Appendix

Appendix description of the model performance metrics.

The c-statistic (i.e., area under the receiver operator curve [AUC]), can be interpreted as the probability that patients with an event receive a higher predicted probability than those without (1-5). The calibration slope measure how well the predicted probability correlates to the observed probability with 1 indicating perfect calibration (1-5). A calibration slope below 1 indicates the model underestimates the risk in dogs with an relatively low observed risk and overestimates the risk in dogs with a relatively high observed risk. A calibration slope above 1 indicates that low predictions are too high and high predicted risk to low. Calibration-in-the-large measures the difference between the mean observed risk and the mean predicted risk, and indicates whether there is any systematic under or over estimation (1-5).

Appendix R script for updating the intercept.

```
# re-estimating the intercept.
fit <- glm(DT$event ~ offset(DT$lp - intercept), family = binomial, data = DT)
(new.intercept <- fit$coef[1])

# DT indicates a dataset with
# the logit(mortality risk) being indicated by “lp”,
# the event (5-month or 1-year mortality) indicated by “event”,
# and the original intercept as “intercept” (this is a single number).
```

Appendix table 1. External validation and updating of a multivariable model predicting 5-month mortality in canines surgically treated for osteosarcoma*.

Variables	Original model		External validation		Update 1		Update 2		Update 3	
	Odds ratio	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$
Calibration slope		0.77(0.55;1.00)		1.15(0.77;1.52)		1.15(0.77;1.52)		1.08(0.73;1.43)		0.83(0.59;1.06)
Calibration in the large		-.0005(-0.13;0.13)		-.1050(-0.29;0.08)		-.0004(-0.16;0.16)		.0014(-0.16;0.16)		0.0004(-0.17;0.17)
AUC (c-statistic)[$]		0.63(0.59;0.67)		0.67(0.61;0.72)		0.67(0.61;0.72)		0.66(0.61;0.71)		0.67(0.61;0.73)
Intercept		-1.2379				-1.3429		-1.4467		-1.3471
Chemotherapy										
No chemotherapy	Reference	0.0000							Reference	0.0000
Cisplatin	0.60(0.31;1.15)	-0.5108							NA	
Carboplatin	0.59(0.32;1.10)	-0.5276							0.33(0.18;0.60)	-1.1087
Doxorubicin	0.52(0.29;0.95)	-0.6539							0.32(0.19;0.54)	-1.1394
Doxorubicin combinations	0.38(0.21;0.68)	-0.9676							NA	
Age (years)	1.03(0.97;1.09)	0.0296							1.01(0.94;1.09)	0.0100
Weight (kg)	1.02(1.00;1.03)	0.0198							1.01(0.99;1.03)	0.0100
Male gender	0.79(0.60;1.05)	-0.2357							0.85(0.60;1.21)	-0.1625
Neutered	0.79(0.54;1.15)	-0.2357							1.09(0.58;2.03)	0.0862
High SALP	1.45(1.08;1.95)	0.3716							1.18(0.78;1.76)	0.1655
Breed										
Other breed	Reference	0.0000							NA	
Rottweiler	0.89(0.58;1.35)	-0.1165								
Golden Retriever	0.86(0.53;1.39)	-0.1508								
Labrador Retriever	0.81(0.48;1.37)	-0.2107								
Greyhound	1.29(0.70;2.37)	0.2546								
Doberman	1.47(0.81;2.69)	0.3853								
Mixed breed	0.73(0.49;1.09)	-0.3147								
Tumor location										
Other	Reference	0.0000							Reference	0.0000
Prox. Humerus	1.54(1.05;2.25)	0.4318							2.04(1.23;3.37)	0.7129
Dist. Femur or Prox. Tibia	0.97(0.65;1.44)	-0.0305							1.42(0.84;2.40)	0.3507
Dist. Radius	0.69(0.46;1.04)	-0.3711							0.74(0.47;1.18)	-0.3011
Monocytes (10^9/L)	NA						1.34(0.87;2.07)	0.2937	1.44(0.92;2.26)	0.3648
Lymphocytes (10^9/L)	NA						0.96(0.75;1.22)	-0.0429	NA	
Breed size	NA									
Mixed									Reference	0.0000
Giant									1.68(0.83;3.42)	0.5188
Large									1.10(0.70;1.73)	0.0953
Medium									0.73(0.25;2.19)	-0.3147

