Chapter 5

Residual mass histology in testicular germ cell cancer: development and validation of a clinical prediction model

Abstract

After chemotherapy for metastatic non-seminomatous testicular germ cell cancer, surgical resection is a generally accepted treatment to remove remnants of the initial metastases, since residual tumour may still be present (mature teratoma or viable cancer cells). In this paper, we review the development and external validation of a logistic regression model to predict the absence of residual tumour. Three sources of information were used. A quantitative review identified six relevant predictors from 19 published studies (996 resections). Second, a development data set included individual data of 544 patients from six centres. This data set was used to assess the predictive relationships of five continuous predictors, which resulted in dichotomization for two, and a log, square root, and linear transformation for three other predictors. The multiple logistic regression coefficients were reduced with a shrinkage factor (0.95) to improve calibration, based on a bootstrapping procedure. Third, a validation data set included 172 more recently treated patients. The model showed adequate calibration and good discrimination in the development and in the validation sample (*c*-statistic 0.83 and 0.82). This study illustrates that a careful modelling strategy may result in an adequate prediction model. Further study of model validity may stimulate application in clinical practice.

Introduction

Metastatic non-seminomatous testicular germ cell cancer can currently be cured by cis-platin based chemotherapy in over 80% of the patients.[1] Metastases are most common in the retroperitoneal lymph nodes in the abdomen. After chemotherapy, remnants of the initial metastases may be detected on computed tomography (CT). Surgical resection may reveal benign tissue as the histology of these residual masses (necrosis or fibrosis), or residual tumour (mature teratoma or viable cancer). Benign tissue indicates that the patient is cured; its resection has no therapeutic value, in contrast to resection of resection of mature teratoma or viable cancer.

In this study, we review the development of a logistic regression model to predict the presence of benign tissue. We describe the validation of the developed model in a more recent cohort of patients. We focus on methodological aspects, and conclude with a discussion on clinical and statistical aspects in this prediction problem.

Patients and methods

Patients

Three sources of information were used. First, a MEDLINE literature search identified 19 studies, published between 1980 and 1993 with information on univariable relationships between patient or tumour characteristics and the histology at resection in a total of 996 patients.[2] Second, an international data set was collected including individual patient data. Patients were treated with chemotherapy for metastatic nonseminomatous testicular germ cell cancer and underwent resection of retroperitoneal lymph nodes after induction chemotherapy. We excluded patients with elevated levels of the serum tumour markers alpha-fetoprotein (AFP) or human chorionic gonadotropin (HCG) at the time of surgery, extragonadal primaries, histologically pure seminoma without elevated prechemotherapy serum tumour markers, or resection after relapse. This selection was motivated by the variation in selection criteria observed in the literature and aimed to identify a more homogeneous group of patients. After this selection, 556 patients were included in this development data set, who were predominantly treated in the 1980s. Further details have been described before.[3] Third, a validation data set was collected, consisting of 172 patients treated during more recent years in the same centres that participated before (n=72) and patients from one other centre (n=100).[4] The outcome considered in the analyses was the histology at postchemotherapy retroperitoneal resection, classified as benign (necrosis/fibrosis) or tumour (mature teratoma or viable cancer cells).

Methods

The development data set was used to specify the prediction model. Missing values were present in 139 of the 556 patients. Twelve patients were excluded from statistical analysis since two or more predictors were missing. Missing values were imputed in 127 patients, based on the correlation with other predictors.[5] Regression models were constructed for the imputation, where a backward stepwise procedure was used to select covariables ($p<0.10$). After imputation, 544 patients were available for multivariable analysis. We first examined the shape of the predictive relationship for continuous predictors. Restricted cubic spline functions were fitted in logistic regression models to obtain flexible and smooth fits.[6] These fits were subsequently approximated with simple transformations. We considered dichotomization and continuous transformations (square, square root, inverse, logarithmic).

