Urologic Oncology: Seminars and Original Investigations 37 (2019) 809.e9–809.e18

UROLOGIC ONCOLOGY

Clinical-Testis cancer

External validation of 2 models to predict necrosis/fibrosis in postchemotherapy residual retroperitoneal masses of patients with advanced testicular cancer

Pia Paffenholz, M.D.[a,*], Tim Nestler, M.D.[a,*], Simon Hoier, M.D.[a], David Pfister[a], Martin Hellmich, M.D.[b], Axel Heidenreich, M.D., Ph.D.[a,c,**]

[a] *Department of Urology, University Hospital Cologne, Cologne, Germany*
[b] *Institute of Medical Statistics and Computational Biology, University of Cologne, Cologne, Germany*
[c] *Department of Urology, Medical University Vienna, Austria*

Received 13 January 2019; received in revised form 12 July 2019; accepted 25 July 2019

Abstract

Objectives: Nonseminomatous testicular germ cell tumors with residual retroperitoneal lesions >1 cm are treated with postchemotherapy retroperitoneal lymph node dissection (pcRPLND). However, up to 50% of patients are overtreated since the histology shows only residual necrosis/fibrosis. We aim to validate the 2 currently best performing prediction models (Vergouwe and Leao) for postchemotherapy residual mass histology.

Methods and materials: We performed a retrospective analysis including 402 patients who underwent a pcRPLND from 2008 to 2015. The study cohort was used to validate the 2 prediction models by Vergouwe and Leao using the published formulas and thresholds.

Results: Using our validation cohort, the Vergouwe model reached a significantly better area under the curve compared to the Leao model (0.760 (confidence interval 0.713–0.807) vs. 0.692 (0.640–0.744), $P = 0.002$) in the prediction of benign histology. At a threshold of >70% for the predicted probability of benign disease, the Leao model revealed that pcRPLND would be avoided in 10.2% of patients with benign disease with an error rate of 3.8% for viable tumor, while the Vergouwe model would avoid pcRPLND in 27.4% of all patients with benign disease with an error rate of 10.1% for viable tumor and 2.9% for teratoma. Adjusting the models to our data had no significant improvement. Limitations include the retrospective design.

Conclusions: The discriminatory accuracy of both models is not sufficient to safely select patients for surveillance strategy instead of pcRPLND. Therefore, further studies including new biomarkers are needed to optimize the accuracy of potential prediction models and to minimize pcRPLND overtreatment.

Keywords: Germ cell cancer; Platinum-based chemotherapy; Postchemotherapy lymph node dissection; Testicular cancer

Abbreviations: AFP, alpha-fetoprotein; AUC, area under curve; CT, computed tomography; hCG, human chorionic gonadotropin; IGCCCG, international germ cell cancer collaborative group; LDH, lactate dehydrogenase; pcRPLND, postchemotherapy retroperitoneal lymph node dissection; TGCT, testicular germ cell tumor

1. Introduction

Metastatic nonseminomatous testicular germ cell tumors (TGCT) with residual retroperitoneal lesions >1 cm should be treated with postchemotherapy retroperitoneal lymph node dissection (pcRPLND) according to current guideline recommendations, while the treatment of residual masses <1 cm is still under debate [1]. However, the histopathological examination of pcRPLND specimens shows viable tumor in only 10% to 15% and teratoma in 40% of all cases, while necrosis/fibrosis is found in 40% to 50% [1–3]. Furthermore, pcRPLND can be a very complex procedure in individual cases displaying a high risk of severe morbidity [4–6]. Thus, up to 50% of patients with necrosis/fibrosis

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

*Authors contributed equally to this study.

**Corresponding author. Tel.: +49 221 478 82108; fax: 004922147882372.
E-mail address: axel.heidenreich@uk-koeln.de (A. Heidenreich).

https://doi.org/10.1016/j.urolonc.2019.07.021
1078-1439/

are potentially overtreated and might unnecessarily suffer from treatment related health problems. The first step to reduce overtreatment was to perform pcRPLND only in patients with residual masses >1 cm, accepting that up to 10% of masses <1 cm experience a progression due to undetected teratoma or viable tumor [7]. However, predicting histology of the residual masses might be the most promising way to identify patients who benefit most from pcRPLND. Therefore, models for the prediction of histology have been proposed based on different clinical parameters [8−15]. Although the most promising models reached a satisfying accuracy (area under the curve (AUC) 0.8), they have not been adopted into clinical practice yet [12−14].

The aim of this study was to validate these models on a large, contemporary cohort of patients and to evaluate if pcRPLND can thus be avoided in a subgroup of patients.

2. Materials and methods

2.1. Study population

Patients treated with pcRPLND at the University Hospitals of Cologne and Aachen were identified in the hospital databases and retrospectively analyzed in a multicenter observational cohort study. We found 619 patients who received a pcRPLND after being treated with a cisplatin-based chemotherapy between 2008 and 2015. According to the required variables of each model, we excluded 217 patients resulting in a cohort of 402 patients who were eligible for both models (exclusion criteria are displayed in Supplementary Fig. 1). Serum tumor markers alpha-fetoprotein, beta human chorionic gonadotropin, and lactate dehydrogenase were measured before chemotherapy and prior to pcRPLND. The largest axial diameter (mm) of retroperitoneal masses before and after chemotherapy treatment was determined using computed tomography images, analyzed by an experienced uroradiologist. Resected masses containing necrosis and/or fibrosis were classified as "benign" disease, those containing mature or immature teratoma were classified as teratoma, and if any germ cell elements were detected they were categorized as viable tumor. All orchiectomy and pcRPLND specimens were evaluated by an experienced uropathologist. We also reviewed further patient and tumor characteristics (Tables 1 and 2). The study complies with the Declaration of Helsinki, local ethics committee approval was obtained (University Hospital of Cologne 17−175).

2.2. Prediction models

The prediction model of Vergouwe et al. [12] is an update of the model previously described by Steyerberg et al. [3] and contains 6 variables (Supplementary Table 1). Leao et al. proposed a prediction model containing 4 different variables (Supplementary Table 1) [13]. The probability for benign histology of both models was calculated according to the published formula by Vergouwe et al. [12] whereas the formula proposed by Leao et al. was given to the author upon request.

