Analysis of sarcoma and non-sarcoma clinical data with statistical methods and machine learning techniques

Kantidakis, G.

9 General discussion

This thesis focused on statistical analyses aimed at improving clinical trial design on behalf of the European Organisation for Research and Treatment of Cancer - Soft Tissue and Bone Sarcoma Group (EORTC - STBSG) and on investigating the potential of survival prediction models with machine learning techniques compared with statistical models. Sarcoma and non-sarcoma clinical data were used to compare the performance of the different prediction models. Results were presented in two different parts.

9.1 Part I: Clinical trials in soft-tissue sarcomas

In **Part** I, modern benchmarks were estimated to design new phase II clinical trials for common histotypes of locally advanced or metastatic soft-tissue sarcoma (STS) patients based on two meta-analyses which originated from an in-house literature review of STS (**Chapters** 2, 3). As an extra step, the prognostic significance of bone metastasis at study entry in STS was investigated to identify high-risk patient populations based on a pooled analysis of five EORTC - STBSG clinical trials (**Chapter** 4).

9.1.1 Designing new phase II trials

The historical benchmarking analysis by Van Glabbeke *et al.* (2002) provided pooled progression-free rates for various STS patients who participated in phase II trials of EORTC - STBSG (see table 1.1) [1]. These have been used to design a large number of new studies (currently more than 420 citations). This is the first gap that this thesis aimed to cover. We performed an extensive in-house literature search to identify all phase II, III, or IV clinical trials of advanced or metastatic STS (from 2003 to 2018). Due to the presence of heterogeneity among clinical trials, we decided to provide an update focusing on the most prevalent STS identified during the review. In **Chapter** 2 [2], we re-evaluated thresholds with a meta-analysis for first-line and pre-treated leiomyosarcoma patients (all or uterine only) using progression-free survival rate (PFSR) as the primary endpoint. Information was acquired from publications and sponsors. Reference values for parameter P_0 (null hypothesis) were calculated as the overall pooled PFSR at 3 and 6 months ((based on inactive and active agents), whereas minimum values to target for parameter P_1 (alternative hypothesis) were calculated using the recommended treatment effect for PFS by the ESMO Magnitude of Clinical Benefit Scale [3]. In **Chapter** 3, another meta-analysis was performed by using the same methodology, for advanced or metastatic liposarcoma and synovial sarcoma patients to provide benchmarks to design new phase II clinical trials with PFSR. Table 9.1 summarises P_0 and P_1 parameters based on the meta-analyses.

A need to raise the bar of thresholds is indicated for the commonest types of STS in future phase II trials by both of our studies, which aligns with the perspective of the American Society of Clinical Oncology [4]. The cost-benefit of new systemic therapies for cancer should be balanced against the societal resources in this era of rapidly rising

Treatment line and analysed group	3 months		6 months	
	Ref (P_0)	Min target (P_1)	Ref (P_0)	Min target (P_1)
First-line uterine LMS	71%	80%	49%	63%
First line all LMS	74%	82%	58%	70%
First-line LPS	69%	79%	56%	69%
First-line SS	74%	82%	56%	69%
Pre-treated uterine LMS	53%	66%	42%	57%
Pre-treated all LMS	48%	62%	28%	44%
Pre-treated LPS	49%	63%	28%	44%
Pre-treated SS	45%	60%	25%	41%

Table 9.1: **Treatment effect (PFSR) for the parameter P_0 (null hypothesis) and the parameter P_1 (alternative hypothesis) of a study**. LMS, leiomyosarcoma; LPS, liposarcoma; SS, synovial sarcoma. The meta-analysis for all LMS excluded trials designed for uterine LMS patients.

healthcare costs.

Strengths and limitations of the meta-analyses

To the best of our knowledge, these are the first attempts to meta-analyse the outcome of patients with advanced/metastatic leiomyosarcoma, liposarcoma, or synovial sarcoma for both first or further lines of treatment. Overall, a key strength of these two studies (**Chapter** 2, **Chapter** 3) is the collection of summary estimates for 1500 leiomyosarcoma (and 421 extra patients from clinical trials designed exclusively for uterine leiomyosarcoma), 1030 liposarcoma, and 348 synovial sarcoma patients from phase II, III, or even IV clinical trials.

Our results cannot be directly compared with the Van Glabbeke study for several reasons: (i) our meta-analyses used patients from phase II, III or IV clinical trials whereas the historical benchmarking analysis used patients who participated in phase II trials only, (ii) the primary endpoint shifted from progression-free rate to PFSR (counting any death as an event), (iii) the 2002 publication exploited individual patient data from the EORTC - STBSG database whereas we used summary estimates retrieved from the publications or the sponsors, (iv) Van Glabbeke *et al.* defined thresholds mixing all STS subgroups in the pre-treated setting, (v) the historical study calculated reference values for drug inactivity (P_0) and separately for activity (P_1) whereas we defined P_0 as the overall pooled rate (based on inactive and active agents) and estimated the values to target for P_1 based on the ESMO Magnitude of Clinical Benefit Scale.

Our studies have some limitations. First, the large majority of the trials were designed for several STS types and were therefore underpowered for specific subgroup analyses (i.e., for leiomyosarcoma). Moreover, effects seen in subgroups may not necessarily be visible in the overall study, thus not prompting an update of the recommendations for standard of care. In addition, some treatments were still in the stage of validation and were not part of the recommended treatments. This might explain the non-significant difference between recommended and non-recommended treatments based on the standard ESMO 2018 or 2021 guidelines [5, 6] for all leiomyosarcoma (first-line or pre-treated), first-line liposarcoma / synovial sarcoma, and pre-treated synovial sarcoma patients. Therefore, we used the overall pooled PFSR to define parameter P_0 in our meta-analyses for the sake of consistency. Secondly, the condition of any meta-analysis that the effect sizes between drugs of the same trial are independent may be violated in the randomised studies, as a random-effect model was used for each treatment regimen. Here, we observed high unexplained overall heterogeneity indicative of a large variation between effect sizes, which may limit our meta-analytic results. Thirdly, PFSRs were calculated based on summary estimates per treatment arm and treatment line that are less reliable than individual patient data but require a smaller amount of time to be collected from the different study sponsors. Moreover, in **Chapter** 3, liposarcomas were addressed as a single disease while it is known that there are three different histologic subtypes (e.g., well differentiated/dedifferentiated,

myxoid, or pleomorphic) that exhibit different clinical behavior and sensitivity to treatments. Yet, in older studies such information might have not been collected at the subtype level.

Towards a histology-tailored research

In the last decades, STS studies were designed based on the one-size-fits-all principle mixing several histologic subtypes. However, more recently research has shifted to a more histology-specific approach to better diversify the eligibility criteria of clinical trials [7–10]. This is also something we have noticed in our literature review (i.e, some studies designed only for uterine leiomyosarcoma / leiomyosarcoma, dedifferentiated liposarcoma / liposarcoma, angiosarcoma). An urgent need remains for the development of individualised treatment plans such as targeted therapy or immunotherapy to move away from the conventional chemotherapy options. Hence, new studies tailoring therapy to specific histological types should be based on modern thresholds for drug activity. With this work, we suggested a new benchmark to aid the design of phase II studies for leiomyosarcoma, liposarcoma, or synovial sarcoma patients using PFSR at 3 or 6 months as the primary endpoint.

The choice of primary endpoint in phase II

In general, the choice of the endpoint in phase II exploratory studies should be tailored on the disease and the drugs under investigation (mechanism of action, potential toxicity). Response-based endpoints defined by RECIST 1.1 [11] - such as the objective response rate - might be appropriate primary endpoints if unambiguous and clinically relevant antitumor activity (tumor shrinkage) is hypothesized by a drug or combination [12]. However, in contrast, if response-based endpoints are not appropriate, PFS (and/or time-to-progression) can be considered the primary endpoint as biological activity is frequently not expected to result in shrinkage of lesions, but rather in stabilisation of disease. In our meta-analysis databases, rather low response rates were observed (frequently $< 15\%$) for the majority of the drugs / drug combinations. This was expected for these STS types (leiomyosarcoma, liposarcoma, synovial sarcoma), as a decrease of tumor volume $> 30\%$ is unlikely with current standard of care for STS. Hence, PFS (or PFSR at 3 and 6 months) is a valuable alternative endpoint for the estimation of the biological antitumor activity of a new treatment (e.g., a targeted therapy) and to justify further investigation in confirmatory phase III trials [1].

The ultimate aim of trials

Clinical research has as an ultimate aim is to provide evidence of improved OS or improved quality of life. Nonetheless, strong surrogacy properties between PFS and OS are questionable based on two meta-analyses of randomised clinical trials with advanced STS [13, 14]. Therefore, PFS carries the risk of misleading conclusions because of erroneous extrapolation of the results, which might lead to exaggerated enthusiasm for a new anticancer therapy [15, 16]. On the other hand, PFS remains an attractive endpoint to identify benefit earlier than OS, and phase II trials are not intended to provide definite proof of the new treatment but rather a justification to further investigation. PFS (or the 3- or 6-month PFSR) can thus be used as the primary endpoint in phase II trials or as futility endpoint in phase III trials, but OS should remain the primary endpoint in phase III trials (whenever possible).