Table 5.1 Predictors for the histology of residual masses

Predictor	*Favourable characteristic*	*Transformation*	χ^2 *– statistics* [a]
Primary tumour histology	Teratoma-negative	-	-
Prechemotherapy AFP	Low values	Normal vs elevated	38 vs 30
Prechemotherapy HCG	Low values	Normal vs elevated	10 vs 14
Prechemotherapy LDH	Higher values	Ln(LDH/normal value)	12 vs 6
Mass size	Small transversal diameter	Square root	109 vs 103
Reduction in size	Large reduction	Linear	137 vs 132

[a] restricted cubic spline function (4 degrees of freedom) versus transformation (1 degree of freedom). The spline function had 5 knots at the 5, 25, 50, 75, and 95 percentiles of the predictor distribution.

A multivariable (or 'multiple) logistic regression model was constructed including all predictors that were identified in the quantitative review. A liberal inclusion of predictors was considered appropriate, since the aim of our model was to provide accurate predictions.[7] A bootstrapping procedure was followed to obtain estimates of internal validity of the model and of a uniform shrinkage factor.[8-10] We drew 200 bootstrap samples with replacement, estimated the regression coefficients in each sample, and evaluated the performance of the models in the bootstrap sample and in the original development sample.

The optimism in the apparent performance in the development sample was indicated by the mean difference in performance between the bootstrap sample and the original sample.[10] The final model was presented as a score chart to facilitate practical application. The score chart was based on rounded values of the shrunk regression coefficients (multiplied by 10).

The validation data set was used to assess the predictive performance of the previously developed model. Calibration, i.e. agreement between observed outcome frequencies and predicted probabilities, was assessed graphically and tested with the Hosmer-Lemeshow test.[11] Discriminative ability was determined with the *c*-statistic, which is equivalent to the area under the ROC curve. For the present paper, calculations were performed with SAS version 6.12 and S-plus version 4.5 software, using the *Design* library.[12]

Table 5.2 Relationships between categorised predictors and the histology at retroperitoneal resection (benign tissue or tumour, n=544 patients)

Predictor	*Favourable characteristic*	*Benign tissue*	*Tumour*
Primary tumour histology	Teratoma-negative	151 (60%)	101 (40%)
Prechemotherapy AFP	Normal	115 (62%)	71 (38%)
Prechemotherapy HCG	Normal	117 (57%)	88 (43%)
Prechemotherapy LDH	Elevated	187 (48%)	206 (52%)
Mass size	0-9 mm	75 (74%)	27 (26%)
	10-19 mm	74 (66%)	39 (34%)
Reduction in size	>= 70%	109 (73%)	41 (27%)
Total	-	245 (45%)	299 (55%)

Results

Six clinical characteristics were identified in the literature as predictive of the histology of residual masses (Table 5.1). The most often identified predictor for benign tissue was the absence of teratoma elements in the primary tumour. Also, prechemotherapy serum tumour marker levels were predictive of benign tissue: low AFP; low HCG; or higher lactate dehydrogenase (LDH) prechemotherapy serum levels. Imaging with CT enables an accurate assessment of the size of the metastases before and after chemotherapy. Both the postchemotherapy mass size and the reduction in mass size were strong predictors, while the prechemotherapy size was unrelated to the residual histology.

Next, transformations were studied for five continuous predictors (Figure 5.1A to 5.1E). Statistical evaluations of the non-linear restricted cubic spline functions and the simple transformations are shown in the final column of Table 5.1. For the serum tumour markers AFP and HCG a dichotomization was considered appropriate, in agreement with previous reports.[13-15] The loss of information for AFP was limited (χ^2 decreased from 38 to 30). Remarkably, dichotomization provided a better fit for HCG (χ^2 increased from 10 to 14). This is due to classification with the local normal values, which implies that the same value may be classified as ‘elevated’ for one patient and as ‘normal’ for another, depending on the hospital. For LDH, mass size, and reduction in mass size, the decrease in fit was small when simple transformations were used instead of spline functions: a logarithmic transformation for standardised LDH values (by dividing through the upper limit of the local normal value); a square root transformation for the postchemotherapy mass size; and a linear transformation for reduction in size.

