Diagnostic Evaluation and Treatment.** All patients referred to the UPT clinic underwent a basic clinical evaluation as previously defined (12). Pathological evaluation consisted first of a review of outside pathology slides if available, with attention to H&E stains as well as special stains. When clinically indicated, after consultation between clinician and pathologist, patients underwent a repeat biopsy for immunohistochemistry or electron microscopy.

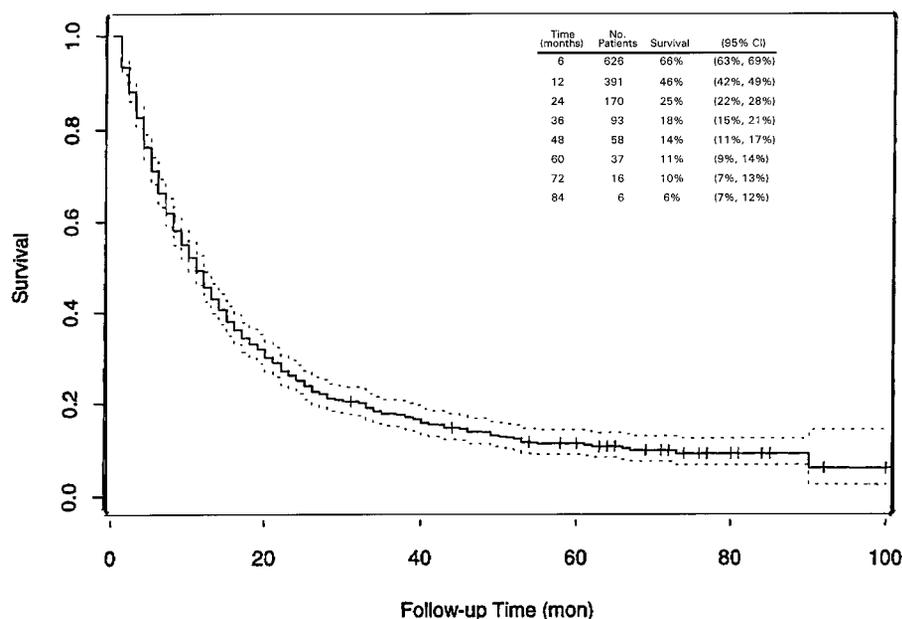


Fig. 1 Kaplan-Meier survival curve of 1000 consecutive patients with UPC. Median survival, 11 months (95% CI, 10–12 months).

The number of involved metastatic organ sites were counted for each patient to provide a crude estimate of tumor burden. In most instances, a positive biopsy was obtained from the most accessible metastatic site, and additional sites of involvement were documented by a physical exam or radiography. For this analysis, a single metastatic organ site was considered to be involved, even if there were multiple individual metastases within that site. Recommendations for therapy were based on the availability of active investigational protocols and the current medical literature (1, 9).

**Statistical Methods.** Patient survival was measured from the time of diagnosis as established by the date of the initial biopsy, and the survival distribution was estimated using the product limit method of Kaplan and Meier (13). Median survival time was computed as the time when the Kaplan-Meier estimate crossed 50%. Confidence limits for the median were computed as the times when the CIs for the Kaplan-Meier estimate crossed 50%. Multivariate analyses of survival were performed using Cox proportional hazards regression analysis (14), and recursive partitioning was referred to as CART. CART analysis was also used to identify optimal cut points in the data and was implemented using a method suggested by Therneau *et al.* (15). In this method, the censored survival data are transformed into a single uncensored data value (the so-called “null martingale residual”), which is used as input into a standard regression tree algorithm (16). This ad hoc method has been shown to perform reasonably well for censored time-to-event data (17). The size of the reported trees was determined based on the results of repeated 10-fold cross-validation (16). In addition to the default tree generated by the CART algorithm, we examined alternative initial splits using systematic inspection (18). Simulations were also computed to assess the frequency of alternative splits (19). A restriction was imposed on the tree construction such that terminal subgroups resulting from any given split must have at least 20 patients. Hazard ratios and corresponding CIs and *Ps*

were computed using the Cox model (14). Analyses were performed using S-PLUS software (Version 3.3, Statistical Sciences, Seattle, WA).