Challenges in rare cancers

Rare cancers account approximately for one-fifth of new cancer cases and patients typically have a poor median survival time [17, 18]. STS - the rare malignancies discussed in this thesis - account for about 1% of all adult tumours [19]. Over the years, more than 100 histologic subtypes have been recognised with widely varying presentation, sensitivity to treatment, and long-term outcomes [7]. Hence, each of these histological types and

subtypes is a rare or even an ultra-rare indication, which brings challenges to design new studies and to build clinical evidence to advance treatment options.

Some important challenges can be summarised as [20]:

- Lack of clinical expertise in hospitals. For instance, it is well-known that management of STS should be carried out in sarcoma reference centers and/or within reference networks that share multidisciplinary expertise [6].
- Smaller incentive by the drug developers due to market limitations. It is very hard to convince a pharmaceutical company to invest money for a histology-specific trial on an ultra-rare cancer subtype (e.g. myxoid liposarcoma in STS).
- A rigorous study design requires a large number of patients. According to conventional methodologies, clinical trials need considerable numbers of patients that are difficult to collect in rare cancers such as STS or their ultra-rare subtypes. Such studies require extensive international collaboration.
- Selection of experimental treatments is often based on inadequate evidence. Consequently, accommodating a higher than average degree of uncertainty is necessary for clinical as well as population-based decision making.
- Randomised controlled trials - which are usually the gold standard - might not be feasible due to limitations in patient availability. This is also a typical problem for sarcoma studies.
- Data collection to perform meta-analysis can be very time-consuming. Our meta-analyses in **Chapter** 2 and **Chapter** 3 provide useful information to aid the design of new single arm phase II exploratory studies for leiomyosarcoma, liposarcoma, and synovial sarcoma patients. For this work, we collected summary estimates from publications and sponsors ($n \geq 10$) which is typically less problematic than collecting individual patient data due to the General Data Protection Regulation (GDPR) restrictions. However, collecting data based on published literature might not be feasible for rarer STS entities. Collection of individual patient data is likely to be needed, which will increase the complexity and the time required to complete such a project.

Methodological recommendations for clinical trials in rare cancer are discussed in the paper by Casali *et al.* [17]. Methodologies should be refined to combine all the available evidence.

9.1.2 Patients with bone metastasis at diagnosis

We were also interested to expand on two previous studies of the EORTC - STBSG [21, 22] to identify high-risk patient populations in clinical trials performed by our group examining patient characteristics. Skeletal metastasis is part of the natural history affecting the prognosis and quality of life of patients with advanced/metastatic STS as a pathological fracture may occur in 20–30% of them together with other skeletal-related events [23]. Hence, in **Chapter** 4 [24], we investigated whether bone metastases at presentation affect overall survival (OS) and progression-free survival (PFS) per treatment line (first versus second line or higher) based on a pooled analysis of five EORTC - STBSG clinical trials [25–29]. Patients were enrolled from April 2003 to June 2015. For the subgroup of patients with bone metastasis presence (n = 140, 13.5%), the strongest metastatic combination was identified between bone and another site (liver, lymph node, lung, soft-tissue, other).

There is an increased occurrence of bone metastasis in advanced-stage cancers [30]. A metastatic bone profile may be part of STS patients' natural history, which negatively affects their prognosis. In this pooled analysis, patients with STS of the extremities, abdomen, thorax, visceral or other sites of primary tumour were included (n = 1034). The unadjusted difference in OS/PFS for patients with or without bone metastasis was statistically significant for first-line treatment (hazard ratios 1.55 [95%-CI: 1.19–2.01] and 1.43 [95% CI: 1.12–1.84] for OS and PSF, respectively). However, this difference was not significant when adjusting for 12 known prognostic factors (hazard ratios 1.33 [95%-CI: 0.99–1.78] and 1.31 [95%-CI: 1.00–1.73]). For second line or further, the unadjusted hazard ratios for OS / PFS were 1.10 (95%-CI: 0.81–1.49) / 1.06 (95% CI: 0.80–1.40) and the adjusted 1.11 (95%-CI: 0.81–1.52) / 1.07 (95%-CI: 0.80–1.43), respectively. An overall worse status is suggested for

patients suffering from bone metastasis in any treatment line, although the effect was not statistically significant.

In our database, 6/1034 patients had exclusively bone metastasis at diagnosis, and therefore a separate analysis of this small subgroup could not be performed. Tentative explanations of this small number could be that (i) bone lesions alone are typically challenging to measure and most trials require a measurable disease to assess response/progression per RECIST 1.1 criteria [11], (ii) bone metastasis at diagnosis is a sign of extensive disease. A limitation of this work is the retrospective exploratory nature. Both randomised and nonrandomised studies were pooled together from the EORTC - STBSG database to increase the statistical power, which is likely to have introduced some selection bias in the population.

A subgroup analysis was performed for 140 patients (13.5%) with bone metastasis at presentation to identify the metastatic organ site combination that is the most detrimental for OS and PFS after adjusting for demographic characteristics, histological entity, tumour grade, site of primary tumour, and time between histological diagnosis and registration/randomisation. A combined bone/lymph nodes metastatic presentation had the worst OS prognosis (hazard ratios 2.97 [1.53, 5.78] for first-line and 1.59 [0.55, 4.54] for second-line or higher treatment). For PFS, bone plus lung metastasis was the most detrimental combination (hazard ratios 2.80 [1.10, 7.09] and 2.45 [0.61, 9.84], respectively). Of note, such combinations were statistically significant for first-line treatment. Due to the limited number of patients per treatment line (77 first-line, 63 second-line or later), results should be interpreted with caution.

Our findings may have some implications on managing advanced/metastatic STS patients with bone metastasis at diagnosis. The impact of skeletal metastases was more profound in first-line treated population. Presence of bone metastasis at study entry was not ascertained (statistically) as a sufficiently important risk factor on its own for first or second line or further to justify stratification in randomised studies for these patients. As individuals with metastatic STS survive these days somewhat longer than 20–25 years ago due to advances in supportive and multidisciplinary care, the prevalence of bone invasion is difficult to be verified and an increase is likely to be observed in the future.

9.1.3 Future perspectives

In this subsection, we provide some recommendations for future research in STS.

The design of new studies

STS subtypes are very heterogeneous rare malignancies with widely varying presentation, sensitivity to treatment, and long-term outcomes [7, 31]. The work provided in **Chapter** 2 and **Chapter** 3 will help to optimise the design of new histology-tailored phase II trials for common histotypes (: leiomyosarcoma, liposarcoma, synovial sarcoma) in advanced or metastatic setting. Further meta-analyses should be designed and performed by using summary estimates or individual patient data for other frequent types such as undifferentiated pleomorphic sarcoma or angiosarcoma. Efficient data sharing could be key to simplify this process, as well as an international collaboration between data centers. We would like to stress the importance of designing robust and efficient phase II studies especially in rare cancers, where there is lack of evidence to support clinical decisions. Our projects have yielded a good historical control database with updated reference values which can be used to design new single arm STS studies (with a sufficient sample size and within a reasonable time-frame). This will lead to a better early evaluation of drug activity as an urgent need remains for the development of personalised treatment plans such as immunotherapy to move away from the conventional chemotherapy options.

Randomised controlled trials are considered to be the gold standard for clinical development, but are challenging to run for simple screening studies in STS mainly due to sample size requirements, and small incentive by pharmaceuticals. Too many randomised controlled phase III trials have recently failed in a rush for success [32]. The development of a new drug might be easier and faster with adaptive randomised trials such as the transformation

of a phase II into a phase III ("seamless phase II/III designs") in case of a positive early stage of a study, or the use of "drop-the-loser" or "play-the-winner" designs [17]. Rare tumors require extremely good and innovative trial design to make optimal use of every patient. Importantly, in order to make progress in this context, practice changing decisions will have to be based on less direct evidence than those in common malignancies and all available evidence should be evaluated. With our meta-analyses, we provided new robust benchmarks to speed up the process through single arm studies which are faster to run, and require fewer patients than randomised phase II trials.

As an example, the Pautier 2015 non-randomised phase II study [33] investigated the potential of adding trabectedin to doxorubicin for first-line advanced uterine or soft-tissue leiomyosarcoma patients. For soft-tissue leiomyosarcoma PFSRs were 90% and 81% at 3 and 6 months (for uterine leiomyosarcoma 87% and 72%), respectively. Therefore, this drug combination even if it was not recommended based on the ESMO 2018 guidelines [5] would qualify according to our updated rules for all or uterine first-line leiomyosarcoma patients (see minimum targets in table 9.1) suggesting a significant and relevant improvement over a standard of care in a prospective randomised phase III trial. As a verification, the results of the phase III trial with 67 uterine leiomyosarcoma and 83 soft-tissue leiomyosarcoma patients are positive [34]. There is a significant increase in PFS compared with the single doxorubicin arm, as well as a benefit in terms of OS. This highlights the pertinence of good historical benchmarks for simple screening phase II studies, which will lead to successful phase III confirmatory trials.

The importance of collaboration and funding

The rarity and heterogeneity of STS highlights the importance of a multidisciplinary approach (i.e., a specialised team of radiologists, pathologists, surgical and medical oncologists) for the management of the disease [7]. A major strength of sarcoma research is collaboration in an academic climate. Progressive multinational collaborative efforts (many sites in several countries) will allow both sufficiently large and sufficiently focused studies to generate high-quality evidence in specific STS patient populations. Such efforts are more than necessary when fighting rare cancers. The EORTC offers an integrated approach to drug development and drug evaluation programs through translational and clinical research of its rich database. The EORTC - STBSG can connect people and sites, as well as start working groups with other national sarcoma societies / groups to reinforce the academic climate and lead to major breakthroughs. The STBSG and other national sarcoma groups should strengthen the relationship with the European Medical Agency (EMA) and U.S. Food and Drug Administration (FDA) and discuss with the regulators regarding the conditions for a new drug approval. In addition, the development of new statistical methodology is more than welcomed to address research questions, and strong collaborations between clinicians and statisticians are key to bring new project ideas into maturity. Overall, in rare cancers such as STS, flexibility and out-of-the-box thinking are required to advance research.