A problem with mass size measurements was that masses smaller than approximately 5 mm were not detectable. We coded mass size in patients without a detectable postchemotherapy mass as 2 mm. With this truncation, the relation between the square root of the mass size and the log odds of benign tissue was reasonably linear.

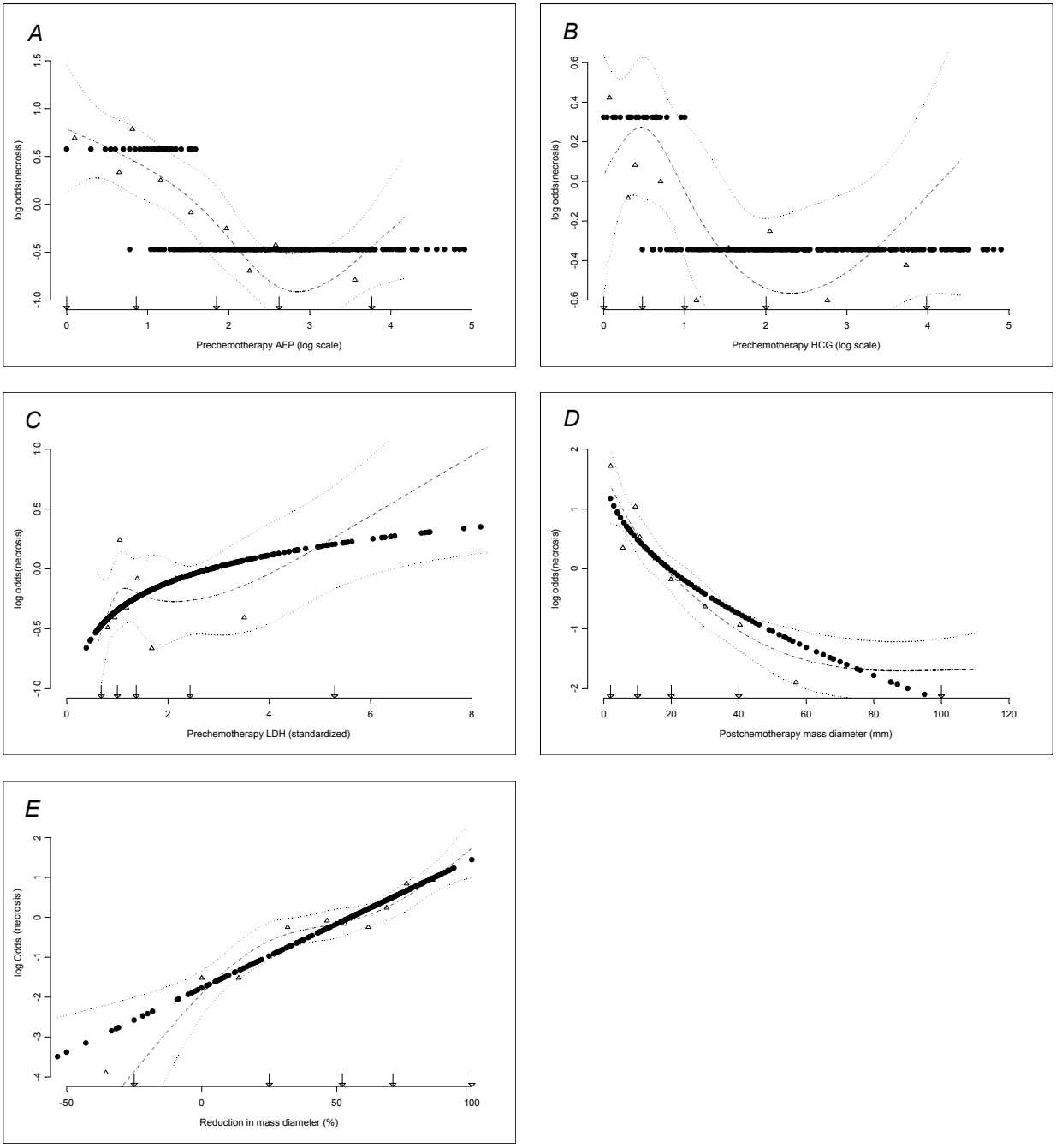

Figure 5.1 Relationships between continuous predictors and benign tissue at resection. A: prechemotherapy AFP levels; B: prechemotherapy HCG levels; C: prechemotherapy LDH levels; D: postchemotherapy mass size; E: reduction in size. Restricted cubic spline functions with 5 knots (arrows) are shown (-.-.-) with 95% confidence intervals. Triangles indicate the observed log odds in grouped patients with similar predictor values. The chosen transformation is indicated with dots for individual predictions.

Table 5.3 Multivariable logistic regression model to predict benign tissue (n=544)

	OR [95%CI] [a]
Primary tumour histology	
Teratoma-negative vs positive	2.5 [1.6 - 3.7]
Prechemotherapy markers	
AFP normal vs elevated	2.5 [1.6 - 3.9]
HCG normal vs elevated	2.2 [1.4 - 3.5]
LDH: $\ln(LDH_{st})$ [b]	2.8 [1.8 - 4.2]
Postchemotherapy mass size	
Sqrt(transversal diameter) [b]	0.74 [.63 - .87]
Change in size	
Per 10% decrease [b]	1.17 [1.1 - 1.3]

[a] Odds Ratio of benign versus tumour with 95% confidence interval
[b] Continuous predictors

Multivariable model

