

Chapter 8

Predicting retroperitoneal histology in postchemotherapy

testicular germ cell cancer:

a model update and multicentre validation with 1094 patients

Abstract

Patients treated with chemotherapy for advanced testicular germ cell cancer may undergo resection of retroperitoneal lymph nodes, that often contain only necrosis or fibrosis (benign tissue). To improve the selection of candidates for resection, we previously developed a logistic regression model with six variables to predict the probability of benign tissue. The aim of the present study was to update the model using new patient data and to study the validity of the updated model. We combined the data of 544 patients that were used before to develop the model with data of 550 new patients and performed a new logistic regression analysis. In total data of 1094 patients were available from five large international centres, a number of smaller centres, and an international randomised clinical trial. The validity of the updated model was studied in individual centres. The agreement between the observed frequencies and the predicted probabilities of the updated model (calibration) was assessed graphically and studied with the Hosmer-Lemeshow test. The discriminative ability was studied with the concordance (*c*)-statistic. The updated model contained only slightly different, though more precise, regression coefficients. Most centres showed good calibration as confirmed by statistically non-significant Hosmer-Lemeshow tests. The *c*-statistics for all centres except one were 0.80 or higher. In conclusion, the updated model was only slightly different from the original model. The validity across a broad spectrum of centres gives confidence in the transportability of the model to support the selection of candidates for resection in new centres.

Introduction

Cis-platin based chemotherapy is a highly effective treatment for patients with advanced nonseminomatous testicular germ cell cancer. Surgical resection after chemotherapy is a standard treatment to remove retroperitoneal lymph nodes. These masses contain necrotic debris or fibrosis in approximately 40% - 50%, teratoma elements in approximately 30% - 40%, and viable tumour in approximately 10% - 20%.^{1,2} Predicting the residual mass

histology before resection would be helpful to determine whether resection should be performed.³⁻⁷ We therefore previously developed a model to predict the probability that the residual mass contains only necrosis and fibrosis (benign tissue) as opposed to teratoma elements or viable tumour (residual tumour).⁸ The prediction model contained six predictors, i.e. the levels of three serum tumour markers, the presence of teratoma elements in the primary tumour, the size of the residual mass measured on CT and the change in mass size induced by chemotherapy.

The validity of the prediction model has been tested in three other patients groups (validation data).⁹⁻¹¹ The patients were treated in other centres or more recently. The strength of the predictors showed some variation in the validation data. Further, the model may need some adjustment for patients from a tertiary referral centre (Indiana University Medical Center [IUMC]). These patients possibly differ in aspects which were not included in the model.

A model based on more patients who are treated in a broader spectrum of centres and treated more recently may be better transportable to new patients. Therefore we aimed to update the original model using all presently available individual patient data from the development and validation studies. Further, the validity of the updated model was studied in individual centres.

Patients and methods

Patients

Patients were treated with chemotherapy for advanced testicular germ cell cancer and had undergone surgical resection of retroperitoneal lymph nodes. Resected masses contained only necrosis or fibrosis (benign tissue), or also teratoma elements or viable cancer (residual tumour). We excluded patients with elevated levels of serum tumour markers alpha-fetoprotein (AFP) or human chorionic gonadotropin (HCG) at the time of surgery, histologically pure seminoma without elevated prechemotherapy serum tumour markers, or with resection after relapse.

The development data (n=544) were mainly derived from patients treated in three large centres (Academic Hospital Groningen [AHG], n=137; Norwegian Radium Hospital [NRH], n=127; and Memorial Sloan Kettering Cancer Center [MSKCC], n=120) and several smaller Dutch centres (n=118) (Table 8.1).⁸ The validation data (n=550) were mainly derived from patients treated in a tertiary referral centre (Indiana University Medical Center [IUMC], n=273)¹⁰ and a large German centre (Klinikum Grosshadern München [KGM], n=98).¹² Further data were derived from an EORTC/MRC trial (n=105) containing only patients with good prognosis disease.¹³ Resection policies varied according to local policies. Ninety percent of the patients in the development data were treated before 1990, while 75% of all patients in the validation data were treated in 1990 or later.

