Prediction of Residual Retroperitoneal Mass Histology After Chemotherapy for Metastatic Nonseminomatous Germ Cell Tumor: Multivariate Analysis of Individual Patient Data From Six Study Groups

By E.W. Steyerberg, H.J. Keizer, S.D. Fosså, D.Th. Sleijfer, G.C. Toner, H. Schraffordt Koops, P.F.A. Mulders, J.E. Messemer, K. Ney, J.P. Donohue, D. Bajorin, G. Stoter, G.J. Bosl, and J.D.F. Habbema

Purpose: To develop a statistical model that predicts the histology (necrosis, mature teratoma, or cancer) after chemotherapy for metastatic nonseminomatous germ cell tumor (NSGCT).

Patients and Methods: An international data set was collected comprising individual patient data from six study groups. Logistic regression analysis was used to estimate the probability of necrosis and the ratio of cancer and mature teratoma.

Results: Of 556 patients, 250 (45%) had necrosis at resection, 236 (42%) had mature teratoma, and 70 (13%) had cancer. Predictors of necrosis were the absence of teratoma elements in the primary tumor, prechemotherapy normal alfa-fetoprotein (AFP), normal human chorionic gonadotropin (HCG), and elevated lactate dehydrogenase (LDH) levels, a small prechemotherapy or postchemotherapy mass, and a large shrinkage of the mass during chemotherapy. Multivariate combination of predictors yielded reliable models (goodness-of-fit tests, $P > .20$), which discriminated necrosis well from other histologies (area under the receiver operating characteristic (ROC) curve, .84), but which discriminated cancer only reasonably from mature teratoma (area, .66). Internal and external validation confirmed these findings.

Conclusion: The validated models estimate with high accuracy the histology at resection, especially necrosis, based on well-known and readily available predictors. The predicted probabilities may help to choose between immediate resection of a residual mass or follow-up, taking into account the expected benefits and risks of resection, feasibility of frequent follow-up, the financial costs, and the patient's individual preferences.

J Clin Oncol 13:1177-1187. © *1995 by American Society of Clinical Oncology.*

SURGICAL RESECTION is a generally accepted treatment for residual retroperitoneal masses after chemotherapy for metastatic testicular nonseminomatous germ cell tumor (NSGCT).[1,2] Resection may reveal necrosis/fibrosis, mature teratoma, or cancer. Because these three histologies are not considered to have a similar necessity of resection, attempts have been made to predict the postchemotherapy histology.[3-5] For patients with a very high likelihood of necrosis, the risk of leaving mature teratoma or cancer unresected might not be balanced by the disadvantages of resection (morbidity, mortality, financial costs).

The decision to perform surgical resection is especially difficult in patients with small residual masses. In some centers, resection is performed if the residual mass exceeds an arbitrarily chosen size (eg, 20 mm[6] or 10 mm[7,8]), with the argument that the probability of cancer or mature teratoma is very low in smaller masses. When no pathologic mass can be detected on the postchemotherapy computed tomography (CT) scan, some surgeons still advocate resection in the following situations: if a teratomatous component is present in the primary tumor[9]; if a prechemotherapy abdominal lymph node metastasis is greater than 30 mm[3]; or if resection is performed as a principle in all patients.[10] The fraction of resected patients varies from approximately 25% (masses > 20 mm[6]) to 86% in accordance with the selection criteria.[10] The extent of surgery is debatable; some surgeons excise only visible abnormal masses,[6] whereas others perform a more extensive retroperitoneal lymph node dissection.[11]

In addition to the size of the residual mass, other characteristics may be considered in the decision to resect a residual mass. Previously recognized predictors of necrosis include the absence of teratoma elements in the primary tumor, prechemotherapy tumor marker levels (alfa-fetoprotein [AFP], human chorionic gonadotropin, [HCG], and lactate dehydrogenase [LDH]), prechemotherapy and postchemotherapy mass size, and shrinkage

Table 1. Characteristics of the Six Participating Study Groups

Study No.	Principal Investigator	References	Study Group	N
1	Toner	3	MSKCC, New York, NY	122
2	Fosså	5,13,14	NRH, Oslo, Norway	127
3	Sleijfer	9,15,20,21	UH, Groningen, the Netherlands	137
4	Mulders	8	UH, Nijmegen, the Netherlands	34
5	Donohue	4	IU, Indianapolis, IN	51
6	Steyerberg	7	UH, Leiden; RCI, Rotterdam; FU, Amsterdam, the Netherlands	85

Abbreviations: MSKCC, Memorial Sloan-Kettering Cancer Center; NRH, Norwegian Radium Hospital; UH, University Hospital; IU, Indiana University; RCI, Rotterdam Cancer Institute; FU, Free University.

during chemotherapy.[3-5,8,9,12-19] The same predictors may help to distinguish cancer from teratoma. In this analysis, our aim is to estimate the probabilities of necrosis, mature teratoma, and cancer in residual retroperitoneal masses based on these well-known and readily available predictors. To obtain a sufficiently large number of patients for statistical analyses, we collected an international data set comprising data from six study groups.

PATIENTS AND METHODS

Inclusion Criteria

The international data set consisted of patients with metastatic nonseminomatous testicular cancer, including patients with pure seminoma and elevated prechemotherapy tumor marker levels, who underwent resection of retroperitoneal residual masses after cisplatin-based induction chemotherapy. Patients with elevated levels of tumor markers AFP or HCG at the time of surgery, with extragonadal tumors, with pure seminoma (normal prechemotherapy AFP and HCG), and those who underwent resection after relapse of tumor after initial chemotherapy were excluded.

Study Descriptions

Individual patient data were retrieved according to a data form that included basic patient identification, year and type of treatment, histology, and information on the following predictors: presence of teratoma elements in the primary tumor, prechemotherapy tumor marker levels (AFP, HCG, LDH), and prechemotherapy and postchemotherapy mass size. Consistency of the data was checked with the participants (agreement with published figures and completion of missing values as far as possible). The studies are numbered 1 to 6 (Table 1) and are briefly described here; details of treatment and resection policy can be found in the original publications. Study no. 1[3] included 122 patients, all of whom fulfilled the inclusion criteria. Study no. 2[5,13,14] included 149 patients, 22 of whom were excluded (11 with extragonadal tumors, 10 with elevated postchemotherapy marker levels (AFP/HCG), and one with both characteristics). Study no. 3[9,15,20,21] included 137 patients. Study no. 4[8] included 49 patients, 15 of whom were excluded (11 with pure seminomas and four with elevated marker levels before surgery). Study no. 5[4] included 80

cases with initial stage B3 disease (palpable prechemotherapy mass > 10 cm). Seventeen patients who underwent salvage chemotherapy for recurrent disease, five patients with pure seminoma, and one patient with an elevated postchemotherapy AFP level were excluded. Of the 51 remaining patients, 50 did not have a prechemotherapy LDH value recorded. Study no. 6[7] included 85 Dutch patients, all of whom fulfilled the inclusion criteria. The European centers participated in consecutive trials of the European Organization for Research and Treatment of Cancer (EORTC) or Medical Research Council (MRC). In total, 556 patients were included in the analysis.