All six predictors were included in a multiple logistic regression model. The odds ratios (OR) corresponding to the regression coefficients are shown in Table 5.3. According to their Wald statistics, the strongest predictors were prechemotherapy LDH and the primary tumour histology. Note that LDH was not a strong predictor in univariable analysis (Table 5.1). This was due to the correlations with the other predictors (especially AFP, HCG and postchemotherapy mass size; Spearman's rank correlations 0.15, 0.21, and 0.34 respectively).

The apparent calibration of model predictions was satisfactory (Figure 5.2A), without clear evidence for a poor fit (Hosmer-Lemeshow statistic p=0.16). The *c*-statistic (or area under the ROC curve) was 0.839 [95% CI 0.81-0.87], indicating good discriminative ability. At a threshold value of 70% for the probability of benign tissue, the true-positive rate (TP, i.e. resection of tumour) was 91% and the false-positive rate (FP, i.e. resection of benign tissue) 51% (Figure 5.3). According to the bootstrapping procedure, the optimism in the *c*-statistic was small (0.01), resulting in 0.83 as the optimism-corrected estimate. The mean slope of the linear predictor was 0.95 in the bootstrap resampling procedure. This estimate was used as a uniform shrinkage factor for all regression coefficients. The intercept was re-estimated such that the sum of the predicted probabilities equalled the number of patients with a benign residual histology.

The final model was presented as a score chart to facilitate practical application (Table 5.4). As an example, we consider a patient with a teratoma-positive primary tumour, elevated AFP and normal HCG before chemotherapy, LDH three times the upper normal level, and a residual mass of 10 mm, measuring 50 mm before chemotherapy (reduction 80%). His score is +0+0+8+11-9+12=+22. The corresponding probability is somewhat less than 80%.