2.3. Statistical analysis

Continuous variables are summarized as median (interquartile range: 25th−75th percentile or minimum − maximum as given in the published data by Leao et al. [13]), categorical variables as *n* (%). For group comparisons, we used either *t* test or chi-square test (standard deviations of literature value had to be approximated based on ranges [min.−max.]. Model equations were applied as mentioned in Supplementary Table 1. To adjust our data to the 2 tested models, adjusted standardized residuals were used to determine significant intragroup variables (Supplementary Tables 2 and 3) and afterward the mentioned model equations were used (Supplementary Table 1). We calculated thresholds of a predicted probability of benign disease, including the threshold of 70% as proposed by Leao and Vergouwe [12,13]. These thresholds were used to identify the maximal percentage of patients that would have correctly avoided pcRPLND due to benign disease and the minimal percentage of those that would have had inappropriately deferred to pcRPLND due to the presence of teratoma or viable tumor. Receiver-operator curves were plotted (sensitivity against 1-specificity) and the area under the curve (AUC) was calculated. We determined the optimal cut point (calculating the product of sensitivity and specificity) and used bootstrapping (1,000 replications) to estimate its variability. A *P*-value of <0.05 was considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 24.0 (Armonk, NY) and StataCorp LLC Stata Statistical Software: Release 15 (College Station, TX).

3. Results

3.1. Patient characteristics

First, we compared the clinical and pathohistological characteristics of Vergouwe's and Leao's study to our patient cohort (Tables 1 and 2).

All patients of our cohort were treated after 1990, while 68.8% of the patients in the Vergouwe model and 10.9% in the Leao model were treated before 1990 ($P < 0.001$).

In contrast to Leao et al., who included all clinical stages, we only analyzed patients with clinical stage II/III disease ($P < 0.001$). Furthermore, we showed significantly more intermediate and poor prognosis patients in our cohort compared to the Leao cohort (53.0% vs. 34.8%, $P < 0.001$). Subgroup analysis of the IGCCCG prognosis group did not reveal any differences in the prediction of benign histology in the Vergouwe and Leao model, especially if comparing

Table 1
Characteristics of the variables used in the Vergouwe model ($n = 1094$) and the Leao model ($n = 184$) as a comparison with our validation cohort ($n = 402$) [12,13].

Parameter	Vergouwe model	Validation cohort for the Vergouwe model	*P* value	Leao model	Validation cohort for the Leao model	*P* value
Histology of orchiectomy specimens, *n* (%)						<0.001
Teratoma	Teratoma elements present 591 (54.0)	Teratoma elements present 146 (36.3)	<0.001	9 (4.9)	58 (14.4)	
Embryonal	–			51 (27.7)	88 (21.9)	
Yolk sac tumor	–			2 (1.1)	13 (3.2)	
Choriocarcinoma	–			2 (1.1)	87 (21.6)	
Mixed	–			120 (65.2)	156 (38.8)	
With teratoma	–			99 (82.5)	88 (56.4)	
Without teratoma	–			21 (17.5)	68 (44.6)	
Histology of pcRPLND specimens, *n* (%)			0.008			<0.001
Benign disease	425 (38.8)	186 (46.3)		44 (24.0)	186 (46.3)	
Teratoma	673 (61.2)	137 (34.1)		123 (66.8)	137 (34.1)	
Viable tumor		79 (19.7)		17 (9.2)	79 (19.7)	
Residual masses before chemotherapy, *n* (%)						
0–19 mm	–	26 (6.5)		–	26 (6.5)	
20–49 mm	–	127 (31.6)		–	127 (31.6)	
50–99 mm	–	162 (40.3)		–	162 (40.3)	
≥100 mm	–	87 (21.6)		–	87 (21.6)	
Residual masses before pcRPLND, *n* (%)			0.026			0.007
0–19 mm	344 (31.4)	106 (26.4)		54 (29.3)	106 (26.4)	
20–49 mm	399 (36.5)	155 (38.6)		88 (47.8)	155 (38.6)	
50–99 mm	215 (19.7)	102 (25.4)		42 (22.8)	141 (35.1)	
≥100 mm	136 (12.4)	39 (9.7)		–	–	
Residual masses before pcRPLND </> 10 mm, *n* (%)			–			–
0–10 mm	–	41 (10.2)		–	41 (10.2)	
>10 mm	–	361 (89.2)		–	361 (89.2)	
Serum tumor markers prechemotherapy, median (IQR)			–			
AFP (kU/l; standard value <5.8 kU/l)	–			61.0 (8–354.5)	225 (21–2000)	<0.001
ß-hCG (U/l; standard value <2.6 U/l)	–			69 (2–1,020)	83 (5–806)	0.160
LDH (U/l; standard value <250 U/l)	–			249 (182–388)	410 (200–786)	<0.001
Serum tumor markers prechemotherapy, *n* (%)					–	
AFP elevated	755 (69.0)	334 (83.1)	<0.001	–		
ß-hCG elevated	716 (65.4)	347 (86.3)	<0.001	–		
LDH elevated 1–2× normal	303 (36.9)	91 (22.6)	0.115	–		
LDH elevated >2× normal	258 (31.4)	166 (41.3)		–		

AFP = alphafFetoprotein; hCG = human chorionic gonadotropin; LDH = lactate dehydrogenase; pcRPLND = postchemotherapy retroperitoneal lymph node dissection.
Note: Continuous variables are presented as median (IQR or min.–max.), categorical variables are given as *n* (%).

Table 2
Patient characteristics of the Vergouwe model ($n = 1094$), the Leao model ($n = 184$) and our validation cohort ($n = 402$) [12,13].