Definitions of Predictors and Histology

The primary tumor histology was defined as teratoma-positive or teratoma-negative,[3,4] according to whether teratomatous elements were present. The AFP, HCG, and LDH tumor marker levels before chemotherapy were classified as elevated or normal by comparison with the normal values of each center. Higher cutoff points and transformations (square root, log) were evaluated for the absolute values of AFP and HCG and for standardized values of LDH (LDH_{st}; LDH value divided by normal value per study).[22] Prechemotherapy and postchemotherapy mass size was measured in transversal direction on CT scan. Shrinkage was calculated as the percentage reduction in size: $100 \cdot (presize - postsize) / presize$. Shrinkage was 100% if the postchemotherapy CT scan did not show any residual mass. Various cutoff values and transformations (square root, log) were assessed.[22]

The histology of the resected material was classified as cancer, mature teratoma, or necrosis according to the worst histologic element. Thus, mature teratoma referred to masses that contained mature teratoma and possibly also necrosis/fibrosis, but no viable cancer cells, and cancer referred to masses that contained viable cancer cells and possibly mature teratoma and/or necrosis.

Missing Values

Of the 556 patients who were included in the analysis, 429 had complete values for all predictors. Missing values were filled in (or imputed) for 115 patients who were missing one value, and the 12 patients with two or more missing values were excluded. Imputation was based on the correlation between each variable with missing values and the other predictors.[23] The correlation was estimated from the 429 complete cases.

The missing values are filled in (or imputed) assuming their absence is random. Regression models for the variables with missing values were estimated from the complete cases. Multiple linear regression models were estimated for the continuous predictors mass size (on the prechemotherapy or postchemotherapy CT scan) and prechemotherapy marker values (AFP, HCG, and LDH_{st}). All continuous predictors had skewed distributions, which became more normally distributed by log-transformation. Independent variables were selected in a stepwise manner, with $P < .05$ for entry of variables and $P < .10$ for removal of variables (SPSS/PC+ software; SPSS Inc, Chicago, IL). A logistic regression model was used to estimate the presence of teratoma elements in the primary tumor. The correlation matrix between the predictors is shown in Table 2. It appears that strong correlations exist between several predictors, and thus imputation based on the values of the other predictors is attractive.

Table 3 shows the independent variables used to impute the missing values as estimated from the complete cases. Missing values were imputed in 115 cases. For the patient with missing primary histology, the predicted probability was .20, and hence the value 0 was imputed

Table 2. Correlations Between Predictors

Correlation	Presize	Postsize	AFP	HCG	LDH$_{st}$	Teratoma
Presize	1.0000					
Postsize	.5679*	1.0000				
AFP	.3100*	.3155*	1.000			
HCG	.2179*	.1997*	.2360*	1.0000		
LDH$_{st}$	.5070*	.2313*	.2047*	.2393*	1.0000	
Teratoma	.1003	.2169*	.0445	.1327†	−.0743	1.0000

NOTE. All continuous predictors are log-transformed.

*Two-tailed significance = .001.

†Two-tailed significance = .01.

(no teratoma). The cases with imputed values were assigned a weight less than the cases with complete values for all predictors (down-weighted). This weight is calculated as $1 - \rho^2_{iy \cdot s}$, where $\rho^2_{iy \cdot s}$ is the partial correlation of the predictor with missing values and y, given the other predictors. The partial correlation was approximated by the ratio of the Wald statistic and −2 times the log likelihood of a base model that contains only the intercept in the logistic regression models predicting necrosis and cancer, respectively.

The results of this analysis were compared with the results obtained when complete cases only were considered.

Statistical Analysis

The histology at resection (necrosis, mature teratoma, or cancer) was predicted using two statistical models. The first model estimated the probability of necrosis by comparing patients with necrosis at resection with patients showing other histologies (teratoma or cancer). The second model aimed to distinguish between cancer and teratoma in the patients who did not have necrosis at resection. This second analysis estimated the ratio of cancer and teratoma, or the relative probability of cancer. The use of these two models supports the clinical notion that the probability of necrosis is of predominant importance for the decision to resect a residual mass and that the ratio of cancer and teratoma is a second consideration.[3]

The probability of necrosis and the relative probability of cancer were related to factors known before resection (predictors). The odds ratio (OR) was used as the measure of association. Relationships between predictors and outcomes were first estimated univariately within each study. If a test for homogeneity indicated no major heterogeneity ($P > .10$) of the relationships, the data were pooled using the Mantel-Haenszel method (EGRET statistical package; Statistics and Epidemiology Research Corporation, Seattle, WA). Pre-

dictors have statistically significant effects ($P < .05$) if the 95% confidence interval (CI) of the OR does not include the value 1. Multivariate logistic regression analysis was used to estimate the probability of necrosis and the relative probability of cancer based on the combination of predictors. Because the aim of this analysis is prediction, all variables contributing information should be included in the models.[24] Therefore, the multivariate analyses included all predictors with P less than .50 in the univariate analysis. Of the three variables related to mass size (prechemotherapy size, postchemotherapy size, and shrinkage), the postchemotherapy size and shrinkage were candidates for use in the multivariate models. Differences between the coefficients of the predictors between the studies were checked by adding interaction terms of the predictors and study number. That the relationships were constant in time was checked by adding interaction terms of the predictors and year of treatment.

Evaluation of Model Performance

Predictive accuracy of the multivariate models can be distinguished in reliability (or calibration) and discrimination. Reliability refers to the amount of agreement between predicted and observed outcomes. For instance, if patients with certain characteristics are predicted to have a 70% chance of necrosis at resection, then 70% of such patients should actually have necrosis at resection. A graphic impression of reliability was obtained by plotting observed frequencies of the outcome (necrosis/cancer) against predicted probabilities. Reliability was tested by the Hosmer-Lemeshow goodness-of-fit test[25] (BMDP module LR; BMDP Statistical Software, Inc, Los Angeles, CA), which evaluates the correspondence between the predicted probabilities of a model and the observed frequencies over groups spanning the entire range of probabilities.