## RESULTS

**General Patient Characteristics.** One thousand four hundred and sixty-one patients were evaluated for suspected UPTs. As defined by our group (13), a primary neoplasm or noncarcinoma cell type (principally lymphoma, sarcoma, or melanoma) was identified in 380 (26%) of the patients initially referred with UPTs. A pathological diagnosis of cancer could not be established in 81 (6%) patients. The demographic and key tumor-related characteristics of the remaining 1000 patients with UPC are outlined in Tables 2 and 3. As compared to our earlier publication on 657 consecutive UPC patients (2), these features have remained consistent as the database has matured. Metastases to lymph nodes were most frequent, followed by liver, lung, or bone metastases. Metastatic involvement of lymph nodes could be further subclassified by anatomical site. Of the 418 patients with nodal metastases, 127 (30%) had mediastinal, 114 (27%) had supraclavicular, 97 (23%) had retroperitoneal, 67 (16%) had cervical, and 63 (15%) had axillary nodal metastases.

Pathological subclassification of the 1000 UPC patients revealed 603 (60%) with adenocarcinoma, 292 (29%) with carcinoma, 62 (6%) with squamous carcinoma, and 43 (4%) with neuroendocrine carcinoma. Histological subclassification based on light microscopy revealed that poorly differentiated tumors were diagnosed in 207 (21%), 146 (15%), 11 (1%), and 2 (<1%) of patients with adenocarcinoma, carcinoma, squamous carcinoma, and neuroendocrine carcinoma, respectively.

**CART Analysis.** The overall survival curve for all 1000 consecutive UPC patients is displayed in Fig. 1. The median survival was 11 months (95% CI, 10–12 months), with only