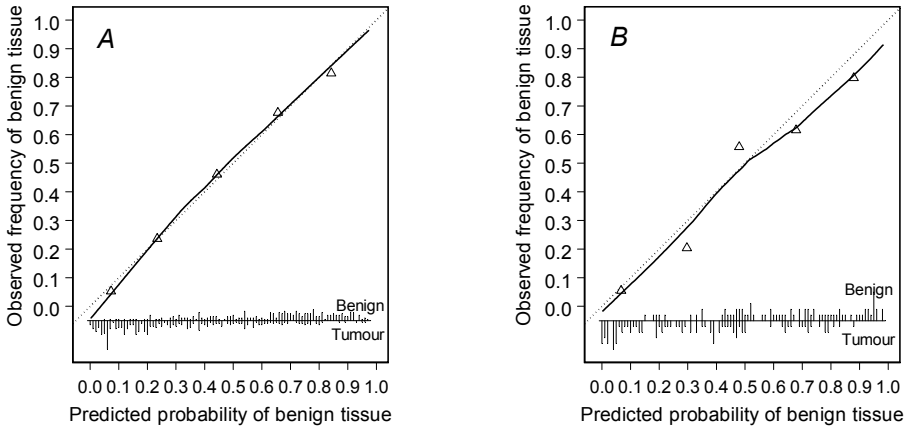

Figure 5.2 Calibration of the prediction model in the development data (A, n=544) and in the validation data (B, n=172). The dotted line indicates perfect calibration. Triangles indicate the probabilities in grouped patients with similar predicted probabilities. A non-parametric, smoothed curve indicates the relation between the observed frequencies and predicted probabilities of benign tissue. Vertical lines indicate the distribution of the predicted probabilities. Line upwards indicate patients with benign tissue, lines downwards tumour.

External validation

The distribution of patient characteristics was very similar in the development and validation data. The prevalence of benign tissue was identical to that in the development data set (45%; 77 of 172 patients). Figure 5.2B shows that the calibration was adequate. The discriminative ability of the prediction model was satisfactory (*c*-statistic 0.824 [0.77-0.88]), although the TP and FP rates were somewhat poorer than in the development data at a threshold value of 70% (85% and 53% respectively, Figure 5.3).

Discussion

We developed and validated a prediction model for the histology of a residual retroperitoneal mass following chemotherapy for metastatic testicular germ cell cancer. Below we discuss clinical and statistical aspects.

Clinical aspects

Clinical decision making on resection is difficult for patients who have completed chemotherapy for metastatic testicular germ cell cancer. Several studies have aimed to identify subgroups of patients with a high likelihood of a benign residual mass, for whom resection would not be beneficial. These subgroups were mostly defined with a limited number of patient characteristics. Examples of subgroups include those with a teratoma-negative tumour, normal serum tumour markers AFP and HCG, and small residual masses (<20 mm)[13] and those with a teratoma-negative tumour and a large reduction in size (>90% in volume)[16]. Local resection policies show a wide variation. The most common policy in Europe is to resect residual masses that are detected on CT as abnormal (>= 10 mm)[14,17]. Effectively, such a policy uses only one predictor (residual mass size), with an arbitrary chosen cut-off (10 mm). Other cut-offs have also been used, e.g. 20 mm in the UK[18]. The extreme policy is to consider all patients as candidates for resection, arguing that the

mortality and morbidity of resection is less important than -even small- risks of missing residual tumour.[13]

Table 5.4 Score chart for the probability of benign tissue after chemotherapy for metastatic testicular germ cell cancer

Predictor	*Value*							*Score*
Primary tumour histology								
Teratoma-negative	+9							.....
Prechemotherapy markers								
Normal AFP	+9							.....
Normal HCG	+8							.....
LDH/normal value	0.6	0.8	1.0	1.5	2.0	3.0	4.5	
Score	-5	-2	0	+4	+7	+11	+15	.....
Postchemotherapy mass size								
Transversal diameter (mm)	2	5	10	20	30	50	100	
Score	-4	-6	-9	-13	-16	-20	-28	.....
Diameter reduction								
100(presize-postsize)/presize	-50	0	50	75	100			
Score	-7	0	+7	+11	+15			.....
Estimated individual probability of benign tissue					Sum score (add)			.....
Sum score	10	15	20	25	30	35	40	
Probability (%)	51	63	74	82	88	93	95	