Discrimination was assessed using receiver operating characteristic (ROC) analysis. The ROC curve is a plot of the true-positive rate (sensitivity) versus the false-positive rate ($1 -$ specificity) evaluated at consecutive cutoff points of the predicted probability. The area under the ROC curve forms a suitable number to summarize the discriminative ability of a predictive model.[26,27] For all possible pairs of patients, the area represents the proportion in which the patients with that outcome (necrosis/cancer) had a higher probability than the patients without the outcome. A useless predictive model, such as a coin flip, would yield an area of .5. When the area is 1.0, the model discriminates perfectly. For our prediction problem, a value greater than .6 may be interpreted as reasonable, greater than .7 as satisfactory, and greater than .8 as good with respect to discriminative ability.

Validity of model performance was distinguished internally and

Table 3. Regression Models for Imputation of Missing Values Estimated From the Complete Cases

Dependent	Independent	r^2	N	wNec	wCan
Presize	AFP, LDH$_{st}$, postsize	.50	7	.981	.997
Postsize	AFP, LDH$_{st}$, presize	.37	23	.985	.997
AFP	HCG, presize, postsize	.15	1	.979	—
HCG	AFP, LDH$_{st}$, teratoma	.13	0	.984	—
LDH$_{st}$	HCG, presize, postsize, teratoma	.27	83	.971	.992
Teratoma	HCG, LDH$_{st}$, postsize	—	1	.974	—

Abbreviations: dependent, (log-transformed) dependent variables; independent, independent variables in the regression equations used to impute missing values; r^2, adjusted multiple correlation coefficient indicating the variance explained by the model; N, number of cases where values were imputed; wNec, weight in logistic regression model predicting necrosis; wCan, weight in logistic regression model distinguishing cancer from teratoma. If shrinkage could not be calculated because of missing presize, the weight of presize was used.

externally. Internal validity indicates whether the results of the analysis hold for the data under study. Internal validity was assessed with bootstrapping techniques.[28] Random bootstrap samples were drawn with replacement from the full sample consisting of all patients (200 replications). Models were estimated on these bootstrap samples and evaluated on the full sample. In this way, the discriminative ability of the models in future but similar patients is estimated. Moreover, bootstrap estimates were used to derive the final predictive models by correcting the logistic regression coefficients for overoptimism.[29]

External validity refers to the validity of the results of this analysis when applied to patients in other centers. To assess external validity, each study was left out of the full sample once. The models were fit on the remaining studies. Discriminative power was tested on the study not included in the fitting procedure (test sample).

RESULTS

Table 4 lists the distribution of patient characteristics in each of the six study groups. Overall, half of the pa-

tients had a teratoma-positive primary tumor histology (53%). AFP, HCG, and LDH tumor marker levels were elevated before chemotherapy in approximately two thirds of all patients (66%, 62%, and 69%, respectively). Half of the patients had a prechemotherapy mass size ≤ 50 mm (55%), a postchemotherapy mass ≤ 20 mm (52%), or a shrinkage in transversal direction during chemotherapy ≥ 50% (57%). A minority of the patients was treated before 1981 (11%). The histology at resection was 45% necrosis, 42% mature teratoma, and 13% cancer. The relative probability of cancer was 70 / (70 + 236) or 23% on average. Study no. 5 contained patients with larger masses compared with the other studies, with less shrinkage during chemotherapy, necrosis in 20% only of the patients and cancer in 35% of the patients.

Table 4. Patient Characteristics per Study Group

	Study No. 1 (n = 122)		Study No. 2 (n = 127)		Study No. 3 (n = 137)		Study No. 4 (n = 34)		Study No. 5 (n = 51)		Study No. 6 (n = 85)		Total (N = 556)	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Primary tumor histology														
Teratoma-positive	46	38	76	60	84	61	16	47	26	51	45	53	293/555	53
Prechemotherapy markers														
AFP														
Elevated	72	59	73	58	97	71	19	59	33	72	65	77	359/548	66
Median (ng/mL)	134		44		48		28		—		121		69	
HCG														
Elevated	67	55	71	56	91	66	22	67	34	74	58	68	343/550	62
Median (IU/L)	2		19		5		15		—		42		11	
LDH														
Elevated	93	76	86	68	93	68	17	52	1	—	32	71	322/465	69
Median (U/L)	357		496		285		542		—		203		399	
Prechemotherapy tumor size														
0-20 mm	9	8	21	17	18	13	1	3	—	0	12	13	61/542	11
21-50 mm	61	54	61	48	68	50	13	38	—	0	34	40	237/542	44
51-100 mm	33	29	34	27	40	29	16	47	10	21	33	39	166/542	31
> 100 mm	9	8	11	9	11	8	4	12	37	79	6	7	78/542	14
Postchemotherapy tumor size														
0-10 mm	30	30	54	43	55	40	9	27	—	0	14	17	162/532	31
11-20 mm	19	19	31	24	25	18	8	24	—	0	30	36	113/532	21
21-50 mm	32	32	26	21	41	30	11	32	—	0	23	27	133/532	25
51-100 mm	17	17	12	9	14	10	4	12	8	16	15	18	70/532	13
> 100 mm	1	1	4	3	2	2	2	6	43	84	2	2	54/532	10
Shrinkage														
≥ 70%	30	33	49	39	50	37	12	35	—	0	14	17	155/521	30
50-69%	21	23	35	28	45	33	9	27	—	0	28	33	138/521	27
30-49%	17	19	15	12	8	6	4	12	3	6	15	18	62/521	12
0-29%	13	14	23	18	19	14	5	15	35	74	25	30	120/521	23
< 0% (increase)	11	12	5	4	15	11	4	12	9	19	2	2	46/521	9
Year of treatment														
1975-1980	16	13	7	6	22	16	3	9	11	22	3	4	62/556	11
1981-1985	83	68	67	53	48	35	21	62	34	67	32	38	285/556	51
1986-1993	23	19	53	42	67	49	10	29	6	12	50	59	209/556	38
Histology at resection														
Necrosis	57	47	66	52	61	45	20	59	10	20	36	42	250/556	45
Mature teratoma	48	39	51	40	70	51	11	32	23	45	33	39	236/556	42
Cancer	17	14	10	8	6	4	3	9	18	35	16	19	70/556	13