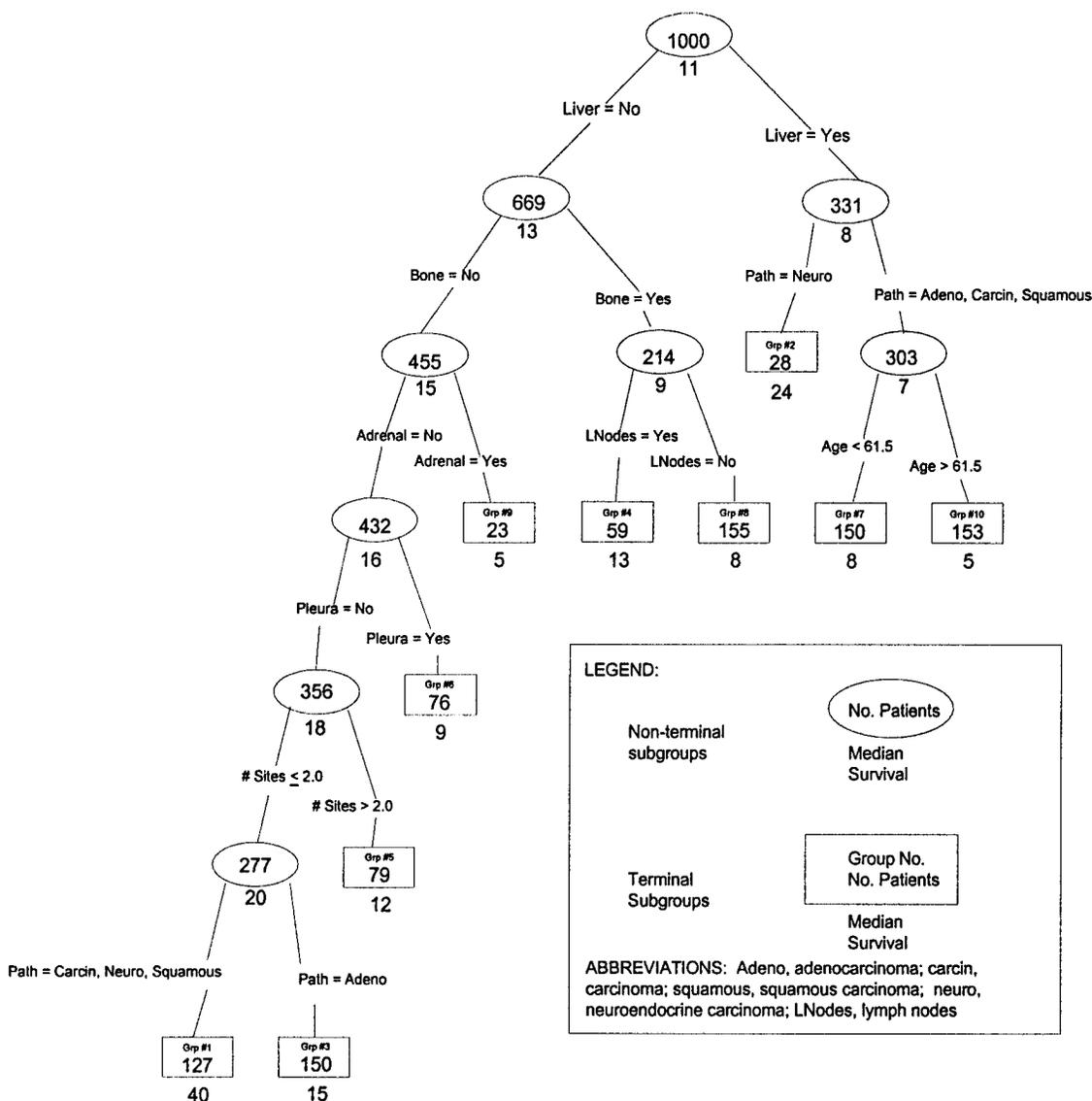


Fig. 2 CART generated with the initial split on the presence or absence of liver metastases (default tree).

11% (95% CI, 9–14%) surviving at 5 years. CART was performed using 26 clinical variables as described in the “Patients and Methods” section. Each tree’s structure depended on the initial split of the patients. A default tree was generated by allowing the CART program to determine the variable with the optimal first split, and two alternative trees were explored through a systematic inspection of alternative splits (burling) and bootstrapping. The results for trees generated on 500 bootstrap samples indicated that liver involvement was chosen as the initial split with a probability of 41%, histology was selected with a 27% probability, and lymph node involvement was selected with a 23% probability. The next highest probability was 3%.

The default tree, therefore, had an initial split on liver involvement, and 10 terminal subgroups were formed. The variables determining the structure of the tree included liver

involvement, bone involvement, adrenal involvement, lymph node involvement, pleural involvement, histology, and number of metastatic sites. The structure of the default tree is presented in Fig. 2, and the corresponding survival curves from the 10 groups generated are presented in Fig. 3. The longest surviving subgroup (group 1) included 127 (12.7%) UPC patients with one or two metastatic organ sites, with nonadenocarcinoma histology, and without liver, bone, adrenal, or pleural metastases. Such patients had a 40-month median survival (95% CI, 22–66 months). A second subgroup (group 2) with a relatively long median survival of 24 months included 28 patients with liver metastases and neuroendocrine carcinoma. One of the shortest surviving subgroups (group 10) included the 153 (15.3%) UPC patients with liver metastases, tumor histologies other than neuroendocrine carcinoma, and age >61.5 years. These patients had a median survival of only 5 months (95% CI, 4–7 months).