Parameter	Vergouwe model	Validation cohort for the Vergouwe model	*P* value	Leao model	Validation cohort for the Leao model	*P* value
Number of patients	1,094	402		184	402	
Period of treatment	1977−1999	2008−2015		1978−2016	2008−2015	
Date of pcRPLND, *n* (%)			<0.001			<0.001
Before 1990	544 (68.8)	0		20 (10.9)	0	
After 1990	235 (30.2)	402 (100.0)		164 (89.1)	402 (100.0)	
Clinical stage (CS), *n* (%)			–			<0.001
CS I (with relapse)	–			25 (13.6)	0	
CS II	–			109 (59.2)	321 (79.9)	
CS III				50 (27.2)	81 (20.1)	
IGCCCG prognosis group, *n* (%)			–			<0.001
Good	–			120 (65.2)	189 (47.0)	
Intermediate	–			43 (23.4)	116 (28.9)	
Poor	–			21 (11.4)	97 (24.1)	
Age at pcRPLND (y), median (min.−max.)	–		–	26 (15−61)	27 (14−57)	0.320
Serum tumor markers postchemotherapy, median (IQR)			–			
AFP (ng/ml = μg/l; standard value <5.8 ng/ml)	–			5 (2−5)	4.0 (2.4−5)	<0.001
ß-hCG (U/l; standard value < 2.6 U/l)	–			1 (1−2)	1.0 (0.1-1.5)	1.000
LDH (U/l; standard value < 250 U/l)	–			196 (167−225)	183.5 (171−266)	<0.001
Type of chemotherapy, *n* (%)						
BEP	–			137 (74.4)	183 (45.5)	
EP	–			6 (3.3)	59 (14.7)	
TIP	–			1 (0.5)	21 (5.2)	
PEI	–			–	37 (9.2)	
VABP/VBP	–			28 (15.2)	–	
VIP	–			2 (1.1)	–	
Unknown	–			10 (5.4)	102 (25.4)	
Surgery time of pcRPLND (min), median (IQR)	–	200 (150−300)		–	200 (150−300)	
Blood loss of pcRPLND (ml), median (IQR)	–	400 (220−975)		–	400 (220−975)	
Length of stay in hospital (days), median (IQR)	–	11 (7.3−14)		–	11 (7.3−14)	

AFP = alpha-fetoprotein; BEP = cisplatin/etoposide/bleomycin; CS = clinical stage; EP = cisplatin/etoposide; hCG = human chorionic gonadotropin; IGCCCG = international germ cell cancer collaborative group; LDH = lactate dehydrogenase; pcRPLND = postchemotherapy retroperitoneal lymph node dissection; PEI = cisplatin/etopside/ifosfamide; TIP = cisplatin/ifosfamide/paclitaxel; VABP = vinblastine/actinomycin/bleomycin/cisplatin, VBP = vinblastine/bleomycin/cisplatin; VIP = vinblastin/ifosfamide/cisplatin.
Note: Continuous variables are presented as median (IQR or min.−max.), categorical variables are given as *n* (%).

good prognosis group vs. intermediate and poor prognosis group (Table 3).

Regarding the pathohistological analysis of the pcRPLND specimens, we had significantly less patients with teratoma compared to the Leao cohort (34.1% vs. 66.8%), while benign disease (46.3% vs. 24.0%) and viable tumor (19.7% vs. 9.2%) were significantly higher in our cohort ($P < 0.001$; Table 2). Compared to the Vergouwe cohort, histological analysis of the pcRPLND specimens demonstrated significantly more patients with benign disease (46.3% vs. 38.8%) in our validation cohort, while we had significantly less patients with teratoma or viable tumor (53.8% vs. 61.2%). Furthermore, the size of residual tumors prior to pcRPLND revealed significantly more patients with small residual masses <19 mm in the Vergouwe cohort compared to our cohort (31.4% vs. 26.4%), and significantly more patients with residual tumors with the size of 50 to 99 mm in our cohort compared to the Vergouwe (19.7% vs. 25.1%) and Leao (22.8% vs. 35.1%) cohort.

3.2. Validation of the prediction models

We next validated the 2 prediction models with our cohort, as displayed in Figs. 1−3. The discriminatory accuracy for the prediction of benign disease was significantly better for the Vergouwe model compared to the Leao model (AUC 0.760 {CI 0.713−0.807] vs. 0.692 [CI 0.640−0.744], $P = 0.002$; Figs. 1 and 3). After adjusting both models to our cohort of patients, the AUC was impaired in Vergouwe's model while it was improved by Leao's model.

Table 3
Subgroup analysis regarding the IGCCCG prognosis group and the size of the residual masses ≤1 cm and >1 cm of the Vergouwe model, the Leao model, and both models adjusted to our cohort of patients in the prediction of benign histology, displayed as AUC of the receiver-operator curves ($n = 402$).

	All patients	IGCCCG 1	IGCCCG 2	IGCCCG 3	IGCCCG 1 + 2	IGCCCG 3	IGCCCG 1	IGCCCG 2 + 3	≤1 cm	>1 cm
Vergouwe model	0.760	0.760	0.787	0.727	0.770	0.727	0.760	0.759	0.580	0.738
Vergouwe model adjusted	0.754	0.741	0.775	0.744	0.753	0.744	0.741	0.763	0.534	0.734
Leao model	0.692	0.705	0.722	0.639	0.711	0.639	0.705	0.683	0.61	0.676
Leao model adjusted	0.718	0.733	0.719	0.686	0.726	0.686	0.733	0.705	0.731	0.692

Note: IGCCCG = international germ cell cancer collaborative group; IGCCCG 1 = good prognosis; IGCCCG 2 = intermediate prognosis; IGCCCG 3 = poor prognosis.

However, there was still a tendency toward a better AUC for the Vergouwe model compared to the Leao model (AUC 0.754 [CI 0.706–0.801] vs. 0.718 [CI 0.668–0.768], $P = 0.091$; Figs. 2 and 3). At the optimal cut point, the Vergouwe model reached a specificity of 71% and a sensitivity of 69% to predict benign disease in our cohort compared to the Leao model, which showed a specificity of 75% and a sensitivity of 62%.

Furthermore, we performed a subgroup analysis to predict benign disease in residual masses ≤1 cm ($n = 41$) and >1 cm ($n = 361$). Residual masses ≤1 cm showed benign disease in 80.5%, teratoma in 14.6%, and viable tumor in 4.9%. The predictive accuracy of both models was impaired in both subgroups (Table 3).