Simultaneously, it is necessary to initiate high value and high-quality company sponsored trials through the interaction with pharmaceutical companies. Major congresses e.g. the European Society of Medical Oncology (ESMO), the European Society for Radiation Oncology (ESTRO) can be used as platforms for this interaction. Next to that, it is also of great importance to secure funding for academic studies and search for national grants or support from industry. The EORTC 1809 STRASS II trial in patients with high-risk retroperitoneal sarcoma belongs to an alternative pipeline of purely academic trials. It addresses a significant unmet need (chemotherapy plus surgery versus surgery only) in this population [35]. The trial recently received support by the Anticancer Fund (ACF) a Belgian foundation of public utility dedicated to expanding the range of treatment options available to cancer patients regardless of commercial value. This shows how essential is the allocation of research funds to areas with limited interest from the profit-driven cancer industry to complement progress.

9.1.4 In conclusion

The first part of this thesis yielded modern benchmarks of efficacy to design new phase II clinical trials for locally advanced or metastatic leiomyosarcomas, liposarcomas, and synovial sarcomas. These meta-analyses were essen-

tial to update and re-evaluate well-established historical thresholds by the EORTC - STBSG. Expanding previous STBSG studies, skeletal metastasis at baseline was found to be detrimental for both OS and PFS in any treatment line with a more profound effect in first-line population, although not statistically significant.

9.2 Part II: Statistical models versus machine learning to predict survival for sarcoma and non-sarcoma clinical data

In **Part** II (**Chapters** 5, 6, 7, 8), the potential of existing and new machine learning (ML) methodologies was explored for survival prediction, and compared with traditional statistical models (SM) for real-life clinical data (small/medium or large sample sizes, low- or high-dimensional settings).

9.2.1 Machine learning versus statistical modelling

Since the last decade, machine learning (ML) has received increased attention in the medical area. The aim of prediction has been of particular interest as part of a growing trend towards personalised medicine [36]. However, concerns have been raised that the employment of ML techniques and artificial intelligence in general is over-hyped in some contexts (e.g., over-fitting the training data, lack of attention towards validation, unsuitable performance measures). Overall, from thousands of publications applying ML to medical data, very few algorithms have meaningfully contributed to clinical care [37]. One of the main reasons is that if a model fails in healthcare the consequences are life-threatening, and thus robust evidence is required [38].

In this thesis, ML methods were explored for prediction of survival data. Due to the presence of censored observations, the extension of ML methods to survival data is not straightforward. Over the years, a variety of methods have been proposed which are adaptations of the ML classifiers (for example random survival forests from random forests, survival neural networks from neural networks [39]). We investigated the potential of these techniques in contrast with conventional statistical benchmarks for time-to-event data and proposed a new extension to the partial logistic artificial neural networks methodology.

A review and critical appraisal

In **Chapter** 5, we performed - to the best of our knowledge - the first ever large-scale review on survival neural networks (SNNs) using prognostic factors for clinical prediction. Our goal was to provide a broad understanding of the literature (1st January 1990 - 31st August 2021). A total of 24 articles were identified based on a global search in PubMed. Relevant manuscripts were classified as methodological/technical (novel methodology or new theoretical model; 13 studies) or applications (11 studies). We discussed how SNNs are employed for prediction in the medical field and how researchers have tried to adapt a classification method to right-censored survival data. There are two methodological trends: either time is added as part of the input features and a single output node is specified, or multiple output nodes are defined for each time interval.

This work was supplemented with a critical appraisal to pinpoint current limitations and identify future research directions. Regarding some general characteristics of the studies, the median total sample size was 920 patients and the median number of predictors was 7 (low-dimensional data). Medical applications were mainly in the field of oncology (73.5%, 25 datasets). The strategy used to address the missing data (if any) was unclear for 9/21 (42.9%) studies (without 3 simulation studies). Major findings included inaccurate model development/validation and poor reporting. In the majority of studies (15, 62.5%), the approach to tune hyperparameters was unclear. The performance criterion for model development was unclear for 6 studies (25.0%). Programming language used for the development of the SNN was unclear in 7 studies (29.2%). We noticed large variability and improper performance measures for survival data, as well as lack of confidence intervals for the predictive measures in 13 of

the 24 studies (54.2%). Calibration plots were available for only 11 studies (45.9%). All in all, 19 studies reported comparisons between Cox models and SNNs from which 15 studies (78.9%) did not consider interaction terms between the predictors.

According to these findings, some general recommendations are provided:

- Complete and transparent reporting of modelling steps and analysis is necessary (e.g., more details on training and test data), to enable reproducibility, and to allow critical appraisal of the results by a wider audience [40, 41].
- Hyperparameter selection and training should be more extensive with the performance criterion for model development clearly reported. A suitable performance measure should take into account the censoring mechanism.
- More attention to model calibration is urgently needed. Calibration should be assessed preferably through calibration plots.
- Larger datasets and/or more predictors are needed for better model development/validation and improved generalisability.
- Comparisons of SNNs with conventional regression models should be made in a fair manner, with the conventional models fully developed and interactions and/or non-linear terms included when appropriate.
- Further methods and guidelines for obtaining confidence intervals are needed. Multiple resampling of all empirical data using bootstrapping can be an advantageous approach when sample size is limited, as it avoids the need to split the data for model development and provides confidence intervals [42].
- Increasing the complexity of a SNN (or a ML prediction model) does not necessarily translate to improved performance on new clinical data. For such survival data, sample size and number of predictors is likely to be insufficient for employing advanced techniques.

These aspects are of great value as suboptimal clinical prediction models with ML or statistical modelling are responsible for research waste [43, 44].

Comparison between methods in different settings

A main objective of this thesis was to compare machine learning (ML) techniques with statistical models (SM) for time-to-event prediction models in the presence of right-censored medical data (liver transplantation, osteosarcoma, and STS).

Chapter 6 [45] provided the first ever study where ML techniques were tested on complex post-transplant liver data (large sample size, high-dimensional setting) from the United States and compared with traditional Cox models. Random survival forests (RSF) [46] and two novel extensions of the partial logistic artificial neural network (PLANN) [47] (neural networks with one hidden or two hidden layers) were applied to retrospective data (n = 62294, p = 97) provided by the Scientific Registry for Transplant Recipients to predict survival. These methods were compared versus three Cox models (with all variables, backward selection and LASSO) [48, 49]. Clinical endpoint was overall graft-survival (time between transplantation and the date of graft-failure or death). To assess the final predictive performance of the models, the concordance index [50], Brier score [51], and Integrated Brier Score (IBS) [52] were used. The strongest prognostic factors were identified for each model. RSF performed better than the Cox models in terms of C-index (0.622 for RSF versus $\leq$ 0.62 for Cox models). This shows the ability of RSF to discriminate better between low and high risk groups of patients. The Brier score was measured at each year for all methods. RSF showed results similar to the Cox models having slightly smaller total prediction error on the test data (IBS 0.182 versus 0.183). The neural networks (IBS 0.180) performed in general better than the Cox models or the RSF and had very similar performance over time. From the three ML techniques, PLANN extended with one hidden layer predicted survival probabilities most accurately. Its calibration was very similar to the Cox model with all variables. The RSF and the PLANN extended with two hidden layers were less calibrated on test data. Special emphasis was given on the interpretation of the models. An indirect comparison was performed to examine which are the most prognostic variables for a Cox model with all variables, a RSF and

the two PLANNs extended. The Cox model with all variables and the PLANNs identified *re-transplantation* as the strongest predictor and *donor age*, *diabetes*, *life support* and *race* as relatively strong predictors. According to RSF, the most prognostic variables were *donor age*, followed by *re-transplantation*, *life support* and *serology status of Chronic hepatitis C virus*.

In **Chapter** 7 [53], SM were compared with ML for non-complex clinical data (small/medium sample size, low-dimensional setting) to investigate a different real-life setting. The dataset originated from a randomised phase III European Osteosarcoma Intergroup study (MRC BO06 / EORTC 80931) that investigated the effect of dose-intense chemotherapy in patients with localised extremity osteosarcoma [54]. A Monte-Carlo simulation study was performed to compare PLANN original [47] or extended (with one hidden layer) [45] with Cox models [48]. Real-life clinical data was mimicked to simulate synthetic data (5 predictors, 250 or 1000 observations) and to address different scenarios (20, 40, 61, or 80% censoring) in the absence of complex functional dependence relationships involving time and covariates. The endpoint of interest was overall survival (time to death from any cause since surgery). Models were evaluated in terms of C-index [50], Brier score at 0-5 years [51], IBS at 5 years [52], and miscalibration at 2 and 5 years in the simulated test datasets. It was shown that SNNs may reach a comparable performance in terms of the C-index, Brier score, or IBS. The standard deviations (over 1000 repetitions) overlapped to a large extent for all scenarios. Predictive performance improved (smaller Brier scores, higher C-indexes) when the sample size increased for all methods. Predictive ability was adequately robust to predefined adverse scenarios on training data (removing patients censored before 2 years, administrative censoring at 5 years). However, the Cox models were usually better calibrated (predicted survival probabilities were closer to the observed). This highlights in particular the relevance of reporting calibration of ML techniques to obtain a neutral comparison with SM. Miscalibration was rather strong for a larger percentage of censoring (less events).