Table 5. Relations of Predictors With the Histology at Resection

	Necrosis (n = 250; 45%)		Teratoma (n = 236; 42%)		Cancer (n = 70; 13%)		Necrosis v Other			Cancer v Teratoma		
	No.	%	No.	%	No.	%	OR	95% CI	P	OR	95% CI	P
Primary tumor histology												
Teratoma-negative	155	50	78	30	29	11	3.35	2.3-5.0	<.001	—*		—*
Teratoma-positive	94	32	158	54	41	14						
Prechemotherapy markers												
AFP												
Normal	116	61	56	30	17	9	2.74	1.9-4.1	<.001	1.05	.52-2.1	.87
Elevated	130	36	176	49	53	15						
HCG												
Normal	119	58	67	32	21	10	2.17	1.5-3.2	<.001	1.17	.61-2.3	.60
Elevated	128	37	166	48	49	14						
LDH												
Elevated	165	51	120	37	37	12	1.69	1.2-2.7	.011	2.62	1.1-6.4	.020
Normal	56	39	78	55	9	6						
Prechemotherapy mass size									Trend			Trend
0-20 mm	35	57	23	38	3	5	1.0	rc	.008	1.0	rc	.16
21-50 mm	120	51	98	41	19	8	.76	.41-1.4		1.32	.31-6.5	
51-100 mm	66	40	74	45	26	16	.51	.29-1.1		2.21	—†	
> 100 mm	24	31	36	46	18	23	.34	—†		1.87	—†	
Postchemotherapy mass size									Trend			Trend
0-10 mm	117	72	38	24	7	4	1.0	rc	<.001	1.0	rc	.46
11-20 mm	62	55	43	38	8	7	.45	.26-.80		1.00	.25-4.0	
21-50 mm	42	32	69	52	22	17	.17	.09-.29		1.27	.42-4.0	
51-100 mm	11	16	47	67	12	17	.05	.03-.17		0.99	.38-5.1	
> 100 mm	10	19	25	46	19	35	.08	—†		2.96	—†	
Shrinkage									Trend			Trend
≥ 70%	114	74	33	21	8	5	1.0	rc	<.001	1.0	rc	.19
50-69%	72	52	51	37	15	11	.42	.25-.72		1.03	.32-3.4	
30-49%	28	45	26	42	8	13	.24	.14-.59		.69	.15-3.0	
0-29%	26	22	69	58	25	21	.09	—†		.66	—†	
< 0% (increase)	—	0	38	83	8	17	.01	—†		.35	—†	

Abbreviation: rc, reference category.

*The ORs were significantly heterogenous between the 6 studies (P = .016).

†The 95% CI could not be calculated because of empty cells in some studies.

Table 5 lists the results of the univariate analyses. ORs for necrosis were reasonably homogeneous between studies; all tests for homogeneity had P greater than .15. All predictors had significant relationships with the finding of necrosis at resection. Patients with a teratoma-negative primary tumor histology had necrosis at resection in 60% of the cases compared with 32% in patients with teratoma-positive primary tumor histology. The following prechemotherapy tumor marker levels were related to the finding of necrosis at resection: normal AFP, normal HCG, or elevated LDH. Smaller prechemotherapy or postchemotherapy masses contained necrosis more often, as well as masses that were reduced largely in size during chemotherapy (large shrinkage). An increase in tumor size during chemotherapy precluded the finding of necrosis (zero of 46 patients).

Cancer could be distinguished from teratoma by a higher prechemotherapy LDH level (Table 5; P = .020). The OR of the primary tumor histology was significantly heterogenous (P = .016). The OR was greater than one in studies no. 1 to 4 and less than one in studies no. 5 and 6. Therefore, the primary tumor histology cannot be used to distinguish between cancer and teratoma. Prechemotherapy AFP and HCG levels were excluded as predictors in the multivariate models (P > .50). Cancer was found in only 4% (seven of 162) of the patients with residual masses ≤ 10 mm and in 5% (eight of 155) of the patients with a shrinkage ≥ 70%.

The multivariate model for necrosis included 544 patients, and 115 of 3,264 (3.5%) of the values were imputed (Table 6). All predictors for necrosis were significant (P < .003), as well as the multivariate model as a whole (P < .0001). Dichotomous characteristics (present/not present) that predict necrosis included a teratoma-negative

Table 6. Results of the Logistic Regression Analysis

	Necrosis v Other (n = 544)		Cancer v Teratoma (n = 299)	
	OR	95% CI	OR	95% CI
Primary tumor histology				
Teratoma-negative v positive	2.46	1.6-3.7	—	
Prechemotherapy marker levels				
AFP normal v elevated	2.49	1.6-3.9	—	
HCG normal v elevated	2.22	1.4-3.5	—	
LDH ln(LDH_st)*	2.76	1.8-4.2	1.58	.93-2.7
Postchemotherapy mass size				
Sqrt(transversal diameter)*	.744	.63-.87	1.17	.99-1.4
Shrinkage				
Per 10% decrease*	1.17	1.1-1.3	1.06	.95-1.2

Abbreviation: ln(LDH_st), natural logarithm of LHD divided by normal value; Sqrt, square root.
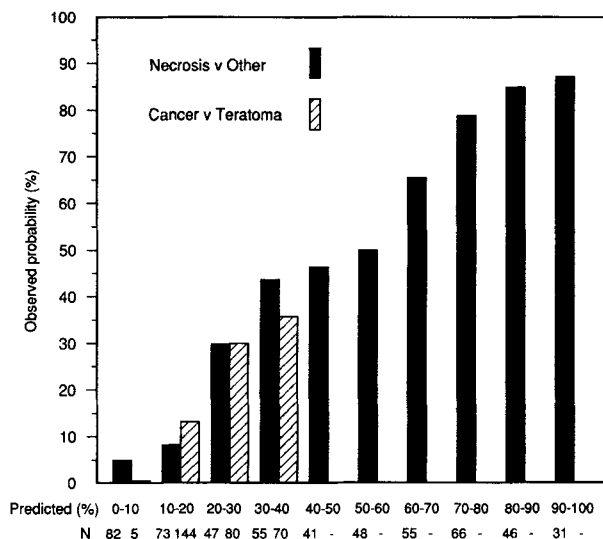*Continuous predictors.

Fig 1. Reliability of the models predicting necrosis and distinguishing cancer from teratoma. Number of patients in each group of the predicted probability is listed after N.

primary tumor histology and normal prechemotherapy AFP and HCG levels. Three other predictors were used as continuous variables. The natural logarithm (ln) appeared to be the optimal transformation of LDH_{st}. The square root of the residual mass size was taken. Shrinkage of the mass during chemotherapy was used as a continuous, untransformed variable, and the OR was calculated per 10% decrease. Thus, a decrease of 10% indicated an approximately 1.2-times-higher probability of necrosis compared with no decrease, simultaneously adjusting for the other predictors.