Table 8.1 Patient groups, numbers of patients and years of treatment (range) in the development and validation data

<i>Patient groups</i>	<i>Development data</i>		<i>Validation data</i>	
	<i>N</i>	<i>Years</i>	<i>N</i>	<i>Years</i>
Academic Hospital Groningen	137	1979 – 93	17	1995
Small Dutch centres	118	1979 – 91	34	1992 – 95
Norwegian Radium Hospital	127	1980 – 90	23	1995
Klinikum Grosshadern München	–	–	98	1980 – 96
Indiana University Medical Center	42	1979 – 85	273	1986 – 99
Memorial Sloan Kettering Cancer Center	120	1977 – 88	–	–
EORTC / MRC trial	–	–	105	1995 – 98
Total	544	1977 – 93	550	1980 – 99

Prediction model

The prediction model to estimate the probability of benign tissue was developed with logistic regression analysis.⁸ Three dichotomous and three continuous predictors were included based on a literature review.¹⁴ The three dichotomous predictors were the presence of teratoma elements in the primary tumour (no versus yes), determined as teratoma differentiated or malignant teratoma intermediate, and the prechemotherapy levels of the serum tumour markers AFP and HCG (normal versus elevated). The three continuous predictors were the prechemotherapy level of the serum tumour marker lactate dehydrogenase (LDH) expressed as the natural logarithm of the standardised value ($\ln[\text{LDH}_{\text{st}}] = \ln(\text{LDH}/\text{upper limit of the normal value})$), the square root of the maximum diameter of the residual mass measured on CT in mm after chemotherapy (sqpost), and the change in diameter per 10% induced by chemotherapy (change10).

For clinical practice, we previously proposed a threshold value of 70%, i.e. patients with predicted probabilities of benign tissue below 70% may be advised to undergo surgery and patients with probabilities above 70% not.^{9,15} This value lies in between implicit threshold values corresponding to more conservative resection practices (around 60%) and more aggressive practices (around 85%).¹⁶

Updating the model

The original model was updated by estimating new multivariable logistic regression coefficients for the six predictors using all patient data (n=1094). The exponent of a logistic regression coefficient equals the odds ratio (OR). For the 273 patients from IUMC included in the validation data, values of LDH were missing and values of mass size and change in mass size were categorically reported instead of continuously.¹⁰ Therefore, multivariable information of the six predictors was available for 821 patients (complete cases). We used an adaptation method to include also the univariable information of the predictors teratoma elements in the primary tumour, and prechemotherapy serum levels of AFP and HCG from

the 273 patients with missing or categorical data for the other three predictors. This method was originally developed to combine literature results, which are often univariable in nature, with multivariable results from individual data.^{17,18} Similarly, we estimated the univariable regression coefficients of the three predictors known for all patients (n=1094). The univariable results were adapted by the difference in the univariable and multivariable regression coefficients as estimated in the complete cases (n=821). The multivariable regression coefficients of the three continuous predictors were only based on the complete cases. The adaptation method was also used to estimate the multivariable coefficients in the validation data (n=550, 277 complete cases).

The updated model was presented as a score chart. To derive the scores, the regression coefficients were multiplied by 10 and subsequently rounded.

Model validity by centre and in time

To study the validity of the updated model by centre and in time, we added the variable centre (categorical) and year of treatment (continuous) as an additional covariable in the multivariable regression model with the six predictors. Several small Dutch centres were grouped and the EORTC/MRC trial was also considered as one centre in this analysis. A statistically significant effect of the covariable centre or year of treatment would indicate variation between centres or in time. This variation would be a consequence of differences in the average frequency of benign tissue (outcome frequency) after adjustment for possible case-mix differences (i.e. differences in the distributions of the predictor values between centres or in time). The data of IUMC were not considered in these analysis, because adjustment for case-mix was not possible for the continuous predictors of the model. We also included interaction terms of centre and year of treatment with the predictors in the multivariable model. Statistically significant interaction would indicate that the predictor effects were different between centres or in time.

