From the Section of Orthopaedics

Department of Molecular Medicine and Surgery

Karolinska Institute

Stockholm Sweden

TURNING DATA INTO DECISIONS

❖

CLINICAL DECISION SUPPORT IN ORTHOPAEDIC ONCOLOGY

Jonathan A. Forsberg, M.D.

Karolinska Institutet

Stockholm 2015

Cover Illustration “harvesting data from the tree” by Patricia Schultz

All previously published papers were reproduced with permission from the publisher

Published and printed by the Karolinska University Press

Box 200, SE-171 77 Stockholm, Sweden

ISBN 978-91-7549-514-9

TURNING DATA INTO DECISIONS—
CLINICAL DECISION SUPPORT IN ORTHOPAEDIC ONCOLOGY

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Jonathan A. Forsberg, M.D.

Principal Supervisor:
Rikard Wedin, M.D. Ph.D.
Associate Professor
Karolinska Institute
Department of Molecular Medicine and Surgery
Division of Orthopaedics

Co-supervisor(s):
Professor Henrik Bauer, M.D. Ph.D.
Karolinska Institute
Department of Molecular Medicine and Surgery
Division of Orthopaedics

Professor Lars Weidenhielm, M.D. Ph.D.
Karolinska Institute
Department of Molecular Medicine and Surgery
Division of Orthopaedics

Opponent:
Nicola Fabbri, MD
Memorial Sloan-Kettering Cancer Center
Weill-Cornell School of Medicine
Department of Surgery
Division of Orthopaedics

Examination Board:
Professor Anders Ekbom, M.D. Ph.D.
Karolinska Institute
Clinical Epidemiology Unit
Department of Medicine

Jan Adolfsson, M.D. Ph.D.
Karolinska Institute
Associate Professor of Urology
Department of Clinical Science, Intervention and Technology

Professor Ola Wahlström, M.D. Ph.D.
Linköping University
Department of Clinical and Experimental Medicine
Division of Orthopaedics

“You can’t cross the sea merely by standing and staring at the water.”

—Rabindranath Tagore

“The theory of probabilities is…nothing but common sense reduced to calculus.”

—Pierre-Simon Laplace

“All models are wrong, some are useful.”

—Box and Draper

“There are three kinds of lies: lies, damned lies, and statistics.”

—Mark Twain

“You’ve got to be kidding…”

—Dawn Forsberg, my dear wife

Excerpts

…successful treatment—and ultimately quality of life—depends on accurate survival estimates, derived not only by orthopaedic surgeons, but also by medical oncologists who facilitate orthopaedic referrals. –p. 8

The result of this work should not be viewed as a single tool, but a methodology that may be applied to any clinical question for which there is prior knowledge in the form of existing, quality data. –p. 53

By "crowd sourcing" data collection and automating the analytics, we ensure each model remains broadly applicable, clinically relevant, and can "evolve" over time. –p. 54

When using an "app," physicians should demand the same level of scrutiny and apply the same healthy skepticism as they do for the literature they read, the implants they select and the medications they prescribe. –p. 56

For my family

Abstract

Background: The treatment of patients with skeletal metastases is predicated on each patient's estimated survival. In order to maximize function and quality of life, orthopaedic surgeons must carefully avoid over- or undertreatment of the disease. Unfortunately, physician estimates are notoriously inaccurate and there are no validated means by which to estimate patient survival in patients with long-bone skeletal metastases. The purpose of this thesis is to apply machine learning (ML) approaches to (1) develop a clinical decision support (CDS) tool capable of estimating survival in patients with operable skeletal metastases, and (2) establish guidelines so that this approach may be used in other relevant topics within the field of orthopaedics.

Methods: We first defined the scope of the problem using data from the Karolinska Skeletal Metastasis Registry. We then developed objective criteria by which to estimate patient survival using data gleaned from the Memorial Sloan-Kettering Skeletal Metastasis Database (n=189). We employed ML techniques to find patterns within the data associated with short- and long-term survival. We chose three and 12 months because they are widely accepted to guide orthopaedic surgical decision-making. We developed an Artificial Neural Network (ANN), a Bayesian Belief Network (BBN), and a traditional Logistic Regression (LR) model. Each resulting model was internally validated and compared using Receiver Operator Characteristic (ROC) analysis. In addition, we performed decision analysis to determine which model, if any, was suited for clinical use. Next, we externally validated the models using Scandinavian Registry data (n=815), and again using data collected by the Società Italiana di Ortopedia e Traumatologia (SIOT) (n=287). We then created a web-based CDS tool as well as the infrastructure to collect prospective data on a global scale, so the models could be improved over time. Finally, we used BBN modeling to describe the hierarchical relationships between features associated with the treatment of high-grade soft tissue sarcomas (STS), and codify this complex information into a graphical representation to promote a more thorough understanding of the disease process.

Results: We found that implant failures in patients with skeletal metastases remain relatively common—even in the revision setting—as patients outlive their implants. On the other hand, perioperative deaths are relatively common, indicating that an estimation of life expectancy should be part of the surgical decision making process. Using ML approaches, we found several criteria that can be used to estimate longevity in this patient population. When compared to other techniques, the ANN model was most accurate, and also resulted in highest net benefit on decision analysis, compared to the BBN and LR models. However, the BBN is the best suited to accommodate missing data, which is common in the clinical setting. The three- and 12-month BBN models were successfully externally validated using the SSMR database (Area under the ROC curve (AUC) of 0.79 and 0.76, respectively), and again using SIOT data (AUC 0.80 and 0.77). In the setting of high-grade, completely excised STS, BBN Modeling identified the first-degree associates of disease-specific survival to be the size of the primary tumor, and the presence and timing of local and distant recurrence.

Conclusions: We successfully developed and validated a CDS tool designed to estimate survival in patients with operable skeletal metastases. In addition, we made this tool available to orthopaedic surgeons, worldwide, at www.pathfx.org. We also created an international skeletal metastasis registry to continue to collect data on patients with skeletal metastases. Within this framework, prognostic models have the capacity to improve over time, as treatment philosophies evolve and more effective systemic therapies become available. These techniques may now be applied to other disciplines, in an effort to turn quality data into decision support tools.

Sammanfattning på Svenska

Bakgrund: Behandlingen av skelettmetastaspatienter baseras på förväntad överlevnad. För att kunna maximera funktion och livskvalitet behöver ortopeden noggrant undvika över- eller underbehandling. Tyvärr är läkarens bedömning av prognosen oftast otillförlitlig och det finns ingen validerad metod att uppskatta överlevnadstiden för en patient med skelettmetastas. Målsättningen med den här avhandlingen är att skapa ett kliniskt beslutsstöd för prognostik av patienter med operabla skelettmetastaser oavsett lokalisation. Därigenom kommer vi också kunna utvärdera nyttan av maskininlärning (machine learning, ML) inom ortopedisk onkologi.

Metoder: Vi började med att definiera storleken av det kliniska problemet genom analys av patientdata från det Karolinska Skelettmetastasregistret. Därefter utvecklade vi objektiva kriterier för att prognostisera patientöverlevnad med hjälp av patientdata från Memorial Sloan Kettering Cancer Centers databas (n=189). Vi använde oss av ML-teknik för att hitta mönster i patientdata associerad till korttidsöverlevnad (<3 månader), respektive långtidsöverlevnad (>12 månader). Vi valde dessa tidpunkter då de är allmänt använda inom ortopedisk onkologi. Härefter skapade vi en artificiell neural nätverksmodell (Artificial Neural Network, ANN), en Bayesiansk nätverksmodell (Bayesian Belief Network, BBN) och slutligen en traditionell logistisk regressionsmodell (LR). Varje modell validerades internt och jämfördes med varandra med hjälp av ROC-analys (Receiver Operator Characteristic). Dessutom utfördes beslutsanalys (decision analysis) för att avgöra vilken av de tre modellerna som var bäst lämpad för kliniskt bruk. Modellerna validerades med hjälp av patientdata från det Skandinaviska Skelettmetastasregistret (n=815) och från Società Italiana di Ortopedia e Traumatologia (SIOT) (n= 287). Därefter konstruerade vi ett internetbaserat kliniskt beslutsstöd och dessutom infrastrukturen för att prospektivt och globalt kunna insamla patientdata för att kontinuerligt uppdatera och förbättra vårt beslutsstöd. Slutligen använde vi BBN för att beskriva de hierarkiska sambanden mellan de olika tumörspecifika variablerna som är viktiga vid behandlingen av mjukdelssarkom. Målet var att koda denna komplexa information till en tydlig grafisk bild som kan användas för att skapa nya hypoteser.

Resultat: Haveri av skelettrekonstruktioner efter både primär- och reoperation av metastaspatienter med lång överlevnad var vanligt. Samtidigt fann vi att det även var vanligt med en mycket kort överlevnad efter operation. När vi använde ML identifierades flera användbara faktorer för att prognostisera överlevnad hos patienter med skelettmetastaser. Vid jämförelse av de olika modellerna visade det sig att ANN gav mer exakta resultat jämfört med BBN och LR. Däremot var BBN kliniskt mer användbar då den fungerade bäst även om viss patientdata saknades vilket ofta är fallet. 3- och 12-månadersmodellen av BBN validerades framgångsrikt med hjälp av patientdata från SSMR (area under receiver operator character curve, AUC, 0.79 resp. 0.76) och från SIOT (AUC: 0.80 resp. 0.77). BBN visade att de viktigaste prognostiska faktorerna för patienter med mjukdelssarkom var primärtumörens storlek samt förekomst och tid till lokalt tumörrecidiv resp. fjärrmetastas.

Slutsatser: Vi har framgångsrikt skapat och utvärderat ett kliniskt beslutsstöd avsett att prognostisera överlevnad hos patienter med operabla skelettmetastaser. Detta verktyg finns nu globalt tillgängligt på internetadressen www.pathfx.org. Vi har också skapat ett internationellt skelettmetastasregister för att fortsätta samla patientdata. Genom detta tillvägagångssätt har prognostiska beslutsstöd, så som det vi nu skapat, möjligheter att förändras över tid i takt med att nya effektivare behandlingar införs. Den här tekniken kan nu tillämpas inom andra dicipliner för att omvandla patientdata till kliniska beslutsstöd, t ex för primärtumörer inom ortopedisk onkologi.

List of Original Papers and Summaries

This thesis is based on the following papers, referenced in the text by their Roman Numerals (I-VI)

Study I

Which Implant Is Best After Failed Treatment for Pathologic Femur Fractures?
Forsberg, J. A., Wedin, R. & Bauer, H.
***Clin Orthop Relat Res* 2013 Mar;471(3):735-40. doi: 10.1007/s11999-012-2558-2.**

The purpose of this study was to evaluate patients with femoral metastases in whom constructs failed to determine (1) the rate of reoperation for any reason; (2) the timing of and most common causes for failure; and (3) incidence of perioperative death and other complications, not requiring surgery. In a cohort of 88 patients, we found that 17 (19%) required reoperation and that material failure was responsible for the overwhelming majority of these. As expected, endoprostheses were more durable with very few treatment failures when compared to plate or intramedullary nail fixation; however, the procedures were associated with higher proportions of perioperative complications. These findings reiterate that patient selection is important when choosing a reconstructive option, but most importantly, careful attention must be paid to each patient's estimated life expectancy.

Study II

Estimating Survival in Patients with Operable Skeletal Metastases: An Application of a Bayesian Belief Network
Forsberg, J. A., Eberhardt, J., Boland, P. J., Wedin, R. & Healey, J. H.
***PLoS ONE* 2011 6: e19956.**

We determined the feasibility of developing Bayesian models for the purpose of estimating survival. We chose a Bayesian approach because it can be used with missing input data, which is common in the clinical setting. For this study, we collected information from 189 patients undergoing surgery for metastatic bone disease involving the axial and appendicular skeleton. We developed two models using fifteen candidate features theorized to be related to survival, each designed to estimate the likelihood of survival at three months and 12 months post surgery. We chose these time points because they are useful for orthopaedic surgical decision-making. For example, short-term survival helps surgeons decide whether to offer surgery, and long-term survival helps surgeons decide whether a more durable implant is necessary. Cross validation demonstrated an AUC of 0.85 and 0.83 for the three- and 12-month models, respectively.

Study III

Treating Metastatic Disease: Which Survival Model Is Best Suited for the Clinic?
Forsberg, J. A., Sjoberg, D., Chen, Q.-R., Vickers, A. & Healey, J. H.
***Clin Orthop Relat Res* 2013 Mar;471(3):843-50. doi: 10.1007/s11999-012-2577-z.**

After the relative success of the BBN models described in Study II, we questioned whether other modeling techniques could be used. To this end, we developed three models designed to estimate survival using similar data, an artificial neural network (ANN), a BBN and a logistic regression (LR) model. We then asked: (1) Which model was most accurate on ROC analysis? And (2) which model performs best on decision curve analysis and is therefore most clinically useful? In doing so, we introduced the orthopaedic community to the concept of Decision Curve Analysis (DCA), which is used to compare models by weighing the relative consequences of false positive or false negative classifications. The results were interesting. The ANN models were most accurate (0.89 and 0.93) for the three- and 12-month models, respectively. In comparison, the BBN was slightly more accurate than the LR model with an AUC of 0.85 and 0.83. In addition, the ANN model resulted in the highest net benefit throughout the entire range of threshold probabilities. Nevertheless, the ANN may not be best suited for clinical use, since it requires complete input data to function. For this reason, we concluded that the BBN model might be better suited for clinical use.

Study IV

External Validation of the Bayesian Estimated Tools for Survival (BETS) Models in Patients with Surgically Treated Skeletal Metastases

Forsberg, J. A., Wedin, R., Bauer, H., Weidenhielm, L., Hansen, B. H., Laitinen, M., Trovik, C., Keller J. Ø., Boland, P. J., Healey, J. H.

BMC Cancer, 12(1), 493-51.

The purpose of this study was to externally validate the BBN models using an independent, international skeletal metastasis registry. In doing so, we demonstrated the utility of the Bayesian approach that retains functionality in the presence of missing input data. In this study, several features were missing, including the surgeon's estimate of survival, which was missing in all 815 records. The BBN models, now referred to as PATHFx correctly estimated three- and 12-month survival in the majority of records, with an AUC of 0.79 and 0.76, respectively.

Study V

External Validation of a Tool for the Estimation of Life Expectancy in Patients with Skeletal Metastases—Decision Analysis and Comparison of Three Major International Patient Populations

Piccioli, A., Spinelli, M.S., Forsberg, J.A., Wedin R., Healey, J.H. Ippolito, V., Daolio, P., Ruggieri, P., Maccauro G., Gabsarrini A., Biagini R., Piana R., Fazioli F., Luzzati, A., Di Martino, A., Nicolosi F., Camnasio F., Rosa, M.A., Campanacci, D.A., Denaro, V., Capanna R.

(Submitted for publication)

This study is similar to the previous external validation study, however, its purpose was to (1) externally validate the PATHFx tool in an Italian patient population and (2) compare the distributions of patients to both the training set (U.S.) and first external validation (Scandinavian) datasets, respectively. Using the data from 287 records, PATHFx proved sufficiently accurate with an AUC of 0.80 and 0.77 for three and 12 months, respectively. There were missing data, which was similar to the previous external validation study. However, both the patient demographics as well as the indications for surgery differed significantly between this patient population, and the training and first validation sets. In addition to being useful in the presence of missing data, this suggests PATHFx has widespread applicability in cultures with differing treatment philosophies to those previously studied.

Study VI

A Probabilistic Analysis of Completely Excised High-Grade Soft Tissue Sarcomas of the Extremity: An Application of a Bayesian Belief Network

Forsberg, J. A., Healey J. H., Brennan M. B.

Ann Surg Oncol 2012; 19(9):2992-3001.

We sought to demonstrate the applicability of Bayesian methodology to an entirely different oncologic scenario by describing the hierarchical relationships between features related to the treatment of soft tissue sarcomas. For this study, we focused only on completely excised, localized high-grade soft tissue sarcomas of the extremity to determine which features were most highly associated with disease specific survival (DSS). In doing so, we highlight BBN modeling as a tool to codify complex hierarchical relationships between features we believe to be representative of cancer biology into clear graphical representations. The model structure revealed first-degree associates of DSS were the size category of the primary tumor; presence of and time to distant recurrence; and presence of and time to local recurrence. We show that Bayesian modeling can be used to gain insight into the interrelationships between features in preparation for the development of clinical decision support tools using more restrictive Bayesian models or other modeling techniques.

TABLE OF CONTENTS

Abbreviations

ANN Artificial Neural Network—a machine learning method used in Study III with powerful pattern recognition and discriminatory ability

AUC Area under the receiver operator characteristic curve, used to assess accuracy of a model

BBN Bayesian Belief Network, a machine learning technique with a graphical output; retains functionality in the presence of missing input data

CDS Clinical Decision Support, the process of basing clinical decisions on objective data

DAG Directed Acyclic Graph

DCA Decision Curve Analysis—a means to compare models and determine which is better suited for the clinical setting

HTML Hypertext Markup Language, used to create web pages and online content

INCA Information Network for Cancer—A national IT platform for managing registries

IQR Interquartile Range

jPDF joint probability distribution function that defines conditional relationships between features in a BBN

KUH Karolinska University Hospital, Stockholm, Sweden

LR Logistic Regression

ML Machine Learning—automates the discovery of conditional relationships as in BBN analysis, and pattern recognition as in ANN analysis

MLP Multilayer Perceptron—the architecture of the artificial neural network used in Study III

MSKCC Memorial Sloan-Kettering Cancer Center, New York, New York, USA, the source of data for Studies II, III and VI

R® Statistical Software, Version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria[1])

RF Random Forest, a machine learning technique that constructs multiple Decision Trees.

ROC Receiver Operator Characteristic, a curve used to measure accuracy of a model

SD Standard Deviation

SIOT Società Italiana di Ortopedia e Traumatologia, the Italian Society of Orthopaedics and Traumatology that facilitated data collection for Study V

SRE Skeletal-Related Event—bone pain requiring palliative radiotherapy, pathologic fractures, spinal cord compression, and orthopaedic surgery in the context of metastatic bone disease.

SSMR Scandinavian Skeletal Metastasis Registry—provided data for Study IV

Introduction

There is perhaps no better application of personalized medicine than in the treatment of cancer. In addition to matching systemic therapy to cellular genotypes and/or phenotypes, we strive to derive personalized estimates of survival to help set patient, family and physician expectations. Since no two tumors are the same, and no two patients present with exactly the same disease burden, these estimates are used to guide the medical, surgical, and sometimes palliative decision-making processes. Unfortunately, our ability as clinicians to answer the question, "Doc, how long have I got?" is generally inaccurate, and better means of prognostication are needed.[2,3]

Patients with skeletal metastases are terminal and estimations of longevity are therefore very important. However, for many reasons, most physicians refuse to prognosticate whenever possible. Dr. Elizabeth Lamont[4] describes this phenomenon nicely. In a study of 311 patients with terminal cancer, physicians were perfectly able to derive estimates of survival in nearly all (97%) cases. However, they communicated their *actual* estimate to only 37%, knowingly misled patients by providing a *different* estimate to 40%, and *withheld* the information entirely from the remaining 23% of their patients. For the group who were misled, estimates were almost always optimistic and the patient would then make decisions based on information twice-removed from reality (Figure 1).[4] On one hand, optimistic survival estimates fuel each cancer patient's need for optimism. However, if *overly* optimistic, this information is likely to have unanticipated consequences.