In **Chapter** 8, the comparison between ML techniques and traditional SM was extended to competing risks (CRs) framework in another simple clinical setting (small/medium sample size, low dimensional data). Our aim was to develop and validate clinical prediction models for CRs with the first ever study of this kind in STS. A dataset with 3826 retrospectively collected patients with extremity STS and nine predictors from the PERsonalised SARcoma Care (PERSARC) Study Group was used. Three ML techniques a) PLANN for CRs (PLANNCR) original [55], b) PLANNCR extended (a new method developed by the authors), and c) RSF for CRs (RSFCR) [56] as well as two statistical models i) cause-specific Cox [48], and ii) Fine-Gray model [57] were compared. The endpoint of interest was time in years between surgery and disease progression (event of interest) or death (competing event). Predictive performance of the methods was assessed for the event of interest and the competing event in 100 validation datasets based on the Area Under the Curve (AUC) and the Brier score at 2, 5, or 10 years (t-year predicted risks evaluation) [58]. Miscalibration (absolute predictive accuracy) was estimated at the same time points [59]. Results showed that ML models have similar performance with SM in terms of Brier score and AUC at 2, 5, and 10 years for disease progression and death (95% confidence intervals overlapped). From the three ML models, predictive ability of PLANNCR extended was usually better than RSFCR and PLANNCR original especially in terms of AUC. This means that PLANNCR extended was able to better discriminate between low and high risk groups of patients. Nevertheless, the SM were frequently better calibrated than the three ML methods. Miscalibration of PLANNCR original and extended was more pronounced for the competing event (death). These findings indicate that more attention to model calibration is urgently needed for ML methods.

Advantages and disadvantages

Pros and the cons of SM and ML techniques for survival analysis are presented below in terms of interpretability, flexibility, and practical utility.

(a) **Interpretability**: Cox model - the well-established statistical benchmark for survival data - offers a straightforward interpretation via hazard ratios, which is very useful for clinicians to make informed decisions. On the contrary, ML techniques usually have limited interpretability. For example, in **Chapter** 6 we used the relative importance and the variable importance methods for neural networks and RSF to extract model interpretation, respectively [45]. However, these cannot indicate whether the effect of a covariate is pro-

tective or not.

(b) **Flexibility**: ML techniques make minimal assumptions and are very flexible as they can model automatically non-linear relationships between variables. On the other hand, SM are less flexible as they make some usually strong assumptions for the modelling process. For instance, the Cox model assumes proportionality of hazards over time and additivity of effects (as any regression model). Hence, any interactions between variables need to be manually pre-specified, which can be problematic in the presence of many variables and multiway interactions.

(c) **Practical utility**: SM such as the Cox model have a fast implementation in popular open-source programming languages such as `R` or `Python`. Then again, ML techniques require a nontrivial implementation time for data pre-processing, tuning of hyperparameters (the larger the number of hyperparameters the more the time and effort needed for model training), and are computationally more intensive to run. Typically, calculation complexity is based on sample size and predictors multiplicity. Moreover, model optimisation of neural networks is a delicate task which requires robust numerical methods and skillful use, else the network might converge in suboptimal minima in the error function [60, 61].

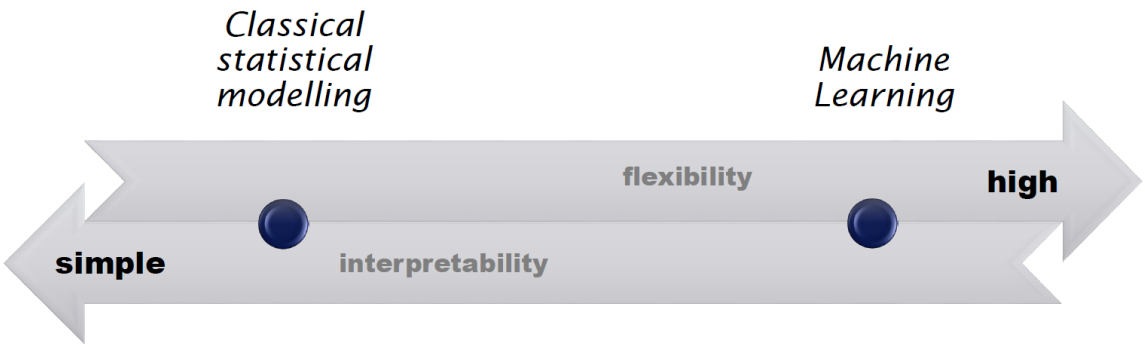

Figure 9.1: The trade-off between model interpretability and flexibility with SM and ML techniques (Wallisch *et al.* 2019 [62]).

Figure 9.1 presents the trade-off between simple interpretability and high flexibility with SM and ML techniques. Classical SM are typically placed on the left side of the arrows (simple interpretability, low flexibility). By adding interactions and non-linear terms, the flexibility of a model can be increased at a cost of interpretability. ML models are usually positioned towards the right side of the arrows (complicated interpretability, high flexibility). However, there are some ML techniques such as survival trees which can be placed near the left side. In general, an increase in model complexity will lead to higher flexibility, but a more complicated interpretability. SNNs developed in recent years usually have more complicated structures and make use of multiple hidden layers (deep learners). It should be noted, however, that increasing the complexity of an ML prediction model does not necessarily translate to improved performance on new clinical data. An increase in the complexity, and by extent flexibility of a network, may produce a model that is too attuned to the training data with poorer generalization to new data (overfitting), with the extra cost of a more limited interpretability.

Key considerations

Table 9.2 presents some important considerations for choosing between SM and ML. A much larger sample size might be necessary when using ML approaches to develop risk prediction models [42]. A main cause for this problem is that the number of predictor (also called ”feature”) parameters considered by ML will usually be substantially larger than that for regression (even with the same set of predictors) as these techniques model automatically second and higher order interaction terms. Therefore, they might actually need ”big data” to ensure small overfitting of the developed models [63]. On the contrary, the sample size of most medical research datasets is

more appropriate for regression models such as the Cox proportional hazards model for time-to-event data. Furthermore, regression models provide a transparent model equation and allow for a straightforward interpretation which is likely to be crucial for clinicians to implement a model going forward in routine care practice.

Statistical models	Machine Learning
Focus on relationships between independent variables	Focus on prediction
Favor additivity of predictor effects	No special emphasis to additivity of effects
Low uncertainty tolerance (confidence intervals, hypothesis testing, assumptions)	High uncertainty tolerance (adaptability, no assumptions)
Small - medium sample size	Large - huge sample size
Small number of predictors	Large number of predictors
Semi-parametric, parametric models	Non-parametric models
Simple interpretability	Complicated interpretability
Low flexibility (small number of interactions pre-specified)	High flexibility (higher order interactions presence is expected)
Low signal to noise ratio (human outcomes, weather forecasting)	High signal to noise ratio (image recognition, playing games)

Table 9.2: **Key considerations for choosing between SM and ML** [64, 65]. The first columns shows when SM might be the better choice for analysis, whereas the second column indicates when ML can be the better choice.

9.2.2 Future perspectives

In this subsection, we discuss about the future of ML for time-to-event data.

Challenges of machine learning

While throughout the years a substantial increase in ML articles in medical research can be observed, there are few algorithms actually implemented in clinical practice [37, 66]. Below, we report some major challenges which need to be addressed prior to the establishment of ML methods.

- **Access to medical data.** Clinical data are (rightfully) challenging to access. An informed consent and ethics committee approval, as well as appropriate handling (anonymisation) are required before data sharing [67]. Limited public data availability also hinders the validation of ML algorithms and their results by other authors. Findings obtained with a set of methodologies are not easily comparable with different datasets (as they reflect other real-life settings).
- **Model complexity in relation to the amount of available data.** Clinical data typically include tens to thousands of patients. This number is usually not sufficiently large for the employment of modern flexible ML techniques such as neural networks which are data hungry [63, 68]. Hence, there is a substantial risk of overfitting (excessive tailoring of algorithms on training data), thus limiting the ability of these models to generalise (accurately perform) on new data. The use of shallower neural networks (one hidden layer) could reduce the risk of overfitting. ML techniques should preferably be employed for large datasets.
- **Clinical validation.** It is the most critical component of performance assessment. The presence of missing data, regional variations in practice, and logistical/infrastructural limitations may complicate model validation [69]. If possible, an external validation (e.g., using population from different centers) should follow a meticulous internal validation. Transparent reporting of ML prediction models is required [40, 44]. Both aspects of model discrimination and calibration should be taken into account for performance evaluation

[41, 70]. Advantages and shortcomings in comparison with traditional benchmarks (SM) should be documented. Recently, researchers are increasingly aware of the scientific rigor needed to be demonstrated prior to the employment of a ML model in clinical practice.

- **Model interpretation.** This is a key obstacle for the integration of ML methods in medicine. For example, (deep) neural networks are called "black boxes" since they develop complex internal functions that are hard to interpret. The extraction of a meaningful model interpretation is critical for clinicians to take informed decisions and to trust these models in routine care. A common metric is needed to directly compare SM with ML. Research in this area will likely be on the spotlight for many years.