The multivariate model for the relative probability of cancer included 299 patients (77 of 897, or 8.6% of values imputed). Prechemotherapy LDH level, postchemotherapy mass size, and shrinkage were used as predictors, resulting in a significant model ($P = .003$). The relationship of the predictors did not vary significantly with study or year of treatment in both multivariate models ($P > .10$).

Reliability of the multivariate models is shown in Fig 1. Overall, the correspondence between observed and expected probabilities is good. In patients with a predicted probability of necrosis greater than 80%, the observed probability was 66 of 77 (86%). The goodness-of-fit tests indicated no lack of fit of the models (necrosis, $P = .59$; cancer, $P = .34$). Discriminative ability of the multivariate models is shown in Fig 2. Clearly, discrimination of necrosis from another histology is much more feasible than discrimination of cancer from teratoma.

Validity of the models is listed in Table 7. Assessment of internal validity indicates that overall discriminative ability of the model for necrosis is expected to be good in similar patients (area, .83). The models for cancer will have less discriminative ability (area, .65).

The external validation procedure mimics the situation that one of the studies would not have been included in the analysis. Table 7 shows that the necrosis model discriminates well in all studies, with some higher discrimination in study no. 4 and some lower discrimination in study no. 6. The second model discriminates reasonably in studies no. 3, 5, and 6.

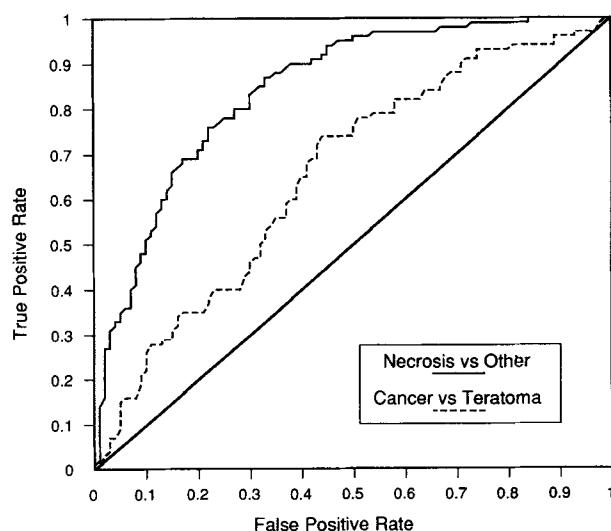When the multivariate analysis is restricted to the pa-

Fig 2. ROC curves of the models predicting necrosis and distinguishing cancer from teratoma, indicating discriminative ability. The areas under the curves are .839 and .661.

tients with complete values only, the necrosis model remains very similar with respect to both the ORs and model performance. However, the model distinguishing between cancer and teratoma has a smaller OR for LDH in the complete case analysis (OR = 1.50 v OR = 1.70; Table 7), and its discriminative ability is inferior to the model with imputation of missing values (area ≈ .60 v area ≈ .65).

Finally, the two multivariate models are presented in a prognostic score chart (Table 8). This score chart is intended to facilitate the estimation of the probabilities of necrosis, mature teratoma, and cancer at resection in clinical practice using the final models. Scores for each predictor were derived from the logistic regression coefficients, which were reduced by a correction for overoptimism (multiplied by .955 in the necrosis model and by .870 in the cancer model), and subsequently multiplied by 10 and rounded to whole numbers. Ten points on the score chart correspond to an OR of e^1 = 2.72; an OR of two (doubling of the odds) is obtained by a score of seven points. Values for continuous predictors are given with such intervals that the scores show small steps, but scores for intermediate values may be estimated by linear interpolation.

For individual patients, the scores corresponding to the values of the predictors can be filled in on the score chart. An individual sumscore consists of the sum of all scores and a constant that represents the score if all values were zero. Figure 3 shows the probabilities corresponding to this sumscore.

The use of the score chart is illustrated with two patients (Table 9). Patient no. 1 had a teratoma-negative primary tumor, normal AFP and HCG levels before chemotherapy, LDH level three times the normal level, and a residual mass of 10 mm, which measured 50 mm before chemotherapy (shrinkage, [50 − 10] / 50; 80%). Patient

Table 7. Internal and External Validation of the Models Predicting Necrosis and Distinguishing Cancer From Teratoma

	Necrosis v Other		Cancer v Teratoma	
	Area Under the ROC Curve	N	Area Under the ROC Curve	N
Internal validation				
Bootstrapping	.83	544	.65	299
External validation				
Study no. 1	.83	121	.58	65
Study no. 2	.85	127	.55	61
Study no. 3	.84	137	.60	76
Study no. 4	.91	33	.53	14
Study no. 5	.85	42	.67	34
Study no. 6	.75	84	.64	49

no. 2 had a teratoma-positive primary tumor, an elevated AFP level, and other characteristics the same as patient no. 1. The sumscores for patient no. 1 are +30 and −12, and for patient no. 2 +12 and −12 (necrosis and cancer, respectively). The corresponding probabilities can be read from Fig 3. Patient no. 1 has a 95% probability of necrosis, leaving approximately 5% for the sum of the probabilities of teratoma and cancer. His relative probability of cancer is approximately 25% (Fig 3). This means that the absolute probability of cancer is approximately 25% · 5% (1.25%), and complementarily, the probability of teratoma is approximately 3.75%. Alternatively, the exact probabilities can be calculated using the formulas under Table 8: necrosis, 94.6%; teratoma, 4.1%; and cancer, 1.3%. The probabilities for patient no. 2 are 75.5%, 18.5%, and 6.0%, respectively.

DISCUSSION

We developed two models to predict the histology of residual retroperitoneal masses in patients who were treated with cisplatin-based chemotherapy for metastatic NSGCT and who obtained normal values of AFP and HCG levels before resection. The first model aimed to predict the finding of necrosis only, whereas the second model was developed to separate cancer from mature teratoma in patients without necrosis. Individual patient data from six study groups were available, providing the largest data set of this type of patients thus far.[3-5,7-9,13-15,20,21]

The model for necrosis consisted of six predictors, that were highly significant when analyzed alone (univariately) or combined (multivariately). Predictors for necrosis were a teratoma-negative primary tumor, normal prechemotherapy AFP or HCG level, elevated prechemotherapy LDH level, a relatively small residual mass, and a large shrinkage of the mass during chemotherapy. The model distinguishing cancer from teratoma consisted of three predictors that together constituted a statistically significant multivariate model. Cancer was found relatively more often if prechemotherapy LDH level was elevated, the residual mass was larger, and if a large shrinkage of the mass occurred during chemotherapy.

