Chapter 6

Validation of a prediction model and its predictors for the histology of residual masses in testicular germ cell cancer

Abstract

The purpose of this study was to validate a prediction model for the histology (benign or tumour) of residual retroperitoneal masses in patients treated with chemotherapy for metastatic nonseminomatous testicular germ cell cancer. We studied 276 patients, who were treated with chemotherapy before retroperitoneal lymph node dissection at Indiana University Medical Center (IUMC) between 1985 and 1999. A previously developed prediction model was modified in order to provide predictions for the IUMC data based on five predictors (alternative model). For these predictors (presence of teratoma elements in the primary tumour, prechemotherapy levels of the serum tumour markers alpha-fetoprotein and human chorionic gonadotropin, size of the residual mass and change in mass size) univariable and multivariable odds ratios were calculated. The alternative model was evaluated by calculating the *c*-statistic and studying the calibration (reliability) of the model. All odds ratios from the univariable and multivariable analyses were in the expected directions. The alternative model had good discriminative ability (*c*-statistic=0.79), but the predicted probabilities for benign tissue were in general too high. In conclusion, this study confirms the predictive ability of formerly identified predictors for the histology of residual retroperitoneal mass in testicular germ cell cancer. However, the previously developed prediction model may need an adjustment for the local overall ratio of benign versus tumour histology to provide reliable predictions for the patients of IUMC.

Introduction

Fortunately, cis-platin based chemotherapy can currently cure the majority of patients with metastatic non-seminomatous testicular germ cell cancer.[1,2] After chemotherapy, remnants of the retroperitoneal masses may be detected on computer tomography (CT). Retroperitoneal lymph node dissection (RPLND) may reveal the histology of these residual masses. The mass can be totally benign (necrosis or fibrosis) or can contain residual tumour (teratoma or viable cancer cells). Resection of benign tissue has no therapeutic value. Therefore, patients with a high probability for benign tissue should preferably not be resected considering the side effects and costs. In contrast, resection is indicated for patients with a high risk for tumour elements in the residual masses.[3,4]

A number of patient characteristics have been identified as predictors of residual mass histology.[5-8] These predictors include the histology of the primary tumour, prechemotherapy serum levels of the serum tumour markers alpha-fetoprotein (AFP), human chorionic gonadotropin (HCG) and lactate dehydrogenase (LDH), size of the residual mass after chemotherapy and change in mass size during chemotherapy. The predictors have been incorporated in a prediction model which has been developed with logistic regression analysis. The model provides predictions of the probability of benign tissue for the individual patient.[9]

Often, variable selection, model development and model validation are performed with the same data, resulting in over-optimistic estimates of model performance. For our model, we used a meta-analysis for the variable selection[10] and two similar cohorts for model development[9] and validation[11] Therefore, the external validity of the model was satisfactory. Here, we study the applicability of the model for more recently treated patients in a single local setting. This reflects the situation in which clinical readers consider to apply a prediction model for their own patients.

Materials and methods

IUMC data

Between January 1985 and May 1999, 459 patients were treated with chemotherapy and underwent RPLND for remnants of retroperitoneal metastases of nonseminomatous testicular germ cell cancer at Indiana University Medical Center (IUMC). In this retrospective study, we excluded patients with elevated levels of the serum tumour markers AFP or HCG at the time of surgery (50), histologically pure seminoma without elevated prechemotherapy serum tumour markers (2), or with resection after relapse (91). Resection after relapse was defined as RPLND performed more than ten months after orchidectomy. From the 316 patients left, 276 had values for all predictors but prechemotherapy LDH. This predictor was known in only 74 patients.

Studied patient characteristics were primary tumour histology (teratoma-positive or teratoma-negative), prechemotherapy serum tumour marker levels of AFP and HCG (normal or elevated), prechemotherapy and postchemotherapy maximum transversal size in millimetres as measured on CT and the histologic outcome of RPLND. Transversal size was documented categorically: normal; questionable mass; definite mass < 2 cm; 2 – 5 cm; 5 – 10 cm; 10 – 15 cm; 15 – 25 cm; > 25 cm. Change in mass size was categorised as reduction, no change or progression. The histology at resection was classified as benign or tumour. Masses classified as benign contained only fibrotic or necrotic tissue. Tumour masses contained teratoma elements, viable cancer or both.