Research directions

In this thesis, we investigated the predictive ability of different ML techniques: RSF [46], PLANN original [47], PLANN extended [45, 53] for right-censored clinical data. Models were compared with Cox proportional hazards models [48] in different real-life settings. The work was also extended for CRs examining RSFCR [56], PLANNCR original [55], and PLANNCR extended versus the cause-specific Cox model and the Fine-Gray model [57]. In the future, research for ML models should also focus on left- and interval-censored data.

Several ML models have been developed and applied to deal with right-censored medical data. Wang *et al.* provide a comprehensive review of commonly used methods in survival analysis [39]. Such ML techniques include survival trees [71], SNNs some of which have been included in our review (**Chapter** 5, e.g. [72, 73]), support vector machines [74], or other ensemble methods such as boosting [75]. It would be interesting to compare the predictive ability of these techniques with traditional statistical benchmarks in a variety of settings (small or large sample size, low or high dimensional data) with actual or synthetic data (simulation studies) to establish their potential role in clinical practice.

Both traditional Cox models and PLANNs allow for the inclusion of time-dependent covariates. Cox models can incorporate these variables in standard software. PLANNs can naturally incorporate time-dependent covariates due to the essential data transformation into a long format for each patient. In the future, it might be useful to compare the predictive ability of these models for time-dependent variables (also between the cause-specific Cox model and PLANNCRs for competing events). Moreover, RSF (for a single event), and RSFCR, Fine-Gray model (for competing events) can be extended to provide dynamic predictions with time-dependent covariates by creating a landmark dataset at a set of landmark time points t_{LM} [52, 76].

Last but not least, a subsequent step after a rigorous evaluation of a new prediction model with ML is its implementation in open source software. Popular programming languages such as `R` or `Python` can be used to integrate new methodologies. Research should also focus on the uniformisation of packages between different software. This will make a larger number of methods widely accessible, and will lead more academics and other interested individuals to engage with them. As a consequence, this competition will promote good standards of practice, and will help with the development of more methodological extensions for survival data.

Fuelling the debate

ML and artificial intelligence will be on the spotlight of medical research for many years. Hence, it is of paramount importance to develop methods and assess their predictive performance against conventional statistical benchmarks. Appropriate performance measures should be selected to evaluate model discrimination / calibration. For survival data the censoring mechanism should be taken into account [52]. A complete and transparent reporting of all modelling steps is required to enable critical appraisal and allow reproducible analysis. Researchers should follow the best practice guidelines and recommendations on clinical prediction models [40, 41, 44, 77, 78].

Another necessary aspect to consider is the intended purpose of a ML prediction model in healthcare. Existing or new methods should address unmet needs such as research, benchmarking, or bedside application. Perhaps a model has some novelty (addressing diseases or outcomes where information is not available), or it shows clinical

usefulness improving discrimination (separation of low- from high-risk patients) compared with current practice [78]. Or perhaps the application of ML techniques is motivated by exploration of the collected medical data to assess linear and additive model assumptions. Being aware of the advantages / disadvantages and key aspects of ML techniques and SM can help the reader to take meaningful decisions regarding the choice of methodology, and provide her with enough stimuli to seek for new developments.

9.2.3 In conclusion

In the second part of this thesis, we performed a review and critical appraisal that shed light on the current state of art of SNNs in medicine with prognostic factors for clinical prediction (January 1990 to August 2021). Furthermore, we extended existing methods in ML and compared ML techniques with SM for prediction in different real-life settings. For complex liver transplantation data (large sample size, many predictors), it was shown that ML techniques can be a useful tool for both prediction (discrimination / calibration) and interpretation in the survival context. In a simulation study performed for synthetic osteosarcoma data in a simple setting (small / medium sample size, small number of predictors), SNNs reached a comparable predictive performance with Cox models but were generally less well calibrated. In another study with extremity STS data in a simple clinical setting with competing events, ML methods were able to reach a comparable performance with traditional regression models but the latter were frequently better calibrated. Therefore, more attention to model calibration is urgently needed for ML.

References

[1] M. Van Glabbeke, J. Verweij, I. Judson, and O. S. Nielsen. Progression-free rate as the principal end-point for phase II trials in soft-tissue sarcomas. *European Journal of Cancer*, 38(4):543–549, 2002. doi: 10.1016/S0959-8049(01)00398-7.

[2] G. Kantidakis, S. Litière, A. Neven, M. Vinches, I. Judson, P. Schöffski, E. Wardelmann, S. Stacchiotti, L. D'Ambrosio, S. Marréaud, W. T. A. van der Graaf, B. Kasper, M. Fiocco, and H. Gelderblom. Efficacy thresholds for clinical trials with advanced or metastatic leiomyosarcoma patients: A European Organisation for Research and Treatment of Cancer Soft Tissue and Bone Sarcoma Group meta-analysis based on a literature review for soft-tissue sarcoma. *European Journal of Cancer*, 154:253–268, 2021. ISSN 18790852. doi: 10.1016/j.ejca.2021.06.025.

[3] N. I. Cherny, U. Dafni, J. Bogaerts, N. J. Latino, G. Pentheroudakis, J. Y. Douillard, J. Tabernero, C. Zielinski, M. J. Piccart, and E. G. E. de Vries. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Annals of Oncology*, 28(10):2340–2366, 2017. doi: 10.1093/annonc/mdx310.

[4] L. M. Ellis, D. S. Bernstein, E. E. Voest, J. D. Berlin, D. Sargent, P. Cortazar, E. Garrett-Mayer, R. S. Herbst, R. C. Lilenbaum, C. Sima, A. P. Venook, M. Gonen, R. L. Schilsky, N. J. Meropol, and L. E. Schnipper. American society of clinical oncology perspective: Raising the bar for clinical trials by defining clinically meaningful outcomes. *Journal of Clinical Oncology*, 32(12):1277–1280, 4 2014. ISSN 15277755. doi: 10.1200/JCO.2013.53.8009.

[5] P. G. Casali, N. Abecassis, H. T. Aro, S. Bauer, R. Biagini, S. Bielack, S. Bonvalot, I. Boukovinas, J. V. M. G. Bovee, T. Brodowicz, J. M. Broto, A. Buonadonna, E. De Álava, A. P. Dei Tos, X. G. Del Muro, P. Dileo, M. Eriksson, A. Fedenko, V. Ferraresi, A. Ferrari, S. Ferrari, A. M. Frezza, S. Gasperoni, H. Gelderblom, T. Gil, G. Grignani, A. Gronchi, R. L. Haas, B. Hassan, P. Hohenberger, R. Issels, H. Joensuu, R. L. Jones, I. Judson, P. Jutte, S. Kaal, B. Kasper, K. Kopeckova, D. A. Krákorová, A. Le Cesne, I. Lugowska, O. Merimsky, M. Montemurro, M. A. Pantaleo, R. Piana, P. Picci, S. Piperno-Neumann, A. L. Pousa, P. Reichardt, M. H. Robinson, P. Rutkowski, A. A. Safwat, P. Schöffski, S. Sleijfer, S. Stacchiotti, K. Sundby Hall, M. Unk, F. Van Coevorden, W. T. A. Van Der Graaf, J. Whelan, E. Wardelmann, O. Zaikova, and J. Y. Blay. Soft tissue and visceral sarcomas: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 29(Supplement_4):iv51–iv67, 2018. ISSN 15698041. doi: 10.1093/annonc/mdy321.

[6] A. Gronchi, A. B. Miah, A. P. Dei Tos, N. Abecassis, J. Bajpai, S. Bauer, R. Biagini, S. Bielack, J. Y. Blay, S. Bolle, S. Bonvalot, I. Boukovinas, J. V. M. G. Bovee, K. Boye, B. Brennan, T. Brodowicz, A. Buonadonna, E. De Álava, X. G. Del Muro, A. Dufresne, M. Eriksson, F. Fagioli, A. Fedenko, V. Ferraresi, A. Ferrari, A. M. Frezza, S. Gasperoni, H. Gelderblom, F. Gouin, G. Grignani, R. Haas, A. B. Hassan, S. Hecker-Nolting, N. Hindi, P. Hohenberger, H. Joensuu, R. L. Jones, C. Jungels, P. Jutte, L. Kager, B. Kasper, A. Kawai, K. Kopeckova, D. A. Krákorová, A. Le Cesne, F. Le Grange, E. Legius, A. Leithner, A. Lopez-Pousa, J. Martin-Broto, O. Merimsky, C. Messiou, O. Mir, M. Montemurro, B. Morland, C. Morosi, E. Palmerini, M. A. Pantaleo, R. Piana, S. Piperno-Neumann, P. Reichardt, P. Rutkowski, A. A. Safwat, C. Sangalli, M. Sbaraglia, S. Scheipl, P. Schöffski, S. Sleijfer, D. Strauss, S. Strauss, K. Sundby Hall, A. Trama, M. Unk, M. A. J. van de Sande, W. T. A. van der Graaf, W. J. van Houdt, T. Frebourg, P. G. Casali, and S. Stacchiotti. Soft tissue and visceral sarcomas: ESMO–EURACAN–GENTURIS Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, 32(11):1348–1365, 2021. ISSN 15698041. doi: 10.1016/j.annonc.2021.07.006. URL https://doi.org/10.1016/j.annonc.2021.07.006.

[7] A. C. Gamboa, A. Gronchi, and K. Cardona. Soft-tissue sarcoma in adults: An update on the current state of histiotype-specific management in an era of personalized medicine. *CA: A Cancer Journal for Clinicians*, 70(3):200–229, 2020. ISSN 0007-9235. doi: 10.3322/caac.21605.