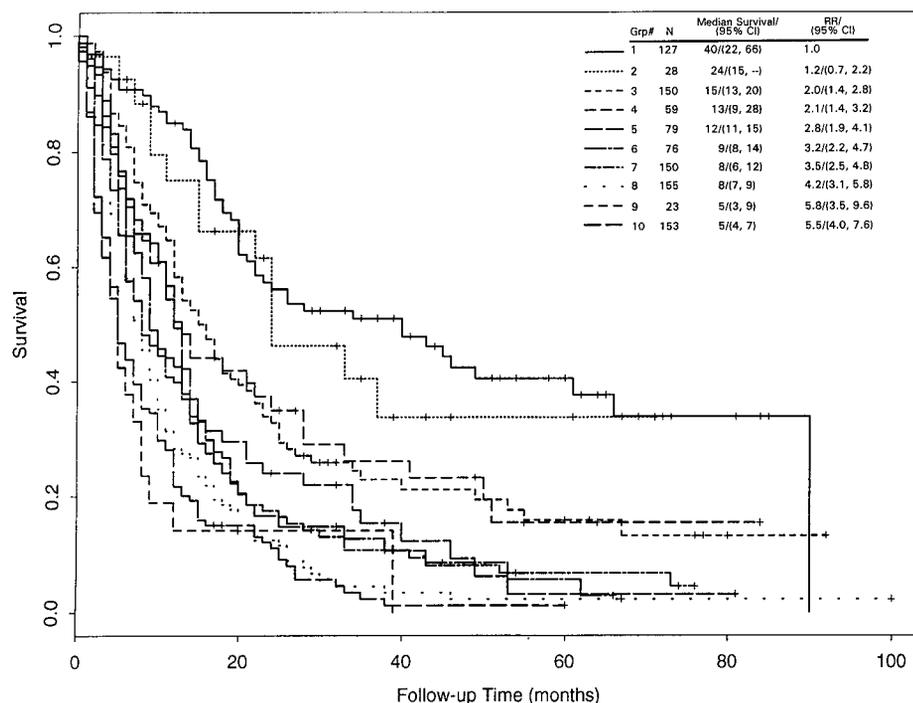


Fig. 3 Kaplan-Meier survival curves of the 10 terminal subgroups generated from the default CART analysis.

A small subgroup of 23 patients (group 9) with adrenal metastases also had a short median survival of 5 months.

To explore other interactions of the clinical variables on the survival of these patients, two alternative trees were created whereby the CART was generated with initial splits on histology and lymph node involvement. As noted previously, these two variables were chosen through a systematic inspection of alternative splits. The prognostic importance of these two variables had also been identified by Cox univariate and multivariate analyses previously conducted on a subgroup of these patients (2). For the first alternative tree, the initial split was made on pathology, but this tree did not generate patient groups that were notably different from the default tree (data not shown).

A second alternative tree was created with the initial split on lymph node involvement (Fig. 4). The structure of this tree was quite distinct from either the default tree or the first alternative tree. For this tree, the best survival (Fig. 5) was in a subgroup of 99 (9.9%) UPC patients with lymph node involvement, one or two total organ sites involved, and nonadenocarcinoma histology (median survival, 45 months; 95% CI not reached). The subgroup with the shortest survival included 117 (11.7%) UPC patients (group 9) with nonneuroendocrine liver metastases but without lymph node involvement (median survival, 5 months; 95% CI, 4–7 months) and 39 patients with adrenal metastases, >2 involved organ-sites, and lymph node metastases (median survival, 5 months; 95% CI, 4–8 months).

The visual inspection of the CARTs' structure suggested that statistically significant interactions between some of the clinical variables and not others were responsible for determining the overall structure of the trees. This was especially apparent at early splits during the recursive partitioning process. For example, using the default CART analysis among the 669 pa-

tients without liver metastases, the hazard ratio for bone metastases versus no bone metastases was 1.7 (95% CI, 1.4–2.1), with  $P < 0.0001$ . But among the 331 patients with liver metastases, the hazard ratio for bone metastases was 1.1 (95% CI, 0.8–1.4), with  $P = 0.73$ . The  $P$  for the statistical interaction between liver and bone was 0.0076. Thus, the effect of bone metastases on survival depends on whether patients have liver metastases. Similarly, we also assessed whether the effect of bone metastases depended on the presence of liver metastases. Among the 711 patients without bone metastases (Table 2), the hazard ratio for liver metastases was 1.9 (95% CI, 1.5–2.2), with  $P < 0.0001$ , whereas among the 289 patients with bone metastases, the hazard ratio for liver metastases was 1.1 (95% CI, 0.9–1.5), with  $P = 0.37$ . The fact that the default CART tree split on bone in the group of patients without liver metastases but not in the group of patients with liver metastases indicated the possibility of a biologically meaningful interaction between liver and bone that was borne out in this study.