Prediction model

The original prediction model has been developed in 544 patients from six European and US centres (development data) and has been described in detail before.[9] The development data contained 51 patients from IUMC; most of them where treated before 1985. Three of these

patients were included in the current analysis. Histology of primary tumour and prechemotherapy serum levels of AFP and HCG were included as dichotomous variables. Prechemotherapy serum level of LDH, postchemotherapy mass size and change in mass size were included as continuous variables (after transformation).

The original model could not directly be applied in the IUMC data, because variables were documented categorically and LDH was often missing. Therefore, we modified the model using the development data which resulted in a model with five predictors (alternative model). The variable LDH was left out and the variables postchemotherapy mass size and change in mass size were categorised.

Effects of predictors

Relations between outcome and predictors as included in the alternative model were estimated univariably and multivariably with logistic regression analysis for both the development data and the IUMC data. The exponent of the parameter estimate is the odds ratio (OR). The OR was used as the measure of association. We hypothesised that the multivariable effects of the predictors in the IUMC data did not differ from the effects we found in the development data. To test this hypothesis, we compared the log likelihood of the alternative model containing parameters estimated in the development data with the log likelihood of a model with the same form containing re-estimated parameters from the IUMC data. The difference was tested with a likelihood ratio test with seven degrees of freedom.

Model performance

The performance of the alternative model was studied with respect to discrimination and calibration (reliability). Discrimination refers to the ability of the model to distinguish a patient with benign histology from a patient with residual tumour. We used the concordance-(*c*) statistic to quantify discriminative ability[12] which is identical to the area under the ROC curve. The *c*-statistic represents the probability that a patient with benign tissue has a higher predicted probability than a patient with residual tumour for a random pair of patients with different histologies.

Calibration indicates to what extent the observed frequencies of patients with benign tissue agree with the predicted probabilities for benign tissue. An impression of the calibration was obtained by graphically plotting the observed frequencies versus the predicted probabilities using the Design library in S-plus.[13] Specifically, the hypothesis was tested whether the intercept was different from zero in a logistic regression model, where all coefficients were fixed at the values of the alternative model (calibration in the large). In addition, calibration was statistically tested by the Hosmer-Lemeshow goodness-of-fit test for validation samples.[14]

Table 6.1 Characteristics of testicular germ cell cancer patients undergoing retroperitoneal lymph node dissection in the IUMC data and in the development data; n (%)

Patient characteristics	*IUMC data*		*Development data*	
Year of treatment				
before 1985	-		278	(51)
1985 – 1989	76	(28)	213	(39)
1990 – 1994	101	(37)	53	(10)
1995 - 1999	99	(35)	-	
Primary tumour histology				
teratoma-negative	105	(38)	252	(46)
teratoma-positive	171	(62)	292	(54)
Prechemotherapy AFP level				
normal	70	(25)	186	(34)
elevated	206	(75)	358	(66)
Prechemotherapy HCG level				
normal	76	(27)	205	(38)
elevated	200	(73)	339	(62)
Postchemotherapy mass size				
0 - 5 cm	121	(44)	412	(76)
5 - 10 cm	82	(30)	121	(22)
> 10 cm	73	(26)	11	(2)
Change in mass size during chemotherapy				
reduction	130	(47)	338	(62)
no change	105	(38)	181	(33)
progression	41	(15)	22	(4)
Residual histology				
benign	76	(27)	245	(45)
tumour	200	(73)	299	(55)

P-values smaller than 0.05 were considered as statistically significant. Calculations were performed with SPSS version 8.02 (SPSS Inc., Chicago, Ill), LogXact version 2.1 (Cytel Software Corporation, Cambridge, MA) and S-plus version 4.5 (MathSoft Inc., Seattle, WA) software.

Results

The distributions of prechemotherapy levels of AFP and HCG and the histology of the primary tumour were similar in the IUMC data and in the development data. The resected residual masses were between 0 and 5 cm in 44% of the patients from IUMC, whereas in the development data 76% of patients had a residual mass between 0 and 5 cm. The change in mass size after chemotherapy also differed between the development and the IUMC data. The percentage of patients with benign tissue at resection was lower in the IUMC data than in the development data (28% and 45% respectively, Table 6.1). From the 200 patients in the IUMC data with residual tumour, 164 (59%) had teratoma and 36 (13%) had cancer.