* In the external validation the original model was applied to an independent dataset without re-estimating any coefficients. In update 1 the original model was again applied to the same independent dataset only now with a re-estimated intercept coefficient. In update 2 the variables monocytes and lymphocytes were added to the model. In update 3 the entire model was re-estimated excluding the variables breed and lymphocytes. NA indicates that this association was not estimated for the respective model. In the original publication lobaplatin and carboplatin where combined, in this validation study only carboplatin was available. Results are presented as odds ratios or the natural logarithm of the odds ratio ($\hat{\beta}$) with 95% confidence intervals in brackets. The calibration slope measures how well observed and predicted risk correlates in the tails and is ideally 1. Calibration-in-the large is the mean difference between observed and predicted risk on the logit scale and measures any systematic over- or underestimation. $ The c-statistics is the proportion of subjects that experienced an event and that received a higher predicted risk than subjects that did not experience an event. Also note that while it is reported in the column ($\hat{\beta}$) it is not actually on the natural logarithmic scale but is a regular proportion bounded by 0 and 1 **Appendix table 2. External validation and updating of a multivariable model predicting 1-year mortality in canines surgically treated for osteosarcoma***

Variables	Original model		External validation		Update 1		Update 2		Update 3	
	Odds ratio	$\hat{\beta}$	*Odds ratio*	$\hat{\hat{\beta}}$	*Odds ratio*	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$	*Odds ratio*	$\hat{\beta}$
Calibration slope		0.82(0.61;1.02)		0.95(0.63;1.28)		0.95(0.63;1.28)		0.90(0.60;1.20)		0.81(0.57;1.04)
Calibration in the large		-.0002(-0.12;0.12)		0.2519(0.10;0.40)		-.0019(-0.15;0.15)		-.0016(-0.15;0.15)		-.0010(-0.15;0.15)
AUC (c-statistic)[$]		0.64(0.60;0.67)		0.62(0.58;0.66)		0.62(0.58;0.66)		0.62(0.58;0.66)		0.65(0.61;0.69)
Intercept		-0.4634				-0.2114		-0.2582		-0.8277
Chemotherapy										
No chemotherapy	Reference	0.0000							Reference	0.0000
Cisplatin	0.86(0.45;1.64)	-0.1492							NA	
Carboplatin	1.00(0.55;1.84)	0.0000							0.70(0.42;1.16)	-0.3600
Doxorubicin	1.01(0.56;1.81)	0.0070							0.55(0.35;0.85)	-0.5972
Doxorubicin combinations	0.65(0.37;1.14)	-0.4326							NA	
Age (years)	1.02(0.97;1.07)	0.0173							1.07(1.00;1.15)	0.0720
Weight (kg)	1.01(1.00;1.02)	0.0127							1.02(1.00;1.03)	0.0163
Male gender	1.08(0.83;1.39)	0.0740							0.86(0.62;1.20)	-0.1464
Neutered	0.84(0.59;1.21)	-0.1690							1.04(0.58;1.85)	0.0367
High SALP	1.58(1.19;2.11)	0.4603							1.23(0.85;1.79)	0.2087
Breed									NA	
Other breed	Reference	0.0000								
Rottweiler	1.18(0.78;1.77)	0.1648								
Golden Retriever	1.06(0.69;1.63)	0.0591								
Labrador Retriever	1.03(0.64;1.66)	0.0282								
Greyhound	0.88(0.50;1.56)	-0.1241								
Doberman	1.59(0.85;2.97)	0.4649								
Mixed breed	0.84(0.60;1.19)	-0.1687								
Tumor location										

Other	Reference	0.0000					Reference	0.0000
Prox. Humerus	2.38(1.64;3.46)	0.8687					2.39(1.44;3.97)	0.8725
Dist. Femur or Prox. Tibia	1.34(0.95;1.91)	0.2960					1.40(0.86;2.30)	0.3385
Dist. Radius	0.79(0.56;1.12)	-0.2356					0.68(0.46;1.02)	-0.3811
Monocytes (10^9/L)	NA			1.27(0.78;2.07)	0.2376		1.36(0.82;2.26)	0.3106
Lymphocytes (10^9/L)	NA			0.95(0.76;1.18)	-0.0557		NA	
Breed size	NA							
Mixed							Reference	0.0000
Giant							2.01(1.06;3.83)	0.6985
Large							1.29(0.87;1.91)	0.2538
Medium							0.83(0.34;2.01)	-0.1853

* In the external validation the original model was applied to an independent dataset without re-estimating any coefficients. In update 1 the original model was again applied to the same independent dataset only now with a re-estimated intercept coefficient. In update 2 the variables monocytes and lymphocytes were added to the model. In update 3 the entire model was re-estimated excluding the variables breed and lymphocytes. NA indicates that this association was not estimated for the respective model. In the original publication lobaplatin and carboplatin where combined, in this validation study only carboplatin was available. Results are presented as odds ratios or the natural logarithm of the odds ratio ($\hat{\theta}$) with 95% confidence intervals in brackets. The calibration slope measures how well observed and predicted risk correlates in the tails and is ideally 1. Calibration-in-the large is the mean difference between observed and predicted risk on the logit scale and measures any systematic over- or underestimation. $ The c-statistics is the proportion of subjects that experienced an event and that received a higher predicted risk than subjects that did not experience an event. Also note that while it is reported in the column ($\hat{\theta}$) it is not actually on the natural logarithmic scale but is a regular proportion bounded by 0 and 1

Appendix table 3. Result from a backward model selection procedure (update 4) for models predicting 5-month and 1-year mortality in canines surgically treated for osteosarcoma*.