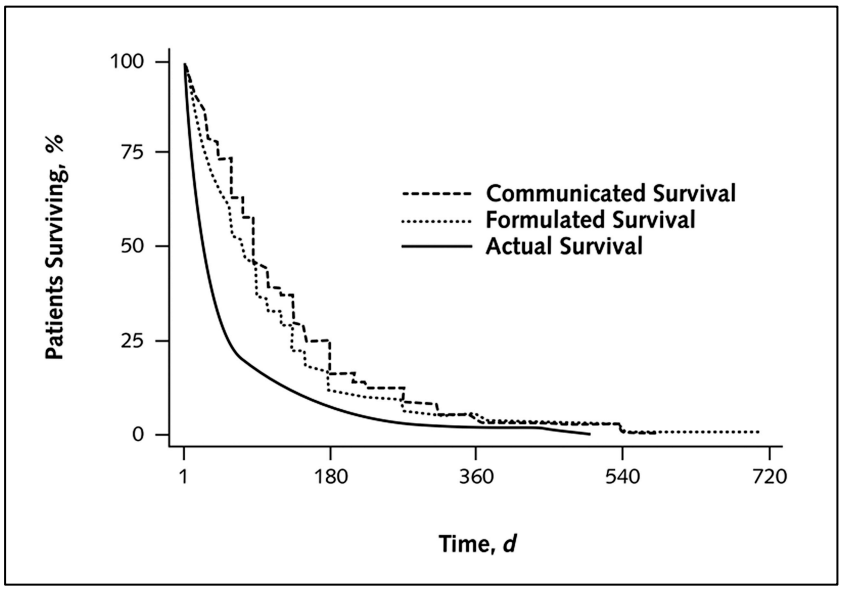


Figure 1. This Kaplan-Meier survival curve demonstrates differences in actual, formulated and communicated survival in 300 terminally ill cancer patients. Reprinted with permission.

At the end of life, patients must balance the reality of dying with a sometimes overwhelming desire for self-preservation. This paradox is important when one considers that maintaining dignity and control is among the top priorities for terminally ill patients.[5,6] Receiving an overly optimistic prognosis from a treating physician (often viewed as an authority figure) can have a dramatic influence on treatment decisions. Patients in this setting are more likely to opt for more aggressive treatment, rather than perhaps more appropriate palliative measures. This, in turn, leads to higher complication rates, specifically readmission to the hospital for >72 hours, at least one

resuscitation attempt, or death while on a ventilator.[7] Clearly, better communication between patients and their physicians could lead to improved shared decision-making regarding end of life care.

For the orthopaedic surgeon, the question of life expectancy may be less philosophical but is certainly no less important. In fact, the goals of orthopaedic surgery in terminally ill patients are to relieve pain and preserve function for the maximum amount of time.[8,9] Surgical considerations in this setting are dependent on three things: The location and physiology of the skeletal metastases, how long the patient is likely to live, and the desired mechanical properties of the implant.[10] Though we often cannot control tumor-specific variables, such as its biomechanical impact, we do have the ability to influence the other two considerations. First, by deriving accurate survival estimates, and thus deciding whether to operate at all, and second, choosing the most appropriate implant when doing so. However, as mentioned above, our ability to accurately predict survival is dangerously inaccurate and becomes even more so at longer intervals.[2] This topic is important since patients with advanced cancer are living longer, and a considerable number of them are living with metastatic bone disease. Specifically, a recent analysis of skeletal metastases in Danish breast cancer patients[11] demonstrated a one- and three-year cumulative survival of 59% and 22%, respectively. A similar study performed in prostate cancer patients[12] observed one- and three-year cumulative survival of 47% and 9%, respectively.

In this setting, skeletal related events (SREs) are an important cause of mortality and disability, adversely affecting quality of life.[13,14] By definition, SREs include radiotherapy to symptomatic lesions, pathologic fracture, spinal cord compression and/or surgical intervention.[15-17] Large, population based studies characterizing the prevalence of SREs in patients with long-bone metastases found that 34-68% of breast cancer patients[11,16,18-21] and 40-44% of men with prostate cancer[12,22] had an SRE during the course of their treatment. Considering breast and prostate cancer account for the majority of skeletal metastases, this translates to millions of *at-risk* patients, worldwide, in which successful treatment—and ultimately quality of life—depends on accurate survival estimates, derived not only by orthopaedic surgeons,[23] but also by medical oncologists who facilitate orthopaedic referrals.

As such, the economic impact of metastatic bone disease is enormous.[11,12,16,22,24] In the United States alone, it is estimated that the 5.3% of cancer patients with skeletal metastases account for a disproportionate 17% of the $74 billion NIH estimated total direct medical cost for cancer.[25] To compound this, we believe the prevalence of metastatic bone disease is underreported, so as the population ages over the coming decades, these estimates are likely to grow.[26]

Skeletal metastases also negatively impact survival, particularly when pathologic fractures occur.[13,15,27] However, the value of surgical stabilization is well known for both impending and complete pathologic fractures. Decreased pain, improved mobilization, ease of nursing care are among the most widely accepted benefits of surgery.[28-39] Unfortunately, reoperation rates range from 3.1% to an astounding 42% for patients who survive more than 1 year after fixation for pathological fracture.[10,40,41] As more patients with skeletal metastases live longer,[25,26] more implants are at risk for failure.[23] As such, the causes of, and treatment options available for, failed constructs have received considerable attention in the literature.[23,40,42,43] The general recommendation is that durable implants such as endoprostheses should be used more liberally. Though the decision to proceed with, or forego, skeletal stabilization is based on the three major criteria listed above, estimating the life expectancy of the patient is arguably the most critical.[44-46]

For example, surgical stabilization or reconstruction is generally not recommended for those patients in whom the recovery and rehabilitation time is longer than their estimated survival,[47] though palliative procedures are sometimes indicated. Those patients with a longer life expectancy (3-12 months) may require less invasive stabilization, thereby minimizing pain and convalescence, to improve mobility and quality of life in their final months. Patients with more favorable survival estimates (>12 months) require durable, but more complicated, reconstructive procedures that should be expected to last many years, in some cases.

Unfortunately, falsely optimistic or pessimistic estimates can have dramatic clinical implications. Falsely optimistic survival estimates may influence patients and clinicians to pursue more aggressive therapies, rather than perhaps more appropriate conservative measures. This approach results in a higher proportion of both major perioperative complications as well as premature death during convalescence.[7] Conversely, falsely pessimistic survival estimates are problematic when surgeons choose a less invasive, less durable implant that lacks sufficient biomechanical durability to outlast the patient. In this setting, implant failures can occur, which require more complicated revision procedures, often at the end of life.[23,40]

Early attempts at estimating survival

Several authors have identified independent predictors of survival in patients with bony metastases, operative or otherwise.[27,45,48-51] These include the specific oncologic diagnosis, subjective Eastern Cooperative Oncology Group (ECOG) performance status,[52] number of bone metastases, presence of visceral metastases,[53] serum hemoglobin,[51] the senior surgeon's estimate of survival,[50] a diagnosis of lung cancer,[51] appendicular bone metastases,[27] the type of reconstructive procedure performed,[49] and time from oncologic diagnosis to total hip arthroplasty (for proximal femoral metastases).[45] Despite the large number of covariates that have been associated with survival in this patient population, there exists no consensus as to which ones should be routinely used. As such, their ability to predict survival as part of a cohesive model is unacceptably inaccurate, at 5-15% in the best of the reported series.[50] Nevertheless, this body of work demonstrated that it was possible to derive generalized estimations of survival based on an individual's disease-related and laboratory parameters. However, more accurate, personalized estimations were not yet possible.

In an attempt to develop a prognostic tool useful for surgical decision-making, Tokuhashi et al.[54] developed a scoring system to categorize postoperative survival into one of three groups: survival < 6 months, > 6 months, or > 1 year. The group collected a series of prognostic variables including, for the first time, the Karnofsky score, a measure of performance status.[55] The authors also documented the number of intra- and extraspinal bone metastases, the number and type of organ metastases, the primary oncologic diagnosis, and the Frankel classification that describes the degree of neurologic impairment. The group then applied their scoring system to 246 additional patients and observed that survival greater than, or less than six months could be reliably estimated by this method.[56] Independent validation produced similar results,[57] however, Tokuhashi's scoring system has limited valued for the orthopaedic community because it applies only to patients with symptomatic spine metastases.

Recognizing the value of a model that could be applied to all patients with skeletal metastases, Nathan and Healey[50] constructed a sliding scale in an attempt to codify the independent predictors listed above. Variables included ECOG performance status, number of bone metastases, presence or absence of visceral metastases, and

serum hemoglobin. Despite widespread interest in a prognostic tool, this particular model remains unvalidated.

Reasons for this are multifactorial. First, by treating the outcome as a continuous variable, the model is designed to provide discrete survival estimates. This is problematic when wide confidence intervals are observed, such as those reported for several of the covariates. These convey a high degree of uncertainty when used to derive a continuous outcome. In addition, nomograms such as this one become less accurate in the setting of dischordant (rare presentations of disease) or incomplete input information. As such, the model correctly estimated survival in a mere 18% of records on internal validation. The authors attempted to improve model performance by focusing on minimum survival estimates, which generated accurate estimates in 61% of records, but only after applying a rather generous 20% margin of error. External validation in a five-patient test set failed, by producing accurate assessments in only two of the records. The results prompted the authors to issue a "wake-up call" to the orthopaedic and oncologic communities, calling for more accurate means of deriving patient survival estimates.[50,58]

Nevertheless, Nathan's data revealed something interesting about the surgeon's estimate. Not only did it account for a large portion of the variance, as evidenced by an R^2 of 0.33, it was clearly superior to any other covariate studied. For the surgeon, estimates of survival are made after reviewing the patient's chart, the imaging studies and carefully examining the patient. Some variables included in this assessment are quantifiable such as performance scores, laboratory analyses, and radiographic findings. Some, however, are subjective, and include how ill the patient appears, the patient's demeanor, and the "gut feeling" the surgeon has after the consultation. In the traditional sense, subjective determinations such as this are unlikely to be useful as individual covariates because they are intangible and almost entirely dependent on the surgeon experience. However, including them in the analyses may give us clues as to which objective clinical information may act as surrogates for this important, albeit subjective assessment. In turn, by applying advanced mathematical techniques, we may objectively analyze and adjust prognostic estimates based on seemingly intangible, subjective data that are otherwise difficult to interpret in a reliable and reproducible manner.

Modeling survival

Recently, there has been a resurgence of interest in and application of machine learning (ML), a constellation of advanced analytical techniques that can be useful in the evaluation of complex relationships, such as those that may exist between oncologic variables. Bayesian classification methodology, also referred to as Bayesian Belief Network (BBN) modeling, is being used with increasing frequency in medicine for several reasons. First, BBN modeling can be used to analyze highly complex data containing nonlinear relationships. Second, this methodology effectively accounts for uncertainty within the data and unlike many statistical approaches, maintains predictive accuracy and robustness in the face of incomplete input information (i.e., missing data).[46] Finally, the BBN codifies complex data into clear, predictive models by incorporating the outcome(s) and covariates into a single, graphical network. In fact, Bayesian statistics have previously been used successfully to estimate the likelihood of survival in a wide variety of oncologic diagnoses.[59-65]

Despite the applicability of the Bayesian method to this problem, it is possible that other statistical techniques may accurately model survival. For instance, artificial neural networks (ANNs), known for their exceptional discriminatory ability, have been

used to risk stratify patients and tumors using gene expression and other pathophysiologic data.[66,67] In addition, regression-based techniques similar to those used to develop the highly successful MSKCC soft tissue sarcoma nomogram,[68] should also be explored. Models using various techniques could be evaluated and compared directly to assess not only which is most accurate, but also to determine which is best suited for clinical use.

Traditionally, model comparisons focus solely on accuracy. In this manner, sensitivity and specificity may be assessed, and the area under the receiver operator characteristic (ROC) curve (AUC) calculated.[69] However, an over-reliance on accuracy is problematic. First, these metrics fail to address consequences associated with a falsely positive or negative result. These are often of unequal importance, particularly in the oncologic setting, in which the consequences of a missed diagnosis generally outweigh the risk of unnecessary testing and treatment. However, for complicated surgical interventions, the risks associated with overtreatment can be quite unacceptable in many cases. As such, the consequences of wrong answers generated by would-be models must be evaluated by decision analysis[70] to determine not only which model is superior but also whether the model(s) are likely to be useful in the clinical setting.[71] Second, each model must remain broadly relevant, over time, a process that requires prospective validation in a variety of centers, with differing patient populations and treatment philosophies. As medical and surgical treatments become more effective for certain diagnoses, survival is likely to improve and, thus, becomes a moving-target. Existing models must then be re-fit, from time to time, to ensure both accuracy and precision are maintained in this changing environment. Each of these topics will be discussed in detail in the *Statistical Considerations* section, below.

Aims of the Thesis

Given that there were no validated means by which to estimate survival in patients with long bone skeletal metastases, we sought to apply ML techniques to create and compare models that could be validated in a variety of settings. In doing so, we acknowledged that the success of any tool designed for this purpose would depend on its applicability to diverse patient populations, its availability to clinicians and researchers worldwide, and its ability to evolve in the face of new therapies that may alter patient prognosis or outcomes.

1. First, we sought to define the scope of treatment failures in patients undergoing surgical treatment for skeletal metastases, while describing how orthopaedic considerations fit within the context of the terminally ill. *(Study I)*

2. Using a well-characterized metastatic disease registry, we explored ML approaches to modeling patient specific data in an effort to estimate the likelihood of postoperative survival. *(Study II)*

3. We then asked which of the models was (1) more accurate using ROC analysis and (2) resulted in highest net benefit using DCA in an effort to determine which, if any, is best suited for clinical use. *(Study III)*

4. After determining which models were most suitable for clinical use, we sought to externally validate them in other patient populations with differing healthcare systems and treatment philosophies from that represented by the MSKCC training set. *(Studies IV and V)*

5. Finally, we asked whether BBN modeling could be used to codify complex information related to the treatment of patients with localized soft tissue sarcomas into clear, graphical models in an effort to stimulate understanding and develop future hypotheses, and thus demonstrate the applicability of this methodology to an entirely different oncologic scenario. *(Study VI)*

Ethical Considerations

Study	Source of Data	Location	Ethical Approval Required?	Ethical Approval Granted / Number
I	Swedish Registry Data	Karolinska University Hospital, Stockholm	Yes	Yes Dnr 2012/272-31/4
II	Metastatic Disease Database	Memorial Sloan-Kettering Cancer Center, NY USA	Yes	Yes #WA0023-11
III	Metastatic Disease Database	Memorial Sloan-Kettering Cancer Center, NY	Yes	Yes #WA0023-11
IV	Scandinavian Registry Data	Karolinska University Hospital Stockholm	Yes	Yes Dnr 2012/272-31/4
V	Prospectively Collected Clinical Data	Rome, Italy	Yes	Prot: 15/13 OSS. ComEt CBM
VI	Soft Tissue Sarcoma Registry Data	Memorial Sloan-Kettering Cancer Center, NY USA	Yes	Yes WA0555-10

Patient Population and Methods

Study I: Which Implant Is Best After Failed Treatment for Pathologic Femur Fractures?

Previous data[23,40] suggested that implant failures were common following the treatment of skeletal metastases. In addition, estimations of survival could be used to help set patient and surgeon expectations, as well as help determine whether a more durable implant was necessary. With this in mind, we asked whether these considerations were applicable to the revision setting, as well as the primary, or index surgical procedure.

We chose the Karolinska Skeletal Metastasis Registry for this purpose because it contains prospectively collected data on patients undergoing surgery for skeletal metastases. In addition to demographic and surgical information, this registry contains detailed information regarding the intra- and postoperative complications for index and revision procedures, if applicable. We then identified all 93 patients who underwent surgery for failed treatment of femoral metastases between 1990 and 2010. Five patients who underwent amputations were excluded, leaving 88 patients who underwent limb salvage procedures using one of three types of implants: plate fixation (PLATE), intramedullary nailing (IMN) or endoprosthesis (EP). As per institutional guidelines, the indications for revision surgery were material failure, implant malposition, progression of disease including local recurrence, or a combination of these.

All salvage procedures were performed at the Karolinska University Hospital. From a treatment standpoint, skeletal metastases of the femur were approached in a manner(s) previously described.[10] If *en bloc* resection was indicated, reconstruction was performed using a modular tumor prosthesis whenever possible. Lesions confined to the femoral neck were treated with standard or long-stem hemiarthroplasty components. In keeping with institutional treatment philosophy, most cases underwent intralesional curettage before stabilization. Antibiotic-impregnated polymethylmethacrylate cement was used whenever possible. It is important to note that despite the pervasive institutional philosophy mentioned above, implants and surgical techniques were not standardized and were chosen at the discretion of the treating surgeon. The following implants were used, due largely to availability. Unreamed, locked femoral nails or cephalomedullary devices (Synthes Stratec Medical, Oberdorf, Switzerland) were used (IMN group). When plate fixation was used, a dynamic hip screw or limited contact dynamic compression plate (Synthes Stratec Medical) was chosen. For those who required prosthetic replacement, the Austin Moore hemiarthroplasty (Corin Medical, Cirencester, UK), Charnley (DePuy, Leeds, UK), and Spectron (Smith & Nephew, Memphis, TN, USA) hip prostheses were used (EP group). Modular tumor prostheses such as the Modular Implant System (METS) (Stanmore, Middlesex UK) and Global Modular Replacement System (GMRS) (Stryker Nordic, Malmö, Sweden) were implanted after en bloc resections of the distal femur. Postoperative radiotherapy was not used following salvage procedure, since most patients had been treated following their index procedure.

As part of the registry, long-term follow-up was conducted at regular intervals in conjunction with the patient's regularly scheduled medical oncology visits. In addition, the research team performed chart reviews at regular intervals to determine if and when patients had complications or required a re-operation. No patients were recalled specifically for this study; all data was obtained from medical records and radiographs via the registry.

We classified surgical complications as described by Dindo et al.[72] The primary outcome was reoperation after salvage treatment. Secondary outcomes were overall survival, perioperative death, defined as death within one month of surgery, and other minor complications that did not require surgery or general anesthesia.

Study II: Estimating Survival in Patients with Operable Skeletal Metastases: An Application of a Bayesian Belief Network.

After institutional review board approval, we searched the MSKCC patient management database (Disease Management System, v.5.2, 1996) for all patients who underwent orthopaedic surgery for skeletal metastases of the spine and extremities between 1999 and 2003. We then reviewed all medical records and imaging studies and used these data, along with several other features, to construct the BBN models. All patients had adequate follow-up to determine 12-month survival.

We selected fifteen candidate features based on their current or historical association with survival in this patient population—those undergoing surgery for skeletal metastases. These included the following: age at the time of surgery, race, gender, calculated glomerular filtration rate (mL/min/1.73 m^2), serum calcium concentration (mg/dL), serum albumin concentration (g/dL), type of fracture (impending or complete pathologic fracture), extent of bone metastases (solitary or multiple), senior surgeon's estimate of survival (postoperatively, in months), presence or absence of visceral metastases, presence or absence of lymph node metastases, prior treatment with chemotherapy (yes or no), preoperative hemoglobin (g/dL, on admission), absolute lymphocyte count (K/μL), and the primary oncologic diagnosis. Diagnoses were classified into three groups in a similar fashion as described by Katagiri et al.,[73] but with modifications. Specifically, we considered breast, renal cell, prostate and thyroid carcinoma, multiple myeloma, and lymphoma as Group 3; sarcomas and other carcinomas as Group 2; and gastric, hepatocellular carcinoma lung and melanoma as Group 1.

We used the following definitions during data curation: An impending pathologic fracture was one in which the degree of bone and/or cortical disruption warranted prophylactic fixation to prevent fracture. A pathologic fracture was one in which a cortical lesion resulted in a change in length, alignment, rotation, or loss of height for spine lesions, as determined by good-quality imaging. The surgeon's estimate of survival was made preoperatively after reviewing the patients' medical records and imaging studies, obtaining a complete medical history, and performing a thorough physical examination. Biopsy-proven or clinically obvious metastases to organs within the chest or abdomen were considered visceral metastases. Only biopsy-proven metastases to the lymph nodes were considered indicative of lymph node involvement. Finally, prior chemotherapy was listed if a patient received chemotherapy for the current, active oncologic diagnosis. A list of candidate features is shown in Table 1, below.