Higher values of prechemotherapy LDH are thus found related with a higher probability of necrosis but also with a higher relative probability of cancer. This implies that the probability of mature teratoma decreases with higher prechemotherapy LDH values. The absolute probability of cancer is lowered by higher LDH values in most patients because the multivariate OR of LDH for cancer is smaller than for necrosis (Table 6). In agreement with previous statements,[3] a higher prechemotherapy LDH level is therefore a fortuitous prognostic sign in patients

with normal levels of tumor markers AFP and HCG after chemotherapy. On the other hand, a higher LDH level has been found to indicate a lower probability of complete response and a worse survival in patients with metastatic disease, analyzing from the start of primary chemotherapy.[10,30,32] It may therefore be postulated that patients with

Table 8. Prognostic Score Chart for the Probability of Necrosis and the Relative Probability of Cancer at Resection of Residual Masses in NSGCT Patients With Normal Levels of Tumor Markers AFP and HCG Before Resection

		Necrosis	Cancer
Primary tumor histology			
Presence of teratoma elements	If negative	+9	—
Prechemotherapy markers			
AFP	If normal	+9	—
HCG	If normal	+8	—
LDH$_{st}$ (LDH/normal value)	0.6	−5	−2
	0.8	−2	−1
	1.0	0	0
	1.2	+2	+1
	1.5	+4	+2
	2.0	+7	+3
	3.0	+11	+4
	4.5	+15	+6
	7.0	+19	+8
Postchemotherapy mass size			
Transversal diameter (mm)	2*	−4	+2
	5	−6	+3
	10	−9	+4
	15	−11	+5
	20	−13	+6
	30	−16	+7
	50	−20	+10
	70	−24	+11
	100	−28	+14
Shrinkage (%)			
100 · presize − postsize/presize	−50	−7	−3
	0	0	0
	50	+7	+3
	75	+11	+4
	100	+15	+5
Constant		−10	−24
Sumscore (add relevant scores)†	—	—	

*If no mass is detectable on the postchemotherapy CT scan, a size of 2 mm is assumed.

†The exact formulas to calculate the sumscores are as follows: sumscore (necrosis), −9.78 + 8.58 · teratoma-negative + 8.70 · AFPnormal + 7.61 · HCGnormal + 9.69 · ln(LDH$_{st}$) − 2.83 · Sqrt(postsize) + .147 · shrinkage; sumscore (cancer), −24.18 + 3.95 · ln (LDH$_{st}$) + 1.36 · Sqrt(postsize) + .053 · shrinkage, where the variables teratoma-negative, AFPnormal, and HCGnormal are 1 if true, 0 if false, ln (LDH$_{st}$) is the natural logarithm of LDH/normal value, postsize is expressed in millimeters, and shrinkage is expressed as percent. The corresponding probabilities are calculated with the following formulas: probability (necrosis), 1/[1 + e$^{-(sumscore(necrosis)/10)}$]; probability (cancer), [1 − probability(necrosis)] · [1/ (1 + e$^{-(sumscore(cancer)/10)}$]; probability (teratoma), 1 − [probability(necrosis) + probability(cancer)].

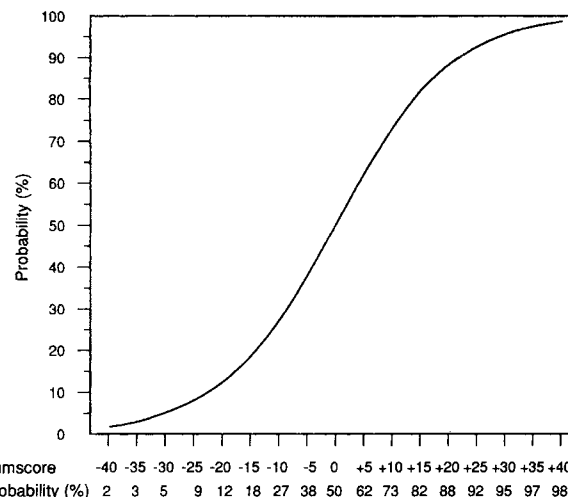| Sumscore	−40	−35	−30	−25	−20	−15	−10	−5	0	+5	+10	+15	+20	+25	+30	+35	+40
Probability (%)	2	3	5	9	12	18	27	38	50	62	73	82	88	92	95	97	98

Fig 3. Predicted probabilities corresponding to the sumscores as calculated with the prognostic score chart (Table 8; see text). For example, a sumscore of +15 corresponds to a probability of 82%.

residual masses who respond to chemotherapy (as indicated by normalization of AFP/HCG levels) form a favorable subgroup of patients with a prechemotherapy high LDH level.

The two models were reliable in that the observed probabilities agreed with predicted probabilities. Internal and external validation procedures showed that necrosis could be discriminated well from other histologies, but that cancer could only reasonably be distinguished from teratoma. It should be realized that the probability of cancer strongly depends on the probability of necrosis, as estimated with the first model. For example, if the estimated probability of necrosis is very high, eg, 95%, the probability of cancer can never be higher than 5%. If we use the average relative probability of cancer (23%) instead of the second model, this results in a risk of cancer of only 23% · 5% (1.2%). Thus, even when no predictors for the ratio of cancer and mature teratoma are used, the risk of cancer will be very low once the probability of necrosis is high.

The observation that mature teratoma cannot be distinguished easily from cancer may be explained partly by the fact that the histologic distinction between cancer and mature teratoma is made less clearly than the distinction between purely benign tissue (necrosis/fibrosis) and other tissue. For example, one study[19] described that the histologic classification changed in a substantial proportion of the patients when reviewed with the Indiana criteria[32] for the diagnosis of cancer. In five patients (12%), the diagnosis changed from cancer to atypia, and in three patients (7%), from other diagnoses to cancer.[19]

Table 9. Application of the Prognostic Score Chart in Two Hypothetical Patients

	Patient No. 1	Patient No. 2	Necrosis$_1$	Necrosis$_2$	Cancer$_1$	Cancer$_2$
Primary tumor histology						
Presence of teratoma elements	Negative	Positive	+9	0	—	—
Prechemotherapy markers						
AFP	Normal	Elevated	+9	0	—	—
HCG	Normal	Normal	+8	+8	—	—
LDH$_u$ (LDH/normal value)	3.0	3.0	+11	+11	+4	+4
Postchemotherapy mass size						
Transversal diameter (mm)	10	10	−9	−9	+4	+4
Shrinkage (%)						
100 · (presize − postsize)/presize	80	80	+12	+12	+4	+4
Constant			−10	−10	−24	−24
Sumscore (add relevant scores)			+30	+12	−12	−12
Probabilities (%)						
Using Fig 3			95	≈75	≈25	≈25
Using formulas						
Necrosis	94.6	75.5				
Teratoma	4.1	18.5				
Cancer	1.3	6.0				