Similar interactions were observed in the second alternative tree initially split on lymph node involvement. Among the 582 patients without lymph node involvement, the hazard ratio for metastatic sites >2 was 1.3 (95% CI, 1.0–1.7), with  $P = 0.031$ . Among the 418 patients with involvement of lymph node sites, the hazard ratio for metastatic sites >2 was 2.1 (95% CI, 1.7–2.7), with  $P < 0.0001$ . Thus, the effect of the number of sites was more pronounced in patients with lymph node involvement.

## DISCUSSION

Considerable emphasis has been placed on the identification of UPC subgroups, with favorable natural histories or

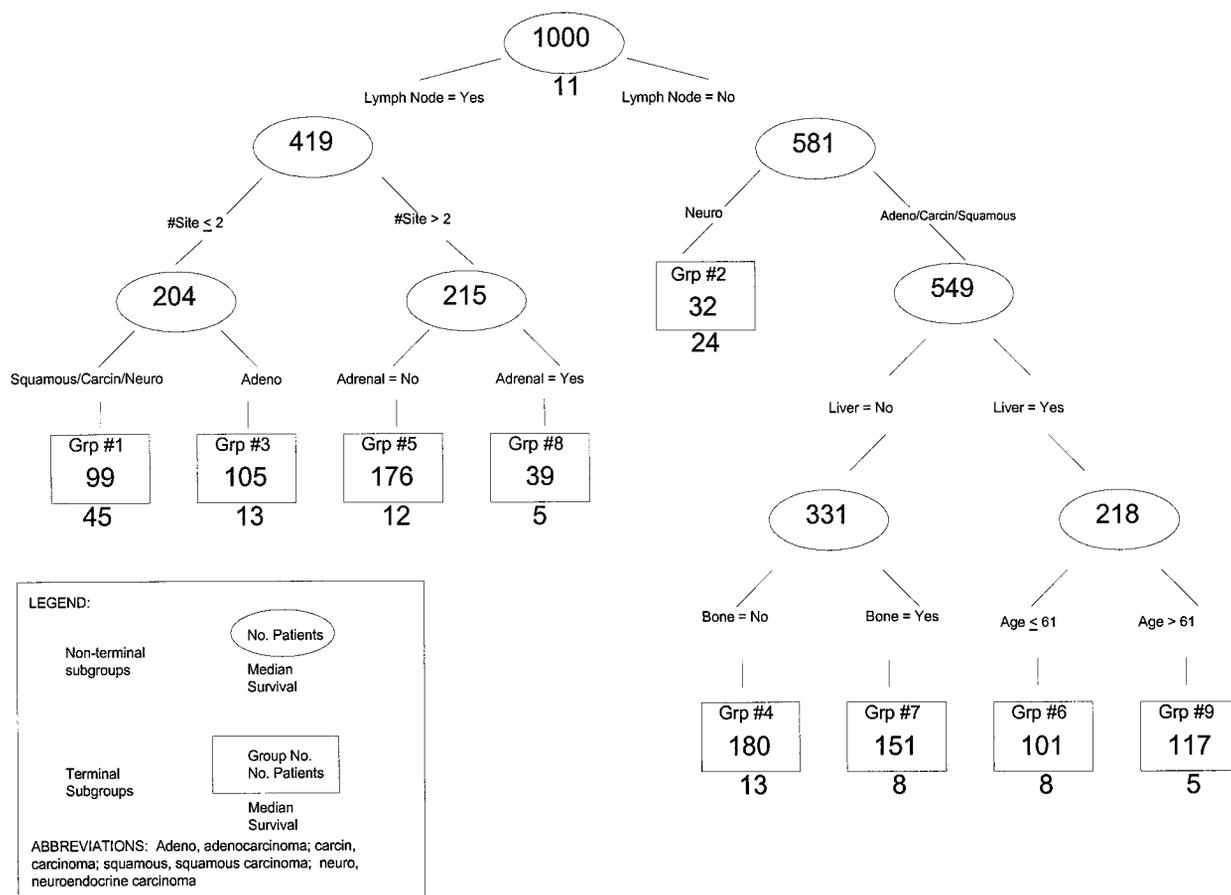


Fig. 4 CART generated with the initial split on the presence or absence of lymph node metastases (second alternative tree).