Using the features described above, we developed two BBN models trained to estimate the likelihood of survival at 3 and 12 months, respectively. These time points were chosen because they are generally accepted to guide surgical decision-making. Briefly, we used commercially available software that automatically learns network structures and priors from the training data. As such, expected prior distributions (the value or values each feature is likely to assume under various circumstances) were not specified a priori.[74,75] This done, features that were redundant, or unrelated were pruned (removed) from the preliminary models to produce the final model. The number of categories each node could assume was also varied in an effort to maximize the number of first and second-degree associates—those features that are

most closely associated with the outcome of interest, survival greater than three or 12 months, respectively. After model development, cross validation was performed to assess overall accuracy. A more detailed description of the Bayesian modeling and cross validation process is described below in *Statistical Considerations*.

Table 1. This table depicts the candidate features for inclusion into the BBN models for Study II

Feature	Model Label	Description	Node States
Survival > 12 months	Survival > 12 months	Overall survival exceeding 12 months	yes, no
Survival > 3 months	Survival > 3 months	Overall survival exceeding 3 months	yes, no
Surgeon's estimate of survival	Surgeon's Estimate of Survival	The senior surgeon's estimate of survival (in months) after obtaining the patient's history, reviewing his or her laboratory and imaging results, and performing a thorough physical examination	<4, 4-9, 9-18, >18
Oncologic diagnosis group	Diagnosis Group	Primary oncologic diagnosis, grouped as follows: 1: lung, hepatocellular, and gastric carcinoma; melanoma 2: sarcoma and other carcinoma, not in Groups 1 or 3 3: breast, prostate, thyroid, and renal cell carcinoma; myeloma; lymphoma	1, 2, 3
ECOG performance status	ECOG Performance Status	Eastern Cooperative Oncology Group performance status, assessed preoperatively by treating physician	≤2, ≥3
Pathologic fracture status	Pathologic Fracture	Indicates whether surgery was performed for an impending or complete pathologic fracture	yes, no
Number of bone metastases	Number of Bone Metastases	Indicates whether the patient had solitary, or multiple skeletal metastases	solitary, multiple
Organ metastases	Visceral Metastases	Presence of metastases to visceral organs, lungs, or brain	yes, no
Lymph node metastases	Lymph Node Metastases	Presence of lymph node metastases	yes, no
Sex	Gender	Patient sex	male, female
Hemoglobin concentration	Hemoglobin	Preoperative hemoglobin concentration (in g/dL), prior to blood transfusion, if applicable	<10.1, 10.1–11.4, 11.4–12.9, >12.9
Absolute lymphocyte count	Absolute Lymphocyte Count	Preoperative absolute lymphocyte count (K/μL) prior to transfusion, if applicable	<0.6, 0.6–1.1, 1.1–1.6, >1.6

Continuous variables are represented as categorical variables. ECOG = Eastern Cooperative Oncology Group Performance Status

Study III: Treating Metastatic Disease: Which Survival Model Is Best Suited for the Clinic?

For this study, we retrospectively reviewed our institution-owned patient management database (Disease Management System, v5.2, 1996). From this database, we identified all 189 patients who underwent surgery for metastatic bone disease at Memorial Sloan-Kettering Cancer Center between 1999 and 2003. No records were excluded and each contained fifteen variables and sufficient follow-up information to establish survival at 12 months after surgery. Features were identical to those collected for Study II, as were the definitions and outcomes (the likelihood of survival at three and 12 months).

We developed BBN, ANN and logistic regression models for this study. Each was constructed using the same data and trained to estimate postoperative survival at three and 12 months. There were no missing data. Each of the models was internally validated using the cross validation techniques then using decision curve analysis (DCA) to determine which model, if any, was best suited for clinical use. Each of these statistical techniques is described briefly here, but in more detail in the *Statistical Considerations* section, below.

The BBN was developed in a manner similar to that described in Study II. Briefly, all 15 variables (features) were considered as candidate features for inclusion in the model. Again, prior distributions were estimated from the training set and not specified *a priori.* The BBN models were trained to discriminate between two possible outcomes (survival at three and 12 months: yes or no). We then performed ten-fold cross validation to assess the accuracy of the models.

We developed the ANN models using the Oncogenomics Online Artificial Neural Network Analysis (OOANNA) system.[76] First, principal component analysis was performed on all 15 candidate features to identify the top 10 linearly uncorrelated variables with the largest variance. The ANN was composed of three layers: an input layer consisting of the 10 principal components identified above, a hidden layer (which may change the relative emphasis placed on data from each of the inputs) with five nodes, and an output layer, which based on information from the hidden layer estimates the most likely outcomes (survival at three and 12 months: yes or no). Leave-one-out cross-validation was performed by training the model on n - 1 (188) records and then testing it on one independent test record.

Finally, for comparison to the two ML techniques described above, we developed a conventional logistic regression model using variables observed to be potentially significant on univariate analysis (oncologic diagnosis, presence of visceral metastasis, preoperative serum hemoglobin concentration, Eastern Cooperative Oncology Group performance status, and the surgeon's estimate of postoperative survival). For this portion of the analysis, we used STATA® 11.0 statistical software (StataCorp LP, College Station, TX, USA). Ten-fold cross-validation was performed.

We then compared each model using a variety of methods. First, calibration curves were created that plotted predicted risk against actual risk to assess the accuracy of the model predictions. Second, we assessed accuracy by calculating the AUC. Third, decision curve analysis[70] was performed in an effort to help quantify the consequences of over- or undertreatment and to determine which model, if any, was better suited for clinical use.

Study IV: External Validation of the Bayesian Estimated Tools for Survival (BETS) Models in Patients with Surgically Treated Skeletal Metastases.

The purpose of this study was to externally validate the BBN model developed in Study II, using an independent, international skeletal metastasis registry. For this, we chose the Scandinavian Skeletal Metastasis Registry (SSMR). The SSMR contains the records of patients with surgically treated metastatic bone disease, at one of eight major Scandinavian referral centers between 1999 and 2009. Each record contains 84 demographic and clinical variables, including most of the preoperative features required to validate the BBN models. The likelihood of survival at three and 12 months was the outcome. Features and definitions were identical to those described in Studies II and III. Though initially called BETS-3 and BETS-12, the models are now collectively referred to as PATHFx.

Although missing data were acceptable, the validation process required that the variables present within the PATHFx models also be present within the validation set. To satisfy this requirement, we converted the Karnofsky performance score, recorded in the SSMR, to the ECOG performance score using a formula described elsewhere.[77] In addition, we converted hemoglobin concentrations, recorded in the SSMR in mmol/L or g/L, to g/dL prior to validation.

We compared the demographics and patient characteristics of the validation set to those of the test set. Next, we applied data contained in the validation set to the PATHFx models using commercially available software (FasterAnalytics, DecisionQ Corp., Washington, DC, USA). The likelihood of postoperative survival at both three and 12 months was estimated for each record, and we performed ROC analysis to assess accuracy. Validation was considered successful if the AUC was greater than 0.70, and was determined *a priori*. Finally, we performed a detailed analysis of incorrect estimations to describe the misclassification rate, in an effort to characterize the potential clinical impact.

Study V: External Validation of a Tool for the Estimation of Life Expectancy in Patients with Skeletal Metastases—Decision Analysis and Comparison of Three Major International Patient Populations.

Following the first successful external validation study (Study IV), we sought to perform an additional external validation in Italian patients. For this, we partnered with the Italian Society of Orthopaedic and Traumatology (Società Italiana di Ortopedia e Traumatologia or SIOT). Within this society, the Bone Metastasis Study Group identified 287 patients from 2010 to 2013 who underwent orthopaedic surgery for skeletal metastases of the spine and extremities. Each record contained 17 demographic and clinical variables, including most of the preoperative features required to validate the PATHFx models, as well as adequate follow-up to determine postoperative survival at 12 months. All definitions and outcomes (the likelihood of survival at three and 12 months) were similar to those described in the previous, Scandinavian, external validation study (Study IV).

External Validation was also performed as described in Study IV and validation was considered successful if the AUC, determined *a priori*, was greater than 0.70. Finally, decision analysis was performed as described in Study III to better characterize if and how the models should be implemented in both Italian and Scandinavian patient populations.

Study VI: A Probabilistic Analysis of Completely Excised High-Grade Soft Tissue Sarcomas of the Extremity: An Application of a Bayesian Belief Network

The previous studies demonstrated that BBN methodology could be used to produce models designed to estimate the likelihood of survival in patients with metastatic bone disease requiring surgery, and thus guide appropriate surgical treatment. In doing so, we were able to codify complex data into clear graphical models. The purpose was to apply this approach to another, entirely different, area within orthopaedic oncology. Study VI is designed to describe the hierarchical relationships between features related to the treatment of completely excised, localized, high-grade soft tissue sarcomas of the extremity. The results could be used to bolster our understanding regarding the conditional relationships between important features, as well as serve as the basis for later hypothesis testing.

For this study, we turned to the Memorial Sloan-Kettering Cancer Center (MSKCC) Sarcoma Database. This database contains prospectively collected clinical, pathologic, and treatment-related variables for all adult patients with primary and recurrent STS treated at MSKCC since 1982. We extracted all patients with high-grade, extremity STS, who had complete resections. I.e. R0 margins. We chose this specific patient population to control for tumor grade and margin status prior to performing the initial probabilistic analysis.

We chose twenty-seven candidate features based on availability, as well as their current clinical or historical association with DSS in patients with high-grade STS. These included: age at the time of index surgery; gender; size, depth, and location of the primary tumor; histology and histologic variant, if applicable; any oncologic procedures done prior to referral; whether the sarcoma was radiation-induced; the patient's home zip code at the time of referral; the surgical service and surgeon of record; need for re-excision following referral; type of surgical procedure performed; presence of bone, nerve, or vascular invasion; whether bone or nerve was resected with the tumor; adjuvant treatment and timing of chemotherapy or radiotherapy; presence of and time to local recurrence (LR); presence of and time to distant recurrence (DR); and death from disease.

Only extremity tumors were considered; that is, those distal to the axillary fold in the upper extremity or those distal to the inguinal ligament in the lower extremity. In addition, we used the following definitions. A sarcoma was considered radiation-induced if it occurred within the radiated field, more than six-months after irradiation, and was histologically dissimilar from the original tumor. Bone adherence/invasion was considered present if, on radiographs or cross-sectional imaging, the tumor exhibited any effect on any bone, including periosteal reaction. Nerve and/or vascular invasion was determined by histology. The presence of and time to DR was determined by imaging. Local recurrences were diagnosed on follow-up by physical examination and/or imaging, and were confirmed by histology. Time to both DR and LR were calculated from the date of initial operation. The candidate features and descriptions of node states are listed in Table 2.

Table 2. This table depicts the 27 candidate features considered for inclusion in the BBN model for Study VI

Candidate Feature	Label	Description	States
Age*	AGE	Patient age, at the time of surgery	CV
Gender	SEX		Male Female
Size*	PRIMARY SIZE CATEGORY	Size category of tumor in maximum dimension	≤5cm 5-10cm >10cm
Depth*	DEPTH	Depth of primary tumor compared to investing fascia of limb	Superficial Deep
Site*	SITE		Upper extremity Lower extremity
Subsite	SUBSITE	Extremity tumors distal to the vertical plane made by the axillary fold and horizontal plane made by the inguinal ligament were considered.	Hand Forearm Elbow Arm Axilla Shoulder Groin Hip Thigh Knee Leg Ankle Foot
Histology*	HISTOLOGY	Final histology following excision, review by three pathologists	MFH/HGPS Synovial Liposarcoma Leiomyosarcoma MPNST Fibrosarcoma Other
Variant*	VARIANT	Histologic variant, if applicable	Monophasic Biphasic…etc.
Presentation status*	PRESENTATION STATUS	Oncologic procedures performed prior to referral (if any)	None Biopsy only Marginal excision Wide excision
Radiation induced	RT INDUCED	Whether the sarcoma is considered radiation induced	Yes No
Referring zip code	FIRST 3 DIGITS ZIP	First three digits of patient's home zip code at the time of referral	CV
Surgeon*	SURGEON CODE	Thirty-one surgeons, listed anonymously	A-EE
Service*	SERVICE CODE	Two surgical services	GMT Orthopaedic Surgery
Re-excision*	RE EXCISION	Whether the patient, upon referral, required a tumor bed excision	Yes No
Procedure*	PROCEDURE	Type of procedure performed	Limb sparing Amputation
Bone invasion*	BONE INVASION		Yes No
Bone resection*	BONE RESECTED		Yes No

Candidate Feature	Label	Description	States
Nerve invasion*	NERVE INVASION		Yes No
Nerve resection*	NERVE RESECTED		Yes No
Vascular invasion	VASC INVASION		Yes No
Chemotherapy			
Pre-op*	PREOP CHEMO		Yes No
Post-op*	POSTOP CHEMO		Yes No
Radiotherapy			
Pre-op*	PREOP RT		Yes No
Post-op*	POSTOP RT		Yes No
Time to local recurrence*	TIME TO LR	In months	CV None
Time to distant recurrence*	TIME TO DR	In months	CV None
Death from disease*	DOD	Whether the patient died of disease, a reflection of disease-specific survival	Yes No

* Denotes those candidate features included in the final model. CV=Continuous variable, MFH/HGPS=malignant fibrous histiocytoma/high-grade pleomorphic sarcoma, MPNST=Malignant peripheral nerve sheath tumor, LR=Local recurrence, DR=distant recurrence

The BBN models were developed in a manner similar to those developed for Studies II and III. Briefly, all 27 features were considered as candidates for inclusion in the model. We imputed values for features in which missing data represented less than 30% of the entire field. This included six features within the training set: bone invasion (missing in 5.4%), bone resection (7.9% missing), nerve invasion (11.6% missing), nerve resection (12.2% missing), vascular invasion (12.4% missing), and repeat excision (27.5% missing). None of the features were pruned because of missing data. Since relatively few patients had LR (14.7%) and/or DR (31.7%), we defined a ‘‘missing’’ value for each of these features as no LR or DR.

We then trained the BBN model to evaluate prior probability distributions in order to develop a classifier to estimate the probability of DSS. The network structure was then portrayed graphically to illustrate the conditional interdependence and hierarchy of the features, and inference tables were calculated to describe the posterior estimates of probability for all possible permutations with respect to the outcome. We then performed ten-fold cross-validation to assess the accuracy and robustness of the final model. Model development and cross validation techniques are described in more detail in the *Statistical Considerations* section, below.

Statistical Considerations

This thesis focuses on applying Bayes' theorem of conditional probability to develop clinical decision support tools. In doing so, we show how this ML technique is used to understand important relationships between features, and how perturbations of each feature influence the probability of an outcome. This is one of the most important qualities of Bayesian networks and allows us to infer the likelihood of "unobserved" outcomes, like the probability of survival, for instance, from information on other features in the network that have been observed.

This approach stands in contradistinction to classical, frequentist statistical analysis. In the classical sense we focus on estimating and testing hypotheses by assessing parameters of a fixed distribution using a sampling of that distribution. This approach is well suited for the analysis and comparison of most biomedical data and even allows one to calculate the likelihood of past and future events—as long as the experimental conditions do not change. However, if the goal is to describe a system in which the experimental conditions cannot, or should not, be replicated, then we should look beyond the frequentist approach in favor of probability theory.

In cancer, therapeutics and treatment philosophies evolve, and ideally improve, over time, and each patient's burden of disease is arguably different than the last. In addition, health care practice and policy differs between cultures, and is often isolated (not controlled for) by experimental design. As such, there is no reason to believe that the conditions (patient and disease demographics, treatment protocols, *etc*.) present in a given cancer center would apply to patients in other centers, worldwide. In this dynamic and variable landscape, we must choose an approach that can accommodate a range of experimental conditions and uncertainty within the data, while describing relevant relationships in a transparent, economical manner. Though many ML techniques can be applied individually, or in combination, the focus of this thesis involved the use of Bayesian statistics.

Bayes' Theorem of Conditional Probability

In the simplest form, Bayes' theorem is represented by the following equation:

$$P(h|B) = \frac{P(B|h) \times P(h)}{P(B)} \qquad (1)$$

By definition, $P(h)$ represents the prior probability of hypothesis h and reflects any background knowledge regarding the likelihood that h is correct. Although this can be derived from expert opinion or large-scale studies, unsupervised ML methods such as those used for this thesis derive h from the training data and were not specified *a priori*. $P(B)$ represents the prior probability of B, which is the probability that the data, B, will be observed. Next, $P(B|h)$ is the probability of observing B, given that h is true. Finally, $P(h|B)$, the posterior probability of h reflects confidence that h remains true after B has been observed. In the above equation $P(B)$ is a normalization factor that will ensure the $P(h|B)$ is a number between 0 and 1. In essence, the formula allows us to estimate the likelihood of h based on the available data (or evidence), B. The term $P(h|B)$ is generally called the "*posterior*" estimate, and equation (1), when written in English becomes:

$$posterior = \frac{prior \times likelihood}{evidence\ based\ normalization\ factor} \qquad (2)$$

In other words, Equation (2) helps us understand that Bayes' theorem is a tool that can be used to update beliefs (in the form of a hypothesis h), in response to the available evidence. This may seem familiar to clinicians and scientists, who do the very same thing when processing information from the literature, conferences, and their clinical practice. However, Bayes' theorem generates the probability that h is true, and also provides us with a measure of confidence in that estimate. This allows clinicians to weigh the results of each estimate, while assimilating it and other sources of clinical information when formulating a treatment plan.

When written in the context of this thesis, we can better understand how Bayes' theorem applies to patients with metastatic bone disease. For instance, if an outcome, phrased as a hypothesis, h, is that *"this patient will survive longer than one year,"* then $P(h|B)$ represents the probability that "*this patient will survive longer than one year*," given the body of knowledge we call "*evidence*," B, which contains information about the patient's disease, the extent of it, laboratory values, functional status, and the physician's own experience.

Complex probabilistic models may be represented as graphical networks to provide the user with an easy to understand visual representation of complex conditional relationships between variables. These provide a convenient way to express assumptions, facilitate the concise representation of the relationships between features, and derive inference from observations.[78] Graphical networks can then depict all of the relevant relationships between variables derived from equation (1) in an intuitive, transparent and comprehensive manner. This is a major advantage over other non-graphical techniques and allows the physician to easily understand the hierarchy and relative importance of each feature.

The structure of the network depends solely on the probabilistic relationships between features. First, conditional dependence for a group of features can be represented mathematically by a joint probability distribution function (jPDF). The resulting jPDF allows one to describe the hierarchical relationships between features in a graphical manner and then calculate the probability of a feature (e.g., three-month postoperative survival) assuming a particular value (yes/no), expressed in terms of the values of two or more features.

The resulting network is called a directed acyclic graph (DAG), wherein relationships between nodes shown in Equation (3) can be described as parent and/or child relationships, depending on which feature directly determines the value of the other. In general, the value of the child node is dependent on the value of the parent node.

$$parent \rightarrow child \tag{3}$$

By the same logic, two features can be considered conditionally independent, particularly in the setting of an intermediate feature. In the following examples of simple Bayesian networks, DAGs are represented in Equations (4-6). Note there is a link, referred to as an "edge," between A and B, and B and C, but not between A and C. Because independence is implied by the absence of an edge. Features A and C are thereby conditionally independent, if the value of B is known.

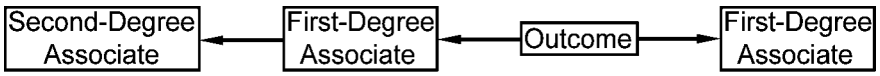
$$A \rightarrow B \rightarrow C \tag{4}$$

$$A \rightarrow B \leftarrow C \tag{5}$$

$$A \leftarrow B \leftarrow C \tag{6}$$

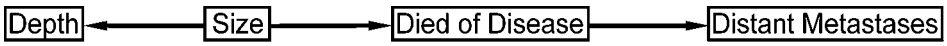
However, if B is the outcome of interest, its value is conditionally dependent on the values of both A and C. Though directionality is important when inferring causality (a child node is dependent on the parent node), it is not a function of the jPDF, and therefore directionality is not used to estimate the likelihood of any feature. As such, the jPDFs for any variables in the network depicted in Equations (4-6) are identical regardless of implied directionality.