Variation in the average outcome frequency of benign tissue between centres would statistically be reflected by variation in the model intercept that can be calculated for each centre separately. Therefore, we studied the possible variation in outcome frequency also with a random-effects method.¹⁹ We estimated, for each centre separately, a model intercept and standard error, given the regression coefficients of the updated model. We tested whether the model intercepts could have been derived from the same Normal distribution, i.e. from the same underlying population.²⁰

Further, the performance of the updated model was studied by centre. Calibration was studied graphically by plotting the observed outcome frequencies against the predicted probabilities.^{21,22} Ideally this relation is described by a straight line with intercept=0 and slope=1. We tested whether the intercept differed statistically significant from 0.²³ The Hosmer-Lemeshow statistic for goodness-of-fit was also estimated.²⁴ Discriminative ability, i.e. whether the model is able to distinguish patients with benign tissue from patients with residual tumour, was determined with the concordance (c)-statistic. This statistic is for

binary outcomes identical to the area under the ROC curve.²⁵ The *c*-statistic of the original model was 0.83 (after correction for optimism).⁸

Results

Table 8.2 shows the relationship of the six predictors with the residual mass histology for the development and validation data. The proportion of patients with benign residual masses was lower in the validation data (33% versus 45% in the development data). All univariable ORs of the two data sets were generally in the same direction (larger or smaller than 1).

Table 8.2 Distribution of the predictors and results of univariable logistic regression analyses in the development and validation data

<i>Predictors</i>	<i>Values</i>	<i>Development data</i>			<i>Validation data</i>		
		<i>Benign/All</i>	<i>%</i>	<i>OR [CI95]</i>	<i>Benign/All</i>	<i>%</i>	<i>OR [CI95]</i>
Teratoma elements in primary tumour	present	94/292	32%	1	52/299	17%	1
	absent	151/252	60%	3.1 [2.2-4.5]	128/251	51%	4.9 [3.4-7.3]
AFP serum level	elevated	130/358	36%	1	95/397	24%	1
	normal	115/186	62%	2.8 [2.0-4.1]	85/153	56%	4.0 [2.7-5.9]
HCG serum level	elevated	128/339	38%	1	113/377	30%	1
	normal	117/205	57%	2.2 [1.5-3.1]	67/173	39%	1.5 [1.0-2.2]
LDH serum level	normal	58/149	39%	1	29/111	26%	1
	1-2 x normal	98/215	46%	1.3 [0.9-2.0]	32/88	36%	1.6 [0.9-3.0]
	>2 x normal	89/180	49%	1.5 [1.0-2.4]	43/78	55%	3.5 [1.9-6.4]
Residual mass size	0-19 mm	153/228	67%	9.0 [4.3-19]	61/116	53%	7.2 [3.4-15]
	20-49 mm	71/184	39%	2.8 [1.3-5.8]	75/215	35%	3.5 [1.7-6.9]
	50-99 mm	11/78	14%	0.7 [0.3-1.8]	33/137	24%	2.0 [1.0-4.3]
	≥100 mm	10/54	20%	1	11/82	15%	1
Change in mass size	increase	0/42	0%	-	0/91	11%	0.4 [0.2-0.7]
	0-49%	54/187	29%	1	65/257	25%	1
	50-69%	75/154	49%	2.3 [1.5-3.7]	54/113	48%	2.7 [1.7-4.3]
	≥70%	116/161	72%	6.3 [4.0-10]	51/89	57%	4.0 [2.4-6.6]
Total		245/544	45%		180/550	33%	