To evaluate the accuracy of both models, we analyzed different thresholds for a predicted probability of benign disease including the threshold of >70%, which was proposed by Vergouwe and Leao. These thresholds identify the percentage of the cohort which would appropriately avoid pcRPLND due to benign histology and the percentage of patients that would have had deferred pcRPLND inappropriately due to the presence of teratoma or viable tumor (Tables 4a and 4b). Using the Vergouwe model, our validation cohort showed that at a threshold of >70% pcRPLND would be avoided in 27.4% of patients with benign disease, but 10.1% of patients harbored viable tumor and 2.9% of patients with teratoma would have been deferred inappropriately (Table 4a). Using the Leao model, our validation cohort showed that at a threshold of >70% pcRPLND would have been avoided in 10.2% of patients with benign disease, but also 3.8% of patients with viable tumor would have been deferred inappropriately (Table 4b). We furthermore identified the thresholds for a predicted probability of benign disease, which show optimal results, namely an error rate of 0% (Tables 4a and 4b). When using the Vergouwe model, our validation cohort showed that no patient would be deferred inappropriately at a threshold of >90% avoiding 2.2% of all pcRPLND with benign results (Table 4a).

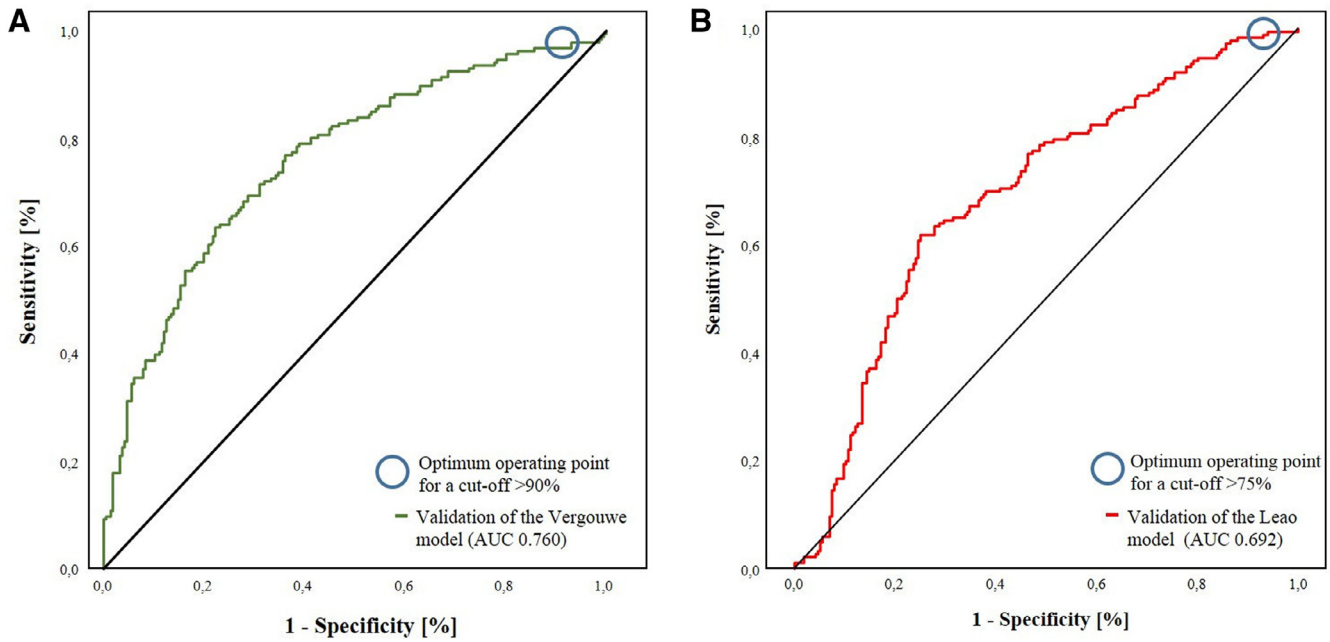

Fig. 1. Receiver-operator curve for benign histology for the (A) Vergouwe model and the (B) Leao model using our validation cohort ($n = 402$). The corresponding optimal cut points in the ROC curves were located at 0.33 (CI 0.21–0.44) for the Vergouwe model and at 0.28 (CI 0.22–0.34) for the Leao model.