For this thesis, we sought to estimate the likelihood of certain outcomes. As the following DAG shown in Equation (7) depicts, features that share an edge with the outcome are *First-Degree Associates.* Those that share an edge with *First-Degree Associates* are *Second-Degree Associates*, regardless of the direction of dependent relationships indicated by the arrows.



(7)

In general, knowledge of a *Second-Degree Associate* does not improve estimates of the outcome, provided the value of the shared *First-Degree Associate* is known (i.e., as long as the value of the *First-Degree Associate* is present in the evidence). In the context of soft tissue sarcomas described in Study VI, we observed these conditional relationships, represented in the following DAG:



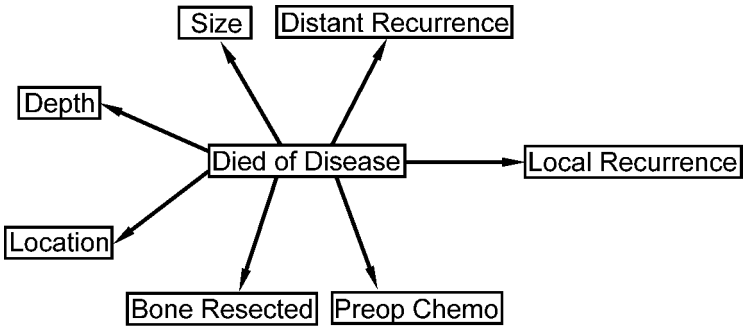
(8)

In this Bayesian network, both *Size* and *Distant Metastases* are first-degree associates of *Died of Disease*. Based on the direction of the edges, this model suggests that both disease-specific survival and metastases are dependent on the size of the tumor at diagnosis. However, because the jPDF does not consider causality, the network structure allows one to infer the probability of *[death from] Disease*, if either the *Size* of the tumor or the presence of *Distant Metastases* are known. In addition, the model structure informs us that the *Depth* of the tumor is conditionally independent of *[death from] Disease*, if the *Size* of the tumor is known. In other words, if our goal is to estimate the likelihood of *[death from] Disease*, knowing the *Depth* of the tumor (a second-degree associate) does not improve this estimate, as long as we are able to measure its *Size* (the shared first-degree associate). One useful characteristic of a Bayesian approach is its ability to predict outcomes even when values for some variables are not provided in the evidence. For example, if a first-degree associate is missing (*Size*) but a second-degree associate is provided (*Depth*), then tumor size can be estimated and the best estimate for the outcome, given the available evidence, can still be calculated. In fact, once the jPDF for a given system is known, one can instantaneously investigate all nodes as potential outcomes of interest. This is helpful, not only when imputing missing data, but also to develop and test additional hypotheses, such as those described in Study VI, wherein we evaluated the effects of tumor size, and timing of local recurrence on disease-specific survival.

For some data, a "Naïve Bayes classifier" may be used. Given a set of clinical data B containing several features $(F_1, F_2...F_n)$, and outcome, Z, naïve Bayes learning assumes each feature is unrelated to all others, except Z. In other words, the influence of each feature on Z is independent of the presence or value of one or more

other features. In addition, this simple Bayes classifier considers each feature to contribute independently to the probability of Z. The following example of a naïve Bayes network shown in Equation (9) assumes each of the features inherent to the treatment of high-grade soft tissue sarcomas contributes to the probability of the outcome: $Died\ of\ Disease$.

Though useful in small datasets, the naïve Bayes classifier seems rather ill suited for use in living systems, since most biologic and physiologic processes are interrelated. In Equation (9), features such as $Size$ and $Depth$ are nearly always related (larger tumors are more commonly deep), and some features like $Distant\ Recurrence$ are typically much more influential than others in terms of disease specific survival. Nevertheless, Naïve Bayes modeling is an efficient screening technique and is commonly used to determine if more thorough modeling is necessary.



(9)

Bayesian Belief Model Development

To mitigate the problems with assuming each feature is independent of the others, the BBN models developed for this thesis were created using a method that makes no assumptions regarding independence. We used commercially available ML algorithms (FasterAnalytics; DecisionQ, Washington, DC) that automatically learn network structures and priors from the training data. By this method, priors—how and under what circumstances the value assumed by one feature depends on the value(s) of other features—were not specified *a priori*.[74,75] Rather, FasterAnalytics assigns a network configuration or structure, computes the strength of the networks robustness or predictive ability, then perturbs the network repeatedly until a maximum is achieved. In this fashion, robustness (r), depicted in Equation (10), is maximized by generating a series of network structures $S=(s1, s2... sn)$, across a continuum. Random starts and stops ensure the true maximum, and not local maxima are achieved.

$$f'(r_S) = 0 \tag{10}$$

It is important to note that the most robust network is a compromise, balancing accuracy with the degree of model complexity that is defined by the user. This subjective input helps prevent undue model complexity that would limit clinical applicability. End users may find models containing scores of features tedious; however, even large models may be useful in cases where inputs are automated, e.g. following genetic sequencing, multiplex assays or if nested within an electronic medical record. By the same token, one must also avoid overfitting the model to the

training data, thereby limiting applicability in other settings.

Overfitting occurs when a modeling process describes the noise (random variation for a variable) within a set of observations, rather than the signal (the features that actually influence the outcome). Overfitting is usually revealed during external validation, when independent data is used to validate the model. If overfit, the model's accuracy, or predictability will be rather poor, in most cases no better than a random guess. We used various techniques to mitigate the risk of overfitting the models generated for this thesis, which will be discussed in more detail, below.

Data Curation and Model Development

First, we accounted for missing data using a passive, truncation-based imputation algorithm.[79] We typically imputed values for features in which missing data comparised less than 30% of the total records. Next, continuous features were converted to categorical ones using an equal-area binning process based on prior distributions learned from the training set.[78] In an effort to balance goodness-of-fit against robustness, we applied a parsimony metric designed to reduce the risk of overfitting.[79] Using an iterated process, unrelated and redundant features were pruned, and thereby removed from the preliminary models to produce the final model. The number of categories each node could assume was also varied systematically in an effort to maximize the number of *First-* and *Second-Degree Associates*.

For the estimation of survival, two models were generated, one for each survival time period (outcome). It is easy to see that different variables could influence survival over a three-month period versus a 12-month period. If both survival periods were included in the same model, different survival estimates could serve as confounders in the same model. Second, by estimating the likelihood of three-month and 12-month survival, we effectively classify patients into three survival estimates: <3 months, 3-12 months, and >12 months. These time points were chosen because they are useful for orthopaedic surgical decision-making, as discussed previously.

Artificial Neural Network Development

We developed the ANN models using the Oncogenomics Online Artificial Neural Network Analysis (OOANNA) system,[76] which uses feed-forward multilayer perceptron (MLP) ANNs. We performed principal component analysis on all fifteen-candidate features to identify the top 10 linearly uncorrelated variables with the largest variance. This was done in an effort to simplify, as well as mitigate overfitting of the model to the training data.

This MLP network was composed of three layers: an input layer consisting of the 10 principal components identified above, a hidden layer (which may change the relative emphasis placed on data from each of the inputs) with five nodes, and an output layer, which based on information from the hidden layer estimates the most likely outcomes (survival at three and 12 months: yes or no). Briefly, data from all 189 study subjects were uploaded into the OOANNA system, which automatically selected the top 10 principal components, for inclusion in the ANN model.

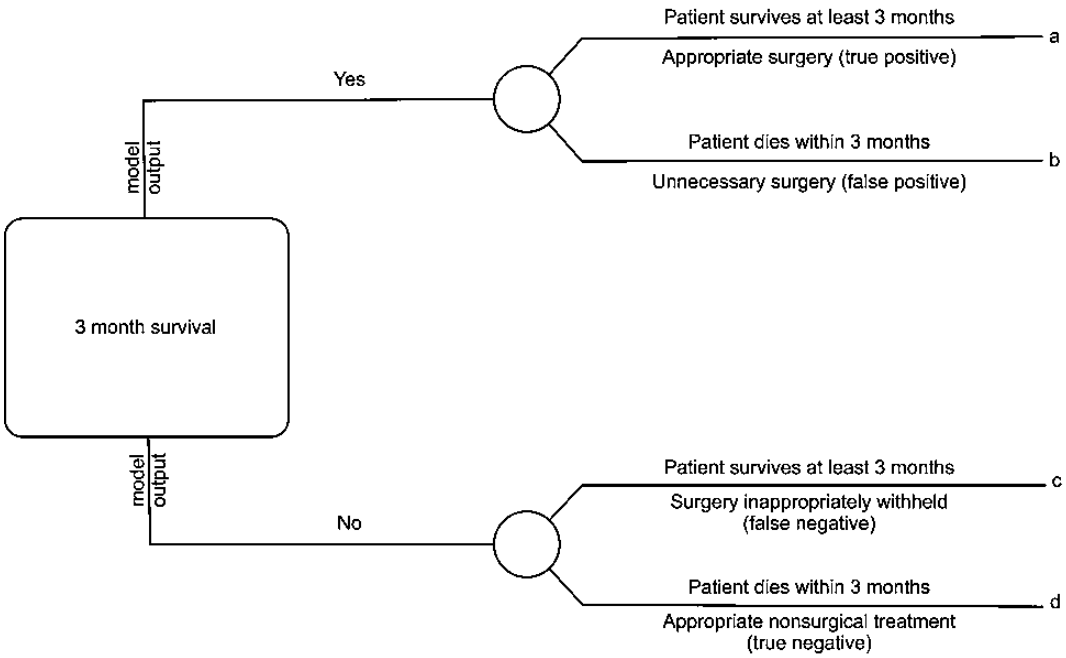
Target Shuffling

In order to ensure the relationships identified by the modeling methods described above are real, and not due to chance alone, target shuffling should be used.[80] For this evaluation, outcomes in the training data are shuffled randomly among each of the records. This process is repeated for 1000 iterations and the modeling

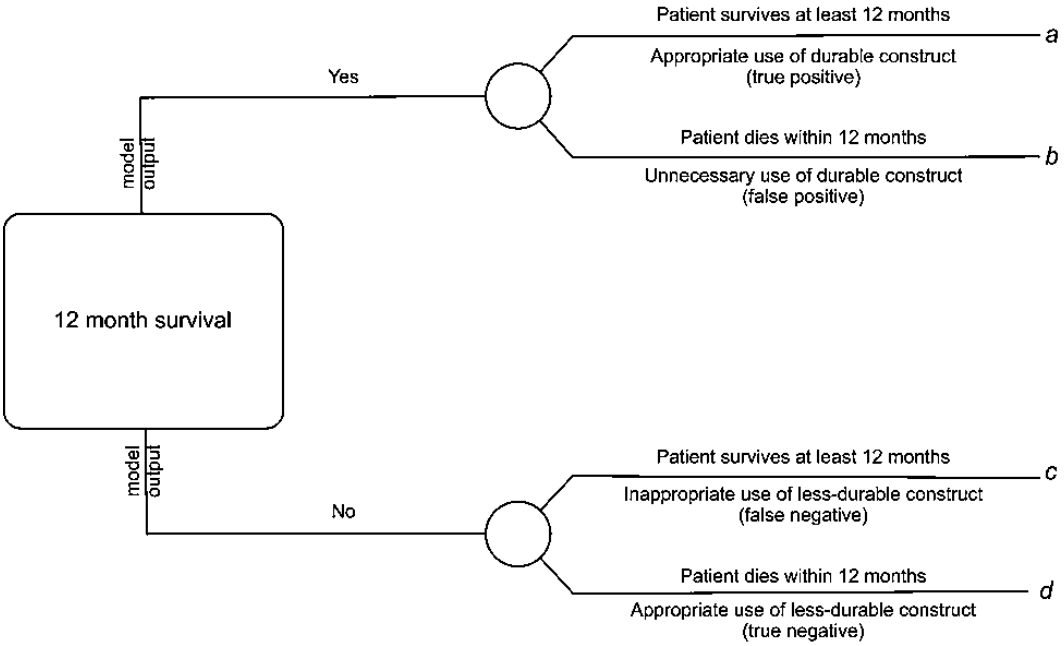
methods are applied to each of these "permuted null distributions." By this method, the random association(s) between survival outcomes, and features contained within the training set are expected to be weak. Iterated cross-validation is then used to evaluate the likelihood of sample bias resulting from a single round of k-fold cross-validation. Each model for each of the 1000 iterations undergoes 10-fold cross validation as described in detail, in *Validation Techniques*, below. The mean AUC for all 1000 iterations is reported. This allows one to calculate the probability that one or more of the modeling techniques described above resulted from chance alone.

Decision Analysis

We performed decision curve analysis[70] in the following manner. In contrast to decision tree analysis, which weighs the possible consequences of several decisions, decision curve analysis helps quantify the consequences of over- or undertreatment of a disease process. When constructing the decision curves, we assumed clinical decisions would be based strictly on the output of each model. For instance, the decision to offer surgery would be based on the likelihood of survival at three months, whereas the choice of implant (more durable or less durable) would be based on the likelihood of survival at 12 months. Each model generates a survival probability p at specific time points after surgery. If the probability is near 1, surgeons may choose to recommend surgery in the case of the three-month model and a more durable implant in the case of the 12-month model. If the probability is near 0, nonsurgical treatment may be recommended in the case of the three-month model or a less invasive/less durable implant in the case of the 12-month model. At some probability between 1 and 0, however, surgeons may have difficulty choosing a treatment method. For this study, we defined the point at which surgeons become indecisive as the threshold probability p_t, where the expected benefit of treatment is equal to the expected benefit of no treatment. The treatment decision trees are depicted for three-month survival in Equation (11), and 12-month survival in Equation (12); in which a, b, c, and d represent values associated with each possible outcome (*Reprinted with kind permission from Springer Sciences + Business Media*) (Study III).



(11)



(12)

For instance, for the three-month models, a - c is defined as the consequence of a false-negative result, withholding surgery in someone who actually survives long enough to benefit (ie, > 3 months, in this case). Similarly, d - b is defined as the consequence of a false-positive result, performing surgery in someone who does not live long enough to benefit (< 3 months). For the 12-month models, the definitions remain the same; however, the clinical impact changes. For 12-month survival, a - c remains the consequence of a false-negative result, but in this case, a less durable implant is inappropriately chosen in a patient who outlives his/her implant and subsequently requires a revision procedure. Similarly, d - b remains the consequence of a false-positive result; however, this time it results in unnecessarily aggressive surgery in a patient who does not live long enough to benefit.

From the decision trees, we derive the following formula as previously described:[70]

$$\left(\frac{a-c}{d-b}\right) = \frac{1-p_t}{p_t} \tag{13}$$

Simply stated, the threshold probability of survival $\mathrm{p_t}$ in which a surgeon decides (a) whether to offer surgery and/or (b) whether a more durable implant is necessary is related to how he/ she weighs the consequences of overtreating or undertreating the patient. By letting the value of a true-positive result be 1, we arrive at the following formula: [81]

$$Net\ benefit = \left(\frac{true\ positive\ count}{n}\right) - \left(\frac{false\ positive\ count}{n}\right) \times \left(\frac{p_t}{1-p_t}\right) \tag{14}$$

In this fashion, each model's net benefit, defined as a three- or 12-month survivor who duly receives an operation and implant commensurate with his/her estimated survival, can be plotted against the entire range of threshold probabilities, acknowledging that clinically-relevant ranges comprise a smaller interval.

Validation Techniques

Cross Validation

In smaller datasets without external validation data (Studies II, III, VI), we used cross validation methods to assess the accuracy of the models and help mitigate the risk of overfitting to the training data. For ten-fold cross validation, we first randomized the data into 10 matching train-and-test sets using R© Version 3.0.2.[1] Each set consisted of a training set composed of 90% of patient records and a test set composed of the remaining 10% of records. Each matching set was unique to ensure there was no overlapping information between sets. Ten models were developed using each training set, and then tested on the unknowns contained within the corresponding test set. The ANN code used for Study III could not be configured for ten-fold cross validation. Instead, we used leave-one-out cross validation, which was accomplished by training the model on n - 1 (188) records and then testing it on one independent test record. In this fashion, the ANN, using the 10 principal components, estimated the likelihood of three- and 12-month survival for each independent test record.

"Holdout" Validation

For larger datasets, it is possible to create a "holdout" set using R© Version 3.0.2.[1] This technique produces training and test sets that are both relatively large. Prior to model development, 25% of records are typically "held out" for later testing, although this proportion may be varied. Models would then developed using the training set, comprised of 75% of the original records. Importantly, the distribution of outcomes should be held constant between the training and test sets, whenever possible.

External Validation

Because cross validation techniques typically overestimate model accuracy, external validation is necessary prior to widespread clinical use. This ensures models are tested in a variety of cultures and settings, with varying patient demographics and treatment philosophies. We were fortunate to have two international external validation studies to support this thesis (Studies IV and V). To perform external validation, data from the independent sources were applied to each model. There were large amounts of missing data, including the surgeon's estimate, as well as the absolute lymphocyte count that were missing in the majority of records. This highlights one of the strengths of the Bayesian method, which retains functionality in the presence of missing input data.

Frequentist analysis

For frequentist analysis, (all studies) we used the following approach. Continuous variables were tabulated and presented as mean (standard deviation), median (interquartile range—IQR) and categorical variables as number (%). The distribution of each continuous variable was compared with the normal distribution using the Shapiro-Wilk test. Equality of variance for continuous variables was determined using the Brown-Forsythe and Levene test. Statistical differences between continuous variables versus the bivariate outcome variables were evaluated using the Mann-Whitney U-test and the post hoc Tukey-Kramer assessment. Categorical

variables were also tabulated and associations compared using Fisher's exact test or chi-square analysis, depending on the number of expected values in the contingency matrix. A two-tailed α of 0.05 was considered statistically significant. We used JMP® Version 9.0.2 (SAS Institute, Inc., Cary, NC, USA), R© Version 3.0.2[1] and STATA® 11.0 statistical software (StataCorp LP, College Station, TX, USA) for statistical estimations.

Figures, Illustrations, and Equations

For figures and illustrations, we used GraphPad Prism® Version 5.0, or Adobe® Illustrator® CS6 Version 16.0.4. We adjusted image size and resolution for some figures using Adobe® Photoshop® CS6 Version 13.0x64. All equations were generated using OmniGraffle Version 5.4.4, or the Equation Editor nested within Microsoft® Word for Mac 2011, Version 14.4.6.

Results

Study I: Which Implant Is Best After Failed Treatment for Pathologic Femur Fractures?

Follow-up was complete for all patients at a median of 8 months (IQR 3, 22). At the time of last follow-up, six patients remained alive with disease; however, most (82) patients had died. Five patients died within the perioperative period, defined as within four weeks of surgery, but there were no intraoperative deaths.

We compared age, sex, location within the femur, the type of initial and final implant used (Table 3). The most common oncologic diagnoses were breast, kidney and myeloma, and the majority of patients had generalized skeletal metastases (Table 4). There were 16 PLATE, 11 IMN, 61 EP implants used.

Table 3. Patient characteristics by treatment group are depicted for 88 patients in whom primary constructs failed.

Patient characteristic	Entire cohort (n = 88)	EP (n = 61)	PLATE (n = 16)	IMN (N = 11)	95% C.I. or prob.
Follow-up (months)	8 (3, 22)	7 (3, 17)	13 (3-30)	15 (1-56)	-2.1, 38.0 -10.6, 36.2 -22.9, 33.1
Age (years)	65 (58,74)	65 (58,74)	68 (61,74)	70 (58-73)	-4.9, 10.2 -8.6, 9.0 -7.6, 13.3
Male	37 (42)	28 (46)	6	3	p=0.47
Location					p<0.001
Femoral neck	10 (11)	8 (13)	2	0	
Peritrochanteric	23 (26)	21 (34)	1	1	
Subtrochanteric	31 (35)	25 (41)	2	4	
Diaphyseal	13 (15)	4 (7)	5	4	
Distal	11 (13)	3 (5)	6	2	
Initial fixation method					p=0.02
Plate	37 (42)	26 (43)	9	2	
Intramedullary nail	26 (30)	17 (28)	3	6	
Endoprosthesis	22 (25)	18 (29)	3	1	
Cement with or without pin Fixation	3 (3)	0	1	2	
Salvage procedures requiring reoperation	17 (19)	4 (7)	8	5	p<0.001
Time to failure of salvage implant (months)	10 (2,14)	7 (3,14)	6(4,12)	12 (10,35)	p<0.001

All values are presented as median (interquartile range) or frequency (percent); EP = endoprosthesis; IMN = intramedullary nail.