[8] N. T. Hoang, L. A. Acevedo, M. J. Mann, and B. Tolani. A review of soft-tissue sarcomas: Translation of biological advances into treatment measures. *Cancer Management and Research*, 10:1089–1114, 2018. ISSN 11791322. doi: 10.2147/CMAR.S159641.

[9] A. M. Frezza, S. Stacchiotti, and A. Gronchi. Systemic treatment in advanced soft tissue sarcoma: What is standard, what is new. *BMC Medicine*, 15(1):1–12, 2017. ISSN 17417015. doi: 10.1186/s12916-017-0872-y.

[10] A. Smrke, Y. Wang, and C. Simmons. Update on systemic therapy for advanced soft-tissue sarcoma. *Current Oncology*, 27(s1):25–33, 2020. ISSN 17187729. doi: 10.3747/CO.27.5475.

[11] E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009. ISSN 09598049. doi: 10.1016/j.ejca.2008.10.026. URL http://dx.doi.org/10.1016/j.ejca.2008.10.026.

[12] L. Seymour, S. P. Ivy, D. Sargent, D. Spriggs, L. Baker, M. J. Ratain, M. L. Blanc, D. Stewart, J. Crowley, J. S. Humphrey, P. West, and D. Berry. The Design of Phase II Clinical Trials Testing Cancer Therapeutics: Consensus Recommendations from the Clinical Trial Design Task Force of the National Cancer Institute Investigational Drug Steering Committee. *Clinical Cancer Research*, 16(6):1764–1769, 2010. doi: 10.1158/1078-0432.CCR-09-3287.The.

[13] M. Savina, S. Litière, A. Italiano, T. Burzykowski, F. Bonnetain, S. Gourgou, V. Rondeau, J. Y. Blay, S. Cousin, F. Duffaud, H. Gelderblom, A. Gronchi, I. Judson, A. Le Cesne, P. Lorigan, J. Maurel, W. van der Graaf, J. Verweij, S. Mathoulin-Pélissier, and C. Bellera. Surrogate endpoints in advanced sarcoma trials: A meta-analysis. *Oncotarget*, 9(77):34617–34627, 2018. ISSN 19492553. doi: 10.18632/oncotarget.26166.

[14] K. Tanaka, M. Kawano, T. Iwasaki, I. Itonaga, and H. Tsumura. Surrogacy of intermediate endpoints for overall survival in randomized controlled trials of first-line treatment for advanced soft tissue sarcoma in the pre- and post-pazopanib era: A meta-analytic evaluation. *BMC Cancer*, 19(1):1–9, 2019. ISSN 14712407. doi: 10.1186/s12885-019-5268-2.

[15] W. D. Tap, R. L. Jones, B. A. Van Tine, B. Chmielowski, A. D. Elias, D. Adkins, M. Agulnik, M. M. Cooney, M. B. Livingston, G. Pennock, M. R. Hameed, G. D. Shah, A. Qin, A. Shahir, D. M. Cronier, R. Ilaria, I. Conti, J. Cosaert, and G. K. Schwartz. Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial. *The Lancet*, 388(10043):488–497, 7 2016. ISSN 1474-547X. doi: 10.1016/S0140-6736(16)30587-6. URL https://pubmed.ncbi.nlm.nih.gov/27291997/.

[16] W. D. Tap, A. J. Wagner, P. Schöffski, J. Martin-Broto, A. Krarup-Hansen, K. N. Ganjoo, C. C. Yen, A. R. Abdul Razak, A. Spira, A. Kawai, A. Le Cesne, B. A. Van Tine, Y. Naito, S. H. Park, A. Fedenko, Z. Pápai, V. Soldatenkova, A. Shahir, G. Mo, J. Wright, and R. L. Jones. Effect of Doxorubicin Plus Olaratumab vs Doxorubicin Plus Placebo on Survival in Patients With Advanced Soft Tissue Sarcomas: The ANNOUNCE Randomized Clinical Trial. *JAMA*, 323(13):1266–1276, 4 2020. ISSN 1538-3598. doi: 10.1001/JAMA.2020.1707. URL https://pubmed.ncbi.nlm.nih.gov/32259228/.

[17] P. G. Casali, P. Bruzzi, J. Bogaerts, J. Y. Blay, M. Aapro, A. Adamous, A. Berruti, J. Bressington, B. Bruzzi, R. Capocaccia, F. Cardoso, J. E. Celis, A. Cervantes, F. Ciardiello, C. Claussen, M. Coleman, S. Comis, S. Craine, D. De Boltz, F. De Lorenzo, A. P. Dei Tos, G. Gatta, J. Geissler, R. Giuliani, E. Grande, A. Gronchi, S. Jezdic, B. Jonsson, L. Jost, H. Keulen, D. Lacombe, G. Lamory, Y. Le Cam, S. Leto di Priolo, L. Licitra, F. Macchia, A. Margulies, S. Marreaud, G. McVie, S. Narbutas, K. Oliver, N. Pavlidis, J. Pelouchova, G. Pentheroudakis, M. Piccart, M. A. Pierotti, G. Pravettoni, K. Redmond, P. Riegman, M. P. Ruffilli, D. Ryner, S. Sandrucci, M. Seymour, V. Torri, A. Trama, S. Van Belle, G. Vassal, M. Wartenberg, C. Watts, A. Wilson, and W. Yared. Rare Cancers Europe (RCE) methodological recommendations for clinical studies

in rare cancers: A European consensus position paper. *Annals of Oncology*, 26(2):300–306, 2015. ISSN 15698041. doi: 10.1093/annonc/mdu459.

[18] Jan Bogaerts, Matthew R. Sydes, Nicola Keat, Andrea McConnell, Al Benson, Alan Ho, Arnaud Roth, Catherine Fortpied, Cathy Eng, Clare Peckitt, Corneel Coens, Curtis Pettaway, Dirk Arnold, Emma Hall, Ernie Marshall, Francesco Sclafani, Helen Hatcher, Helena Earl, Isabelle Ray-Coquard, James Paul, Jean Yves Blay, Jeremy Whelan, Kathy Panageas, Keith Wheatley, Kevin Harrington, Lisa Licitra, Lucinda Billingham, Martee Hensley, Martin McCabe, Poulam M. Patel, Richard Carvajal, Richard Wilson, Rob Glynne-Jones, Rob McWilliams, Serge Leyvraz, Sheela Rao, Steve Nicholson, Virginia Filiaci, Anastassia Negrouk, Denis Lacombe, Elisabeth Dupont, Iris Pauporté, John J. Welch, Kate Law, Ted Trimble, and Matthew Seymour. Clinical trial designs for rare diseases: Studies developed and discussed by the International Rare Cancers Initiative. *European Journal of Cancer*, 51(3):271–281, 2015. ISSN 18790852. doi: 10.1016/j.ejca.2014.10.027.

[19] M. E. Kallen and J. L. Hornick. The 2020 WHO classification: What's new in soft tissue tumor pathology? *American Journal of Surgical Pathology*, 45(1):1–23, 1 2021. ISSN 15320979. doi: 10.1097/PAS.0000000000001552. URL https://pubmed.ncbi.nlm.nih.gov/32796172/.

[20] K. S. Panageas. Clinical trial design for rare cancers - why a less conventional route may be required. *Expert review of clinical pharmacology*, 8(6):661–663, 11 2015. ISSN 17512441. doi: 10.1586/17512433.2015.1088382. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4724195/.

[21] L. H. Lindner, S. Litière, S. Sleijfer, C. Benson, A. Italiano, B. Kasper, C. Messiou, H. Gelderblom, E. Wardelmann, A. Le Cesne, J. Y. Blay, S. Marreaud, N. Hindi, I. M. E. Desar, A. Gronchi, and W. T. A. van der Graaf. Prognostic factors for soft tissue sarcoma patients with lung metastases only who are receiving first-line chemotherapy: An exploratory, retrospective analysis of the European Organization for Research and Treatment of Cancer-Soft Tissue and Bone Sarcoma. *International Journal of Cancer*, 142(12):2610–2620, 2018. ISSN 10970215. doi: 10.1002/ijc.31286.

[22] E. Younger, S. Litière, A. Le Cesne, O. Mir, H. Gelderblom, A. Italiano, S. Marreaud, R. L. Jones, A. Gronchi, and W. T. A. van der Graaf. Outcomes of Elderly Patients with Advanced Soft Tissue Sarcoma Treated with First-Line Chemotherapy: A Pooled Analysis of 12 EORTC Soft Tissue and Bone Sarcoma Group Trials. *The Oncologist*, 23(10):1250–1259, 2018. ISSN 1083-7159. doi: 10.1634/theoncologist.2017-0598.

[23] B. Vincenzi, A. M. Frezza, G. Schiavon, D. Santini, P. Dileo, M. Silletta, F. Bertoldo, G. Badalamenti, G. G. Baldi, S. Zovato, R. Berardi, M. Tucci, J. Whelan, R. Tirabosco, A. P. Dei Tos, and G. Tonini. Bone metastases in soft tissue sarcoma patients: A survey of natural, prognostic value, and treatment. *Clinical sarcoma research*, 3(1):1–5, 2013. ISSN 0732-183X. doi: 10.1200/jco.2012.30.15{_}suppl.10063.