responsiveness to therapy. Although some progress in identifying such subgroups has been made, most previous analyses have been conducted using small numbers of patients identified retrospectively. In some cases, recent analyses using larger numbers of consecutively evaluated patients have not confirmed the positive impact on survival of historically accepted patient subsets, such as poorly differentiated carcinomas and poorly differentiated adenocarcinomas (3).

The challenges presented by the UPC population highlight two related but distinct goals in prognostic factor studies: (a) to identify the covariate structure (e.g., find independent prognostic factors); and (b) to identify prognostic subgroups. CART marries these objectives nicely by constructing subgroups directly on the covariates. In our previous work identifying prognostic factors that influence UPC survival, we relied on Cox univariate and multivariate analyses (2). Although we clearly identified important prognostic factors, we experienced problems with the bedside utility of this type of data. The principal difficulty was that because UPC patients presented with variable patterns of good and bad prognostic factors, it was difficult to use the Cox-based data to estimate survival for an individual patient. This made it important and challenging to integrate the available prognostic information into patient management.

The analyses conducted in this study demonstrated that the

variables reported to be important using the Cox univariate and multivariate technique were consistently applied by the CART program to segregate patients into groups with similar clinical features and survival. For example, we previously reported that clinical variables, such as hepatic involvement, number of metastatic organ sites, lymph node involvement, and tumor histology were statistically significant independent prognostic factors (2). In each of the three trees generated, these variables were used by the program to generate the best splits of patients into groups with differing survival times. In other instances, clinical variables previously reported through univariate analysis to be statistically correlated with survival (such as bone metastases or age) were similarly used by the CART algorithm to generate groups of patients with differing survival times. The fact that each of these approaches used similar clinical variables to stratify patient survival confirms their clinical importance and supports the validity of the CART analysis.

Interestingly, each CART analysis identified a subset of patients with adrenal metastases that experienced very poor survival. In addition, the default tree (initial split on liver involvement) and alternative tree 1 (initial split on pathology) identified a subset of patients with pleural metastases with a median survival of 9 months. These subsets of UPC patients have not been previously described, suggesting that CART was