Table 4. This table describes the oncologic diagnosis by median survival and treatment group.

Cancer diagnosis	Median survival in months (IQR)	Overall	EP	PLATE	IMN
Breast	11 (5,20)	27	18	5	4
Kidney	7 (3,38)	21	13	6	2
Myeloma	18 (5,43)	16	11	1	4
Prostate	7 (4,9)	9	6	2	1
Lung	3 (2,6)	8	7	1	0
Other	3 (2,22)	7	6	1	0
Diagnosis category					
I	3 (2,6)	8	7	1	0
II	3 (2,22)	5	4	1	0
III	11 (4,30)	75	50	14	11
Extent of metastases					
Solitary skeletal	13 (5,39)	27	18	4	5
Multiple skeletal	9 (4,22)	41	33	4	4
Generalized	5 (2,22)	20	10	8	2

IQR= interquartile range EP = endoprosthesis; IMN = intramedullary nail; Diagnosis category I=lung; II=other; III=Breast, Kidney, Myeloma, Prostate

Of the 88 patients requiring salvage treatment after failed reconstructions, 17 required reoperation (Table 3). The demographic information including median age, duration of follow-up, gender distribution and the proportion of perioperative deaths were similar between groups. As expected, EPs were performed most commonly for proximal (neck, peri- and subtrochanteric) lesions; plate fixation for distal femoral (metaphyseal and diaphyseal) lesions; and IMNs were for those in the diaphyseal and subtrochanteric regions. We observed a difference ($p < 0.001$) in the proportion of failures across treatment groups: EP (7%), IMN, (45%), and PLATE (50%) (Figure 2). Using logistic regression, we were able to discern the effect of the salvage implant on treatment failure after controlling for age, diagnosis, and location (chi-square = 7.92, DF = 2, $p = 0.019$) Specifically, the EP group had a lower chance of treatment failure than the PLATE (OR, 0.10; 95% CI, < 0.001 to 0.27) group. The observed odds ratios between EP and IMN and between IMN and PLATE with regard to treatment failure were 0.08 (0.01–1.11) and 0.13 (0.006–2.17), respectively.

Failure of salvaged implants, requiring reoperation (grade III complication[72]) occurred at a median time of 10 months (IQR 2, 14). The most common cause was material failure (n = 15) followed by progression of disease (n = 1) or a combination of these (n = 1). Considering the location of the tumor, failures in the diaphyseal region were most common (n = 8) followed by subtrochanteric (n = 5), peritrochanteric (n = 2), and distal femoral (n = 2). There were no failures after salvage in the femoral neck region, ostensibly since these were routinely excised and reconstructed with an EP. There were no reoperations for dislocations or infections and no pathologic fractures.

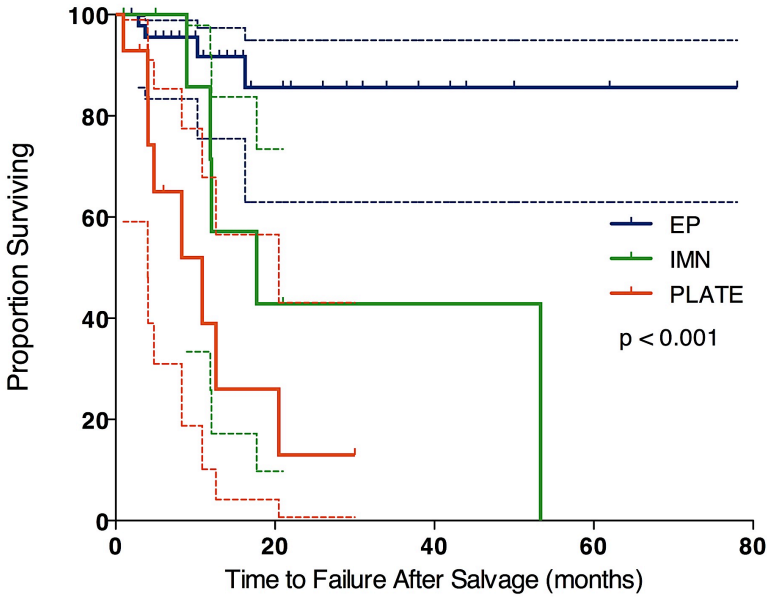


Figure 2. This Kaplan-Meier survival curve depicts the time to reoperation after salvage treatment, grouped by the type of implant. EP = endoprosthesis; IMN = intramedullary nail; PLATE = plate fixation including screw and side-plate devices. 95% C.I. boundaries are reported for each group.

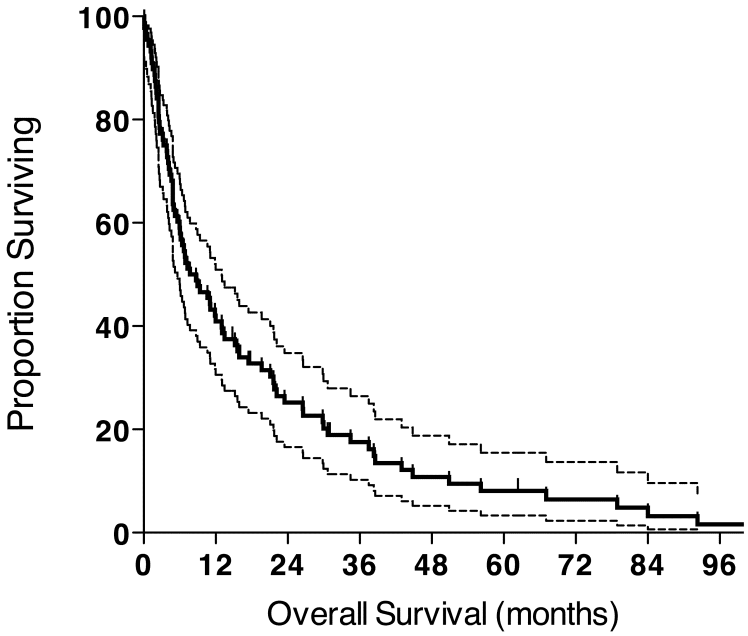


Figure 3. This Kaplan-Meier survival curve illustrates overall survival (OS) of the entire cohort after salvage treatment for failed femoral implants. Note that the median OS is approximately eight months (IQR 3-22).

Seventeen patients had perioperative complications that did not require surgery. These included prosthetic dislocations (grade IIIa, n = 6), superficial wound infections (grade I, n = 8) as well as systemic (medical) illness (grade II, n = 2) or a combination of these (n = 1). Though implant survival was clearly superior compared to the IMN or PLATE groups, the proportion of perioperative complications was higher (p = 0.04) in the EP group. The overall survival for the entire patient population is shown in Figure 3.

Study II: Estimating Survival in Patients with Operable Skeletal Metastases: An Application of a Bayesian Belief Network.

We identified the records of 189 consecutive patients suitable for analysis. Median follow-up was 8 months (IQR 2, 22). This was adequate to determining overall survival at 12 months after surgery. Median patient age was 62 years (IQR 54, 72). Most patients were women (55%), and white, non-Hispanic (85%). Most also had visceral metastases (60%), multiple skeletal metastases (71%), and prior systemic therapy (56%). Most patients were in oncologic Group 3 with more favorable diagnoses (55%), followed by Group 1 (27%), the least favorable diagnoses and then Group 2 (18%). As shown in Figure 4, 58 patients (31%) survived less than three months, 53 (28%) survived 3-12 months, and 78 (41%) survived more than 12 months.

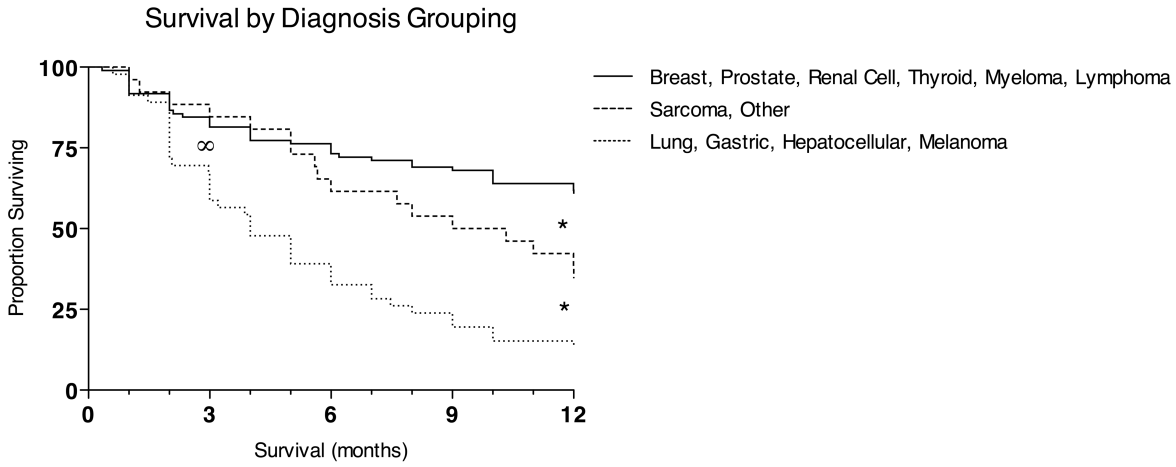


Figure 4. This Kaplan-Meier curves illustrates overall survival (OS) for each patient, by diagnosis group. The OS in Group 1 was significantly lower than that that in Groups 2 and 3 at the three-month time point$^{\infty}$($p<0.0001$, log-rank test). Overall survival was also significantly different between all groups at the 12-month time point* ($p<0.0001$, log-rank test).

First-degree associates differed between the two models. In the three-month model (Figure 5), the *Surgeon's Estimate of Survival*, preoperative *Hemoglobin* concentration, preoperative *Absolute Lymphocyte Count*, complete *Pathologic Fracture*, and *ECOG Performance Status* were first-degree associates of survival. In the 12-month model (Figure 6), only the *Surgeon's Estimate of Survival*, preoperative *Hemoglobin* concentration, *Number of Bone Metastases*, and the oncologic *Diagnosis Group* were first-degree associates of survival. In the three-month model, the oncologic *Diagnosis Group* and the presence of *Visceral Metastases* were first-degree associates of the *Surgeon's Estimate of Survival.* This indicates that if the *Surgeon's Estimate of Survival* is unknown, knowledge of the oncologic *Diagnosis Group* and whether the patient has *Visceral Metastases* can serve as acceptable surrogates. In the 12-month model, however, *ECOG Performance Status* and presence of *Visceral Metastases* were first-degree associates of the *Surgeon's Estimate of Survival*, and represent acceptable surrogates in this setting.

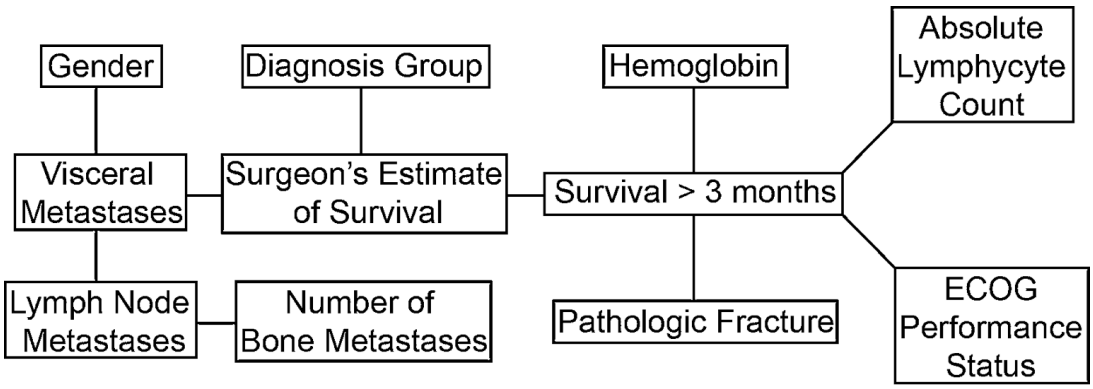


Figure 5. This is a BBN model depicting the relationships between features associated with three-month survival. As shown, there are five first-degree associates of three-month survival. These include the *Surgeon's Estimate of Survival*, preoperative *Hemoglobin* concentration, preoperative *Absolute Lymphocyte Count*, *ECOG Performance Status*, and the presence of a *Pathologic Fracture*. The network structure indicates that the primary oncologic *Diagnosis Group* and the presence of *Visceral Metastases* are both first-degree associates of the *Surgeon's Estimate of Survival*.

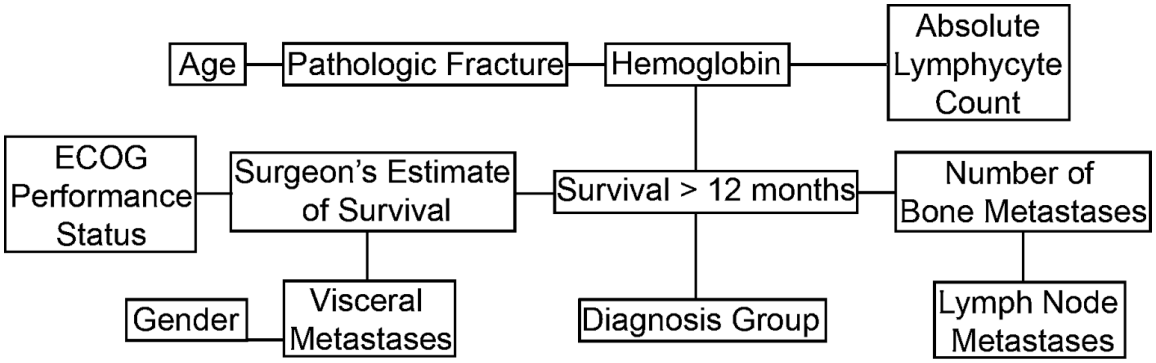


Figure 6. This is a BBN model depicting the relationships between features associated with 12-month survival. There are four first-degree associates of 12-month survival. These include the *Surgeon's Estimate of Survival*, preoperative *Hemoglobin* concentration, the *Number of Bone Metastases*, and the primary oncologic *Diagnosis Group*. In contradistinction to the three-month model, *ECOG performance status* and the presence of *Visceral Metastases* are first-degree associates of the *Surgeon's Estimate of Survival*.

Cross-validation produced a mean AUC of 0.85 (95% CI: 0.80–0.93), and 0.83 (95% CI: 0.77–0.90) for estimating the probability of postoperative survival at three and 12 months, respectively. We generated inference tables for the ten most likely combinations (Tables 5 and 6). The probability of three-month survival for the ten-most common scenarios ranged from 3.3– 99.2%, while the probability of 12-month survival varied from 0.8–93.6%. In all, there were 256 and 128 potential permutations for the three- and 12-month models respectively.

Table 5. Posterior estimates of survival at three months, ten most frequent cases

	First-Degree Associates					Outcome	
Expected Frequency	ECOG	Absolute lymphocyte count (K/µL)	Completed Pathologic Fracture	Hemoglobin (g/dL)	Surgeon's estimate of survival (months)	Probability of Survival >3 months	
						No	Yes
2.0%	≥ 3	< 0.6	Yes	< 10.1	< 4	96.7	3.3
1.3%	≥ 3	< 0.6	No	< 10.1	< 4	91.1	8.9
1.7%	≥ 3	< 0.6	Yes	10.1–11.4	< 4	95.3	4.7
1.2%	≥ 3	< 0.6	No	10.1–11.4	< 4	87.6	12.4
1.1%	≥ 3	0.6–1.1	Yes	< 10.1	< 4	94.8	5.2
1.0%	≥ 3	0.6–1.1	Yes	10.1–11.4	< 4	92.7	7.3
0.9%	≤ 2	< 0.6	Yes	< 10.1	< 4	89.5	10.5
0.9%	≥ 3	< 0.6	Yes	11.4–12.9	< 4	86.5	13.5
0.8%	≤ 2	1.1–1.6	No	> 12.9	4–9	0.8	99.2
0.8%	≤ 2	< 0.6	Yes	10.1–11.4	< 4	85.6	14.4

Table 6. Posterior estimates of survival at twelve months, ten most frequent cases

	First-Degree Associates				Outcome	
Expected Frequency	Number of bone metastases	Diagnosis Group	Hemoglobin (g/dL)	Surgeon's estimate of survival (months)	Probability of Survival >12 months	
					No	Yes
3.1%	Multiple	3	< 10.1	< 4	94.4	5.6
3.1%	Multiple	3	10.1-11.4	< 4	93.3	6.7
3.0%	Multiple	1	< 10.1	< 4	99.2	0.8
2.9%	Multiple	3	>12.9	9-18	16.2	83.8
2.9%	Multiple	1	10.1-11.4	< 4	99.1	0.9
2.5%	Multiple	3	10.1-11.4	4-9	75.0	25.0
2.4%	Multiple	3	< 10.1	4-9	78.4	21.6
2.2%	Multiple	3	11.4-12.9	< 4	80.7	19.3
2.1%	Multiple	3	10.1-11.4	9-18	49.3	50.7
2.0%	Solitary	3	11.4-12.9	9-18	6.4	93.6

Study III: Treating Metastatic Disease: Which Survival Model Is Best Suited for the Clinic?

The ANN models were the most accurate, with AUCs of 0.89 (95% CI, 0.84–0.94) and 0.93 (95% CI, 0.89–0.96), for the three- and 12-month models, respectively. The BBN and logistic regression models performed similarly, with AUCs of 0.85 (95% CI, 0.79–0.91) and 0.83 (95% CI, 0.77–0.89); and 0.84 (95% CI, 0.77–0.90) and 0.83 (95% CI, 0.78–0.89), respectively. The results of accuracy metrics are listed in Table 7.

On decision curve analysis, all models demonstrated a net benefit, indicating each could be used clinically, rather than assume all patients or no patients will survive longer than three or 12 months, respectively. All three-month models performed similarly; however, there were subtle differences among them (Figure 7). Any differences noted by this method are thought to be clinically important.