Table 6.2 Univariable analysis of association between predictors and histology of residual mass in patients from IUMC

Predictors	*Benign tissue (n)*	*Tumour (n)*	*Odds Ratio*	*95% Confidence Interval*
Primary tumour histology				
teratoma-negative	52	53		
teratoma-positive	24	147	6.0	3.4 – 10.7
Prechemotherapy serum markers				
AFP normal	35	35		
elevated	41	165	4.0	2.3 – 7.2
HCG normal	26	50		
elevated	50	150	1.6	0.9 – 2.8
Postchemotherapy mass size				
0 – 5 cm	47	74	1.0 [a]	
5 – 10 cm	19	63	0.5	0.3 – 0.9
> 10 cm	10	63	0.3	0.1 – 0.5
Change in mass size during chemotherapy				
reduction	48	82	1.8	1.0 – 3.1
no change	26	79	1.0 [a]	
progression	2	39	0.2	0.04 – 0.7

[a] Reference category

Table 6.2 shows the number of patients with benign tissue or with tumour for the different predictors along with the univariable ORs and the 95% confidence interval. The multivariable ORs were in the same direction as the ORs calculated in the development data (Table 6.3).

However, some were closer to 1 (prechemotherapy level of HCG and postchemotherapy mass size) and others were larger than in the development data (histology of the primary tumour and prechemotherapy level of AFP). Overall, the multivariable ORs were not significantly different in the IUMC (p=0.12).

The alternative model, as estimated in the development data, was well able to indicate patients with residual tumour in the IUMC data (*c*-statistic = 0.79). A graphical impression of the discriminative ability of the alternative model is shown in Figure 6.1A. The median of the predicted probabilities of benign tissue was 18% for patients with tumour and 56% for patients with benign tissue. Figure 6.1B shows the general agreement between the observed frequencies and the predicted probabilities (calibration). Most predictions were somewhat too high. This is expressed by the curve being below the ideal curve which represents equality of predictions and observations. The OR of the difference between the

observed curve and the ideal curve was 0.82 (p=0.21). This is the OR for the prevalence of benign tissue in the IUMC data versus the development data after correction for the fact that patients had more severe characteristics in the IUMC data, e.g. larger residual masses. Without correction for patient characteristics the OR was 0.46 (p=0.001). This implies that a substantial part of the observed difference in prevalence of benign tissue (28% versus 45%) could be explained by the characteristics considered in the alternative model. Overall, the goodness-of-fit was reasonable, as is indicated by the insignificant Hosmer-Lemeshow test (p=0.54).

Discussion

This study shows that previously identified predictors for the histology of residual retroperitoneal masses in patients with nonseminomatous testicular germ cell cancer are also relevant for patients from Indiana University Medical Center (IUMC). A previously developed prediction model containing these predictors was validated after some modification. The model could distinguish benign tissue from tumour.

The present data contained more patients with severe disease. This was shown by the number of residual masses larger than 10 cm and the number of patients with mass enlargement after chemotherapy. Furthermore, fewer patients had benign tissue (28% in IUMC data versus 45% in development data).[9] The IUMC treats more often severe patients as it is a tertiary referral hospital, whereas the hospitals participating in the previous studies included secondary referral centres.[9,11]

Table 6.3 Multivariable odds ratios (95% confidence intervals) of the predictors as included in the alternative model

Predictors	*IUMC data n=276*		*Development data n=544*	
Primary tumour histology				
teratoma-negative vs. positive	5.0	(2.7 – 9.4)	2.7	(1.8 – 4.0)
Prechemotherapy serum markers				
AFP normal vs. elevated	3.7	(1.8 – 7.8)	2.5	(1.7 – 3.9)
HCG normal vs. elevate	1.2	(0.6 – 2.6)	1.8	(1.2 – 2.7)
Postchemotherapy mass size				
0 – 5 cm	1.0[a]		1.0[a]	
5 – 10 cm	0.8	(0.4 – 1.7)	0.3	(0.2 – 0.6)
> 10 cm	0.8	(0.3 – 2.1)	0.6	(0.1 – 3.1)
Change in mass size during chemotherapy				
reduction	1.8	(0.9 – 3.4)	2.4	(1.5 – 3.8)
no change	1.0[a]		1.0[a]	
progression	0.2	(0.04 – 0.9)	0.2[b]	