OR: odds ratio; CI95: 95% confidence interval

Table 8.3 shows the multivariable ORs corresponding to the regression coefficients of the original model and of the updated model, together with the ORs as estimated in the validation data. The ORs for the presence of teratoma elements in the primary tumour and normal AFP serum level were somewhat higher in the validation data compared with the original model (4.5 versus 2.5, and 4.4 versus 2.5, respectively), while the ORs for normal HCG serum level and standardised LDH serum level were somewhat lower (1.4 versus 2.2, and 1.9 versus 2.7, respectively). The other two ORs were more similar (residual mass size: 0.79 versus 0.75, and change in mass size: 1.13 versus 1.16), although small differences in effect for continuous variables may become substantial for high predictor values. The ORs estimated for the updated model were in between the ORs of the original model and the estimates in the validation data and hence only slightly different from the original model. Further, the estimates were more precise, as reflected in smaller confidence intervals.

Table 8.4 shows the variation in the average outcome frequency of benign tissue between centres and in years of treatment. The ORs were estimated univariably and after adjustment of the six predictors. The univariable ORs for centre and year of treatment showed clear variation ($p=0.003$ and $p=0.002$, respectively). Most of the variation could be explained by a difference in case-mix, since the case-mix adjusted ORs were close to 1 (p -values > 0.20). The random effect method also did not indicate variation between centres.

Interaction with centre was univariably found for change in mass size ($p=0.04$). The OR for change in mass size was near 1 for the EORTC/MRC trial compared to the overall OR of 0.86. Change in mass size ($p=0.07$) and LDH serum level ($p=0.04$) showed some univariable differences in effect over time. When the interactions were tested multivariably, none of the differences were statistically significant (all p -values > 0.20), implying that it was reasonable to assume similar predictor effects across centres and in time.

Table 8.3 Results of the multivariable logistic regression analyses in the development data (original model), validation data and the two data sets combined (updated model); odds ratio [95% confidence interval]

<i>Predictors</i>	<i>Development data</i> <i>n=544</i>	<i>Validation data</i> <i>n=550</i>	<i>Combined data</i> <i>n=1094</i>
Teratoma elements absent	2.5 [1.6 – 3.8]	4.5 [2.3 – 8.8]	3.1 [2.3 – 4.2]
AFP normal	2.5 [1.6 – 3.9]	4.4 [2.2 – 8.7]	3.0 [2.2 – 4.3]
HCG normal	2.2 [1.4 – 3.4]	1.4 [.68 – 2.9]	2.0 [1.4 – 2.9]
LDH transformed	2.7 [1.8 – 4.0]	1.9 [1.2 – 2.9] ^a	2.3 [1.7 – 3.0] ^b
Postchemo mass size	.75 [.64 – .88]	.79 [.66 – .94] ^a	.76 [.68 – .86] ^b
Change in mass size per 10%	1.16 [1.05 – 1.29]	1.13 [1.03 – 1.23] ^a	1.15 [1.08 – 1.23] ^b

^a Estimates based on patients with complete data, $n=277$

^b Estimates based on patients with complete data $n=821$

Table 8.4 Average outcome frequencies per centre and time period with corresponding odds ratios

<i>Effects</i>	<i>Benign/All (%)</i>	<i>Univariable OR [CI95]</i>	<i>Case-mix adjusted OR [CI95]</i>
<i>Centre</i>			
Academic Hospital Groningen	70/154 (46%)	1	1
Small Dutch centres	66/152 (43%)	0.92 [0.59 – 1.4]	1.1 [0.63 – 1.9]
Norwegian Radium Hospital	78/150 (52%)	1.3 [0.83 – 2.0]	1.1 [0.64 – 1.9]
Klinikum Grosshadern München	45/98 (46%)	1.0 [0.61 – 1.7]	1.0 [0.54 – 1.9]
Indiana University Medical Center	84/315 (27%)	0.44 [0.29 – 0.65]	-
Memorial Sloan Kettering Cancer Center	55/120 (46%)	1.0 [0.63 – 1.6]	0.61 [0.34 – 1.1]
EORTC/MRC trial	27/105 (26%)	0.42 [0.24 – 0.71]	0.92 [0.47 – 1.8]
<i>Year of treatment</i>			
before 1986	170/341 (50%)	2.8 [1.6 – 4.7]	1.1 [0.54 – 2.1]
1986 – 1990	86/203 (42%)	2.0 [1.2 – 3.6]	0.90 [0.45 – 1.8]
1991 – 1995	63/152 (41%)	2.0 [1.1 – 3.5]	0.85 [0.41 – 1.8]
1996 – 1999	22/83 (27%)	1	1