Table 7. This table depicts the accuracy metric of each of the models generated for Study III

Model	AUC (95%C.I.)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)
3-Month †					
ANN	0.89 (0.84-0.94)	100	5	70	100
BBN	0.85 (0.79-0.91)	97	24	74	78
LOGIT	0.83 (0.77-0.89)	98	17	73	83
12-Month ‡					
ANN	0.93 (0.89-0.96)	60	60	88	92
BBN	0.84 (0.77-0.90)	63	63	83	89
LOGIT	0.83 (0.78-0.89)	57	59	83	89

AUC= Area Under the Receiver Operator Characteristic Curve. Sensitivity and specificity were generated using clinically-relevant cut-points of 0.1† and 0.4‡

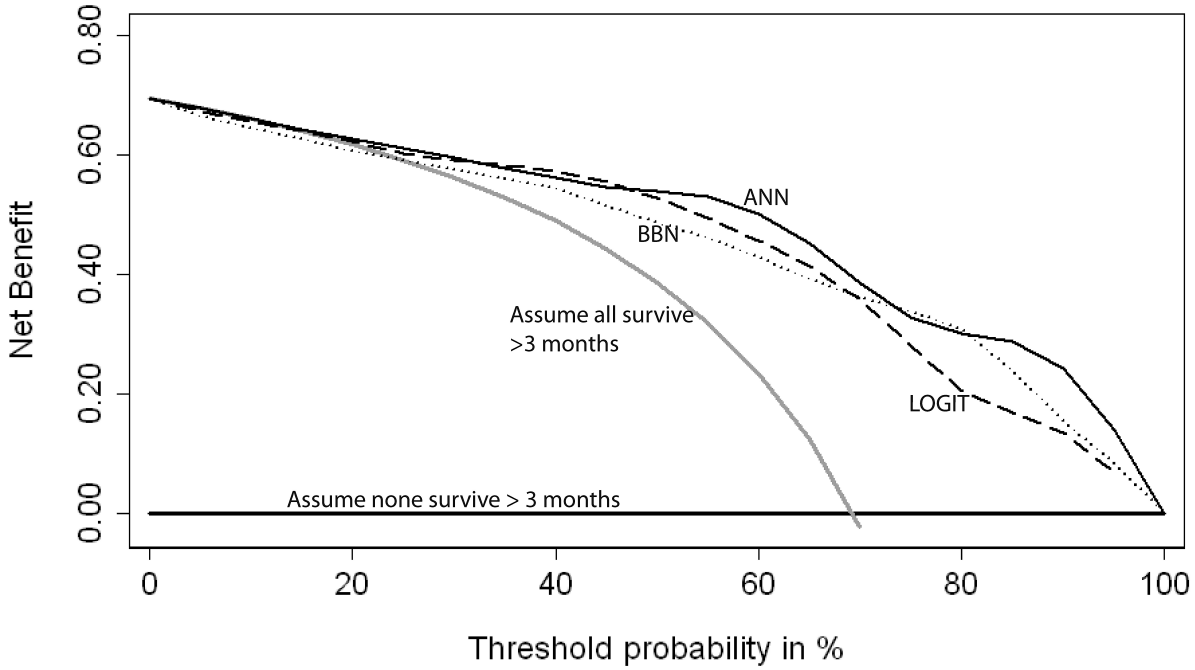


Figure 7. This graph represents Decision Curve Analysis of the three-month models, plotting net benefit versus threshold probability of three-month survival. Net benefit is defined as a three-month survivor who duly receives an operation commensurate with his/her estimated survival. The y-axis and the maximum value of the net benefit is dependent on the proportion of patients that survive at least three months after surgery. All models (ANN, BBN, logistic regression [LOGIT]) performed similarly and could be used clinically rather than assume all patients (or no patients) will survive longer than three months after surgery.

Regarding the 12-month models (Figure 8), the ANN produced the highest net benefit across all threshold probabilities. The BBN and logistic regression models performed similarly. At both three- and 12-month time points, the ANN performed best at or near the threshold probability of 0.5, corresponding to a 50% probability of survival at each time point.

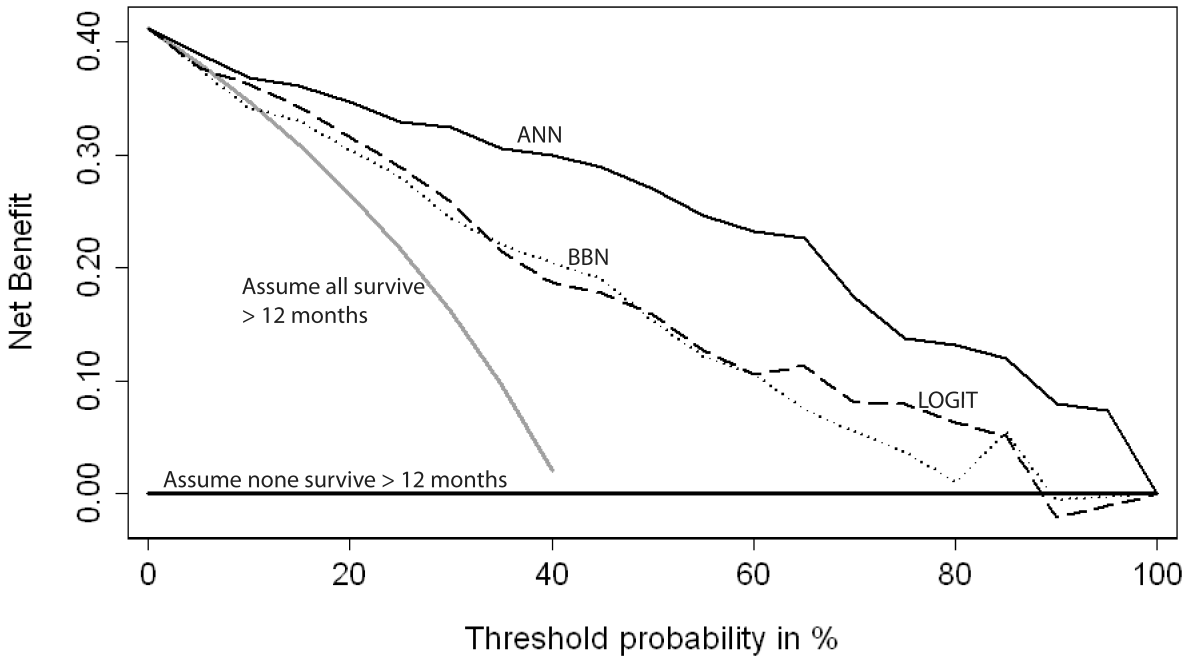


Figure 8. This graph represents Decision Curve Analysis of the 12-month models, plotting net benefit versus threshold probability of 12-month survival. Net benefit is defined as a 12-month survivor who duly receives an implant commensurate with his/her estimated survival. The y-axis and the maximum value of the net benefit is dependent on the proportion of patients that survive at least 12 months after surgery. All models (ANN, BBN, logistic regression [LOGIT]) resulted in positive net benefit, indicating they could be used clinically rather than assume all patients (or no patients) will survive longer than 12 months after surgery. Note the ANN outperformed the other models across all threshold probabilities, including the clinically useful threshold probability of 50% estimated survival at 12 months.

Study IV: External Validation of the Bayesian Estimated Tools for Survival (BETS) Models in Patients with Surgically Treated Skeletal Metastases.

The external validation set contained eight-hundred fifteen (815) records, each with follow-up information to determine survival at three and 12 months postoperatively. The demographic and clinical features differed from those described in the training set (Tables 8 and 9). Specifically, we observed significant differences ($p < 0.05$) in the following features: *Age at surgery*, oncologic *diagnosis group*, presence of *visceral* and *lymph node metastases*, *number of bone metastases*, presence of a complete *pathologic fracture*, *ECOG performance status*, and 12-month mortality. We observed non-significant differences in *gender*, preoperative *hemoglobin* concentration, *absolute lymphocyte count*, and three-month mortality. Most of the features contained missing data, including the *surgeon's estimate of survival* (missing in 100% of records) *absolute lymphocyte count* (84.8% missing), and *lymph node metastases* (61.7% missing). Each of these is a first- or second-degree associate of survival.

Table 8. This table summarizes the comparison of categorical features between the training and validation sets

Feature		Training set (n = 189)		Validation set (n = 815)			p
		No. of patients	%	No. of patients	%	% Missing	
Gender	male female	85 104	45 55	369 446	45.3 54.7	0	0.91
Diagnosis group	1.0 2.0 3.0	52 34 103	27 18 55	173 74 567	21.3 9.2 69.1	0.4	0.001*
Visceral metastases	yes no	114 75	60 40	325 441	39.8 54.1	6.1	<0.0001*
Lymph node metastases	yes no	36 153	19 81	169 143	20.7 17.5	61.8	<0.0001*
Number of bone metastases	solitary multiple	55 134	29 71	123 666	15.1 81.7	3.2	<0.0001*
Pathologic fracture	complete impending	84 105	44 56	614 196	75.3 24.0	0.7	<0.0001*
ECOG performance status	0,1,2 3,4	93 96	49 51	558 257	68.5 31.5	0	<0.0001*
Survival > 3 months	yes no	129 60	68 32	557 258	68.3 31.7	0	0.78
Survival > 12 months	yes no	79 110	42 58	241 574	29.6 70.4	0	0.002*

ECOG=Eastern Cooperative Oncology Group; % Missing=the proportion of unknown or missing data within the validation set.
*Distributions are significantly different between the training and validation sets by the chi-square method.

Table 9. This table summarizes the comparison of continuous features between the training and validation sets

Feature		Training set (n = 189)	Validation set (n = 815)	% Missing	p
Age at surgery (years)	Mean SD Median IQR	62.4 13.7 62.7 54.4, 72.2	66.3 12.8 67.0 58.0, 76.0	0	0.0002*
Hemoglobin (mg/dL)	Mean SD Median IQR	11.5 1.9 11.4 10.1, 12.9	11.5 3.5 11.3 10.3, 12.6	0.6	1.0
Absolute lymphocyte count (K/µL)	Mean SD Median IQR	1.2 1.3 1.0 0.6, 1.5	1.2 0.74 1.2 0.8, 1.6	83.8	0.48
Surgeon's estimate of survival (months)	Mean SD Median IQR	10.3 8.6 6.0 4.0, 12.0	N/A	100	N/A

SD=standard deviation; IQR=interquartile range; N/A=not applicable; % Missing=the proportion of unknown or missing data within the validation set.
*Distributions are significantly different between training and validation sets by two-tailed Student's *t*-test.

For the accuracy assessment, we used a cut point of 0.5, representing a 50% probability of survival. In turn, we correctly classified three-month survival in 633 of 815 (77.7%) patients, and 12-month survival in 555 of 815 (68.1%) records. On ROC curve analysis, the AUCs were 0.79 and 0.76 for the three-month and 12-month models, respectively. When compared with the cross-validation AUCs of 0.86 and 0.83, presented in Study II, we observe a 0.07-point degradation in each model. Though nontrivial, we believe this reduction in accuracy to be acceptable in an external validation setting, particularly in the presence of considerable amounts of missing data.

The majority of incorrect, or misclassifications were optimistic, by overestimating survival in most cases. Specifically, the three- and 12-month models misclassified a total of 182 (22.3%) and 260, (31.9%) records, respectively, as shown in Figures 9 and 10. Of the 182 records misclassified by the three-month model, 57 (31%) were underestimates (patients lived longer than predicted) and 125 (69%) were overestimates (patients did not live as long as predicted). It is important to note that the models are not designed to differentiate between patients who die of disease from those who succumb to other unrelated perioperative complications. However, the majority (70%) of patients in which three-month survival was overestimated lived longer than 1 month after surgery (Table 10). As such, surgery may have still been appropriate for most patients for whom survival was overestimated. Therefore, 125, or 15.3% of the entire validation set, represent the theoretical maximum proportion of patients that may have otherwise been treated non-operatively at the end of life. Of the 260 records incorrectly classified by the 12-month model, 198 (76%) were overestimates and 62 (24%) were underestimates. In this latter case, implants may lack sufficient durability to out last the patient. Thus, 62, or 7.6% represents the theoretical maximum proportion of implants at risk for failure. However, the majority (69%) of patients in whom 12-month survival was underestimated survived less than two years after surgery with even fewer (<10%) living longer than three years after surgery (Table 11).

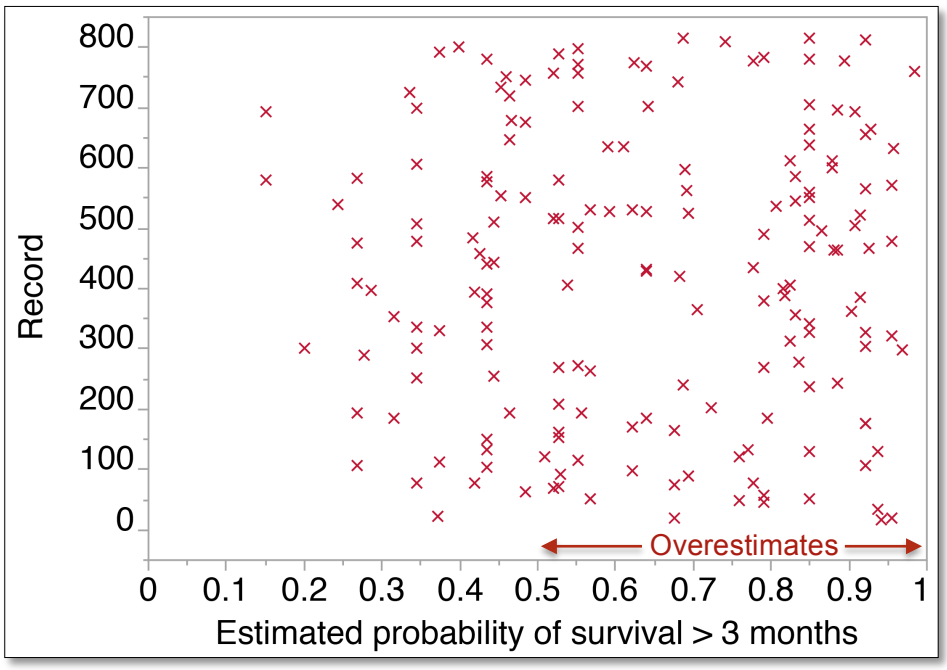


Figure 9. This plot characterizes all *misclassifications* made by the three-month model in study IV. As shown, 69% of misclassifications (15.3% of the total validation set) were optimistic, and three-month survival overestimated. These patients did not live as long as the model predicted, and 15.3% represents the theoretical maximum proportion of patients for whom surgery may have been unnecessary.

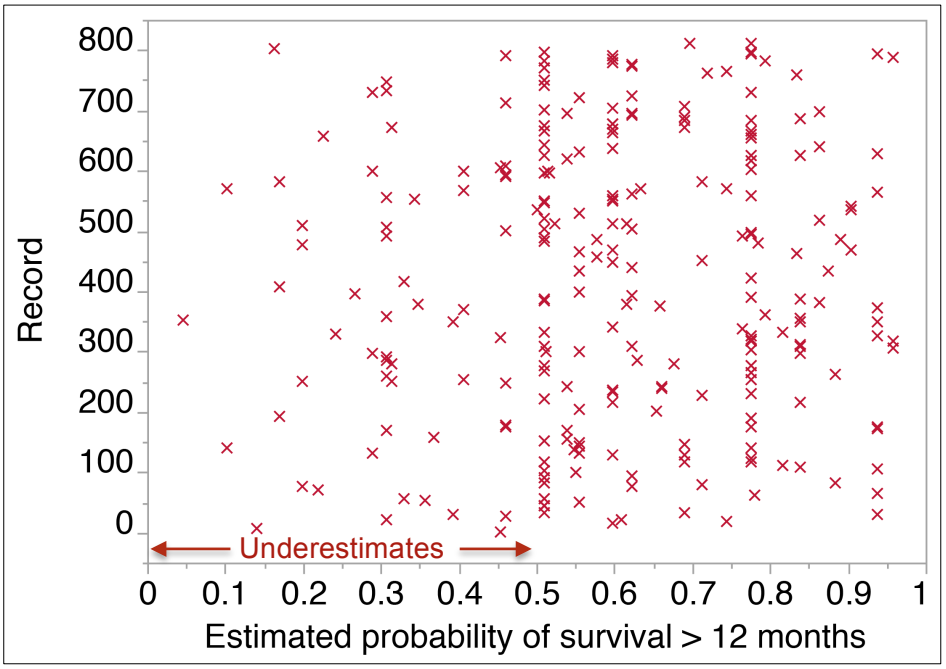


Figure 10. This plot characterizes all *misclassifications* made by the 12-month model in study IV. Though most were optimistic, 24% of misclassifications (7.6% of the total validation set) were underestimates. These patients lived longer than the model predicted, and 7.6% represents the maximum proportion of cases at risk for implant failure if less durable constructs were used.

Table 10. An analysis of *overestimates* by the three-month model indicates the theoretical number of patients who may have been spared surgery at the end of life. However, a minority of patients in whom survival was overestimated lived less than one month following surgery.

Actual Survival	Number of patients	Proportion of overestimates n=125 (%)	Proportion of total validation set n=815 (%)
<1 month	38	30	4.7
1-2 months	44	35	5.4
2-3 months	43	34	5.3

Table 11. An analysis of *underestimates* by the 12-month model indicates the theoretical number of patients at risk for implant failure, if less durable implants were used. As expected, the number of patients (or implants) at risk diminishes considerably over time.

Actual Survival	Number of patients	Proportion of underestimates n=62 (%)	Proportion of total validation set n=815 (%)
1-2 years	43	69	5.3
2-3 years	14	23	1.7
>3 years	5	8	0.6

Study V: External Validation of a Tool for the Estimation of Life Expectancy in Patients with Skeletal Metastases—Decision Analysis and Comparison of Three Major International Patient Populations.

This external validation set contained two hundred eighty seven (287) records with adequate follow-up information to establish survival at three and 12 months postoperatively. None of the records were excluded.

Using a cut point of 0.5, representing a 50% probability of survival, PATHFx correctly classified three-month survival in 253 of 287 (88%) patients, and 12-month survival in 199 of 287 (69%) patients. On ROC curve analysis, the AUCs were 0.80 and 0.77, respectively, for the three- and 12-month models, respectively. Incorrect classifications by PATHFx were more likely optimistic, than pessimistic.

On Kaplan-Meier analysis, the median survival of patients in the Italian validation set was longer at 12 months (95% C.I. 9-14 p=0.005) compared to 8 months (95% C.I. 6-11) and 7 months (95% C.I. 6-8) for the training set and first validation set, respectively.

As expected, the demographic and clinical features of patients in the Italian validation set differed from those observed in the training set (U.S.), and previous external validation set as shown in Tables 12 and 13. Several features differed significantly ($p < 0.05$) including, presence of *visceral* and *lymph node metastases*, *number of bone metastases*, and *three-* and *12-month survival*. Nonsignificant differences were observed in *age at surgery*, *gender*, preoperative *hemoglobin* concentration, *absolute lymphocyte count*, oncologic *diagnosis group*, *pathologic fracture* status, *ECOG performance status*, and the *surgeon's estimate of survival*.

Table 12. This table summarizes the comparison of categorical features between the training (U.S.), first validation (Scandinavian) and second validation (Italian) datasets.

Feature		**Training set n=189**		**Scandinavian set n=815**			**Italian set n=287**				
		No.	%	No.	%	% Missing	No.	%	% Missing	vs. training set p.	vs. Scandinavian set p.
Gender	male female	85 104	45 55	369 446	45.3 54.7	0	120 167	42 58	0	0.50	0.31
Oncologic diagnosis group	1.0 2.0 3.0	52 34 103	27 18 55	173 74 567	21.3 9.2 69.1	0.4	63 44 173	23 16 62	2	0.42	0.007*
Organ metastases	yes no	114 75	60 40	325 441	39.8 53.9	6.3	91 161	36 64	12	0.0001*	0.08
Lymph node metastases	yes no	36 153	19 81	169 143	20.8 17.6	61.6	96 146	40 60	16	0.0001*	0.0007*
Number of bone metastases	solitary multiple	55 134	29 71	123 666	15.2 81.4	3.4	139 144	49 51	1	0.0001*	0.0001*
Pathologic fracture	yes no	84 105	44 56	614 196	75 24.1	0.9	143 131	52 48	5	0.08	0.0001*
ECOG performance status	0,1,2 3,4	93 96	49 51	558 257	68.3 31.7	0	123 106	54 46	20	0.39	0.0001*
Survival > 3 months	yes no	129 60	68 32	557 258	68.2 31.8	0	267 20	93 7	0	0.0001*	0.0001*
Survival > 12 months	yes no	79 110	42 58	241 574	29.8 70.2	0	181 106	63 37	0	0.0001*	0.0001*

Abbreviations: ECOG, Eastern Cooperative Oncology Group; % Missing, the proportion of unknown or missing data within the validation set;

* Proportions are significantly different between training and validation sets by the Chi-square method.

When compared to the previous validation set, most features differed significantly ($p < 0.05$), except *gender*, preoperative *hemoglobin* concentration, *absolute lymphocyte count* and the presence of *visceral metastases*. Similar to the previous external validation set, most features contained missing data, also summarized in Tables 12 and 13. Notable features included the *surgeon's estimate of survival* (missing in 87%), *absolute lymphocyte count* (missing in 23%), and *ECOG performance status* (missing in 20%), all of which are important first- or second-degree associates of survival.