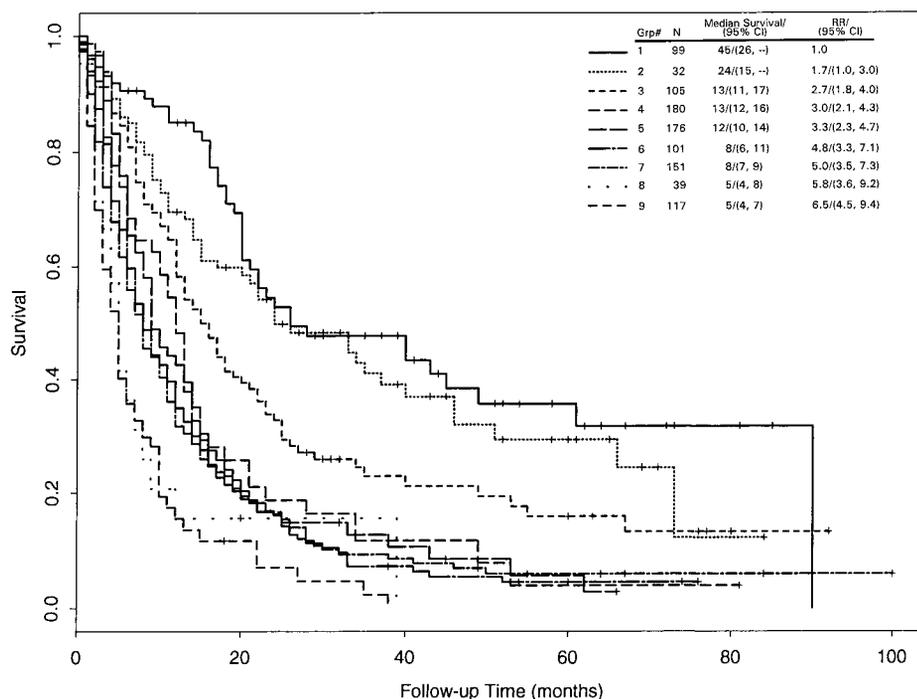


Fig. 5 Kaplan-Meier survival curves of the nine terminal subgroups generated from the second alternative CART analysis.

also able to identify novel patient subsets that may require special treatment strategies.

Although this analysis did not specifically seek to compare CART to other prognostic factor methodologies, an advantage for CART is that it can identify prognostic subgroups that are clinically useful because they are based on simple combinations of clinical characteristics. In contrast to traditional regression methods (e.g., Cox proportional hazards regression), which compute a prognostic index as a weighted average of the patient's characteristics (i.e., an algebraic formula), CART constructs groups based on logical combinations of patient characteristics. Thus, the prognostic subgroups are based directly rather than indirectly on the patient characteristics. Another advantage is the simple, intuitive nature of the CART algorithm (i.e., find the best split by examining all possible splits in all available variables, form subgroups based on this split, repeat in each subgroup). Understanding the essential elements of this process does not require great statistical sophistication, yet the trees often capture much of the relevant covariate structure of the data, including complex interactions and nonlinearities that traditional methods can only handle with much effort. Because it recursively looks for covariate structure within patient subgroups, local covariate effects (i.e., when a covariate has a certain prognostic relationship in one patient subgroup but other relationships in other subgroups) can be easily identified. For example, among the 331 patients with liver metastases (right branch on first split on default tree), a split on pathology was performed with neuroendocrine patients split off from the others. However, for the last split on the left (227 patients without liver, bone, adrenal, or pleural metastases and only 1 or 2 total metastatic sites), pathology was used for a split in which the adenocarcinoma patients were split from the other patients.

Thus, the pathology variable was used in different ways in different parts of the tree.

Two negative aspects of CART deserve mention: (a) because its algorithm performs hundreds of statistical comparisons during the construction of the tree,  $P$  values that may be computed comparing identified subgroups are difficult to interpret (i.e., the overall type I error rate is corrupted by all of the preliminary comparisons). Thus, before accepting this model, validation must be performed on an independent data set. (b) CART may not capture modest, global linear effects because it must approximate the linear effect with a series of splits (i.e., a step function), and quite likely, the individual splits would not be statistically significant.

Finally, CART is a simple method for dissecting complex clinical issues such as those presented by UPC patients. By using CART in our practice with UPC patients, we are able to rapidly develop an estimate of the survival probability of an individual patient based simply on clinical features that are readily apparent on completion of the work-up. Future clinical trials of patients with UPC should prospectively examine the ability of the prognostic information obtained from CART to facilitate precise clinical decision-making. Further, these data can be used to identify relatively homogeneous UPC patient populations with similar survival times for analysis of novel therapeutic interventions. This technique may also be readily applicable to other complex clinical data sets.

## REFERENCES

1. Abbruzzese, J. L., and Raber, M. N. Unknown primary carcinoma. In: M. D. Abeloff, J. O. Armitage, A. S. Lichter, and J. E. Niederhuber (eds.), *Clinical Oncology*, pp. 1833-1845. New York, NY: Churchill Livingstone, 1995.