OR: odds ratio; CI95: 95% confidence interval

Figure 8.1 shows the performance of the updated model in the different patient groups (AHG, the small Dutch centres, NRH, KGM, MSKCC, and the patients from the EORTC/MRC trial). The model performance could not be assessed for the IUMC data, because of the missing values for LDH serum levels. The calibration was satisfactory for three of the four large centres (AHG, NRH, and KGM) and the small Dutch centres. Particularly in the decision area around 70%, observed frequencies and predicted probabilities of benign tissue were in agreement. Predicted probabilities for patients from MSKCC however were in general too high. The deviation was statistically significant (intercept of calibration line = -0.48; $p=0.03$). In contrast, the Hosmer-Lemeshow test was not statistically significant ($p=0.17$). The calibration for the patients of the EORTC/MRC trial were reasonable for higher predictions, but low predicted probabilities showed disagreement with the observed frequencies (Hosmer-Lemeshow test: $p<0.01$). The discrimination was in general good. For all centres except the EORTC/MRC trial, the c -statistic was 0.80 or higher.

Figure 8.2 shows the updated model in a score chart format. As an example, we consider a patient with no teratoma elements in the primary tumour (11 points), normal serum level of prechemotherapy AFP (11 points) and elevated serum level of HCG (0 points), standardised LDH serum level of 3 times normal (9 points) postchemotherapy mass size of 20 mm (-3 points) and a reduction in mass size of 50% (7 points). He has a sumscore of 35, which corresponds with a probability of benign tissue of 81%. Application of the original model resulted in an predicted probability of 79% for this patient.