The present study demonstrated that 34 (12%) records were misclassified by the three-month model. Of these, survival was overestimated in 13 records, representing the theoretical maximum number of patients who may have been spared surgery at the end of life. Of the 88 (31%) records misclassified by the 12-month model, survival was underestimated in 44 cases. This too represents the theoretical maximum proportion of patients at risk for implant failure if a less durable implant were used. Decision Curve Analysis indicated that for the Scandinavian set, PATHFx may be used, rather than assume all patients or no patients survive greater than three or twelve months, respectively (Figure 11). For the Italian set, use of the 12-month model is likely to improve outcomes, rather than assume all patients or no patients survive 12 months. However, because of the exceptionally high proportion of patients who survive three months, DCA revealed that it is better for an Italian orthopaedic surgeon to assume all patients will survive more than three months, rather than use the three-month model (Figure 12).

Table 13. This table summarizes the comparison of continuous features between the training (U.S.), first validation (Scandinavian) and second validation (Italian) datasets.

		Training set n=189	**Scandinavian set n=815**		**Italian set n=287**			
Feature				% Missing		% Missing	vs. Training set p.	vs. Scandinavian set p.
Age at surgery (years)	Mean SD Median IQR	62.4 13.7 62.7 54.4, 72.2	66.3 12.8 67 58, 76	0	63.1 11.7 64 56, 72	0	0.54	<0.005*
Hemoglobin (mg/dL)	Mean SD Median IQR	11.5 1.9 11.4 10.1, 12.9	11.5 3.5 11.3 10.3, 12.6	0.6	11.5 1.4 12 11,13	10	0.83	0.90
Absolute lymphocyte count (K/µL)	Mean SD Median IQR	1.2 1.3 1.0 0.6, 1.5	1.2 0.74 1.2 0.8, 1.6	83.8	1.3 0.50 1.5 1.0, 2.0	23	0.59	0.40
Senior surgeon's estimate of survival (months)	Mean SD Median IQR	10.3 8.6 6.0 4.0, 12.0	N/A	100	11.2 7.0 10 5, 20	87	0.56	N/A

SD=standard deviation; IQR=interquartile range; N/A=not applicable.
*Distributions are significantly different between training and validation sets by two-tailed Student's *t*-test.

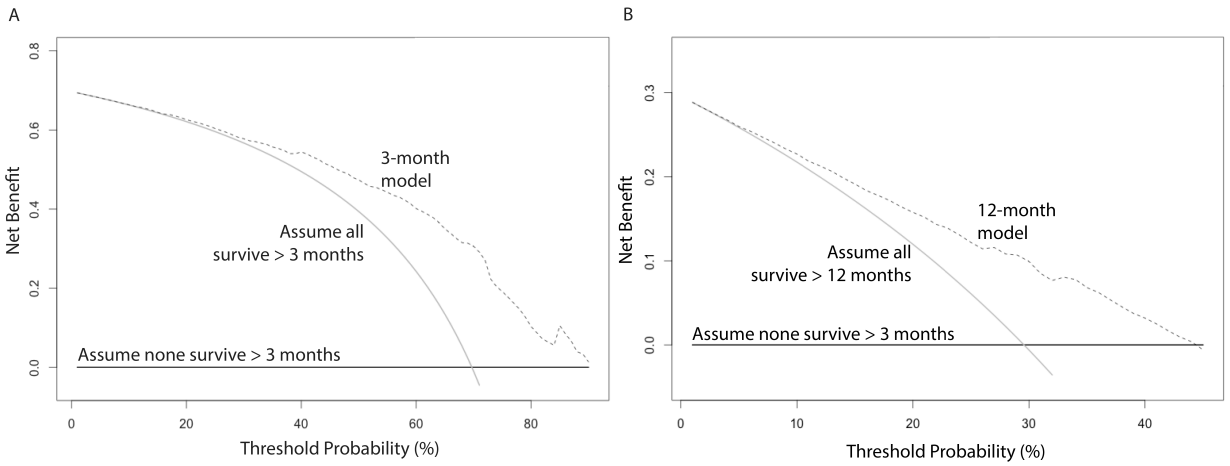


Figure 11. These decision curves depict the net benefit of the three-month (A) and 12-month (B) models, when applied to the Scandinavian external validation set. Net benefit is defined as a three- or 12-month survivor who duly receives an operation and implant commensurate with his/her estimated survival. As shown, each of the models should be used; rather than assume all patients, or none of the patients will survive greater than three or 12 months, respectively.

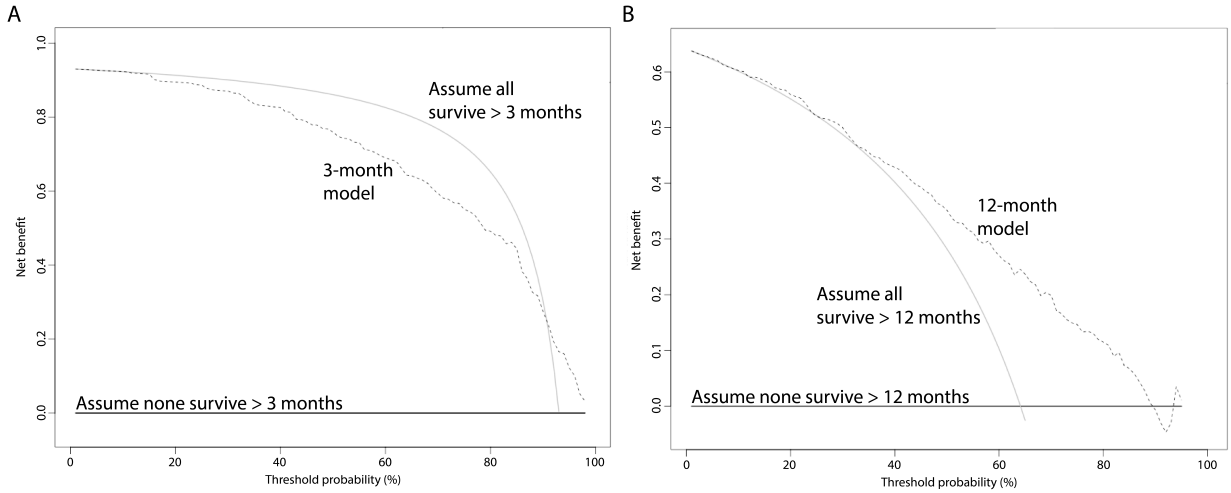


Figure 12. These decision curves depict the net benefit of the three-month (A) and 12-month (B) models, when applied to the Italian external validation set. Net benefit is defined as a three- or 12-month survivor who duly receives an operation and implant commensurate with his/her estimated survival. It is important to note that nearly all (93%) patients referred for orthopaedic intervention survived longer than three months. As a result, DCA of the 3-month model (A) indicates that one could achieve better outcomes by assuming all patients will survive greater than 3 months rather than using the three-month model. This highlights the importance of decision analysis, even for relatively accurate models such as this one, with an AUC of 0.80 on external validation. The results of DCA of the 12-month model (B) indicate it should be used, rather than assume all patients, or none of the patients will survive greater than 12 months.

Study VI: A Probabilistic Analysis of Completely Excised High-Grade Soft Tissue Sarcomas of the Extremity: An Application of a Bayesian Belief Network

We identified the records of 1318 patients meeting the inclusion criteria. No records were excluded. The clinical characteristics and demographics of the patients are shown in Table 14. Briefly, the median age was 54 years (IQR 38, 58). Most patients were male (55.2%), and lower-extremity lesions (73.1%) predominated. Tumor size was divided relatively equally, with those less than 5 cm in 35.5%, 5–10 cm in 32.5%, and greater than 10 cm in 31.6% of cases. The distribution of histologic subtypes included malignant fibrous histiocytoma or high-grade pleomorphic sarcoma (39.6%), synovial sarcoma (15.8%), liposarcoma (12.7%), leiomyosarcoma (10.8%),

malignant peripheral nerve sheath tumor (3.9%), and fibrosarcoma (2.6%). DR occurred in 31.8% of patients, at a median of 11 months (IQR 5.0, 24.0). Overall survival for the entire cohort was 54.4%, and DSS was 73.5%, at a median follow-up of 39.9 months (IQR 14.6, 96.8).

Table 14. Patient characteristics and demographics are depicted from 1318 patients with completely excised, high-grade, extremity soft tissue sarcomas.

Feature	No	%	Median	IQR
Age			54	38, 58
Gender				
Male	728	55.2		
Female	590	44.8		
Size category				
≤ 5 cm	468	35.5		
5-10 cm	428	32.5		
> 10 cm	416	31.6		
unknown	6	0.5		
Depth				
Superficial	265	20.1		
Deep	1053	79.9		
Site				
Upper extremity	354	26.9		
Lower extremity	964	73.1		
Subsite				
Hand	37	2.8		
Forearm	99	7.5		
Elbow	26	2.0		
Arm	91	6.9		
Axilla	33	2.5		
Shoulder	68	5.2		
Groin	43	3.3		
Hip	12	0.9		
Thigh	607	46.0		
Knee	73	5.5		
Leg	149	11.3		
Ankle	22	1.7		
Foot	58	4.4		
Histology, variant				
MFH/HGPS	522	39.6		
Pleomorphic	200	38.3		
Myxofibrosarcomatous	171	32.8		
Giant Cell	10	1.9		
Inflammatory	2	0.4		
NOS	139	26.6		
Synovial sarcoma	208	15.8		
Monophasic	136	65.4		
Biphasic	70	33.7		
NOS	2	0.9		
Liposarcoma	168	12.7		
Myxoid/round cell	83	49.4		
Pleomorphic	57	33.9		
Dedifferentiated	21	12.5		
NOS	7	4.2		
Leiomyosarcoma	142	10.8		
MPNST	51	3.9		
Fibrosarcoma	34	2.6		
Other	193	14.6		
Presentation status				
No prior treatment	260	19.7		
Biopsy only	555	42.1		
Marginal excision	371	28.1		
Wide excision	132	10.0		
Radiation induced				
Yes	19	1.4		
No	1299	98.6		
Zip code upon referral				
First three digits			112	087, 125

Feature	No	%	Median	IQR
Surgeon				
A	318	24.2		
B	245	18.6		
C	146	11.1		
D	123	9.3		
E	116	8.8		
F	103	7.8		
G	102	7.8		
Other	165	12.4		
Surgical service				
GMT	798	60.6		
Orthopaedic surgery	512	38.8		
Other	8	0.6		
Tumor bed excision				
Yes	457	34.7		
No	498	37.8		
Missing	363	27.5		
Procedure				
Amputation	106	8.0		
Limb-sparing surgery	1212	92.0		
Bone invasion				
Yes	51	3.9		
No	1196	90.7		
Missing	71	5.4		
Bone resected				
Yes	183	13.9		
No	1031	78.2		
Missing	104	7.9		
Nerve invasion				
Yes	27	2.1		
No	1137	86.3		
Missing	154	11.6		
Nerve resected				
Yes	156	11.8		
No	1001	78.0		
Missing	161	12.2		
Vascular invasion				
Yes	73	5.5		
No	1081	82.0		
Missing	164	12.4		
Chemotherapy				
Preop	164	12.4		
Postop	202	15.3		
None	952	72.2		
Radiotherapy				
Preop	51	3.9		
Postop	579	43.9		
None	688	52.2		
Local recurrence	194	14.7		
Time to LR			15	6, 29.5
Distant recurrence	419	31.8		
Time to DR			11	5, 24
Death from disease	349	26.5		
Follow-up (months)			39	14.6, 96.8

MFH/HGPS=malignant fibrous histiocytoma/high-grade pleomorphic sarcoma, MPNST=Malignant peripheral nerve sheath tumor LR=Local recurrence, DR=distant recurrence

Bayesian analysis revealed hierarchical associations between several features. As Figure 13 demonstrates, there are three first-degree associates of the outcome of interest, labeled "*DOD*" (Died of Disease). They are, the *size category* of the primary tumor, *time to- and presence of DR*, and *time to- and presence of LR*. The second-degree associates of DSS are, the anatomic *site* and *depth* of the tumor, any oncologic (surgical) treatment prior to referral, and whether the patient received neoadjuvant *chemotherapy*.

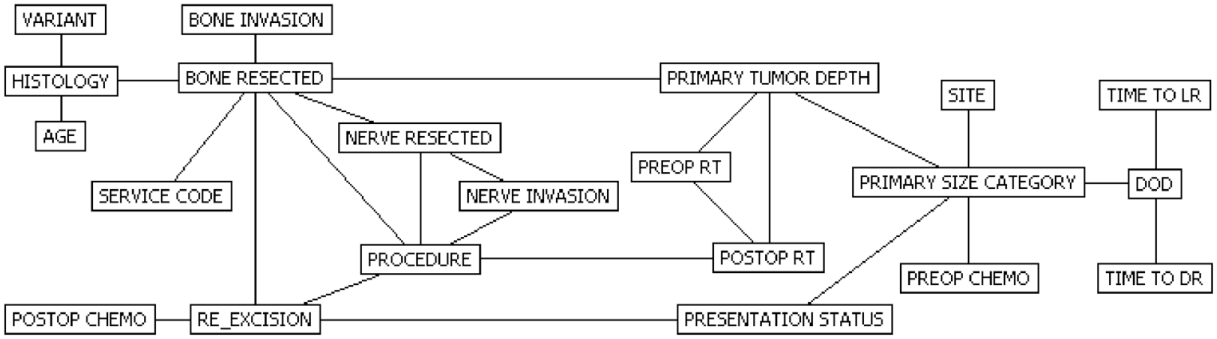


Figure 13. This is a BBN model depicting a comprehensive list of features, and their relationships. There are three first-degree associates with the outcome, *Died of Disease* (*DOD*) including presence of-, and time to distant recurrence (*TIME TO DR*); presence of- and time to local recurrence (*TIME TO LR*), and size of the tumor (*PRIMARY SIZE CATEGORY*)

On cross-validation, considering DSS as the outcome, ROC curve analysis demonstrated an AUC was 0.94 (95% C.I. 0.93–0.96). We generated inference tables based on the three first-degree associates, and the fifteen most common clinical scenarios are shown in Table 15, out of 144 total potential permutations.

Table 15. This table summarizes the probability of death from disease for the 15 most common clinical scenarios, out of 144 total potential permutations.

Probability of case based on training data (%)	Primary size category (cm)	Time to LR (mo)	Time to DR (mo)	Predicted probability of death from disease (%)
24.5	≤5	No LR	No DR	0.1
18.6	5-10	No LR	No DR	0.4
14.6	>10	No LR	No DR	0.7
2.4	>10	No LR	≤4	83.3
2.3	>10	No LR	14-28	83.9
2.3	>10	No LR	9-14	85.3
2.2	>10	No LR	>28	75.0
2.1	>10	No LR	4-9	84.1
1.9	5-10	No LR	≤4	72.4
1.8	5-10	No LR	>28	61.2
1.8	5-10	No LR	14-28	73.3
1.8	5-10	No LR	9-14	75.4
1.6	5-10	No LR	4-9	73.5
1.4	≤5	No LR	>28	35.8
1.3	≤5	No LR	≤4	48.1

Abbreviations: LR, local recurrence; DR, distant recurrence.

For patients in which tumors recurred locally, BBN model demonstrated a difference in survival based on the size of the primary tumor. We then generated case-specific examples of LR for each of the three size categories, which are summarized in Table 16. Importantly, the probability of death from disease was 28.6% for tumors less than 5 cm, but increased to 52.5% and 67.9% for tumors 5–10 cm and greater than 10 cm, respectively.

Table 16. This table depicts the association between the size category of the primary tumor and disease- specific survival, in locally recurrent cases.

LR	Size category of primary tumor	Predicted probability of death from disease (%)	Change in probability above baseline (%)
N/A	N/A	26.8	0
Yes	≤ 5 cm	28.6	+1.9
Yes	5-10 cm	52.5	+25.7
Yes	> 10 cm	67.9	+41.2

We also demonstrated an association between timing of LR and DSS For this, we also generated five case-specific examples in which LR occurred prior to and after 18 months. As shown in Table 17, if LR occurred prior to 18 months after surgery, the probability of DR was 59.6–68.2% and the likelihood of death from disease was 55.8–65.9%. However, if LR occurred 18 months after surgery, the likelihood of DR and death from disease was substantially less at 37.4–39.8% and 29.7–32.4%, respectively.

Table 17. This table depicts the time-dependent association between local recurrence and disease-specific survival.

LR	Time to LR	Predicted probability of death from disease (%)	Change in probability above baseline (%)
N/A	N/A	27.5	0
Yes	≤ 5 months	55.8	29.1
Yes	5-11 months	67.5	40.8
Yes	11-18 months	65.9	39.2
Yes	18-37 months	32.4	5.8
Yes	> 37 months	29.7	3.0

Discussion

We successfully developed a reliable, objective, clinical decision support tool by applying advanced ML techniques. In addition, we demonstrated how ML techniques such as Bayesian Networks may be used to represent the complex relationships germane to the treatment of patients with cancer. In the process of developing and vetting these tools within the oncologic community, we emphasize the importance of rigorous data curation, external validation in a variety of settings, and decision analysis to define whether, and how each prospective model should be used in clinical practice. The process of turning data into decisions, as we have done, can be now be applied to other areas within oncology—and medicine in general—for the purpose of improving outcomes while containing costs. Importantly, is not sufficient to identify whether a particular treatment or technology should be used, but rather, to whom it should be prescribed.[82] Decision support tools designed to provide answers to this question must therefore be developed.

We showed that Bayesian belief network models enable the clinician to assimilate complex high-dimensionality data sets into usable and individualized prognostic information.[83] In the papers that comprise this thesis, we demonstrate this approach is well suited for the purpose of estimating survival in patients with operative skeletal metastases using common demographics and laboratory information (Study II). We then proved them to be widely applicable, by externally validating them in two international patient populations (Studies IV and V). In addition, we explored techniques designed to assess accuracy, as well as clinical utility (Studies III and V). In doing so, we highlight the importance of decision analysis prior to recommending if and how the models should be deployed in a given population. Finally, we showed that ML techniques can be useful to describe important relationships between features in other settings, by focusing on patients with completely excised high-grade soft tissue sarcomas (Study VI).

As we become accustomed to smartphones and continuous Internet connectivity, applications known as “apps” take center stage. The medical community is inundated with “apps” designed to make life easier by providing reference and anatomical information on devices such as computers, tablets and smartphones. Unfortunately, “apps” often convey an air of sophistication, which is undeserved in many cases. “Apps” are not necessarily clinical decision support (CDS) tools. In fact, those that have not undergone the rigorous evaluation described in this thesis should be used with extreme caution and skepticism. Our ability to adequately and accurately apply personalized medicine depends on judicious use of CDS tools, which is the next most logical step toward optimizing outcomes. Doing so helps avoid complications associated with over- or undertreatment, provided that the tools we use have been developed and vetted properly.

We chose to use Bayesian modeling for three reasons. First, the technique generates a graphical depiction of all features, including the outcome(s), in a single model. In this fashion, even the most complex hierarchical relationships can be represented clearly and transparently. Second, because Bayesian networks encode information about the relationships between features in a jPDF, resulting models remain functional in the presence of missing data. This is a major advantage over other modeling techniques, in which correlations between input variables are not encoded. The ability to function in the presence of missing input data is an advantage in the clinical setting, because treatment decisions are often based upon incomplete information. Finally, the BBN method generates not only an estimation of likelihood, but also provides the user with a quantitative assessment of the quality of evidence supporting each estimation. This allows clinicians to weigh the results

provided by BBN models like PATHFx—just as they do for other diagnostic tests—while assimilating them and other sources of information to make personalized, evidenced-based, decisions.

In the process of developing PATHFx, we better characterized the role of the physician's input. The BBN revealed that the *oncologic diagnosis*, *ECOG performance status* and the presence of *visceral metastases* were first-degree associates of the *surgeon's estimate*, which remained the single best predictor of outcome (Study II, Figures 5 and 6). The inclusion of the *surgeon's estimate* in a cohesive model is controversial. However, it seems to carry a great deal of importance, as noted by our results and others[84,85] including Glare et. al,[2] who systematically reviewed physician survival estimates and called for their inclusion in subsequent prognostic models.

