

Original article manuscript 2

Chemotherapy Effectiveness and Mortality Prediction in Surgically Treated Osteosarcoma Dogs: a Validation Study.

A F Schmidt^{a,b,c,d,*}, M Nielen^c, S J Withrow^e, L E Selmic^e, J H Burton^{e,f}, O H Klungel^{a,b}, R H H Groenwold^{a,b}, J Kirpensteijn^g.

^a. Julius Center for Health Sciences and Primary Care, University Medical Centre Utrecht, P.O. Box 85500, 3508 GA Utrecht, the Netherlands.

^b. Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, P.O. Box 80082, 3508 TB Utrecht, The Netherlands.

^c. Department of Farm Animal Health, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 7, Utrecht 3584 CL, The Netherlands.

^d. Institute of Cardiovascular Science, Faculty of Population Health, University College London, London WC1E 6BT, The United Kingdom.

^e. Flint Animal Cancer Center, Department of Clinical Sciences, College of Veterinary Medicine and Biomedical Sciences, Colorado State University, Fort Collins, Colorado, The United States of America.

^f. University of California, Davis, Department of Surgical and Radiological Sciences, School of Veterinary Medicine, Davis California, The United States of America.

^g. Department of Clinical Sciences of Companion Animals, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 8, Utrecht 3584 CM, The Netherlands. Current address: Hills Pet Nutrition, 400 SW 8th Ave, Topeka, KS 66603, USA

* Corresponding author: Tel.: +44 20 3549 5625.

E-mail address: amand.schmidt@ucl.ac.uk (A.F.Schmidt).

Target journal: Preventive veterinary medicine.

Running title: predicting OS mortality.

Word count text: 3219

Word count abstract: 209

Number of references: 36

Number of tables: 5

Number of figures: 2

(Web)appendices: 2

Abstract

Canine osteosarcoma is the most common bone cancer and an important cause of mortality and morbidity in large purebred dogs. Previously we constructed two multivariable models to predict a dog's 5-month and 1-year mortality risk after surgical treatment for osteosarcoma. According to these models, dogs with a relatively low risk of 5-month mortality benefited most from additional chemotherapy treatment. In the present study, we externally validated these results using an independent cohort study of 794 dogs. External performance of our prediction models showed some disagreement between observed and predicted risk, mean difference: -0.11 (95% confidence interval [95%CI]-0.29; 0.08) for 5-month risk and 0.25 (95%CI 0.10; 0.40) for 1-year mortality risk. After updating the intercept, agreement improved: -0.0004 (95%CI -0.16; 0.16) and -0.002 (95%CI -0.15; 0.15). The chemotherapy by predicted mortality risk interaction (P-value = 0.01) showed that the chemotherapy compared to no chemotherapy effectiveness was modified by 5-month mortality risk: dogs with a relatively lower risk of mortality benefited most from additional chemotherapy. Chemotherapy effectiveness on 1-year mortality was not significantly modified by predicted risk (P-value = 0.28). In conclusion, this external validation study confirmed that our multivariable risk prediction models can predict a patient's mortality risk and that dogs with a relatively lower risk of 5-month mortality seem to benefit most from chemotherapy.

MeSH/keywords: Canine; Clinical Prediction Rule; Chemotherapy, Adjuvant; Personalized Medicine; Oncology, Bone tumour.

Introduction

Surgically treated dogs with appendicular osteosarcoma (OS), a malignant tumor of mesenchymal origin that produces osteoid, have a median survival time of 5 months (Brodey and Abt, 1976; Cooley and Waters, 1997; McNeill et al., 2007; Norrdin et al., 1989; Ru et al., 1998; Spodnick et al., 1992; Straw and Withrow, 1996). For the average patient, previous studies have shown that additional chemotherapy can extend median survival beyond these 5 months (Bailey et al., 2003; Chun et al., 2005; Chun et al., 2000; Straw et al., 1991; Vail et al., 2002).

Recently, using an Individual Patient Data Meta-Analysis (IPDMA), we constructed a multivariable prediction tool, predicting a dog's risk of mortality at 5 months and 1 year after receiving surgical treatment for OS (Schmidt et al., 2013). This tool predicts mortality risk based on a patient's age, weight, gender, neuter status, serum alkaline phosphatase (SALP) level, breed, and tumor location. In a nested study, we explored whether chemotherapy effectiveness differed between dogs with a different predicted risk (i.e., subgroup analysis; **Manuscript 1**). Results showed that chemotherapy (compared to no chemotherapy) was most effective in dogs with a relatively low predicted risk. This implies that perhaps dogs with a lower predicted risk of mortality should be preferentially treated with additional chemotherapy. Combining information on a dog's mortality risk with a personalized estimate of treatment effect can aid veterinary professionals to better tailor treatment to a dog's needs, which is relevant in terms of extending survival, healthcare costs and quality of life.