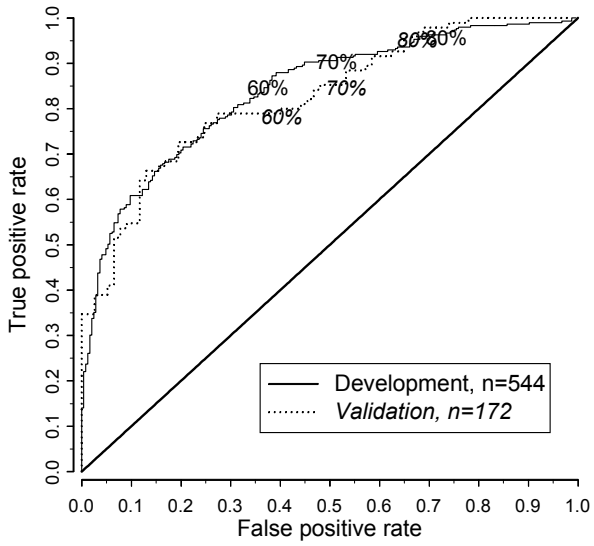

Figure 5.3 Receiver operating characteristic (ROC) curve for the prediction model as evaluated in the development data (n=544) and in the validation data (n=172). Threshold values for the probability of benign tissue (range: 60-80%) are indicated.

Our prediction model combines all well-known predictors in a regression formula, which can be used to calculate the individual probability of benign tissue. The calculation can be facilitated by presentation in a simple score chart (as in Table 5.4), a nomogram, or a table. The latter representation would require that continuous variables are categorised, implying some loss of information.

For clinical decision making, a threshold has to be chosen for the probability of benign tissue above which patients are candidates for observation rather than resection. We considered two approaches to support the choice of this threshold.

First, we compared the performance of the regression model with current resection policies.[19] We hereto entered the predictors considered in a policy in a logistic regression model. The implicit threshold for the policy was calculated as the probability corresponding to the threshold used for one or more clinical variables. For example, the policy to resect all detectable masses had an implicit threshold of 61%, since 61% was the probability corresponding to filling in 10mm for the postchemotherapy mass size in a logistic regression model with mass size as the single predictor. Most current resection policies had implicit thresholds in the range of 60 to 80%.[19]

Second, we performed a decision analysis.[20] A major difficulty in this analysis was that the risks associated with leaving masses with tumour unresected, such as growth and relapse, needed to be quantified. By definition, no direct empirical data can be obtained for these risks, since the only reliable way to classify residual masses as benign or tumour is to perform a resection. Some indirect evidence might come from follow-up of unresected patients, but no studies with sufficient detail and numbers of patients are yet available. We therefore asked ten experts to estimate relapse and survival rates within five years. We found that the probability of benign tissue needed to be extremely high to make resection not beneficial (>95%). A threshold of around 70% implies that resection is only offered to patients with around 2 years gain in life-expectancy, which may be considered a relatively large benefit.[21]

These findings may support the choice of 70 or 80% as a threshold for the probability of benign tissue. The model did classify around 20% of the patients of both data sets in this category. Note that the percentage can more accurately be assessed in a data set of unselected patients, i.e. not only those undergoing resection, but all fulfilling the selection criteria (e.g. normal serum tumour markers after complete chemotherapy). Furthermore, note that individual circumstances and patient preferences may lead to another threshold.

Currently, the prediction model has not yet widely been applied. In the participating centres the use varies from routine application for all patients to incidental application in surgically difficult cases. Partly, this may be explained by general aspects of decision support tools. For example, a more explicit decision making process is required compared to following a simple resection policy which is also plausible (e.g. resection of a detectable mass). Further, clinicians may be sceptical about the validity of the predicted probabilities. Indeed, many currently available regression models may be unreliable, especially when no correction was made for possible overfitting in small data sets.[22] Further testing of our model

may confirm the validity of the developed model, but may also point at a need for adjustment to local circumstances. We hope that the current trend towards 'evidence-based medicine' will encourage physicians to apply prediction models in clinical practice.