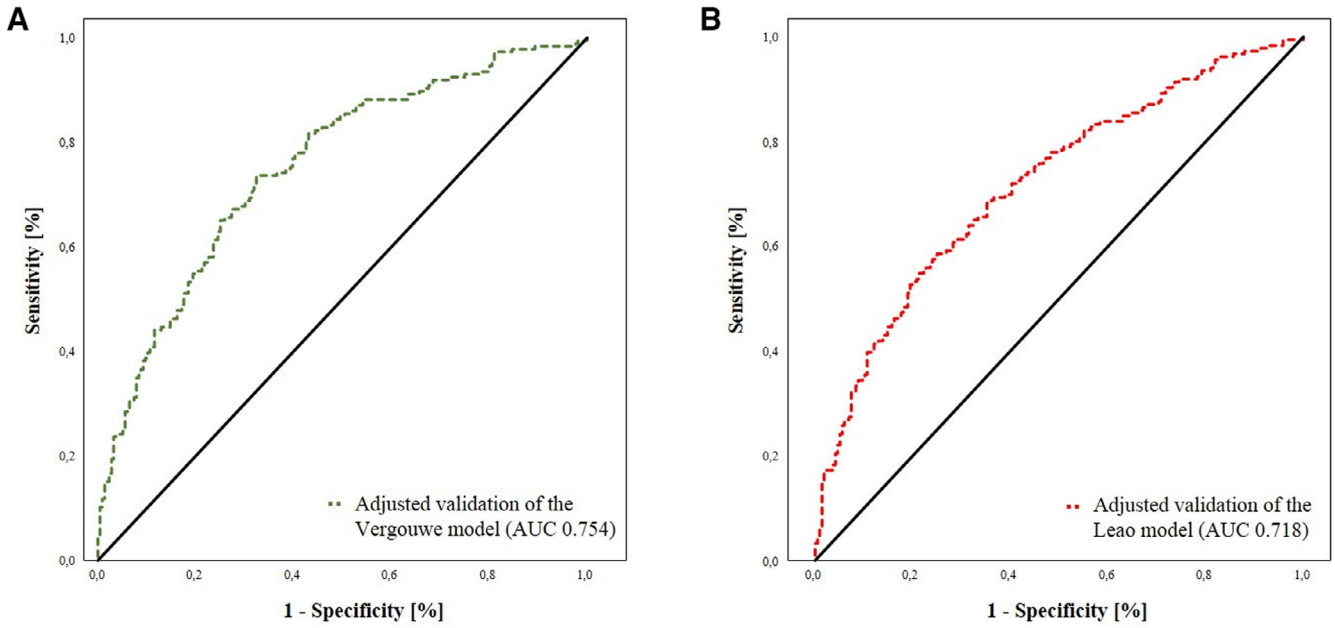

Fig. 2. Receiver-operator curve for benign histology for the (A) Vergouwe model and the (B) Leao model adjusted to our validation cohort ($n = 402$).

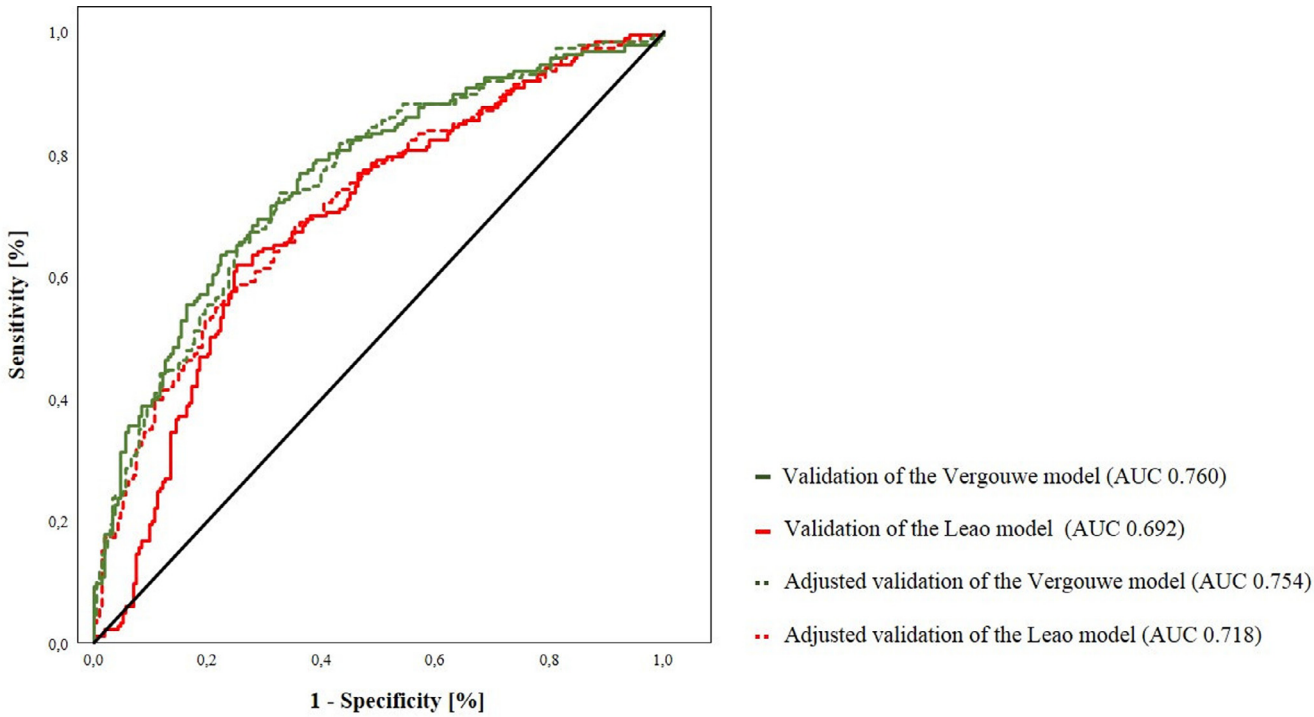

Fig. 3. Receiver-operator curve for benign histology as a comparison for the Vergouwe model and the Leao model using our validation cohort (continuous line) and after adjusting our data to each model (dashed line).

The Leao model revealed an error rate of 0% at a threshold of >75% in our cohort, avoiding pcRPLND in 2.7% of all patients with benign disease (Table 4b).

As a last step, we analyzed the impact of teratoma in the orchiectomy specimen and shrinkage of the retroperitoneal masses after chemotherapy in order to identify a subgroup of patients which might be observed and thus avoid pcRPLND using the prediction models of Leao and Vergouwe (Supplementary Table 4). In the subgroup of patients with a high risk of viable tumor in the pcPRLND specimens, namely patients with teratoma in the orchiectomy specimen and no shrinkage of the retroperitoneal masses after chemotherapy ($n = 27/402$; 7%), histopathological evaluation of the pcRPLND specimen revealed viable

Table 4a
Different thresholds to consider the optimal cohort that maximizes appropriately avoided pcRPLND and includes a minimum of RPLND which would be deferred inappropriately according to the Vergouwe model [12].

Predicted probability of benign disease cut-off	Benign disease *n* (%)	Teratoma *n* (%)	Viable tumor *n* (%)	Avoiding pcRPLND appropriately (% of benign disease)	Avoiding pcRPLND inappropriately (% of all teratoma)	Avoiding pcRPLND inappropriately (% of all viable tumor)
>90%	4 (100)	0 (0)	0 (0)	2.2	0	0
>85%	15 (93.8)	0 (0)	1 (6.3)	8.1	0	1.3
>80%	26 (98.7)	2 (6.9)	1 (3.4)	14.0	1.5	1.3
>75%	41 (85.4)	3 (6.3)	4 (8.3)	22.0	2.2	5.1
>70%	51 (81.0)	4 (6.3)	8 (12.7)	27.4	2.9	10.1
>65%	68 (78.2)	9 (10.3)	10 (11.5)	36.6	6.6	12.7
>60%	82 (74.5)	16 (14.5)	12 (10.9)	44.1	11.7	15.2
>50%	121 (68.4)	32 (18.1)	24 (13.6)	65.1	23.4	30.4

Note: pcRPLND = postchemotherapy retroperitoneal lymph node dissection.