In the present study, we validate these findings using an independent cohort study collected at the Flint Animal Cancer Center at Colorado State University (Selmic et al., 2014). Specifically, we first applied our previously developed "original" prediction model to these

external data and determined model performance. Second, we validated the differential chemotherapy effectiveness between dogs with different baseline mortality risks.

Materials and Methods

The external validation of the prediction models and the chemotherapy subgroup-specific effects were evaluated using a subset of the Colorado State University cohort study (Selmic et al., 2014); data were collected based on a retrospectively review of electronic medical records. For the current analyses, dogs were eligible if they received surgical treatment (amputation or limb-spare) for OS. Because of the relatively rare occurrence, 49 dog receiving cisplatin/carboplatin, cisplatin or any other kind of (combination) chemotherapy were excluded. Patients were also excluded if they received radiation therapy (n = 133), had a zero or negative follow-up time (n = 15, measured from date of surgery to date of last contact), had an erroneous date of metastasis (after the date of death, n = 9), there was confirmed or a suspicion of metastasis at baseline (n = 16), received pamidronate (n = 9) or were small purebred dogs (n = 5). Exclusion criteria were identical to our discovery paper ([Manuscript 1](#)), with the slight difference that (to prevent small exposure categories) dogs with cisplatin or doxorubicin combination therapy were excluded. Data were collected based on medical records, hence routine (scintigraphy based) staging information was not always available. Additionally, we emphasize that sample size was determined in an opportunistic manner, without formal sample size calculations; because of the retrospective nature of this cohort, this did not impact patient safety.

For the 794 remaining patients, baseline data were available on age (years), weight (kg), gender (0 female, 1 male), neuter status (0 intact, 1 neutered), high serum alkaline phosphatase (SALP) defined as above 140 IU/dL, continuous monocytes count (10^9 cells/L),

continuous lymphocyte counts (10^9 cells/L), breed (0 other breed, 1 Rottweiler, 2 Golden Retriever, 3 Labrador Retriever, 4 Greyhound, 5 Doberman, 6 mixed breed) and tumor location (0 other, 1 proximal humerus, 2 distal femur or proximal tibia, 3 distal radius). Additionally, we recorded whether a dog received chemotherapy (0 no chemotherapy, 1 carboplatin or 2 doxorubicin) and if it was alive at 5 months and 1 year (0 alive, 1 dead).

On average, 11% percentage of these variables were missing, the percentage missingness per variable was: 1-year mortality 6.05%, 5-month mortality 2.90%, chemotherapy 27.83%, age 0.13%, weight 0.13%, gender 0.00%, neuter status 0.00%, high SALP 9.57%, monocytes 18.89%, lymphocytes 18.89%, breed 0.00%, and tumor location 1.39%. Univariable analysis showed that missingness was dependent on observed variables (available upon request) indicating that a complete case analysis excluding missing observations would be biased (Groenwold et al., 2012; Rubin, 1976). Instead, we used the dependence between the missing observations and observed variables to impute missing values (Rubin, 1976) using the `aregImpute` algorithm from the `Hmisc` package version 3.14-5 (Harrell, Jr. and Dupont, 2013). This algorithm was implemented using 5 burn-in iterations, predictive mean matching and 100 bootstrap samples to determine the (non) linear relationship between the continuous predictor variables and the missing values. To correct for the inherent underestimation of the variance, 15 imputed datasets were created (i.e., multiple imputation) (White and Carlin, 2010); results of the 15 imputed datasets were pooled using Rubin's rules (Little and Rubin, 2002; Marshall et al., 2009).