Statistical aspects

A major issue in prediction modelling is the choice of covariables in the model. In our study, we first explicitly reviewed the literature for potential predictors and identified six. Some were not recognised as such in individual studies because of the limited number of patients analysed, e.g. normal prechemotherapy AFP and HCG levels. A limitation of such a review is that the focus is on simple univariable relationships, thus ignoring correlations between predictors. In the multivariable model we however found that all six predictors were statistically significant and had independent prognostic relevance. For selection of covariables, one might well argue that any predictive information should be incorporated, independent of statistical significance.[7,10] Alternatively, backward stepwise selection might be applied with a liberal inclusion of covariables in the model, i.e. a p-value of 0.50.[23]

A related issue in prediction modelling is the coding of continuous covariables. We used restricted cubic splines to provide smoothed estimates of the predictive relationships. With five knots, four degrees of freedom are given to the covariables providing sufficient flexibility to fit complex patterns.[6] However, the flexibility may be too large, leading to a too close fitting of idiosyncracies in the data set rather than true patterns. One might therefore penalise the non-linear terms in the spline more than the linear term,[10] or as we did, try to approach the non-linear spline with a simple transformation.[6] In the same spirit, Royston recently proposed to select a parametric function by comparison with a cubic smoothing spline as the reference curve.[24]

Once the model is specified, we estimate the regression coefficients. This estimation usually does not consider the modelling process that preceded it, leading to too extreme estimates.[25,26] The degrees of freedom effectively used in the modelling process may be far larger than the degrees of freedom in the final model.[27] Further, it has been shown that even fully pre-defined models require some shrinkage in the coefficients to provide reliable predictions.[8,9] We estimated the required shrinkage with a bootstrapping procedure, which is relatively easy with current software and computer power. We did not include the model specification phase in the bootstrap procedure, which might have led to a slightly smaller shrinkage factor than the current estimated of 0.95. For comparison, we calculated the shrinkage factor with a heuristic formula: (model χ^2 – degrees of freedom)/ model χ^2. [8,9] For our model, model χ^2 was 211.6 with 6 degrees of freedom, leading to 0.97 as estimate for the shrinkage factor.

The developed model performed well on external validation. In general, we might expect that the combined effect of covariables would be reasonably comparable between largely similar populations, once the estimated regression coefficients were adequately shrunk.[28] Adjustment may then only be required for the intercept in the model, e.g. caused by general changes in treatment.[29]

Our model focused on the distinction between benign and tumour masses, which is clinically the most important. However, missing viable cancer is generally considered more serious than missing mature teratoma. Therefore, a model that predicts three categories might be desirable. Such a model might be constructed in several ways. First, one might think of a polytomous logistic model. It would be natural to make benign tissue the reference category, leading to models which predict the odds of mature teratoma against benign histology and viable cancer against benign histology. One might specify the predictors for both models separately, with separate analysis of the predictive relationship of continuous variables and inclusion of different sets of the six candidate predictors. The clinical application of the models would be harder than the presently developed model, since probabilities would be estimated based on two models, and construction of a one dimensional score chart would not be possible anymore. Alternatively, one might naively develop two simple models: one for mature teratoma against other tissue and one for cancer against other tissue. This would however lead to probabilities that sum to more than 100% for some combinations of predictors. We previously chose a third option, i.e. to develop a submodel to distinguish cancer from teratoma.[3] In the development data set, the average odds of cancer against mature teratoma was 1:3, and three covariables could be used to individualise that odds. The resulting model could be presented in the score chart, the sumscore referring to the conditional probability of cancer. This choice was predominantly motivated by practical applicability of the model, and further statistical research might elaborate on the pros and cons of alternative approaches to outcomes with three or more categories.

In conclusion, we developed and externally validated a prediction model to estimate the probability of benign tissue in residual masses after chemotherapy. Further validation is necessary to convince physicians about the practical usefulness of the model and to identify potential needs for adjustment. The explicit use of various sources of information (literature, development data, validation data) and concern about overfitting and overoptimism may also be relevant in other prediction problems.