Table 4b
Different thresholds to consider the optimal cohort that maximizes appropriately avoided pcRPLND and includes a minimum of pcRPLND which would be deferred inappropriately according to the Leao model [13].

Predicted probability of benign disease cut-off	Benign disease *n* (%)	Teratoma *n* (%)	Viable tumor *n* (%)	Avoiding pcRPLND appropriately (% of benign disease)	Avoiding pcRPLND inappropriately (% of all teratoma)	Avoiding pcRPLND inappropriately (% of all viable tumor)
>90%	1 (100)	0 (0)	0 (0)	0.5	0	0
>85%	1 (100)	0 (0)	0 (0)	0.5	0	0
>80%	1 (100)	0 (0)	0 (0)	0.5	0	0
>75%	5 (100)	0 (0)	0 (0)	2.7	0	0
>70%	19 (86.4)	0 (0)	3 (13.6)	10.2	0	3.8
>65%	34 (79.1)	1 (2.3)	8 (18.6)	18.3	0.7	10.1
>60%	71 (75.5)	9 (9.6)	14 (14.9)	38.2	6.6	17.7

Note: pcRPLND = postchemotherapy retroperitoneal lymph node dissection.

tumor in 19/27 (70%) of all cases, but benign disease in 8/27 (30%) of all cases. The corresponding AUC for the prediction of histology were 0.636 (CI 0.061−0.885) for the Vergouwe model and 0.473 (CI 0.281−0.992) for the Leao model. For the remaining patients ($n = 375/402$; 93%), we calculated an AUC of 0.403 (CI 0.332−0.474) for the Vergouwe model and 0.487 (CI 0.412−0.562) for the Leao model. Next, we analyzed the subgroup of patients with a low risk of viable tumor in the pcRPLND specimens, namely the subgroup of patients with no teratoma and shrinkage of retroperitoneal masses after chemotherapy ($n = 186/402$; 46%) and revealed benign disease in the pcRPLND specimen in 118/186 (63%) of all cases, but viable tumor in 68/186 (37%) of all patients. The AUC was 0.326 (CI 0.230−0.422) for the Vergouwe model and 0.559 (CI 0.328−0.564) for the Leao model. For the remaining patients ($n = 216$) we calculated an AUC of 0.463 (CI 0.366−0.561) for the Vergouwe model and 0.559 (CI 0.462−0.656) for the Leao model.

4. Discussion

Residual retroperitoneal masses >1 cm after cisplatin-based chemotherapy of metastasized nonseminomatous TGCT are frequently managed by pcRPLND. However, a relevant treatment-related morbidity can be associated with this procedure, which is unnecessary in up to 50% of cases with only benign disease in the pathological findings [1]. Therefore, different models were proposed to predict benign histology of residual masses since these patients would not benefit from pcRPLND and therefore could undergo a frequent follow-up [3,12−15]. To our knowledge, only the Vergouwe model was externally validated with a small cohort of 52 patients and revealed a sensitivity of 33% at a specificity of 100% in the prediction of benign disease [16]. Thus, our study is the largest external validation study, and the first which compares 2 of the currently best performing prediction models by Vergouwe et al. and Leao et al. to predict benign histology with a contemporary patient cohort [12,13].

Although several clinical prediction models have been published, they have not yet been adopted into clinical practice [3,12−15,17]. According to our study, discriminatory accuracy as well as sensitivity and specificity of the Vergouwe and Leao model in the prediction of benign histology were not sufficient to safely select patients for follow-up instead of pcRPLND. Remarkably, both models did not perform better for good prognosis patients. For using a prediction model in clinical routine, it is required that almost

no patients with viable tumor will be scheduled to surveillance inappropriately. Leao et al. showed in their study that a threshold for a predicted probability of benign disease of >70% would fulfill these criteria, as 11% of all patients with benign disease could avoid pcRPLND with an error rate of <1% [13]. However, the Leao model performed significantly worse using our validation cohort: At the recommended threshold of 70%, 10.2% of all patients with benign histology would be spared from pcRPLND, but the error rate was 3.8%. Using the proposed threshold of >70% for the Vergouwe model, our validation cohort showed that pcRPLND would be avoided in 27.4% of patients with benign residuals, but the error rate would be unacceptable with 10.1% for viable tumor and 2.9% for teratoma. Taken together, the validated prediction models did not reveal convincing results in the prediction of benign histology in pcRPLND, especially not for the thresholds which were reported with the models.

Most patients of the Vergouwe and even several patients of the Leao study were treated before 1990 in contrast to our patients which were treated starting from 2008. Although bleomycin, etoposide, cisplatin chemotherapy has been the gold standard for treatment of TGCT, treatment options have changed during the last decades, favoring less intensive treatment options. Secondly, pcRPLND was performed in all residual masses, independent of its size according to prior guideline recommendations. Furthermore, the recommended treatment did not differ between marker positive and marker negative nonseminoma in clinical stage IIA–B before 2005 and nowadays also surveillance of marker-negative masses in clinical stage IIA can be performed [1]. Consequently, these therapy modifications might have resulted in different pathohistological findings of the pcRPLND specimen in the different time period and the proposed models might not unhesitatingly be transferable to more recent collectives like ours [1]. To stress the alienability of the models, we adjusted both models to our data. However, the discriminatory accuracy of both models did not improve substantially. In line with our finding, a study by Albers et al. with a more-modern series of patients did not succeed in building a clinically relevant prediction model [15].