Data analysis: prediction model validation

Based on the logistic regression version of our previous derived prediction model (Schmidt et al., 2013), a patient's 5-month and 1-year risk of mortality was calculated by

summing the product of their baseline variables and the relevant coefficients (Table 1); please note that because dogs with combination doxorubicin or cisplatin therapy were excluded, the coefficients for these therapies become redundant. Formally, their predicted $\text{logit}(\text{mortality risk})$ was calculated using equation 1:

$$\text{logit}(\text{mortality risk}) = \text{logit}(\hat{p}_i) = \hat{\beta}_0 + \sum_{j=1}^J \hat{\beta}_j x_{ij} \text{ [equation 1]}$$

Where i represent the i th individual and j the j th variable presented in table 1, $\hat{\beta}_j$ the natural logarithm of the odds ratio for 5-month mortality and x the variable status after surgical treatment. The predicted $\text{logit}(1\text{-year mortality risk})$ was estimated by replacing $\hat{\beta}_j$ by $\hat{\theta}_j$. Finally, the mortality risk was calculated by taking the inverse of the predicted $\text{logit}(\text{mortality risk})$, Table 1. Note that for ease of notation, we will often drop the “predicted” from $\text{logit}(\text{mortality risk})$, however unless stated otherwise this always refers to an estimate from equation 1.

Obviously, these calculations are only relevant for future patients if we can assume the model to be correctly specified (i.e., describe the relationship between the predictors and outcome sufficiently). To evaluate the models predictive performance in this independent validation study we calculated the c-statistic, calibration slope and calibration-in-the-large (Harrell, Jr. et al., 1996). Calibration was also graphically assessed by plotting the mean observed risk per deciles on the y-axis and the predicted risk on the x-axis (i.e., a graphical representation of the Hosmer-Lemeshow goodness-of-fit test) (Harrell, Jr. et al., 1996; Steyerberg, 2009; Steyerberg et al., 2010). Please, see the Appendix for a description of the metrics interpretability.

Besides this simple external validation, the prediction models were corrected for any systematic difference between observed and predicted risk (i.e., calibration-in-the-large $\neq 0$) by re-estimating the intercept in “Update 1” (Moons et al., 2012; Steyerberg, 2009). Such recalibration can be readily applied in clinical practice using a relatively small number of events (e.g., 30) (Steyerberg, 2009). To aid clinicians in updating the model to their local setting a computer code is provided in the Appendix.

Data analysis: estimating chemotherapy effectiveness

After determining the external performance of our prediction models (predicting 5-month and 1-year mortality risk), we assessed whether the effect of “any chemotherapy” (carboplatin or doxorubicin) compared to no chemotherapy in preventing mortality differed between patients with different predicted risks of mortality. To explore consistency, all analyses are repeated for carboplatin compared to no chemotherapy and doxorubicin compared to no chemotherapy at 5-month and 1-year mortality.

This approach to tailor chemotherapy effects was previously described in detail in [Manuscript 1](#), which we briefly repeat here. To get an estimate of the risk of mortality if the patient remained untreated with chemotherapy, we re-calculated the logit(mortality risk) by setting (possibly contrary to the fact) the chemotherapy variable to “no chemotherapy” in equation 1. Second, to test whether chemotherapy effects differed between patients with different logit(mortality risks) a product term (i.e., the product of the variables chemotherapy and logit(mortality risk)) was added to a logistic regression model. This model regressed the mortality variable (5 months or 1 year) on chemotherapy, the logit(mortality risk), the product term and the potential confounders: age, weight, sex, neuter status, SALP, monocytes, lymphocytes, breed and tumor location. Significance of this product term was tested using a

Wald based interaction test (Schmidt et al., 2014a). The chemotherapy effect for a patient with a specific logit(mortality risk) was calculated by adding the chemotherapy coefficient to the interaction term coefficient times the logit(mortality risk) (**Manuscript 1**).

To increase precision of the interaction effect, this was combined with the estimate of the previous study (**Manuscript 1**). Results were pooled using the inverse variance weighted estimator and between study heterogeneity was tested using the Q-statistic (Higgins and Thompson, 2002; Schmidt et al., 2014a)

Data analysis: sensitivity analyses

To explore the robustness of our approach the following three sensitivity analyses were performed:

Linearity of the continuous variables monocytes, lymphocytes, logit(mortality risk) and the chemotherapy by logit(mortality risk) with the outcome at 5 months and 1 year was assessed using restricted cubic splines with 5 knots (Harrell, Jr., 2001). No significant deviations from linearity could be found. Additionally, graphically exploring linearity showed that there was some deviation for the monocytes and lymphocytes (showing a slight sinus pattern) however, these could be approximated by a linear term. Previously, the linearity of age, weight and SALP was assessed using the same approach (Schmidt et al., 2013). In this analysis, SALP was observed to be non-linearly related with the outcome, because of this non-linear relationship SALP was dichotomized.