[24] G. Kantidakis, S. Litière, H. Gelderblom, M. Fiocco, I. Judson, W. T. A. van der Graaf, A. Italiano, S. Marréaud, S. Sleijfer, G. Mechtersheimer, C. Messiou, and B. Kasper. Prognostic Significance of Bone Metastasis in Soft Tissue Sarcoma Patients Receiving Palliative Systemic Therapy: An Explorative, Retrospective Pooled Analysis of the EORTC-Soft Tissue and Bone Sarcoma Group (STBSG) Database. *Sarcoma*, 2022:1–13, 4 2022. ISSN 1369-1643. doi: 10.1155/2022/5815875. URL https://www.hindawi.com/journals/sarcoma/2022/5815875/.

[25] S. Sleijfer, I. Ray-Coquard, Z. Papai, A Le Cesne, M. Scurr, P. Schöffski, F. Collin, L. Pandite, S. Marreaud, A. De Brauwer, M. Van Glabbeke, J. Verweij, and J. Y. Blay. Pazopanib, a multikinase angiogenesis inhibitor, in patients with relapsed or refractory advanced soft tissue sarcoma: A phase II study from the European organisation for research and treatment of cancer-soft tissue and bone sarcoma group (EORTC Study 620. *Journal of Clinical Oncology*, 27(19):3126–3132, 2009. ISSN 0732183X. doi: 10.1200/JCO.2008.21.3223.

[26] P. Schöffski, I. L. Ray-Coquard, A. Cioffi, N. B. Bui, S. Bauer, J. T. Hartmann, A. Krarup-Hansen, V. Grünwald, R. Sciot, H. Dumez, J. Y. Blay, A. Le Cesne, J. Wanders, C. Hayward, S. Marreaud, M. Ouali, and P. Hohenberger. Activity of eribulin mesylate in patients with soft-tissue sarcoma: A phase 2 study in four

independent histological subtypes. *The Lancet Oncology*, 12(11):1045–1052, 2011. ISSN 14702045. doi: 10.1016/S1470-2045(11)70230-3.

[27] W. T. A. van der Graaf, J. Y. Blay, S. P. Chawla, D. W. Kim, B. Bui-Nguyen, P. G. Casali, P. Schöffski, M. Aglietta, A. P. Staddon, Y. Beppu, A. Le Cesne, H. Gelderblom, I. R. Judson, N. Araki, M. Ouali, S. Marreaud, R. Hodge, M. R. Dewji, C. Coens, G. D. Demetri, C. D. Fletcher, A. P. Dei Tos, and P. Hohenberger. Pazopanib for metastatic soft-tissue sarcoma (PALETTE): A randomised, double-blind, placebo-controlled phase 3 trial. *The Lancet*, 379(9829):1879–1886, 2012. ISSN 1474547X. doi: 10.1016/S0140-6736(12)60651-5.

[28] I. Judson, J. Verweij, H. Gelderblom, J. T. Hartmann, P. Schöffski, J. Y. Blay, J. M. Kerst, J. Sufliarsky, J. Whelan, P. Hohenberger, A. Krarup-Hansen, T. Alcindor, S. Marreaud, S. Litière, C. Hermans, C. Fisher, P. C. W. Hogendoorn, A. P. Dei Tos, and W. T. A. van der Graaf. Doxorubicin alone versus intensified doxorubicin plus ifosfamide for first-line treatment of advanced or metastatic soft-tissue sarcoma: A randomised controlled phase 3 trial. *The Lancet Oncology*, 15(4):415–423, 2014. ISSN 14745488. doi: 10.1016/S1470-2045(14)70063-4.

[29] B. Bui-Nguyen, J. E. Butrynski, N. Penel, J. Y. Blay, N. Isambert, M. Milhem, J. M. Kerst, A. K. L. Reyners, S. Litière, S. Marréaud, F. Collin, and W. T. A. van der Graaf. A phase IIb multicentre study comparing the efficacy of trabectedin to doxorubicin in patients with advanced or metastatic untreated soft tissue sarcoma: The TRUSTS trial. *European Journal of Cancer*, 51(10):1312–1320, 2015. ISSN 18790852. doi: 10.1016/j.ejca.2015.03.023.

[30] B. Yücel, M. G. Celasun, B. Öztoprak, Z. Hasbek, S. Bahar, T. Kaçan, A. Bahçeci, and M. M. Şeker. The negative prognostic impact of bone metastasis with a tumor mass. *Clinics*, 70(8):535–540, 2015. ISSN 18075932. doi: 10.6061/clinics/2015(08)01.

[31] X. H. Du, H. Wei, P. Zhang, W. T. Yao, and Q. Q. Cai. Heterogeneity of Soft Tissue Sarcomas and Its Implications in Targeted Therapy. *Frontiers in Oncology*, 10:1–3, 2020. ISSN 2234943X. doi: 10.3389/fonc.2020.564852. URL https://www.frontiersin.org/article/10.3389/fonc.2020.564852.

[32] D. B. Fogel. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary Clinical Trials Communications*, 11(August):156–164, 2018. ISSN 24518654. doi: 10.1016/j.conctc.2018.08.001. URL https://doi.org/10.1016/j.conctc.2018.08.001.

[33] P. Pautier, A. Floquet, C. Chevreau, N. Penel, C. Guillemet, C. Delcambre, D. Cupissol, F. Selle, N. Isambert, S. Piperno-Neumann, A. Thyss, F. Bertucci, E. Bompas, J. Alexandre, O. Collard, S. Lavau-Denes, P. Soulié, M. Toulmonde, A. Le Cesne, B. Lacas, and F. Duffaud. Trabectedin in combination with doxorubicin for first-line treatment of advanced uterine or soft-tissue leiomyosarcoma (LMS-02): A non-randomised, multicentre, phase 2 trial. *The Lancet Oncology*, 16(4):457–464, 2015. ISSN 14745488. doi: 10.1016/S1470-2045(15)70070-7.

[34] P. Pautier, A. Italiano, S. Piperno-Neumann, C.M. Chevreau, N. Penel, D. Cupissol, P. Boudou Rouquette, F. Bertucci, C. Balleyguier, V. Lebrun-Ly, J. Blay, E. Kalbacher, C. Delcambre, E. Bompas, O. Collard, N. Isambert, C. Guillemet, M. Rios, M. Sundqvist, and F. Duffaud. LBA59 - LMS-04 study: A randomised, multicenter, phase III study comparing doxorubicin alone versus doxorubicin with trabectedin followed by trabectedin in non-progressive patients as first-line therapy, in patients with metastatic or unresectable leiomyo. *Annals of Oncology*, 32(suppl_5):S1283–S1346, 2021. URL https://oncologypro.esmo.org/meeting-resources/esmo-congress/lms-04-study-a-randomised-multicenter-phase-iii-study-comparing-doxorubicin-alone-versus-doxorubicin-with-trabectedin-followed-by-trabectedin-in.

[35] EORTC 1809 STRASS II Trial In Retroperitoneal Sarcoma Receives Support By The Anticancer Fund 2022 - EORTC : EORTC. URL https://www.eortc.org/blog/2020/04/28/eortc-1809-strass-ii-trial-in-retroperitoneal-sarcoma-receives-support-by-the-anticancer-fund/.

[36] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*, 19(1):1–18, 2019. doi: 10.1186/s12874-019-0681-4.

[37] R. C. Deo. Machine learning in medicine. *Circulation*, 132(20):1920–1930, 2015. ISSN 15244539. doi: 10.1161/CIRCULATIONAHA.115.001593.

[38] Ira S. Hofer, Michael Burns, Samir Kendale, and Jonathan P. Wanderer. Realistically integrating machine learning into clinical practice: A road map of opportunities, challenges, and a potential future. *Anesthesia and Analgesia*, 130(5):1115–1118, 2020. ISSN 15267598. doi: 10.1213/ANE.0000000000004575.

[39] P. Wang, Y. Li, and C. K. Reddy. Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 51(6):1–36, 2019. doi: https://doi.org/10.1145/3214306.

[40] G. S. Collins, J. B. Reitsma, D. G. Altman, and K. G. M. Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMC Medicine*, 13(1), 1 2015. ISSN 17417015. doi: 10.1186/s12916-014-0241-z. URL http://www.biomedcentral.com/1741-7015/13/1.

[41] P. Dhiman, J. Ma, C. A. Navarro, B. Speich, G. Bullock, J. A. A. Damen, S. Kirtley, L. Hooft, R. D. Riley, B. Van Calster, K. G. M. Moons, and G. S. Collins. Reporting of prognostic clinical prediction models based on machine learning methods in oncology needs to be improved. *Journal of Clinical Epidemiology*, 138: 60–72, 2021. ISSN 18785921. doi: 10.1016/j.jclinepi.2021.06.024. URL https://doi.org/10.1016/j.jclinepi.2021.06.024.

[42] R. D. Riley, J. Ensor, K. I. E. Snell, F. E. Harrell, G. P. Martin, J. B. Reitsma, K. G. M. Moons, G. Collins, and M. Van Smeden. Calculating the sample size required for developing a clinical prediction model. *The BMJ*, 368(March):1–12, 2020. ISSN 17561833. doi: 10.1136/bmj.m441. URL http://dx.doi.org/doi:10.1136/bmj.m441.

[43] E. W. Steyerberg. *Clinical prediction models: A Practical Approach to Development, Validation, and Updating*. Springer, 2nd edition, 2019. doi: https://doi.org/10.1007/978-3-030-16399-0. URL https://www.springer.com/gp/book/9783030163983.