Residual tumor resection is mandatory in all nonseminoma patients with residual masses >1 cm after chemotherapy according to the current guidelines [1]. This is underlined by the fact that in our cohort 20.2% of all patients with residual masses >1 cm showed viable tumor and 34.1% had teratoma. However, the validated models did not reach reliable results in the prediction of benign histology in this subgroup of patients. As overtreatment and thus potential unnecessary surgery-related health problems in patients undergoing pcRPLND is the main issue, another approach to reduce therapy related morbidity was the introduction of modified template pcRPLND for selected patients revealing excellent recurrence free survival [18,19]. Furthermore, the Vergouwe cohort included a significantly higher number of smaller lesions. One can assume that the subgroup analysis with residual masses ≤1 cm might perform better because it is known that up to 70% harbor fibronecrotic tissue [1,2]. However, the discriminatory accuracy of both models to predict benign histology in small lesions <1 cm was considerably worse in our study.

Our analysis furthermore revealed that the patients included in both models differed significantly regarding clinical and pathohistological parameters. In detail, the Leao study included a significantly higher proportion of patients with teratoma in the orchiectomy specimen and consequently in the pcRPLND specimen (67.8%) and thus a lower proportion of patients with benign disease compared to our data and those of other studies [1,2,4–6,15,20–22]. These prior studies revealed teratoma in only 40%, viable tumor in 10% to 15%, and benign disease in 40% to 50% in the histopathological examination of pcRPLND specimens of all patients [2,4–6,15,20–22]. Given the fact that Leao's cohort of patients did not represent this distribution of pathohistological results expected to be found in pcRPLND, especially regarding the percentage of teratomas, the prediction model of Leao should be used with caution in clinical practice. Although our cohort of patients included significantly more intermediate and poor risk patients, we showed a higher percentage of benign disease in the pcRPLND specimens compared to the Leao study. This might reflect the fact that Leao also included patients with a relapse after clinical stage I. It might furthermore be due to different effects of chemotherapy including numbers of cycles or treatment delay due to toxicity. In addition, Leao's model included less clinical parameters compared to the Vergouwe model (4 vs. 6 variables), and its discriminatory accuracy and sensitivity was worse in our independent validation cohort, as expected if variables are reduced [12,13]. Furthermore, we analyzed the impact of teratoma in the orchiectomy specimen and shrinkage of the retroperitoneal masses after chemotherapy in order to identify a subgroup of patients which might be observed and thus avoid pcRPLND using the prediction models of Leao and Vergouwe. However, none of these subgroup analyses led to an improvement of the prediction of the models' performance.

The performance of prediction models could be improved by combining clinical data with further parameters that have been shown to predict histology as miRNA and radiomics. For serum miRNA-371a-3p, a prior study demonstrated the potential to predict viable tumor after chemotherapy prior pcRPLND with a good AUC of 0.874 (CI 0.774–0.974) [23]. However, miRNA-371a-3p is not expressed by teratoma. Therefore, this miRNA could not be used as the only parameter for scheduling patients for pcRPLND. Nevertheless, it was lately shown that another miRNA, the miRNA-375, is highly expressed in teratoma [24]. Furthermore, our research group indicated the potential of radiomics on computed tomography to predict histology in retroperitoneal lymph node metastasis prior pcRPLND, as they achieved a

classification accuracy of 0.81 with a sensitivity of 88%, a specificity of 72%, and a positive predictive value of 78% in an independent validation cohort [25]. Combining miRNAs and radiomics, resulting in a "multiomics" approach, might predict histology even more reliably. However, prospective studies are required to confirm the clinical usefulness of both parameters and its combination.

Although our study is the largest external validation of the currently best performing models to predict benign histology with a contemporary patient cohort, the limitation of this study is its retrospective design and incomplete data of several patients who had to be excluded for this validation, which may even overestimate the performance of the tested models.

5. Conclusions

According to our analysis, none of the prediction models showed promising results as benign histology could not be predicted with sufficient accuracy to safely avoid pcRPLND. Therefore, pcRPLND remains the standard approach in patients with residual masses >1 cm in axial diameter of metastatic nonseminomatous TGCT. Consequently, there is an urgent need for parameters that assist in patient stratification for pcRPLND.

Supplementary materials

Supplementary material associated with this article can be found in the online version at https://doi.org/10.1016/j.urolonc.2019.07.021.

References

[1] Albers P, Albrecht W, Algaba F, Bokemeyer C, Cohn-Cedermark G, Fizazi K, et al. Guidelines on testicular cancer: 2015 update. Eur Urol 2015;68:1054–68. https://doi.org/10.1016/j.eururo.2015.07.044.

[2] Carver BS, Shayegan B, Serio A, Motzer RJ, Bosl GJ, Sheinfeld J. Long-term clinical outcome after postchemotherapy retroperitoneal lymph node dissection in men with residual teratoma. J Clin Oncol 2007;25:1033–7. https://doi.org/10.1200/JCO.2005.05.4791.

[3] Steyerberg EW, Keizer HJ, Fosså SD, Sleijfer DT, Toner GC, Schraffordt Koops H, et al. Prediction of residual retroperitoneal mass histology after chemotherapy for metastatic nonseminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. J Clin Oncol 1995;13:1177–87. https://doi.org/10.1200/JCO.1995.13.5.1177.

[4] Heidenreich A, Haidl F, Paffenholz P, Pape C, Neumann U, Pfister D. Surgical management of complex residual masses following systemic chemotherapy for metastatic testicular germ cell tumours. Ann Oncol 2017;28. https://doi.org/10.1093/annonc/mdw605.

[5] Winter C, Pfister D, Busch J, Bingöl C, Ranft U, Schrader M, et al. Residual tumor size and IGCCCG risk classification predict additional vascular procedures in patients with germ cell tumors and residual tumor resection: a multicenter analysis of the German Testicular Cancer Study Group. Eur Urol 2012;61:403–9. https://doi.org/10.1016/j.eururo.2011.10.045.

[6] Nini A, Konieczny M, Winter C, Lusch A, Krauspe R, Albers P. Surgical management and outcomes of patients with bone metastases in germ cell tumors: a case series. Urol Oncol 2018;36:82.e1–5. https://doi.org/10.1016/j.urolonc.2017.10.016.