As described above, we explored external performance of our “original” prediction models for 5-month and 1-year mortality risk, without and with updating the intercept. As

previously indicated (Schmidt et al., 2013), however, performance could perhaps be improved by adding more variables or recoding variables. To explore this we added the variables monocytes and lymphocytes to the model in “Update 2”. Additionally, to compare performance of our “original” model to a model optimally tailored to the current dataset we re-estimate the entire model with addition of the monocytes variable and a recoded breed variable (mixed breed, giant purebred, large purebred and medium purebred) in “Update 3”. Finally, to determine if reducing the model might improve performance, we performed a backward selection procedure using a P-value criterion of 0.30 (“Update ”). To correct for model optimism, model performance metrics for updates 1-4 were calculated in 200 bootstrap samples (Harrell, Jr. et al., 1996; Steyerberg, 2009; Steyerberg et al., 2010). Model performance of updates 2-4 was very similar to the performance of Update 1 and the external validation (see Appendix Tables 1-3).

In the current cohort study chemotherapy was not allocated randomly. Thus it is likely that dogs receiving chemotherapy had a better prognosis than dog not receiving chemotherapy, which would bias our results. Besides adjusting for measured confounders as described above, we explored this further by repeating all analyses regarding the chemotherapy by logit(mortality risk) interaction excluding patients dying in the first 30 days (a similar analysis was conducted in [Manuscript 1](#)), excluding 23% and 9% of the mortality events at 5-month and 1-year. Results were very similar and are presented in Appendix Table 4.

All tests were applied using a significance level of 0.05 (unless stated otherwise), estimates are presented as odds ratios (OR) with 95% confidence intervals (95%CI). Analyses

were carried out using the R statistical package for windows version 3.1.1 (R Development Core Team, 2013).

Results

Baseline characteristics are presented in Table 2. The median follow-up time was 241 days with 50% of the follow-up being between 146 days and 472 days. During follow-up 163 dogs did not receive any chemotherapy, 172 received carboplatin, 238 doxorubicin and for 221 patients the chemotherapy regime received was not recorded. Except for higher SALP levels, dogs not receiving chemotherapy were similar to dogs receiving carboplatin or doxorubicin therapy.

Results: prediction model validation

The external validation of the original prediction model for 5 months without any updating showed a slightly better c-statistic than in the original derivation data (0.67 95%CI 0.61; 0.72), Table 3. The calibration slope (1.15 95%CI 0.77; 1.52) and the calibration-in-the-large (-0.1050 95%CI -0.29; 0.08) indicated that there was slight model misspecification in tail areas and that the model systematically underestimated the risk (Figure 1). Updating the intercept corrected the systematic underestimation; calibration-in-the-large (-0.0004 95%CI -0.16; 0.16).

At 1 year the c-statistic (Table 4) was 0.62 (95%CI 0.58; 0.66), indicating that it was difficult to discriminate between patient experiencing an event and those, which remained event free. While the model misspecification in the tail areas was small (calibration slope: 0.95 95%CI 0.63; 1.28), there was considerable overestimation, indicated by a calibration-in-the-large of

0.2519 (95%CI 0.10; 0.40). Updating the intercept resulted in an almost perfect calibrated model (Figure 1).

Results: chemotherapy by mortality risk

Given the good performance of the original models we estimated the logit(mortality risk) under no chemotherapy treatment. The median and range logit(mortality risk) under no chemotherapy were: -0.60 (-1.75; 0.85) for 5 months and 0.26 (-0.72; 2.14) for 1 year. Transformed to the risk scale this becomes: 0.35 (0.15; 0.70) and 0.56 (0.33, 0.90).

For a patient with a 0.50 predicted risk of dying at 5 months [i.e., a logit(5-month mortality) of 0.00] the effect of “any chemotherapy” compared to no chemotherapy on preventing mortality was OR 0.45 (95% 0.25; 0.81; Table 5). The interaction effect showed there was considerable difference between patients with a different mortality risk: OR 1.89 (95%CI 0.83; 4.33). To get a more precise estimate results were pooled with those from **Manuscript 1**, resulting in an interaction effect OR 2.31 (95%CI 1.18; 4.53; Table 5). While there was considerable difference in effectiveness in the current population, the majority of patients benefitted from chemotherapy (Figure 2).