The final models were presented in a prognostic score chart, the use of which was illustrated in two patients, both with residual masses of 10 mm. Taking into account all prognostic factors, the probability of necrosis was very high (95%) in the first patient, and somewhat lower (76%) in the second patient. The probabilities for cancer and teratoma were 1.3% and 4.1% for the first patient, and 6% and 18% for the second patient, respectively. Should both patients undergo laparotomy? Or more generally, can we define thresholds for the decision to resect a residual mass? Should resection not be performed if the probability of necrosis is greater than 80%, or is the threshold as high as 90%? And what should we accept as the risk of cancer?

These thresholds need to be determined by the balance between the expected benefits and risks of resection. The risks of resection include long-term morbidity caused by surgery, especially ejaculation problems, which depend on the size and location of the residual mass and the extent of surgery.[6,20,33-35] Short-term morbidity consists of the hospital stay and complications such as hemorrhage, renal failure, and lymphocele.[34] The mortality of laparotomy is low, presumably less than 1% in most experienced centers, but risk estimates for individual patients may be higher.

If mature teratoma or cancer is present in the residual mass, the patient is expected to benefit from resection. The prognosis then is generally favorable, with 5-year relapse-free survival greater than 85% after resection of mature teratoma, and 50% to 80% after resection of cancer.[6-8,16,19,36-38] Resection of viable cancer cells is usually

followed by two additional cycles of chemotherapy. It has been suggested that this postresection chemotherapy preferably should be a different regimen than that given before resection.[3,7] Yet an expectant policy after complete resection of cancer has also been followed (Horwich A, personal communication, September 1992).

If resection is not performed, masses that contain mature teratoma may start to grow during a follow-up of months or even years (growing teratoma syndrome).[39] Resection may then be more complicated than it would have been shortly after the end of chemotherapy. Although residual mature teratomas have a less abnormal karyotype than the primary tumor, a risk of malignant transformation has been reported.[40-44] Leaving masses with residual cancer unresected involves a serious risk. However, it is uncertain how many of the patients with histologically viable cancer cells will eventually relapse. It may be hypothesized that the risk of relapse also depends on the extensiveness of cancer cells in the residual mass, with a low probability of relapse if only small foci of malignancy remain. No data are currently available regarding this relationship. When relapse of malignancy occurs after leaving mature teratoma or cancer unresected, salvage chemotherapy is administered. However, these regimens have rather limited efficacy (approximately 25%[1,2,45,46]), and late relapses frequently have a high degree of chemotherapy resistance.

Given these uncertainties with the benefits of resection, it is difficult to indicate thresholds for the probability of necrosis and cancer. These thresholds also may depend on country- or center-specific circumstances such as the

feasibility of frequent follow-up visits with high-quality CT scanning of the abdomen.[5] If frequent follow-up is impossible, any residual mass should be resected. Finally, in a health-care environment with limited resources, financial costs of surgery and subsequent hospital stay may provide argument against resection in patients with a high probability of necrosis or a low probability of cancer.

In conclusion, the histology at resection, especially that of necrosis, can be predicted with high accuracy; the models predict reliably and discriminate rather well. The models cannot exclude the presence of cancer, but the probability is very low in some patients. The predicted probabilities are easily calculated with the prognostic score chart and may help to choose the optimal treatment, taking into account the potential benefits, morbidity, and mortality of resection, feasibility of frequent follow-up, the financial costs, and the patient's individual preferences.

ACKNOWLEDGMENT

REFERENCES

1. Einhorn LH: Treatment of testicular cancer: A new and improved model. J Clin Oncol 8:1777-1781, 1990

2. Peckham M: Testicular cancer. Rev Oncol 1:439-453, 1988

3. Toner GC, Panicek DM, Heelan RT, et al: Adjunctive surgery after chemotherapy for nonseminomatous germ cell tumors: recommendations for patient selection. J Clin Oncol 8:1683-1694, 1990

4. Donohue JP, Rowland RG, Kopecky K, et al: Correlation of computerized tomographic changes and histological findings in 80 patients having radical retroperitoneal lymph node dissection after chemotherapy for testis cancer. J Urol 137:1176-1179, 1987

5. Fosså SD, Qvist H, Stenwig AE, et al: Is postchemotherapy retroperitoneal surgery necessary in patients with nonseminomatous testicular cancer and minimal residual tumor masses? J Clin Oncol 10:569-573, 1992

6. Hendry WF, A'Hern RP, Hetherington JW, et al: Para-aortic lymphadenectomy after chemotherapy for metastatic non-seminomatous germ cell tumours: Prognostic value and therapeutic benefit. Br J Urol 71:208-213, 1993

7. Steyerberg EW, Keizer HJ, Zwartendijk J, et al: Prognosis after resection of residual masses following chemotherapy for metastatic nonseminomatous testicular cancer: A multivariate analysis. Br J Cancer 68:195-200, 1993

8. Mulders PFA, Oosterhof GON, Boetes C, et al: The importance of prognostic factors in the individual treatment of patients with disseminated germ cell tumours. Br J Urol 66:425-429, 1990

9. Gelderman WAH, Schraffordt Koops H, Sleijfer DTh, et al: Results of adjuvant surgery in patients with stage III and IV nonseminomatous testicular tumors after cisplatin-vinblastine-bleomycin chemotherapy. J Surg Oncol 38:227-232, 1988

10. Aass N, Klepp O, Cavillin-Ståhl E, et al: Prognostic factors in unselected patients with nonseminomatous metastatic testicular cancer: A multicenter experience. J Clin Oncol 9:818-826, 1991

11. Donohue JP, Schraffordt Koops H, Hendry WF, et al: Surgery in advanced disease for testicular cancer, in Newling DWW, Jones WG (eds): EORTC Genitourinary Group Monograph 7: Prostate Cancer and Testicular Cancer. New York, NY, Wiley-Liss, 1990, pp 305-312

12. Steyerberg EW, Keizer HJ, Stoter G, et al: Predictors of residual mass histology following chemotherapy for metastatic nonseminomatous testicular cancer: A quantitative overview of 996 resections. Eur J Cancer 30A:1231-1239, 1994