[7] Kollmannsberger C, Daneshmand S, So A, Chi KN, Murray N, Moore C, et al. Management of disseminated nonseminomatous germ cell tumors with risk-based chemotherapy followed by response-guided postchemotherapy surgery. J Clin Oncol 2010;28:537–42. https://doi.org/10.1200/JCO.2009.23.0755.

[8] Mao Y, Hedgire S, Prapruttam D, Harisinghani M. Predictive models for lymph node metastases in patients with testicular germ cell tumors. Abdom Imaging 2015;40:3196–205. https://doi.org/10.1007/s00261-015-0526-5.

[9] Onozawa M, Kawai K, Yamamoto T, Hinotsu S, Tsukamoto S, Hattori K, et al. Clinical parameters that predict histology of postchemotherapy retroperitoneal lymph node mass in testicular cancer. Int J Urol 2004;11:535–41. https://doi.org/10.1111/j.1442-2042.2004.00832.x.

[10] Li XD, Guo SJ, Chen SL, Liu ZF, Dong P, Zhang ZL, et al. Clinical outcome of postchemotherapy retroperitoneal lymph node dissection and predicting retroperitoneal histology in advanced nonseminomatous germ cell tumours of the testis. Zhonghua Wai Ke Za Zhi 2017;55:603–7.

[11] Spiess PE, Brown GA, Pisters LL, Liu P, Tu SM, Evans JG, et al. Viable malignant germ cell tumor in the postchemotherapy retroperitoneal lymph node dissection specimen: can it be predicted using clinical parameters? Cancer 2006;107:1503–10. https://doi.org/10.1002/cncr.22181.

[12] Vergouwe Y, Steyerberg EW, Foster RS, Sleijfer DT, Fosså SD, Gerl A, et al. Predicting retroperitoneal histology in postchemotherapy testicular germ cell cancer: a model update and multicentre validation with more than 1000 patients. Eur Urol 2007;51:424–32. https://doi.org/10.1016/j.eururo.2006.06.047.

[13] Leão R, Nayan M, Punjani N, Jewett MAS, Fadaak K, Garisto J, et al. A new model to predict benign histology in residual retroperitoneal masses after chemotherapy in nonseminoma. Eur Urol Focus 2018. https://doi.org/10.1016/j.euf.2018.01.015.

[14] Steyerberg EW, Gerl A, Fossá SD, Sleijfer DT, de Wit R, Kirkels WJ, et al. Validity of predictions of residual retroperitoneal mass histology in nonseminomatous testicular cancer. J Clin Oncol 1998;16:269–74. https://doi.org/10.1200/JCO.1998.16.1.269.

[15] Albers P, Weissbach L, Krege S, Kliesch S, Hartmann M, Heidenreich A, et al. Prediction of necrosis after chemotherapy of advanced germ cell tumors: results of a prospective multicenter trial of the German Testicular Cancer Study Group. J Urol 2004;171:1835–8. https://doi.org/10.1097/01.ju.0000119121.36427.09.

[16] Punjani N, Power N, Vanhie JJ, Winquist E. Validation of a prediction model for avoiding post-chemotherapy retroperitoneal lymphadenectomy in patients with metastatic nonseminomatous germ cell cancer. Can Urol Assoc J 2016;10:260–3. https://doi.org/10.5489/cuaj.3558.

[17] Toner GC, Panicek DM, Heelan RT, Geller NL, Lin SY, Bajorin D, et al. Adjunctive surgery after chemotherapy for nonseminomatous germ cell tumors: recommendations for patient selection. J Clin Oncol 1990;8:1683–94. https://doi.org/10.1200/JCO.1990.8.10.1683.

[18] Heidenreich A, Pfister D, Witthuhn R, Thüer D, Albers P. Postchemotherapy retroperitoneal lymph node dissection in advanced testicular cancer: radical or modified template resection. Eur Urol 2009;55:217–26. https://doi.org/10.1016/j.eururo.2008.09.027.

[19] Cho JS, Kaimakliotis HZ, Cary C, Masterson TA, Beck S, Foster R. Modified retroperitoneal lymph node dissection for post-chemotherapy residual tumour: a long-term update. BJU Int 2017;120:104–8. https://doi.org/10.1111/bju.13844.

[20] Beck SDW, Foster RS. Long-term outcome of retroperitoneal lymph node dissection in the management of testis cancer. World J Urol 2006;24:267–72. https://doi.org/10.1007/s00345-006-0060-8.

[21] Heidenreich A., Daneshmand S., Albers P., Fossa S.D., Kollmannsberger C., Krege S., et al. Contemporary management of postchemotherapy testis cancer2012;62:867–76.

[22] Oldenburg J, Alfsen GC, Lien HH, Aass N, Waehre H, Fossa SD. Postchemotherapy retroperitoneal surgery remains necessary in patients with nonseminomatous testicular cancer and minimal residual tumor masses. J Clin Oncol 2003;21:3310–7. https://doi.org/10.1200/JCO.2003.03.184.

[23] Leão R, van Agthoven T, Figueiredo A, Jewett MAS, Fadaak K, Sweet J, et al. Serum miRNA predicts viable disease after chemotherapy in patients with testicular nonseminoma germ cell tumor. J Urol 2018;200:126–35. https://doi.org/10.1016/j.juro.2018.02.068.

[24] Shen H, Shih J, Hollern DP, Wang L, Bowlby R, Tickoo SK, et al. Integrated molecular characterization of testicular germ cell tumors. Cell Rep 2018;23:3392–406. https://doi.org/10.1016/j.celrep.2018.05.039.

[25] Nestler T, Baeßler B, Pinto dos Santos D, Pfister DJKP, Maintz D, Heidenreich A. Predicting vital retroperitoneal residual tumors of metastatic testicular tumor patients after chemotherapy using radiomics. In: Abstr 527, Genitourin Cancers Symp 2019, San Fr 2019. https://doi.org/10.1200/JCO.2019.37.7_suppl.527.