For a patient with a 0.50 risk of 1-year mortality, chemotherapy was slightly less effective in preventing mortality OR 0.57 (95%CI 0.35; 0.91; Table 5). Furthermore, the interaction effect OR of 1.26 (95%CI 0.60; 2.63) was closer to the 1.00 indicating a smaller difference in effectiveness. Combining results with those from the previous study (**Manuscript 1**) resulted in a non-significant interaction effect: OR 1.39 (95CI% 0.76; 2.54; Table 5). Figure 2, indeed shows a small difference in chemotherapy effectiveness. This indicates that perhaps the interaction effect is redundant. Excluding the interaction effect resulted in a

chemotherapy effect that was equal for all patients regardless of their predicted risk: OR 0.60 (95%CI 0.39; 0.92).

Overall, the effects of any chemotherapy against no chemotherapy were comparable to the effects of carboplatin or doxorubicin against no chemotherapy (Table 5). The heterogeneity between the current study and the results from (Manuscript 1) were non-significant (available upon request).

Discussion

In this study, we described the external validation of two multivariable models predicting 5-month and 1-year mortality risk in dogs surgically treated for osteosarcoma (OS). Additionally, we reproduced a previously reported chemotherapy by predicted risk interaction effect, indicating that a dog's predicted risk of mortality modifies the effectiveness of chemotherapy.

Results showed that both prediction models (for 5-month and 1-year mortality risk) generalized well to the current independent dataset. For the 1-year mortality prediction model recalibration of the intercept (Update 1) was needed to correct for differences in mortality incidence. This recalibration is common when applying a prediction model to a new setting and can easily be implemented in clinical practice using the R code provided in the Appendix. Compared to the previous publication (Schmidt et al., 2013) (original model in Tables 3 and 4), the discriminative ability (c-statistic between 0.62; 0.67) of both models was similar but modest. This indicates that the models have difficulty discriminating between subjects experiencing an event and those that did not. In part, this was caused by the fact that most patients had a similar risk. The good calibration (agreement between observed and predicted

risk) indicates that the clustering of risk is a characteristic of the patient population, not a modelling error. Given the modest discriminative ability but good calibration, these risk prediction models are perhaps best used for identifying patients at a high- or low risk of mortality, not for indicating which patient will actually die.

The chemotherapy by predicted risk (strictly speaking the logit of the predicted risk) interaction showed that chemotherapy was more effective in preventing 5-month mortality in lower risk patients (risk cut off 0.52). Our findings imply that short term (i.e., 5 months) effectiveness of chemotherapy depends on a dogs predicted risk of 5-month mortality. Contrary to our previous study ([Manuscript 1](#)), most patients benefitted from additional chemotherapy. This difference in overall chemotherapy effectiveness was caused by a lower baseline risk in this population compared to the previous study. This also demonstrates that in the presence of a chemotherapy interaction (for 5-month mortality) the overall average effect estimate is inappropriate, instead an individualized chemotherapy effect estimate should be used (Schmidt et al., 2014b). Due to this interaction a different fraction of patients will benefit from chemotherapy depending on the population specific predicted risk distribution. If every dog in a particular population benefits from chemotherapy (e.g., because the entire population consists of low risk patients), individualizing treatments becomes less important. Potentially, however, in a setting where each dog benefits, difference in chemotherapy effectiveness may still be of interest from a health economic perspective. At 1 year, we did not find an interaction (similar to the previous study, [Manuscript 1](#)), however, due to the wide confidence interval around the interaction effect, we cannot exclude that such an interaction does exist (Schmidt et al., 2016). Based on the current evidence it seems most appropriate, however, to use the main effect of chemotherapy compared to no chemotherapy, estimated in a model without an interaction, of OR 0.60 (95%CI 0.39; 0.92) for all patients regardless of their

predicted risk. Finally, as in the discovery paper (Manuscript 1), we recognize that the above described interaction tests ignore the uncertainty in the predicted logit(mortality), underestimating the variance. To adjust for this, following the same approach as in our previous paper (Manuscript 1), we bootstrapped the original dataset refitting the prediction model (2,000 bootstrap sample per imputed dataset), re-estimating the predicted risk and the interaction effect. Unexpectedly this led to smaller confidence intervals than when ignoring the first stage uncertainty: interaction OR 1.89 (95%CI 1.20; 2.97) and OR 1.26 (95%CI 0.98; 1.63), for 5 months and 1-year mortality. Possibly this decrease in variance is related to the fact that bootstrapping the original prediction models increases its external performance (i.e., fits external data better), decreasing differences between observed and predicted risk (Harrell, Jr., 2001). However, as shown above, at least for the 5-months prediction model, the predicted risk already fitted the observed risk reasonably well. Clearly, this issue of how to incorporate the first stage uncertainty needs further theoretical consideration.