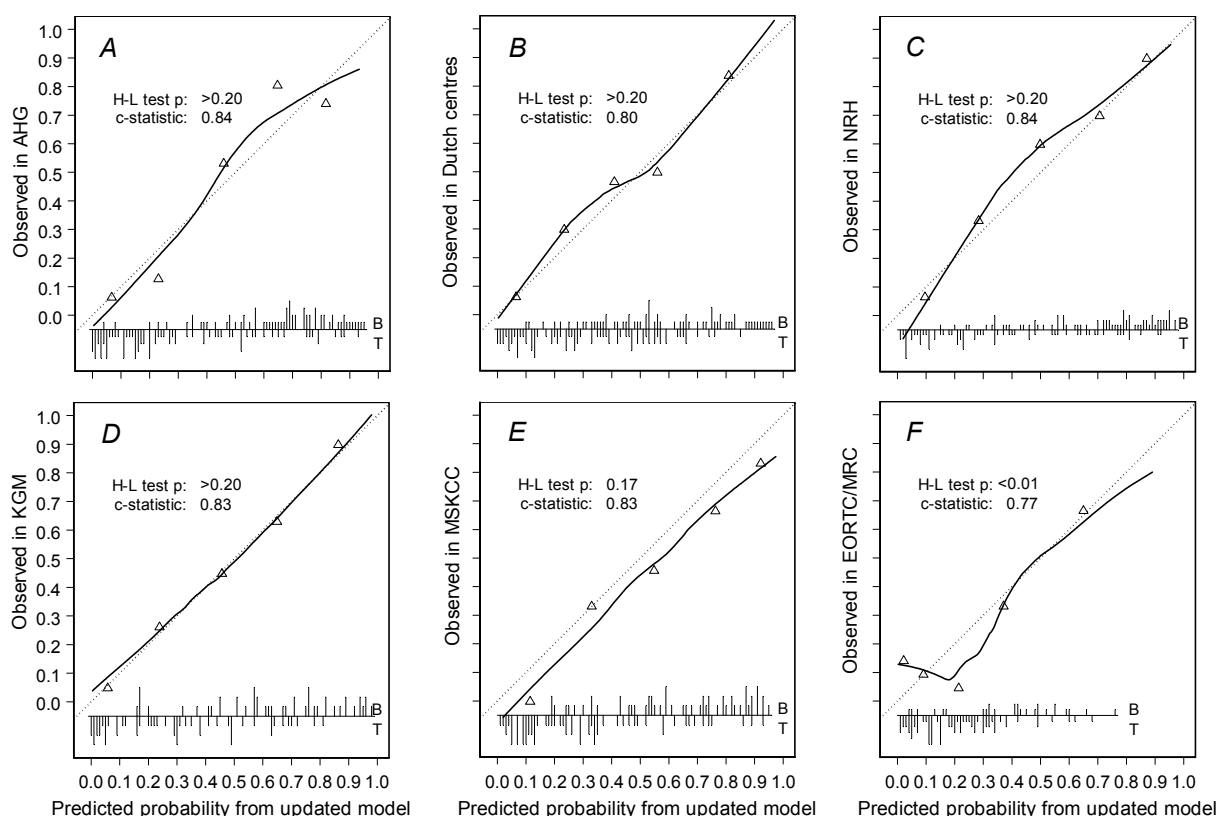


Figure 8.1 Performance of the updated model based on all available data ($n=1094$ patients) in six centres: AHG (A, $n=154$), Dutch centres (B, $n=152$), NRH (C, $n=150$), KGM (D, $n=98$), MSKCC (E, $n=120$), and patients from an EORTC/MRC trial (F, $n=105$). The intercept of the calibration line was tested whether it was statistically significant different from 0. Further, calibration was tested with the Hosmer-Lemeshow statistic, with small p -values indicating miscalibration. Discrimination was quantified by the c -statistic. We note that the calibration was adequate except in MSKCC (predictions too high with the intercept different from 0, $p=0.03$) and in the EORTC/MRC trial (miscalibration, indicated by the Hosmer-Lemeshow test, $p<0.01$). Discrimination was generally good with c -statistics higher than 0.8. B = benign tissue; T = residual tumour.

Discussion

In this paper we updated a prediction model for residual retroperitoneal mass histology in nonseminomatous testicular germ cell cancer. The regression coefficients were only slightly changed, but more precise. As a consequence, the score chart was also slightly different from the one presented before.⁸ The updated model performed well in most of the patient groups from a wide range of centres. This is an important finding, since we may now have more confidence in the transportability of the model to new patients.

We did not find variations between centres, when adjusted for case-mix as defined by the distributions of the predictors in the model. Indeed, we found adequate calibration in most centres. However, a systematic deviation in the predicted probabilities was found for the patients from MSKCC (intercept -0.48). On average, the predicted probabilities were too high. We previously found such a result in patients from IUMC, though not statistically significant.¹⁰ The systematic deviation implies that an important predictor was not included

Predictor	Value	Score	Patient					
Teratoma elements absent in primary tumour	yes	11						
Prechemotherapy AFP level	normal	11						
Prechemotherapy HCG level	normal	7						
Prechemotherapy LDH _{st} level (=LDH/normal value)								
	Value	0.5 0.7 1 1.5 2 3 4 5						
	Score	-6 -3 0 3 6 9 11 13						
Transversal diameter (mm)								
	Value	2 5 10 15 20 30 50 70 100						
	Score	5 3 0 -1 -3 -6 -10 -14 -18						
Change in mass size (=100*[presize-postsize]/presize)								
	Value	-50% -25% 0% 25% 50% 75% 100%						
	Score	-7 -4 0 4 7 11 14						
			----- +					
Total sumscore								
Total sumscore	0 5 10 15 20 23 26 29 32 35 38 40 45 50 55							
Probability (%)	11 18 26 37 49 56 63 70 76 81 85 88 92 95 97							