[a] Reference category
[b] The confidence interval could not be calculated

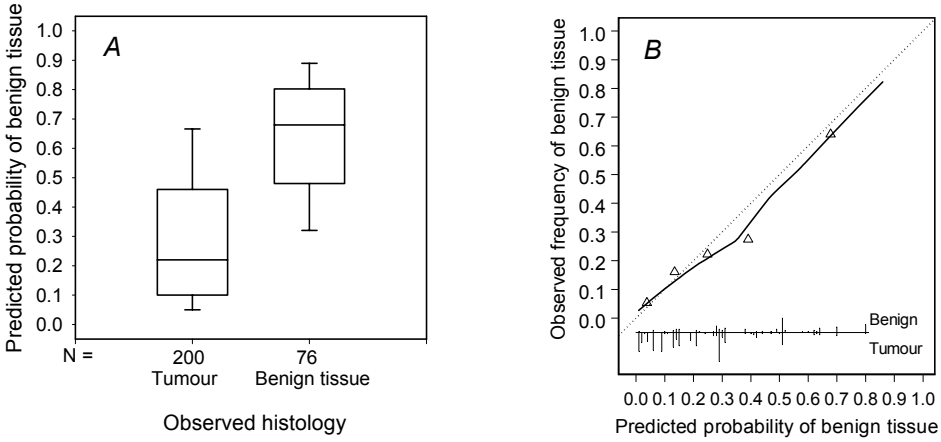

Figure 6.1 Boxplot (A) showing the distribution of the predicted probabilities with the alternative model by histology at resection in the IUMC data. The boxes indicate the median, 0.25 and 0.75 percentiles; the whiskers indicate the 0.10 and 0.90 percentiles.
Calibration of the alternative model in the IUMC data (B). The non-parametric smoothed curve (——) reflects the relation between observed frequencies and predicted probabilities. Perfect calibration would give a line through the origin with slope =1.0 (…). Triangles indicate the observed frequencies of benign tissue in grouped patients with similar predicted probabilities. Vertical lines indicate the distribution of predicted probabilities. Line upwards indicate benign tissue, lines downwards residual tumour.

The original prediction model[9] is based on six predictors: histology of primary tumour, prechemotherapy levels of AFP, HCG and LDH, postchemotherapy mass size and change in mass size after chemotherapy. These characteristics have shown their predictive value for the residual histology in many patient series.[10] The associations between these predictors and the outcome were also shown in the present patients. Unfortunately, LDH could not be studied because of too many missing values. The directions of all ORs were the same compared to the previous studies. The confidence intervals were large due to a smaller number of patients.

From the patients with progression in mass size during chemotherapy two had benign tissue (Table 6.2). It is unclear how benign tissue could progress during chemotherapy. Since we could not specify the percentage of progression, it is also possible that those masses progressed only minimally. In the development data none of the patients with progression in mass size had benign tissue. Thus progression in mass size is a strong predictor for residual tumour, even if the postchemotherapy mass is small. We therefore recommend resection for all patients with increased masses on CT.

We could not apply the original model directly in the IUMC data, because variables were documented categorically and LDH was often missing. Unfortunately, this is a realistic problem, since prediction models can contain variables which are not available for retrospective series. If the physician nevertheless wants to implement that particular model, either all necessary variables have to be collected prospectively or an alternative model - if available - may be used. Here, the model was modified using the development data to make

validation with the IUMC data possible (alternative model). Predictions of the alternative model were very similar to predictions of the original model (Pearson's correlation coefficient=0.86). The discrepancy between the predictions of the two models was predominantly caused by the absence of LDH in the alternative model.

In the IUMC data the prevalence of benign tissue was about half that of the development data (OR=0.46). This could be the result of the more severe patient characteristics in the IUMC data (e.g. larger masses). Indeed, after correction for these characteristics the OR was 0.82, i.e. closer to 1 but still somewhat below 1. The difference in prevalence after correction (not statistically significant) might have been caused by different associations between the probability of benign tissue and predictors, or by characteristics which were not included in the model. Although some associations between the predictors and the probability for benign tissue seemed to be smaller and some seemed to be larger than in the development data, overall there were no significant differences between the two data sets. The comparable discriminative ability in the two data sets (*c*-statistic 0.78 and 0.79) also indicates that the difference in prevalence is not caused by dissimilar associations. Therefore, the difference in prevalence has to be attributed to differences in patient characteristics which were not included in the model, such as referral pattern.