Our finding that chemotherapy effectiveness depends on a second variable has also been shown in other studies. An example would be the synergistic effect between immunotherapy and chemotherapy (MacEwen and Kurzman, 1996; Vail et al., 1995). To the best of our knowledge, we are the first to show that chemotherapy effectiveness may depend on multiple variables.

The current study has some important limitations. First, in the current cohort study chemotherapy was not randomly allocated. While results were adjusted for potential confounding, it seems likely that there is still remaining bias due to residual or unobserved confounding. For example, decisions on euthanasia may partially dependent on the type of chemotherapy prescribed (e.g., due to differences in side effects), and almost certainly on the

decision to not use further chemotherapy. Instead of focussing on all-cause mortality (as done here), one might be tempted to focus on “naturally” occurring death and censor (using a survival model) or disregard (using logistic regression) dogs who were euthanized. However, confounding bias would likely remain and even in its absence this would introduce bias due to competing risks (Satagopan et al., 2004), an issue we addressed in Schmidt et.al (, 2013). On the other hand, because the decisions around chemotherapy may not only depends on a patients’ life expectancy but also on an owner’s willingness to pay, some degree of randomness might be expected. Nevertheless, we expect there to be some degree of residual confounding (by unmeasured variables affecting both chemotherapy decision and life expectancy (Hernan and Robins, 2006)) in the estimates presented, and therefore, we hope that our findings might lead to the initiation of new, or re-analysis of historical, randomized clinical trials (RCTs) further exploring the validity of our results.

Second, as reported, some observations were missing. Instead of focusing on complete observation, which (by ignoring dependencies) leads to selection bias, missing values were imputed (Groenwold et al., 2012; White and Carlin, 2010). Third, no correction for multiple testing was applied. Given the modest amount of tests and that similar findings were reported in a second independent study (Manuscript 1), we feel that such a correction is redundant. Fourth, we emphasize that this is only a single validation study, based on a sample collected in a university hospital in the USA, hence results may differ from a primary care setting in another region. Before implementing the prediction model in a clinical setting, we suggest to validate the model anew, using for example historical medical records and if need recalibrate the model using the code provided in Appendix 1. Based on a recent publication from Collins, Ogundimu and Altman (, 2015) it seems that such a validation study needs at the very least 100 events to provide moderately precise and unbiased estimates of model performance (i.e., calibration and discrimination).

Nevertheless, given the performance in this single external validation study we would expected, even in settings where the predicted risk does not agree with the underlying true risk, that ranking of high- versus low risk patients would still be possible using these models. Because of this we have included an excel spreadsheet to aid in calculating a patient's risk of mortality (see Appendix II). To reiterate, unless a local validation study is performed, we would not expected the predicted risk from this spreadsheet program to match the true risks. Therefore, we are hesitant to provide thresholds to categorize high-, moderate- and low-risk patients, instead we suggest practitioners use this tool to help **rank** patients to identify **relatively** high- or low-risk patients (compared to the "average" patient encounters in their practice). After multiple validation studies, perhaps a consensus could be reached on risk thresholds. For example, in human cardiovascular heart disease (CHD), risk threshold were decided by consensus in guideline groups based on information on external validation of CHD prediction models [e.g., the Framingham (D'Agostino, Sr. et al., 2008)], treatment efficacy (e.g., statins) and safety, and cost-effectiveness (Hingorani and Hemingway, 2011; Hingorani and Psaty, 2009). Due to the mentioned lack of randomization we have not included the interaction with chemotherapy in the spreadsheet, we feel that this is best included after (historical) RCT data have confirmed this interaction.

Conclusions

Based on our results, we conclude that a dog's risk of mortality, after surgical treatment for osteosarcoma, can be predicted using the models presented. Dogs with a lower predicted risk of 5-month mortality seem to benefit most from additional chemotherapy.

Conflict of interest statement

None of the authors of this paper has a financial or personal relationship with other people or organisations that could inappropriately influence or bias the content of the paper.

Author contributions

AFS, JK, RHHG and MN contributed to the idea and design of the study. SJW, LES and JHB collected and shared the data. AFS performed the analyses and drafted the manuscript. JK, RHHG, OHK, JHB, SJW, LES and MN provided guidance during initial planning of the paper and during critical revision. AFS had full access to all of the data and takes responsibility for the integrity of the data presented.

Funding

This work was supported by Research Focus Areas funding of the Utrecht University, which is a collaboration between the faculties of medicine, science, and veterinary medicine. Supported in part by Flint Animal Cancer Center at Colorado State University. The study sponsors were not involved in the design, collection, analysis or writing of this manuscript.