The precise formula is as follows:

$$\text{linear predictor} = -1.20 + 1.13(\text{teratoma-negative}) + 1.11(\text{AFPnormal}) + 0.72(\text{HCGnormal}) + 0.82(\ln[\text{LDH}_{\text{st}}]) - 0.27(\text{sqrt}[\text{postsize}]) + 0.14(\text{change})$$

with the variables teratoma-negative, AFPnormal, and HCGnormal 1 if true and 0 if false;

$\ln[\text{LDH}_{\text{st}}]$ the natural logarithm of LDH/upper limit of the normal value range;

$\text{sqrt}[\text{postsize}]$ the square root of postchemotherapy transverse diameter expressed in millimetres;

and change the change per 10% in mass size during chemotherapy: $((\text{presize}-\text{postsize})/\text{presize}) \times 10$.

The probability of benign tissue is then: $\text{probability} = 1/(1 + e^{-\text{linear predictor}})$.

Figure 8.2 Score chart for the probability of benign tissue at resection of residual masses in NSGCT patients with normal serum tumour markers AFP and HCG before resection. The chart was based on the updated model and enables the estimation of the probability for individual patients.

in the model. Perhaps the referral pattern of the US centres included in this study is different from the European centres. This may result in a different group of patients treated in the US centres. Adjusting the model intercept may be a solution to deal with the systematic deviation. The adjustment can also be achieved in the score chart by subtracting 5 points (i.e. $\text{intercept} = -0.48 * 10 = -4.8$, rounded 5).²⁶

The discrepancy between the statistically non-significant overall results for variation between centres and the statistically significant difference in the average outcome for the patients from MSKCC is related to a more philosophical question. Should a prediction model estimate good overall predictions, or should the model estimate particularly good predictions for individual centres or individual patients? Since a prediction model is used to predict individual patient outcomes, one may argue for the latter. However, the sample size usually limits efforts to estimate subgroup effects reliably. Therefore, only the overall effects are often studied. On the other hand, one might argue that each centre is unique. This would automatically imply that variation does exist between centres. We did may not have identified such variation because the differences between the centres were too small. The clinical relevance of the differences will then probably be small too.

The updated model may improve the current resection practices by supporting patient selection after chemotherapy for metastatic testicular germ cell cancer,. For instance, the Norwegian Radium Hospital (NRH) policy is to operate all patients except for those with masses smaller than 20 mm, teratoma-negative primary tumours, and normal AFP and HCG serum levels.⁴ This resection policy resulted in 31% of the operated NRH patients having predicted probabilities of benign tissue above 70%. Consequently, a number of the NRH patients with high probabilities for benign tissue might be selected not to undergo possible unnecessary resection at an acceptable risk of missing residual tumour.

Alternatively, patients with small residual masses, which are not resected under most of the current European resection policies may benefit from resection, when they have a teratoma-positive primary tumour, elevated serum tumour marker levels or a relatively small reduction in mass size after chemotherapy. For example, 84 of the 241 patients (35%) enrolled in the EORTC/MRC trial with unresected residual masses smaller than 10 mm had predicted probabilities of benign tissue lower than 70%.¹¹ These patients may likely benefit from resection.

In conclusion, the updated prediction model for residual mass histology may improve predictions for future patients from the same or from other centres. These predictions will be helpful in the selection of candidates for resection.