If a residual mass has a high predicted probability for benign tissue, unnecessary resection of the mass can be avoided. Unfortunately, nearly all predicted probabilities for the IUMC data were lower than 80%. Therefore, application of the model in the selected patients has little clinical value. The model might have clinical value to select patients with residual tumour who did not have surgery under the present policy.

If we believe that the deviation in predicted probabilities in the current data is systematic, it would be sensible to adjust the prediction model for patients from IUMC to provide lower predictions for benign tissue. Technically, this can be achieved by decreasing the intercept. The possible need for adjusting the model shows that if a new centre differs clearly from the development data, in aspects like patient selection, referral pattern and treatment, the model should be tested with recently treated patients before implementation. As illustrated, this can be done with retrospective data. Once the (adjusted) model has been implemented, it is necessary to be alert for possible changes in the population which may affect the model performance.

In conclusion, five formerly recognised predictors for the histology of residual retroperitoneal masses after chemotherapy for metastatic testicular germ cell cancer had predictive ability in patients from Indiana University Medical Center. A previously developed prediction model for the histology could be tested after some modification. The discriminative ability of the model was good, but predictions of the alternative model were too high. This illustrates that it is advisable to test the model with retrospective data before implementation in a new centre, if the patients in this centre differ from the development data.

References

1. Einhorn LH: Treatment of testicular cancer: A new and improved model. J Clin Oncol 9: 1777, 1990
2. Mead GM, Stenning SP, Parkinson MC et. al.: The second medical research council study of prognostic factors in nonseminomatous germ cell tumours. J Clin Oncol 10: 85, 1992
3. Donohue JP, Birhle R, Foster RS: Evolving concepts in surgical management of testis cancer. Cancer Treat Res 59: 89, 1992
4. Fox EP, Weathers TD, Williams SD et. al.: Outcome analysis of patients with persistent nonteratomous germ cell tumour in postchemotherapy retroperitoneal lymph node dissections. J Clin Oncol 11: 1294, 1993
5. Toner GC, Panicek DM, Heelan RT et. al.: Adjunctive surgery after chemotherapy for nonseminomatous germ cell tumours: recommendations for patient selection. J Clin Oncol 8: 1683, 1990
6. Donohue JP, Rowland RG, Kopecky K et. al.: Correlation of computerised tomographic changes and histological findings in 80 patients having radical retroperitoneal lymph node dissection after chemotherapy for testis cancer. J Urol 137: 1176, 1987
7. Fosså SD, Qvist H, Stenwig AE et. al.: Is postchemotherapy retroperitoneal surgery necessary in patients with nonseminomatous testicular cancer and minimal residual tumour masses? J Clin Oncol 10: 569, 1992
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, 1990
9. Steyerberg EW, Keizer HJ, Fosså SD et. al.: Prediction of residual retroperitoneal mass histology following chemotherapy for metastatic nonseminomatous germ cell tumour: multivariate analysis of individual patient data from six study groups. J Clin Oncol 13: 1177, 1995
10. 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, 1994
11. Steyerberg EW, Gerl A, Fosså SD et. al.: Validity of predictions of residual retroperitoneal mass histology in nonseminomatous testicular cancer. J Clin Oncol 16: 269, 1998
12. Harrell FE Jr, Califf RM, Pryor DB et. al.: Evaluating the yield of medical tests. JAMA 247: 2543, 1982
13. Harrell FE Jr: Design: S-plus functions for biosstatistical/epidemiological modelling, testing, estimation, validation, graphics, prediction, and typesetting by storing enhanced model design attributes in the fit. Programs available at internet. `http://lib.stat.cmu.edu/DOS/S/Harrell/`.1997
14. Hosmer DW, Lemeshow S.: Assessing the fit of the model. Applied logistic regression. New York, NY: John Wiley & Sons Inc, chapter 5: p.171-73, 1989