Prior postings and presentations

This study and its results were neither previously published, nor presented at conferences.

Appendix A: Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi: ...

References

- Bailey, D., Erb, H., Williams, L., Ruslander, D., Hauck, M., 2003. Carboplatin and doxorubicin combination chemotherapy for the treatment of appendicular osteosarcoma in the dog. *Journal of Veterinary Internal Medicine* 17, 199-205.
- Brodey, R.S., Abt, D.A., 1976. Results of surgical treatment in 65 dogs with osteosarcoma. *Journal of the American Veterinary Medical Association* 168, 1032-1035.
- Chun, R., Garrett, L.D., Henry, C., Wall, M., Smith, A., Azene, N.M., 2005. Toxicity and efficacy of cisplatin and doxorubicin combination chemotherapy for the treatment of canine osteosarcoma. *Journal of the American Animal Hospital Association* 41, 382-387.
- Chun, R., Kurzman, I.D., Couto, C.G., Klausner, J., Henry, C., MacEwen, E.G., 2000. Cisplatin and doxorubicin combination chemotherapy for the treatment of canine osteosarcoma: a pilot study. *Journal of Veterinary Internal Medicine* 14, 495-498.
- Collins, G.S., Ogundimu, E.O., Altman, D.G., 2015. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat. Med.*
- Cooley, D.M., Waters, D.J., 1997. Skeletal neoplasms of small dogs: a retrospective study and literature review. *Journal of the American Animal Hospital Association* 33, 11-23.
- D'Agostino, R.B., Sr., Vasan, R.S., Pencina, M.J., Wolf, P.A., Cobain, M., Massaro, J.M., Kannel, W.B., 2008. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 117, 743-753.
- Groenwold, R.H., Donders, A.R., Roes, K.C., Harrell, F.E., Jr., Moons, K.G., 2012. Dealing with missing outcome data in randomized trials and observational studies. *American Journal of Epidemiology* 175, 210-217.
- Harrell, F.E., Jr., 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.
- Harrell, F.E., Jr., Dupont, C., 2013. *Hmisc: Harrell Miscellaneous*. R package.
- Harrell, F.E., Jr., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat. Med.* 15, 361-387.
- Hernan, M.A., Robins, J.M., 2006. Estimating causal effects from epidemiological data. *J. Epidemiol. Community Health* 60, 578-586.
- Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539-1558.
- Hingorani, A.D., Hemingway, H., 2011. How should we balance individual and population benefits of statins for preventing cardiovascular disease? *BMJ* 342, c6244.
- Hingorani, A.D., Psaty, B.M., 2009. Primary prevention of cardiovascular disease: time to get more or less personal? *JAMA* 302, 2144-2145.

- Little, R.J.A., Rubin, D.B., 2002. *Statistical Analysis with Missing Data*. Wiley-Blackwell.
- MacEwen, E.G., Kurzman, I.D., 1996. Canine osteosarcoma: amputation and chemoimmunotherapy. *Veterinary Clinics of North America: Small Animal Practice* 26, 123-133.
- Marshall, A., Altman, D.G., Holder, R.L., Royston, P., 2009. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology* 9, 57.
- McNeill, C.J., Overley, B., Shofer, F.S., Kent, M.S., Clifford, C.A., Samluk, M., Haney, S., Van Winkle, T.J., Sorenmo, K.U., 2007. Characterization of the biological behaviour of appendicular osteosarcoma in Rottweilers and a comparison with other breeds: a review of 258 dogs. *Veterinary and Comparative Oncology* 5, 90-98.
- Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G., Woodward, M., 2012. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 98, 691-698.
- Norrdin, R.W., Powers, B.E., Torgersen, J.L., Smith, R.E., Withrow, S.J., 1989. Characterization of osteosarcoma cells from two sibling large-breed dogs. *American Journal of Veterinary Research* 50, 1971-1975.
- R Development Core Team, 2013. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ru, G., Terracini, B., Glickman, L.T., 1998. Host related risk factors for canine osteosarcoma. *The Veterinary Journal* 156, 31-39.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63, 581-592.
- Satagopan, J.M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., Auerbach, A.D., 2004. A note on competing risks in survival data analysis. *British Journal of Cancer* 91, 1229-1235.
- Schmidt, A.F., Groenwold, R.H., Knol, M.J., Hoes, A.W., Nielen, M., Roes, K.C., de B.A., Klungel, O.H., 2014a. Exploring interaction effects in small samples increases rates of false-positive and false-negative findings: results from a systematic review and simulation study. *J. Clin. Epidemiol.* 67, 821-829.
- Schmidt, A.F., Hoes, A.W., Groenwold, R.H., 2014b. Comments on 'The use of propensity scores and observational data to estimate randomized controlled trial generalizability bias' by Taylor R. Pressler and Eloise E. Kaizar, *Statistics in Medicine* 2013. *Stat. Med.* 33, 536-537.
- Schmidt, A.F., Klungel, O.H., Nielen, M., de Boer A., Groenwold, R.H.H., Hoes, A.W., 2016. Tailoring Treatments using Treatment Effect Modification. *Pharmacoepidemiol. Drug Saf.*
- Schmidt, A.F., Nielen, M., Klungel, O.H., Hoes, A.W., de Boer, A., Groenwold, R.H., Kirpensteijn, J., 2013. Prognostic factors of early metastasis and mortality in dogs with appendicular osteosarcoma after receiving surgery: An individual patient data meta-analysis. *Preventive Veterinary Medicine* 112, 414-422.
- Selmic, L.E., Burton, J.H., Thamm, D.H., Withrow, S.J., Lana, S.E., 2014. Comparison of carboplatin and doxorubicin-based chemotherapy protocols in 470 dogs after amputation for treatment of appendicular osteosarcoma. *J. Vet. Intern. Med.* 28, 554-563.