References

1. Fox EP, Weathers TD, Williams SD, et al: Outcome analysis for patients with persistent nonteratomatous germ cell tumor in postchemotherapy retroperitoneal lymph node dissections. *J Clin Oncol* 11:1294-9, 1993
2. Fosså SD, Aass N, Ous S, et al: Histology of tumor residuals following chemotherapy in patients with advanced nonseminomatous testicular cancer. *J Urol* 142:1239-42, 1989
3. Toner G, Panicek D, Heelan R, et al: Adjunctive surgery after chemotherapy for nonseminomatous germ cell tumors: recommendations for patient selection. *J Clin Oncol* 8:1683-1694, 1990
4. 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* 10:569-73, 1992
5. Gelderman WA, Schraffordt Koops H, Sleijfer DT, et al: Results of adjuvant surgery in patients with stage III and IV nonseminomatous testicular tumors after cisplatin-vinblastine-bleomycin chemotherapy. *J Surg Oncol* 38:227-32, 1988
6. Mulders P, Oosterhof G, 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
7. Donohue J, Rowland R, 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
8. Steyerberg EW, Keizer HJ, Fosså SD, et al: Prediction of residual retroperitoneal mass histology following chemotherapy for metastatic nonseminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. *J Clin Oncol* 13:1177-1187, 1995
9. 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
10. Vergouwe Y, Steyerberg EW, Foster RS, et al: Validation of a prediction model and its predictors for the histology of residual masses in nonseminomatous testicular cancer. *J Urol* 165:84-88, 2001
11. Vergouwe Y, Steyerberg EW, de Wit R, et al: External validity of a prediction rule for residual mass histology in testicular cancer: an evaluation for good prognosis patients. *Br J Cancer* 88:843-847, 2003
12. Gerl A, Clemm C, Schmeller N, et al: Advances in the management of metastatic non-seminomatous germ cell tumours during the cisplatin era: a single-institution experience. *Br J Cancer* 74:1280-5, 1996
13. de Wit R, Roberts J, Wilkinson P, et al: Final analysis demonstrating the equivalence of 3 BEP vs. 4 cycles and the 5 day schedule vs. 3 days per cycle in good prognosis germ cell cancer. An EORTC/MRC phase III study. *J Clin Oncol* 19:1629-1640, 2001
14. 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
15. Steyerberg EW, Marshall PB, Keizer HJ, et al: Resection of small, residual retroperitoneal masses after chemotherapy for nonseminomatous testicular cancer. A decision analysis. *Cancer* 85:1331-1341, 1999
16. Steyerberg EW, Keizer HJ, Habbema JD: Prediction models for the histology of residual masses after chemotherapy for metastatic testicular cancer. ReHiT Study Group. *Int J Cancer* 83:856-9, 1999
17. Steyerberg EW, Kievit J, de Mol Van Otterloo JC, et al: Perioperative mortality of elective abdominal aortic aneurysm surgery. A clinical prediction rule based on literature and individual patient data. *Arch Intern Med* 155:1998-2004, 1995
18. Steyerberg EW, Eijkemans MJ, Van Houwelingen JC, et al: Prognostic models based on literature and individual patient data in logistic regression analysis. *Stat Med* 19:141-60, 2000
19. Hardy R, Thompson S: A likelihood approach to meta-analysis with random effects. *Stat Med* 16:619-629, 1996
20. van Houwelingen HC: Validation, calibration, revision and combination of prognostic survival models. *Stat Med* 19:3401-3415, 2000
21. Cox DR: Two further applications of a model for binary regression. *Biometrika* 45:562-565, 1958
22. Harrell FE Jr, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 15:361-387, 1996
23. Miller ME, Hui SL: Validation techniques for logistic regression models. *Stat Med* 10:1213-1226, 1991
24. Lemeshow S, Hosmer DW: Applied logistic regression. New York, Wiley, 1989
25. Harrell FE Jr, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. *JAMA* 247:2543-2546, 1982
26. van Houwelingen HC, Thorogood J: Construction, validation and updating of a prognostic model for kidney graft survival. *Stat Med* 14:1999-2008, 1995