13. Fosså SD, Aass N, Ous S, et al: Histology of tumor residuals following chemotherapy in patients with advanced nonseminomatous testicular cancer. J Urol 142:1239-1242, 1989

14. Fosså SD, Ous S, Lien HH, et al: Post-chemotherapy lymph node histology in radiologically normal patients with metastatic nonseminomatous testicular cancer. J Urol 141:557-559, 1989

15. Gelderman WAH, Schraffordt Koops H, Sleijfer DTh, et al: Treatment of retroperitoneal residual tumor after PVB chemotherapy of nonseminomatous testicular tumors. Cancer 58:1418-1421, 1986

16. Tait D, Peckham MJ, Hendry WF, et al: Post-chemotherapy surgery in advanced non-seminomatous germ-cell tumours: The significance of histology with particular reference to differentiated (mature) teratoma. Br J Cancer 50:601-609, 1984

17. Pizzocaro G, Salvioni R, Pasi M, et al: Early resection of residual tumor during cisplatin, vinblastine, bleomycin combination chemotherapy in stage III and bulky stage II nonseminomatous testicular cancer. Cancer 56:249-255, 1985

18. Bracken RB, Johnson DE, Frazier OH, et al: The role of surgery following chemotherapy in stage III germ cell neoplasms. J Urol 129:39-43, 1983

19. Harding MJ, Brown IL, Macpherson SG, et al: Excision of residual masses after platinum based chemotherapy for non-seminomatous germ cell tumours. Eur J Cancer Clin Oncol 25:1689-1694, 1989

20. Nijman JM, Schraffordt Koops H, Kremer J, et al: Gonadal function after surgery and chemotherapy in men with stage II and III nonseminomatous testicular tumors. J Clin Oncol 5:651-656, 1987

21. De Graaf WE, Oosterhuis JW, Van der Linden S, et al: Residual mature teratoma after chemotherapy for nonseminomatous germ cell tumors of the testis occurs significantly less often in lung than in retroperitoneal lymph node metastases. J Urogen Pathol 1:75-81, 1991

22. Harrell FE, Lee KL, Pollock BG: Regression models in clinical studies: Determining relationships between predictors and response. J Natl Cancer Inst 80:1198-1202, 1988

23. Little RJA: Regression with missing X's: A review. J Am Stat Assoc 87:1227-1237, 1992

24. Spiegelhalter DJ: Probabilistic prediction in patient management and clinical trials. Stat Med 5:421-433, 1986

25. Hosmer DW, Lemeshow S: Applied Logistic Regression. New York, NY, Wiley, 1989

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

27. SAS Institute Inc: SAS Technical Report P-200, SAS/STAT Software: CALIS and LOGISTIC Procedures, Release 6.04. Cary, NC, SAS Institute, 1990, pp 194-195

28. Efron B: Estimating the error rate of a prediction rule: Improvement on cross-validation. J Am Stat Assoc 78:316-331, 1983

29. Van Houwelingen JC, Le Cessie S: Predictive value of statistical models. Stat Med 9:1303-1325, 1990

30. Bosl GJ, Geller NL, Cirrincione C, et al: Multivariate analysis of prognostic variables in patients with metastatic testicular cancer. Cancer Res 43:3403-3407, 1983

31. Bajorin D, Katz A, Chan E, et al: Comparison of criteria for assigning germ cell tumor patients to "good risk" and "poor risk" studies. J Clin Oncol 6:786-792, 1988

32. Davey DD, Ulbright TM, Loehrer PJ, et al: The significance of atypia within teratomatous metastases after chemotherapy for malignant germ cell tumours. Cancer 59:533-539, 1987

33. Fosså SD, Kreuser ED, Roth GJ, et al: Long-term side effects after treatment of testicular cancer, in Newling DWW, Jones WG (eds): EORTC Genitourinary Group Monograph 7: Prostate Cancer and Testicular Cancer. New York, NY, Wiley-Liss, 1990, pp 321-330

34. Horwich A: Abdominal surgery postchemotherapy, in Horwich A (ed): Testicular Cancer: Investigation and Management. London, United Kingdom, Chapman & Hall, 1991, pp 278-280

35. Donohue JP, Foster RS, Rowland RG, et al: Nerve-sparing retroperitoneal lymphadenectomy with preservation of ejaculation. J Urol 144:287-292, 1990

36. Mead GM, Stenning SP, Parkinson MC, et al: The second medical research council study of prognostic factors in nonseminomatous germ cell tumors. J Clin Oncol 10:85-94, 1992

37. Geller NL, Bosl GJ, Chan EYW: Prognostic factors for relapse after complete response in patients with metastatic germ cell tumors. Cancer 63:440-445, 1989

38. Fox EP, Weathers TD, Williams SD, et al: Outcome analysis for patients with persistent nonteratomous germ cell tumor in postchemotherapy retroperitoneal lymph node dissections. J Clin Oncol 11:1294-1299, 1993

39. Logothetis CJ, Samuels ML, Trindade A, et al: The growing teratoma syndrome. Cancer 50:1629-1635, 1982

40. Castedo SMMJ, de Jong B, Oosterhuis JW, et al: Chromosomal changes in mature residual teratomas following polychemotherapy. Cancer Res 49:672-676, 1989

41. Ulbright TM, Loehrer PJ, Roth LM, et al: The development of non-germ malignancies within germ cell tumors: A clinicopathologic study of 11 cases. Cancer 54:1824-1833, 1984

42. Ahlgren AD, Simrell CR, Triche TJ, et al: Sarcoma arising in a residual testicular teratoma after cytoreductive chemotherapy. Cancer 54:2015-2018, 1984

43. Ahmed T, Bosl GJ, Hajdu SI: Teratoma with malignant transformation in germ cell tumors in men. Cancer 56:860-863, 1985

44. Molenaar WM, Oosterhuis JW, Meiring A, et al: Histology and DNA contents of a secondary malignancy arising in a mature residual lesion six years after chemotherapy for a disseminated nonseminomatous testicular tumor. Cancer 58:264-268, 1986

45. Dearnaley DP, Horwich A, A'Hern R, et al: Combination chemotherapy with bleomycin, etoposide and cisplatinum (BEP) for metastatic testicular teratoma: Long-term follow-up. Eur J Cancer 27:684-691, 1991

46. Dimipoulos MA, Amato RJ, Logothetis CJ: Predictive factors for effective salvage therapy of nonseminomatous germ cell tumors of testis. Urology 38:351-354, 1991