[44] G. S. Collins and K. G. M. Moons. Reporting of artificial intelligence prediction models. *The Lancet*, 393(10181):1577–1579, 4 2019. ISSN 1474547X. doi: 10.1016/S0140-6736(19)30037-6. URL http://www.thelancet.com/article/S0140673619300376/fulltexthttp://www.thelancet.com/article/S0140673619300376/abstracthttps://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)30037-6/abstract.

[45] G. Kantidakis, H. Putter, C. Lancia, J. de Boer, A. E. Braat, and M. Fiocco. Survival prediction models since liver transplantation - comparisons between Cox models and machine learning techniques. *BMC Medical Research Methodology*, 20(1):1–14, 12 2020. ISSN 14712288. doi: 10.1186/s12874-020-01153-1.

[46] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841–860, 2008. ISSN 1932-6157. doi: 10.1214/08-AOAS169. URL https://projecteuclid.org/journals/annals-of-applied-statistics/volume-2/issue-3/Random-survival-forests/10.1214/08-AOAS169.short.

[47] E. Biganzoli, P. Boracchi, L. Mariani, and E. Marubini. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. *Statistics in medicine*, 17(10):1169–1186, 1998. doi: 10.1002/(sici)1097-0258(19980530)17:10<1169::aid-sim796>3.0.co;2-d.

[48] D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–220, 1972. URL http://www.jstor.org/stable/2985181.

[49] R. Tibshirani. The lasso method for variable selection in the cox model. *Statistics in Medicine*, 16(4):385–395, 1997. ISSN 02776715. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3.

[50] F. E. Harrell, K. L. Lee, and D. B. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*, 15(4): 361–387, 1996. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

[51] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545, 1999. doi: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5. URL http://www.ncbi.nlm.nih.gov/pubmed/10474158.

[52] J. C. van Houwelingen and H. Putter. *Dynamic prediction in clinical survival analysis*. CRC Press, 1st edition, 2012. ISBN 9781439835333. URL https://www.crcpress.com/Dynamic-Prediction-in-Clinical-Survival-Analysis/van-Houwelingen-Putter/p/book/9781439835333.

[53] G. Kantidakis, E. Biganzoli, H. Putter, and M. Fiocco. A Simulation Study to Compare the Predictive Performance of Survival Neural Networks with Cox Models for Clinical Trial Data. *Computational and Mathematical Methods in Medicine*, 2021:1–15, 2021. ISSN 1748-670X. doi: 10.1155/2021/2160322.

[54] I. J. Lewis, M. A. Nooij, J. Whelan, M. R. Sydes, R. Grimer, P. C. W. Hogendoorn, M. A. Memon, S. Weeden, B. M. Uscinska, M. Ven Glabbeke, A. Kirkpatrick, E. I. Hauben, A. W. Craft, and A. H. M. Taminiau. Improvement in histologic response but not survival in osteosarcoma patients treated with intensified chemotherapy: A randomized phase III trial of the european osteosarcoma intergroup. *Journal of the National Cancer Institute*, 99(2):112–128, 2007. ISSN 14602105. doi: 10.1093/jnci/djk015.

[55] E. Biganzoli, P. Boracchi, F. Ambrogi, and E. Marubini. Artificial neural network for the joint modelling of discrete cause-specific hazards. *Artificial Intelligence in Medicine*, 37(2):119–130, 2006. doi: 10.1016/j.artmed.2006.01.004.

[56] H. Ishwaran, T. A. Gerds, U. B. Kogalur, R. D. Moore, S. J. Gange, and B. M. Lau. Random survival forests for competing risks. *Biostatistics*, 15(4):757–773, 2014. ISSN 14684357. doi: 10.1093/biostatistics/kxu010.

[57] J. P. Fine and R. J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999. ISSN 1537274X. doi: 10.1080/01621459.1999.10474144.

[58] P. Blanche, C. Proust-Lima, L. Loubère, C. Berr, J. F. Dartigues, and H. Jacqmin-Gadda. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*, 71(1):102–113, 2015. ISSN 15410420. doi: 10.1111/biom.12232.

[59] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan. Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology*, 21(1):128–138, 1 2010. ISSN 10443983. doi: 10.1097/EDE.0b013e3181c30fb2. URL https://pubmed.ncbi.nlm.nih.gov/20010215/.

[60] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006. ISBN 978-0-387-31073-2.

[61] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2nd edition, 2009. ISBN 978-0-387-84857-0. doi: 10.1007/978-0-387-84858-7. URL http://link.springer.com/10.1007/978-0-387-84858-7.

[62] C. Wallisch, A. Agibetov, G. Dorffner, D. Dunkler, and G. Heinze. *Statistical modelling or machine learning: Interpretability vs. flexibility?* 40th Annual Conference of the International Society for Clinical Biostatistics, 2019. URL https://kuleuvencongres.be/iscb40/images/iscb40-2019-e-versie.pdf.

[63] T. Van Der Ploeg, P. C. Austin, and E. W. Steyerberg. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology*, 14(1):1–13, 2014. ISSN 14712288. doi: 10.1186/1471-2288-14-137.

[64] F. E. Harrell Jr. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Springer, 2nd edition, 2015. ISBN 978-3-319-19425-7. doi: https://doi.org/10.1007/978-3-319-19425-7. URL http://www.springer.com/series/692.

[65] F. E. Harrell Jr. Road Map for Choosing Between Statistical Modeling and Machine Learning | Statistical Thinking. URL https://www.fharrell.com/post/stat-ml/.

[66] N. H. Shah, A. Milstein, and Steven C. Bagley. Making Machine Learning Models Clinically Useful. *JAMA*, 322(14):1351–1352, 10 2019. ISSN 1538-3598. doi: 10.1001/JAMA.2019.10306. URL https://pubmed.ncbi.nlm.nih.gov/31393527/.

[67] R. Cuocolo, M. Caruso, T. Perillo, L. Ugga, and M. Petretta. Machine Learning in oncology: A clinical appraisal. *Cancer Letters*, 481(February):55–62, 2020. ISSN 18727980. doi: 10.1016/j.canlet.2020.03.032.

[68] A. L. Beam and I. S. Kohane. Big data and machine learning in health care. *JAMA - Journal of the American Medical Association*, 319(13):1317–1318, 2018. ISSN 15383598. doi: 10.1001/jama.2017.18391.

[69] M. Nagy, N. Radakovich, and A. Nazha. Machine Learning in Oncology: What Should Clinicians Know? *JCO Clinical Cancer Informatics*, (4):799–810, 2020. ISSN 24734276. doi: 10.1200/cci.20.00049. URL https://doi.org/10.1200/CCI.20.00049.

[70] E. Christodoulou, J. Ma, G. S Collins, E. W. Steyerberg, J. Y. Verbakel, and B. Van Calster. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, 110:12–22, 2019. ISSN 18785921. doi: 10.1016/j.jclinepi.2019.02.004.

[71] I. Bou-Hamad, D. Larocque, and H. Ben-Ameur. A review of survival trees. *Statistics Surveys*, 5:44–71, 1 2011. ISSN 1935-7516. doi: 10.1214/09-SS047. URL https://projecteuclid.org/journals/statistics-surveys/volume-5/issue-none/A-review-of-survival-trees/10.1214/09-SS047.full.

[72] K. Liestol, P. K. Andersen, and U. Andersen. Survival analysis and neural nets. *Statistics in Medicine*, 13 (12):1189–1200, 1994. doi: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.4780131202.

[73] P. J. G. Lisboa, H. Wong, P. Harris, and R. Swindell. A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. *Artificial Intelligence in Medicine*, 28(1):1–25, 2003. ISSN 09333657. doi: 10.1016/S0933-3657(03)00033-2.

[74] F. M. Khan and V. B. Zubek. Support Vector Regression for Censored Data (SVRc): A Novel Tool for Survival Analysis. *2008 Eighth IEEE International Conference on Data Mining*, pages 863–868, 2008. doi: 10.1109/ICDM.2008.50.

[75] T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006. ISSN 14654644. doi: 10.1093/biostatistics/kxj011.

[76] M. A. Nicolaie, J. C. van Houwelingen, T. M. de Witte, and H. Putter. Dynamic prediction by landmarking in competing risks. *Statistics in Medicine*, 32(12):2031–2047, 2013. ISSN 02776715. doi: 10.1002/sim.5665.

[77] R. F. Wolff, K. G. M. Moons, R. D. Riley, P. F. Whiting, M. Westwood, G. S. Collins, J. B. Reitsma, J. Kleijnen, and S. Mallett. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, 170(1):51–58, 2019. ISSN 15393704. doi: 10.7326/M18-1376.

[78] D. E. Leisman, M. O. Harhay, D. J. Lederer, M. Abramson, A. A. Adjei, J. Bakker, Z. K. Ballas, E. Barreiro, S. C. Bell, R. Bellomo, J. A. Bernstein, R. D. Branson, V. Brusasco, J. D. Chalmers, S. Chokroverty, G. Citerio, N. A. Collop, C. R. Cooke, J. D. Crapo, G. Donaldson, D. A. Fitzgerald, E. Grainger, L. Hale, F. J. Herth, P. M. Kochanek, G. Marks, J. R. Moorman, D. E. Ost, M. Schatz, A. Sheikh, A. R. Smyth, I. Stewart, P. W. Stewart, E. R. Swenson, R. Szymusiak, J. L. Teboul, J. L. Vincent, J. A. Wedzicha, and D. M. Maslove. Development and Reporting of Prediction Models: Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Critical Care Medicine*, 48(5):623–633, 2020. ISSN 15300293. doi: 10.1097/CCM.0000000000004246.