Spodnick, G.J., Berg, J., Rand, W.M., Schelling, S.H., Couto, G., Harvey, H.J., Henderson, R.A., MacEwen, G., Mauldin, N., McCaw, D.L., ., 1992. Prognosis for dogs with appendicular osteosarcoma treated by amputation alone: 162 cases (1978-1988). *Journal of the American Veterinary Medical Association* 200, 995-999.

Steyerberg, E.W., 2009. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, New York.

Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 21, 128-138.

Straw, R.C., Withrow, S.J., 1996. Limb-sparing surgery versus amputation for dogs with bone tumors. *Veterinary Clinics of North America: Small Animal Practice* 26, 135-143.

Straw, R.C., Withrow, S.J., Richter, S.L., Powers, B.E., Klein, M.K., Postorino, N.C., LaRue, S.M., Ogilvie, G.K., Vail, D.M., Morrison, W.B., ., 1991. Amputation and cisplatin for treatment of canine osteosarcoma. *Journal of Veterinary Internal Medicine* 5, 205-210.

Vail, D.M., Kurzman, I.D., Glawe, P.C., O'Brien, M.G., Chun, R., Garrett, L.D., Obradovich, J.E., Fred, R.M., III, Khanna, C., Colbern, G.T., Working, P.K., 2002. STEALTH liposome-encapsulated cisplatin (SPI-77) versus carboplatin as adjuvant therapy for spontaneously arising osteosarcoma (OSA) in the dog: a randomized multicenter clinical trial. *Cancer Chemotherapy and Pharmacology* 50, 131-136.

Vail, D.M., MacEwen, E.G., Kurzman, I.D., Dubielzig, R.R., Helfand, S.C., Kisseberth, W.C., London, C.A., Obradovich, J.E., Madewell, B.R., Rodriguez, C.O., Jr., ., 1995. Liposome-encapsulated muramyl tripeptide phosphatidylethanolamine adjuvant immunotherapy for splenic hemangiosarcoma in the dog: a randomized multi-institutional clinical trial. *Clinical Cancer Research* 1, 1165-1170.

White, I.R., Carlin, J.B., 2010. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statist. Med.* 29, 2920-2931.

Figure captions

Figure 1. Calibration plot comparing predicted probabilities to observed probabilities.

[Figure 1]

Perfect calibration is indicated by the solid diagonal line. The dotted line indicates the non-parametric line going through the decile specific estimates (triangles). The solid curve indicates the smoothed polynomial loess curve between the predicted risk and the event indicator. The spike histogram indicates the frequencies of the predicted risk.

Figure 2. Estimated effect of any chemotherapy compared to no chemotherapy on 5-month (left) and 1-year (right) mortality incidence in surgically treated dogs with osteosarcoma.

[figure 2]

Figure shows the odds ratio (OR) of any chemotherapy treatment (solid line) with 95% confidence intervals (dotted lines) for dogs with different predicted risks of 5-month mortality. The horizontal solid line indicates a neutral OR of 1.00. At the bottom a spike histogram is given, corresponding to the patient frequencies of the x-axis measurement.