References

1. Bosl GJ, Motzer RJ: Testicular germ-cell cancer. N Engl J Med 337:242-53, 1997
2. Steyerberg EW, Keizer HJ, Stoter G, et al: Predictors of residual mass histology following chemotherapy for metastatic non-seminomatous testicular cancer: a quantitative overview of 996 resections. Eur J Cancer 30A:1231-9, 1994
3. Steyerberg EW, Keizer HJ, Fosså SD, et al: Prediction of residual retroperitoneal mass histology after chemotherapy for metastatic nonseminomatous germ cell tumour: multivariate analysis of individual patient data from six study groups. J Clin Oncol 13:1177-87, 1995
4. Steyerberg EW, Gerl A, Fosså SD et al: Validity of predictions of residual retroperitoneal mass histology in nonseminomatous testicular cancer. J Clin Oncol 16:269-74, 1998
5. Little RJA: Regression with missing X's: a review. J Am Stat Assoc 87:1227-1237, 1992
6. Harrell FE Jr, Lee KL, Pollock BG: Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 80:1198-202, 1988
7. Spiegelhalter DJ: Probabilistic prediction in patient management and clinical trials. Stat Med 5:421-33, 1986
8. Copas JB: Regression. prediction and shrinkage. J R Stat Soc B 45:311-354, 1983
9. Van Houwelingen JC, Le Cessie S: Predictive value of statistical models. Stat Med 9:1303-25, 1990

10. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361-87, 1996
11. Lemeshow S, Hosmer DW, Jr.: A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 115:92-106, 1982
12. Harrell FE, Jr.: Design: S-plus functions for biostatistical/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/`, 1999
13. 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 10:569-73, 1992
14. Mulders PF, Oosterhof GO, Boetes C, et al: The importance of prognostic factors in the individual treatment of patients with disseminated germ cell tumours. Br J Urol 66:425-9, 1990
15. Tait D, Peckham MJ, Hendry WF, et al: Post-chemotherapy surgery in advanced non-seminomatous germ-cell testicular tumours: the significance of histology with particular reference to differentiated (mature) teratoma. Br J Cancer 50:601-9, 1984
16. 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 137:1176-9, 1987
17. Jansen RL, Sylvester R, Sleyfer DT, et al: Long-term follow-up of non-seminomatous testicular cancer patients with mature teratoma or carcinoma at postchemotherapy surgery. Eur J Cancer 27:695-8, 1991
18. 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 10:85-94, 1992
19. Steyerberg EW, Keizer HJ, Habbema JD: Prediction models for the histology of residual masses after chemotherapy for metastatic testicular cancer. Int J Cancer 83:856-9, 1999
20. Steyerberg EW, Marshall PB, Keizer HJ, et al: Resection of small, residual retroperitoneal masses after chemotherapy for nonseminomatous testicular cancer: a decision analysis. Cancer 85:1331-41, 1999
21. Wright JC, Weinstein MC: Gains in life expectancy from medical interventions--standardizing data on outcomes. N Engl J Med 339:380-6, 1998
22. Altman DG, Royston P: What do we mean by validating a prognostic model? Stat Med 19:453-73, 2000
23. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, et al: Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. Stat Med 19:1059-1079, 2000
24. Royston P: A strategy for modelling the effect of a continuous covariate in medicine and epidemiology. Stat Med 19:1831-47, 2000
25. Chatfield C: Model uncertainty, data mining and statistical inference. J Roy Statist Soc A 158:419-466, 1995
26. Steyerberg EW, Eijkemans MJC, Habbema JDF: Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. J Clin Epidemiol 52:935-942, 1999
27. Ye J: On measuring and correcting the effects of data mining and model selection. JASA 93:120-131, 1998
28. Van Houwelingen HC, Thorogood J: Construction, validation and updating of a prognostic model for kidney graft survival. Stat Med 14:1999-2008, 1995
29. van Houwelingen HC: Validation, calibration, revision and combination of prognostic survival models. Stat Med 19:3401-15, 2000