Variables	5 month			1 year		
	Odds ratio	$\hat{\beta}$	*P-value*	*Odds ratio*	$\hat{\theta}$	*P-value*
Calibration slope		0.90(0.65;1.15)			0.84(0.59;1.08)	
Calibration in the large		0.0003(-0.17;0.17)			-.0032(-0.16;0.15)	
AUC (c-statistic)[$]		0.68(0.62;0.73)			0.65(0.61;0.69)	
Intercept		-1.2997			-0.7848	
Chemotherapy			Overall: <0.01			Overall: 0.03
No chemotherapy	Reference	0.0000		Reference	0.0000	
Cisplatin	NA			NA		
Carboplatin	0.33(0.18;0.61)	-1.1102	<0.01	0.70(0.42;1.16)	-0.3614	0.16
Doxorubicin	0.32(0.19;0.54)	-1.1466	<0.01	0.55(0.36;0.86)	-0.5950	0.01
Doxorubicin combinations	NA			NA		
Age (years)	NA			1.07(1.00;1.15)	0.0711	0.04
Weight (kg)	1.02(1.01;1.03)	0.0187	0.01	1.01(1.00;1.03)	0.0141	0.08
Male gender	0.79(0.56;1.11)	-0.2367	0.18	NA		
Neutered	NA			NA		
High SALP	NA			1.25(0.86;1.81)	0.2204	0.25
Breed categorization 1	NA			NA		
Other breed						
Rottweiler						
Golden Retriever						
Labrador Retriever						
Greyhound						
Doberman						
Mixed breed						
Tumor location			Overall: 0.07			Overall: <0.01
Other	Reference	0.0000		Reference	0.0000	
Prox. Humerus	2.03(1.23;3.33)	0.7062	0.01	2.37(1.43;3.94)	0.8792	<0.01
Dist. Femur or Prox. Tibia	1.44(0.86;2.41)	0.3623	0.17	1.42(0.87;2.31)	0.3487	0.16
Dist. Radius	0.76(0.48;1.21)	-0.2701	0.25	0.68(0.46;1.01)	-0.3859	0.06
Monocytes (10^9/L)	1.46(0.94;2.28)	0.3792	0.09	1.36(0.82;2.26)	0.3082	0.23
Lymphocytes (10^9/L)	NA			NA		
Breed categorization 2	NA					Overall: 0.11
Mixed				Reference	0.0000	
Giant				2.11(1.12;3.98)	0.7464	0.02
Large				1.29(0.87;1.91)	0.2515	0.21
Medium				0.80(0.33;1.92)	-0.22251	0.61

* In the original publication lobaplatin and carboplatin where combined, in this validation study only carboplatin was available. Furthermore, patients did not received cisplatin or doxorubicin combination therapy hence no estimates are provided. NA indicates that this association was not estimated in the final model. Variables were retained if the p-value was equal to or smaller than 0.30. Results are presented as odds ratios or the natural logarithm of the odds ratio ($\hat{\beta}$ *or* $\hat{\theta}$) with 95% confidence intervals in brackets. The calibration slope measures how well observed and predicted risk correlates in the tails and is ideally 1. Calibration-in-the large is the mean difference between observed and predicted risk on the logit scale and measures any systematic over- or underestimation. $The c-statistics is the proportion of subjects that experienced an event and that received a higher predicted risk than subjects that did not experience an event. Also note that while it is reported in the columns ($\hat{\theta}$ *or* $\hat{\beta}$) it is not actually on the natural logarithmic scale but is a regular proportion bounded by 0 and 1

Appendix table 4. Sensitivity analysis including patients surviving the first month. Treatment effect estimates of any chemotherapeutics compared to no chemotherapy on 5-month and 1-year mortality.

	Any chemotherapy for 5 month mortality	Any chemotherapy for 1 year mortality
Main treatment effect	0.64 (0.34; 1.21)	0.69 (0.44; 1.09)
Interaction effect	1.95 (0.74; 5.10)	1.25 (0.58; 2.68)

Results presented as odds ratios (95 % confidence intervals) with no chemotherapy as the reference group. The interaction effect describes by how much the treatment effect estimates change with on unit increase in the logit(predicted risk). All models were adjusted for the covariables age, gender, weight, neuter status, SALP, breed, tumour location, monocytes and lymphocytes; no model selection was performed.

Reference List

(1) Harrell FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. 1st ed. New York: Springer; 2001.

(2) 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 1996 Feb 28;15(4):361-87.

(3) Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. New York: Springer; 2009.

(4) Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-38.

(5) Wolbers M, Koller MT, Witteman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology 2009 Jul;20(4):555-61.