2. Abbruzzese, J. L., Abbruzzese, M. C., Hess, K. R., Raber, M. N., Lenzi, R., and Frost, P. Unknown primary carcinoma: natural history and prognostic factors in 657 consecutive patients. *J. Clin. Oncol.*, *12*: 1272–1280, 1994.
3. Lenzi, R., Hess, K. R., Abbruzzese, M. C., Raber, M. N., Ordoñez, N., and Abbruzzese, J. L. Poorly differentiated carcinoma and poorly differentiated adenocarcinoma of unknown origin: favorable subsets of patients with unknown primary carcinoma? *J. Clin. Oncol.*, *15*: 2056–2066, 1997.
4. Ayoub, J-P., Hess, K. R., Abbruzzese, M. C., Lenzi, R., Raber, M. N., and Abbruzzese, J. L. Unknown primary tumors metastatic to liver. *J. Clin. Oncol.*, *16*: 2105–2112, 1998.
5. Ellerbroek, N., Holmes, F., Singletary, E., Evans, H., Oswald, M., and McNeese, M. Treatment of patients with isolated axillary nodal metastases from an occult primary carcinoma consistent with breast origin. *Cancer (Phila.)*, *66*: 1461–1467, 1990.
6. Lenzi, R., Kim, E. E., Raber, M. N., and Abbruzzese, J. L. Detection of primary breast cancer presenting as metastatic carcinoma of unknown primary origin by <sup>111</sup>In-pentetreotide scan. *Ann. Oncol.*, *9*: 213–216, 1998.
7. Strnad, C. M., Grosh, W. W., Baxter, J., Burnett, L. S., Jones, H. W., III, Greco, F. A., and Hainsworth, J. D. Peritoneal carcinomatosis of unknown primary site in women: a distinctive subset of adenocarcinoma. *Ann. Intern. Med.*, *111*: 213–217, 1989.
8. Wang, R. C., Goepfert, H., Barber, A. E., and Wolf, P. Unknown primary squamous cell carcinoma metastatic to the neck. *Arch. Otolaryngol. Head Neck Surg.*, *116*: 1388–1393, 1990.
9. Hainsworth, J. D., and Greco, F. A. Treatment of patients with cancer of an unknown primary site. *N. Engl. J. Med.*, *329*: 257–263, 1993.
10. Hainsworth, J. D., Johnson, D. H., and Greco, F. A. Cisplatin-based combination chemotherapy in the treatment of poorly differentiated carcinoma and poorly differentiated adenocarcinoma of unknown primary site: results of a 12-year experience. *J. Clin. Oncol.*, *10*: 912–922, 1992.
11. Hainsworth, J. D., Johnson, D. H., and Greco, F. A. Poorly differentiated neuroendocrine carcinoma of unknown primary site: a newly recognized clinicopathologic entity. *Ann. Intern. Med.*, *109*: 364–371, 1988.
12. Abbruzzese, J. L., Abbruzzese, M. C., Lenzi, R., Hess, K. R., and Raber, M. N. Analysis of a diagnostic strategy for patients with suspected tumors of unknown origin. *J. Clin. Oncol.*, *13*: 2094–2103, 1995.
13. Kaplan, E. L., and Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, *53*: 457–481, 1958.
14. Cox, D. R. Regression models and life tables. *J. R. Stat. Soc.*, *34*: 187–220, 1972.
15. Therneau, T., Grambsch, P., and Fleming, T. Martingale based residuals for survival models. *Biometrika*, *77*: 147–160, 1990.
16. Venables, W. N., and Ripley, B. D. Modern applied statistics with S-Plus. *In: Tree-Based Methods*, pp. 329–347. New York, NY: Springer-Verlag, 1994.
17. LeBlanc, M., Crowley, J. Relative risk trees for censored survival data. *Biometrics*, *48*: 411–425, 1992.
18. Clark, L. A., and Pregibon, D. Tree-based models. *In: J. M. Chambers and T. J. Hastie (eds.), Statistical Models in S-Plus*, pp. 377–499. Pacific Grove, CA: Wadsworth & Brooks/Cole, 1992.
19. Efron, B., and Tibshirani, R. J. An introduction to the bootstrap. New York, NY: Chapman & Hall, 1993.