Critics of the *surgeon's estimate* cite the potential lack of reproducibility and generalizability of this variable between surgeons and institutions, particularly in those less experienced in the treatment of metastatic disease. To address this concern, we externally validated PATHFx using an 815-patient test-set from an international registry (Study IV). Importantly, this test-set did not contain the *surgeon's estimate* among the list of covariates. I.e. it was missing from 100% of records. PATHFx, nevertheless, performed accurately despite the missing data. This illustrates the flexibility and sophistication of the Bayesian modeling approach by encoding the experience of orthopaedic oncologists (Drs. Healey, Boland, Morris and Athanasian) who provided data for the training set within the jPDF (Study II). This allows end-users who may be less experienced than those listed above to derive accurate survival estimates simply by including the relevant and objective second-degree associates of survival (*oncologic diagnosis, ECOG performance status* and the *presence of visceral metastases*). The Bayesian framework accommodates this approach even though the most important first-degree associate, the *surgeon's estimate*, is not specified.

The nature of misclassifications by each model observed on external validation deserves discussion. For Study IV, the three-month models misclassified 182 (22.3%) of Scandinavian records. Of these, 57 (31%) were underestimates (patients lived longer than predicted) and 125 (69%) were overestimates (patients did not live as long as predicted). As noted above, the clinical impact of each of these cases is not equivalent, which highlights the importance of decision analysis prior to recommending if and how the models are to be used clinically. Still, these metrics do not discern which patients succumb prematurely to unrelated perioperative complications. Therefore, 125, or 15.3% of the entire validation set may seem high, since these patients may have been treated non-operatively at the end of life. However, if we consider an acceptable margin of error, as Nathan and Healey did,[50] we note the majority (70%) of patients in which three-month survival was overestimated lived longer than 1 month after surgery. If we then consider survival between 2 and 3 months, as "acceptable", the proportion of clinically significant overestimates made by the three-month model shrinks to 82 of 815 or 10%. It is important to mention that a fraction of these patients for whom surgery was the best option, succumbed to unrelated perioperative complication(s). If we consider that between 6 and 23% of patients as reported herein (Studies I and III) and elsewhere[8,86,87] die within six weeks of surgery, then the clinical impact of such overestimates may fall decidedly within the acceptable norm. In fact, decision analysis (Study V, Figure 11A) supported this finding by demonstrating that the use of the three-month model would result in better outcomes, rather than assume all Scandinavian patients would survive 3 months.

However, efforts to improve very short-term survival estimates are underway. In fact, there is a growing body of evidence suggesting systemic inflammatory

mediators may provide useful prognostic information in patients with metastatic disease. In a general sense, a number of circulating cytokines and chemokines are associated with either favorable or unfavorable[88-92] outcomes in several oncologic diagnoses. For instance, elevated tumor necrosis factor-alpha (TNF-α) and interleukin-10 have been associated with unfavorable survival in patients with metastatic breast and lung cancer respectively.[93,94] Likewise, procalcitonin, an important inflammatory protein and precursor to TNF-α has been correlated with progression of liver metastases, as has C-reactive protein and interleukin-6 (IL-6).[95] In fact, elevated IL-6, a strong indicator of systemic inflammation has been implicated in the metastatic progression of several diagnoses including carcinoma of the breast, stomach, lung, colon, renal cell, prostate, and melanoma.[88,91-93,95-98] This information suggests that systemic inflammation may be relevant in many forms of advanced cancer. In the same manner that basic demographics and laboratory information may be used to discriminate short from long-term survivors, an inflammatory profile may be useful to discriminate between short (three-month) and very short-term (<1 month) survivors. With this in mind, we may further improve the accuracy of PATHFx in short-term survivors by including one or more inflammatory proteins.

The considerations regarding misclassifications in long-term survivors are complementary. Of the 260 records incorrectly classified by the 12-month model, 198 (76%) were overestimates and 62 (24%) were underestimates. In the latter case, patients live longer than expected, and implants may lack sufficient durability to out last the patient. Thus, 62, or 7.6% represents the theoretical maximum proportion of implants at risk for failure. However, the majority (69%) of patients in whom 12-month survival was underestimated survived less than two years after surgery with very few (<10%) surviving more than three years. If we apply a margin of error and consider those surviving 1-2 years as “acceptable, then 19 of 815 patients, or 2.3% of the entire validation set, may be considered the theoretical maximum proportion of implants that would be at risk of failure, if less durable constructs were used. Further, DCA highlights the clinical utility of the 12-month model in Scandinavian patients over a broad range of threshold probabilities (Study V, Figure 11B)

In order to improve long-term estimates, a new approach to data collection may be used. Current models focus on estimating the likelihood of survival using a single time-point—the preoperative evaluation. However, risk profiles change over time, particularly for those with metastatic disease. Models designed to provide long-term estimates should be designed generate conditional estimates of survival.[99,100] That is, those that can be used at various time-points throughout treatment. An analysis of this kind accommodates the changing risk profile associated with the natural history of the disease; however, more data points over years of follow-up must be obtained. It would also be useful to include not only patients with operable skeletal metastases, but also those in which lesions are to be treated non-operatively.

In order to be widely accepted, however, prognostic models must be applicable to patients in a variety of settings and cultures. To test this, we evaluated PATHFx in a second external validation study (Study V), performed in thirteen Italian referral centers. The results demonstrated that the models are indeed generalizable, not only to the Scandinavian external validation and the U.S. model development populations, but to an Italian one, as well. This is particularly noteworthy given the significant differences in treatment philosophies between the institutions that provided data for each portion of the study.

Though access to care is similar between Scandinavians and Italians, the distribution of patients undergoing surgery for metastatic bone disease was quite different. In fact, over 93% of Italian patients survived more than three months, and 63% survived longer than one year. These proportions are considerably greater than

those represented in either the Scandinavian or training set, and likely indicates key differences in patient selection between cultures. These differences are unexpected given the similar proportions of patients with pathologic fractures, favorable diagnoses, and good performance when compared to the training set. Still these observed differences might be explained by referral patterns among the Italian centers. In general, Italian oncologists refer patients with excellent prognoses for orthopaedic consultation. In patients with more extensive disease or otherwise less favorable prognoses, surgery may be deemed unsuitable in the eyes of the oncologist, which obviates the need for an orthopaedic opinion. However, this practice may exclude patients with relatively short survival estimates that may benefit from less invasive stabilization or palliative procedures.[29,32,34] In addition, nearly half of Italian patients included in this study presented with a solitary skeletal metastasis. This was also unexpected, given that this proportion is much higher than both the training and previous validation sets. Though it could represent more effective disease surveillance practices than those in Scandinavia or the U.S., given the differences in referral patterns discussed above, it is more likely that Italian patients with less favorable prognoses were not referred for surgical management. This is likely to be true, especially in the setting of impending pathologic fractures. Finally, the Italian external validation study was designed to provide a cross section of treatment practices across all of Italy. Had we sampled only tertiary referral centers such as the Rizzoli Institute, or the Campus Bio-Medico University Hospital, the proportion of cases done in the palliative setting would have arguably been higher. Nevertheless, despite major differences in the US and Italian patient populations, PATHFx remained quite accurate with AUC of 0.8 and 0.77 for the three- and 12-month models, respectively. This suggests that as a CDS tool, PATHFx may be broadly applicable to the European, as well as the U.S. model development, populations.

However, measures of accuracy are not alone sufficient to ensure models are applicable and beneficial to specific patient populations. Though one may consider an AUC of 0.8 to be sufficiently accurate, decision analysis helps illustrate the clinical impact of applying the three-month model to an Italian patient population in which virtually every patient survives three months. Following DCA, we observe that at threshold probabilities less than 15%, the model is equivalent to one in which all patients are expected to survive greater than three months. At thresholds > 90%, the three-month model should result in better outcomes. However, at thresholds between 15% and 90%, an Italian orthopaedic surgeon is better off treating patients as if all will survive more than three months, rather than use the three-month model. In the latter case, using 50% threshold probability as an example, an erroneous underestimate may prompt the surgeon to withhold surgery from one in ten patients in whom it was otherwise indicated.

This paradoxical effect is directly related to the proportion of three-month survivors in the Italian validation set. Approximately 68% of patients in the U.S. and Scandinavian populations were alive three months after surgery. In these patients, DCA demonstrated that the three-month model should be used, rather than assume all patients or no patients would survive three months. Importantly, the model was trained and tested on populations of which 68% survived three months, then applied to an Italian one in which 93% survived. With this relatively “healthy” terminally ill population in mind, one may question the need for a three-month model, entirely. Stated another way, a model designed to estimate the likelihood of three-month survival may be of limited utility when nearly every patient survives three months. However, measures of accuracy such as ROC analysis do not characterize this effect, and only after performing DCA are we able to provide specific recommendations as to how the models should be used clinically. By including thirteen Italian centers, we

expect this to be an accurate sampling of the Italian patient population. Still, it is possible that a sampling bias occurred and centers specializing in palliative orthopaedic treatment were underrepresented in this analysis. In addition, referral patterns and treatment philosophies are known to change over time. We expect this to be the case in Italy, given the accepted benefits of surgical intervention for metastatic bone disease, even in the palliative setting. Thus, these analyses should be repeated, particularly if the proportion of three-month survivors approaches that observed in the US or Scandinavian patient populations.

The methodology of turning data into decisions

As patients, physicians and health care payers demand more accuracy and efficiency, we must make use of existing data. Turning these data into viable CDS tools requires a rigorous series of assessments, comparisons and analyses that have been outlined in this thesis. The result of this work should not be viewed as a single tool, but a *methodology* that may be applied to any clinical question for which there is *prior knowledge* in the form of existing, quality data.

Population-based registries have long been the source of big data for orthopaedic surgeons. These databases are being queried in increasing frequency and are used largely to track failures, make comparisons, and evaluate various associations. In this fashion, clinical questions can be answered using traditional frequentist statistics. However, if the goal is to make better decisions based on the available data, one must employ a variety of statistical modeling methods. That is not to say that one should abandon traditional frequentist methods such as regression models. We know that regression-based techniques may be perfectly capable of codifying data into successful CDS tools.[68] Rather, traditional statistical techniques, which remain familiar to most clinicians and scientists, may be developed along side and compared to other, more computer-intensive methods such as BBN, ANN and RF modeling.

The ability to function in the presence of missing data is critical to widespread acceptance of CDS tools. This is especially important because the increasing amount of diagnostic information available to clinicians is at odds with cost containment desired by managed care systems.[101] Furthermore, diagnostic and treatment philosophies vary by country and by region, outcome and functional assessments are not translated into all languages, and there is no consensus as to which outcome assessments should be used for many oncologic conditions. As a result, we must balance comprehensive variable collection with swift usability if the goal is to design a CDS tool for worldwide clinical use. In order to do that, clinical information must first be collected from a variety of international centers with differing diagnostic and treatment philosophies. Once collected, the truly beneficial diagnostic information can then be filtered using most efficient means from the (often extensive and expensive) noise using the data analytic techniques described above.

This process has proven to be useful in the setting of orthopaedic oncology. As with any statistical methods, applicability in other settings may be limited. As such, it is critically important when applying ML approaches to other clinical problems, that we choose relevant features with proven, or theoretical associations with the outcomes of interest. Once this is done, data can be explored in an effort to describe the relationships between features, as we have done in Studies II and VI. Omitting this process risks generating models that do not contain the conditional information needed for the models to be useful in the presence of missing information.[102]

As we have shown, ML techniques such as BBN and ANN modeling can be very useful in codifying large amounts of data. However, it is vitally important that

these algorithms are not left to perform their duties purely unsupervised. To ensure the information remains clinically relevant, clinicians and scientists must spend a considerable amount of time identifying relevant data that can be organized and curated prior to applying ML techniques. This is necessary since the largest volumes of clinical data exist within institutional, regional or national registries.[103] However, most were created decades ago, without the infrastructure to support built-in analytics necessary to perform unsupervised ML.

One exception is the Information Network for Cancer (INCA) that provides a common, platform for several current registries. The INCA framework allows for web-based access from a variety of centers, while also accommodating an investigator-supervised quality control step, critical to the success of any registry. In addition, structured data from other registries may be used to "cross populate" fields, that speeds data collection and standardizes quality. Importantly, INCA interfaces well with R®, and HTML which allows for real time analytics such as traditional frequentist statistics, survival analysis without the need for data export(s).

We designed an international skeletal metastasis registry within INCA, to be used by participating cancer centers around the world. Because the data curation and quality control are performed up front, we ensure that all data contained within the registry are suitable for analysis. This, in turn, enables one to embed analytic tools to process information in real time.

The process begins with any clinical question that can be posed in the context of probabilities. For instance, one may seek to estimate the probability of venous thromboembolism, wound infection, aseptic loosening, or even a major complication after a hip arthroplasty. Variables such as clinical and laboratory information are collected and deposited into a central, secure, web-enabled registry. A quality control step is performed first to ensure all data contained within the registry is suitable for analysis. Modern analytical techniques using R® Statistical Software[1] and Shiny[104] by RStudio provide investigators with a "dashboard" containing automated, customizable, metrics. In addition to reporting enrollment figures, one may assess distributions and comparisons of each feature—including, but not limited to univariate, multivariate and Kaplan-Meier survival analysis—between institutions, in real time. Next, a variety of modeling techniques can be employed in parallel to classify individual outcomes of interest, and followed by direct comparisons of model accuracy and net benefit by decision analysis. In order to ensure the relationships identified by the modeling methods are real and not due to chance alone, a target shuffling technique is used. This allows one to calculate the probability that one or more of the models resulted from chance alone. At this point, users are presented with the most appropriate model(s) based on the results of target shuffling, as well as decision analysis. The clinician completes the process by considering the threshold probability associated with an individual clinical scenario.

This registry will eventually be available to cancer centers worldwide. By "crowd sourcing" data collection and automating the analytics, we ensure each model remains broadly applicable, clinically relevant, and can "evolve" over time. This framework also allows for the inclusion of newer or otherwise additional features, theorized to be related to a particular outcome of interest. The effect of each can then be evaluated in terms of model accuracy and clinical utility using ROC and DCA, respectively.

Conclusions

1. Treatment failures in patients undergoing surgical treatment for skeletal metastases are relatively common. This reiterates the importance of considering each patient's estimated survival not only during the index procedure, but also in subsequent revision procedures, where the risk of medical complications is higher.
2. We successfully developed BBN and ANN models capable of estimating the likelihood of survival at two time-points useful for orthopaedic surgical decision-making (three months and 12 months post surgery).
3. Receiver Operating Characteristic analysis demonstrated the ANN was more accurate than the BBN and LR models. Similarly, DCA suggested the ANN resulted in higher net benefit across the broadest range of threshold probabilities. However, because the ANN functions only in the presence of complete input data—something that is not always present in the clinical setting, or in external validation sets—the BBN models may actually be better suited for clinical use.
4. We successfully externally validated the BBN models in two separate patient populations using data collected from Scandinavia and Italy. In addition, decision analysis indicated that PATHFx could improve outcomes in Scandinavia, and the 12-month model could improve outcomes in Italy. However, even the most accurate three-month model may not result in better outcomes in the Italian patient population.
5. Finally, using registry data from Memorial Sloan-Kettering Cancer Center, we demonstrated how the Bayesian Belief Network could be used to codify complex information related to the treatment of patients with localized soft tissue sarcomas. In doing so, we produced a clear, graphical model of relevant features and described the relationships between them. Importantly, the study demonstrated the applicability of Bayesian methodology to an entirely different oncologic scenario.

This thesis describes the process of turning clinical and registry data into decisions. We successfully developed a clinically useful decision support tool, externally validated it in two separate patient populations, and made it available to surgeons, worldwide on www.pathfx.org. In doing so, we highlight the importance of data curation, model development, external validation and decision analysis, prior to widespread clinical use. We also demonstrated the utility of the Bayesian statistics in describing the complex relationships between variables inherent to the treatment of soft tissue sarcoma patients.

After completing the work described above, we created the infrastructure for a worldwide registry to ensure these models remain both current and applicable to a variety of cultures and centers. As such, the knowledge and experience gained from this thesis may have a direct benefit on future orthopaedic oncology patients by generating the data necessary to both improve these existing clinical decision support tools and develop new ones. Though we limited the scope of this work to orthopaedic oncology, we expect the *process* of turning data into decisions to be applicable to other topics in orthopaedics surgery and medicine, in general.

Suggested Guidelines

A word of caution: "Apps" are not necessarily clinical decision support tools.
We are in the middle of a health care revolution. Big data including demographics, molecular markers, and physiologic indicators are being codified by advanced techniques in information technology. The result is an explosion in the number of prognostic models that have been applied to a variety of clinical problems.[105] Given that mobile or otherwise interconnected applications are ubiquitous in modern society, physicians may be tempted to confuse finger-tip availability and relative ease of use with a tool that has been properly vetted for clinical use. It would seem that tech-savvy doctors are abandoning their healthy skepticism that has been ingrained by years of journal clubs, academic medicine and/or clinical practice. However, this should not be the case. When using an "app," physicians should demand the same level of scrutiny and apply the same healthy skepticism as they do for the literature they read, the implants they select and the medications they prescribe.

Specific Recommendations

1. Each prospective CDS tool must undergo the following, prior to being recommended for clinical use: In addition to **measures of accuracy**, prospective models must also undergo **decision analysis** to ensure net benefit is conferred, and the model is suitable for clinical use in the intended patient population. Each model should then undergo **external validation** in a variety of centers with differing cultures, patient populations, and treatment philosophies. In settings where these considerations remain stable over time, validation studies may be done retrospectively. However, if the goal is to evaluate new features or improve the model(s), over time, prospective studies will be necessary.
2. Clinical decision support tools should be designed to accommodate uncertainty, whenever possible. Clinicians are often faced with making decisions based on incomplete information. As such, models that require all input features to be present in order to function properly may be of limited value. This is the fundamental reason we chose a method based on Bayes' theorem of conditional probabilities, however, other methods and imputation algorithms could function in this setting. As such, there are exceptions to this recommendation. Some decision support tools are designed to be used with the output of multiplex assays or microarrays,[66,67] which usually result in complete datasets.
3. Mature CDS tools should be incorporated into the electronic medical record. As we maximize efficiency, the *objective* is to improve outcomes, while containing costs. Redundant processes that require time and energy are under scrutiny, in part because they are unpopular with the clinical staff and are prone to error. An example of this is re-entering clinical data into a stand-alone CDS tool or web page. Stand-alone tools detract from efficiency and must be eliminated if they do not support the *objective* of improving outcomes and containing costs, or *automated* if they do. As such, incorporating mature CDS tools into the electronic medical record is absolutely essential for the healthcare enterprise interested in increasing efficiency by basing decisions on objective, quality, data.

4. Clinical decision support tools are no substitute for good clinical judgment and experience. As the name suggests, decision support models are designed to provide objective data on which an independent practitioner may base a decision. Similar to laboratory and imaging tests, clinicians may vary the degree of emphasis he or she places on each result based on a pre-test probability,[106] derived largely on his/her clinical suspicion and level of experience.
5. Caution should be used when entering the Surgeon's Estimate into PATHFx. Although it contains many important subjective features that cannot be quantified, this assessment demands a certain level of experience. If surgeons are unsure about whether an estimate is appropriate, he/she should select "unknown." Doing so will maintain accuracy of the model, while not introducing undue bias, as demonstrated in both external validation studies used to support this thesis. (Studies IV and V)

Future Directions

The worldwide registry developed as part of this thesis allows us to evaluate other features theorized to be associated with short and long term survival. By adhering to the recommendations listed above, our goal is to develop timely, useful, clinical decision support tools for use in a variety of settings.

The first project, entitled "Improving a Bayesian Model's Survival Estimates in Patients Needing Surgery for Bone Metastases" will evaluate whether SF-36 data and/or serum inflammatory cytokines and chemokines are capable of estimating survival in patients with skeletal metastases. The systemic inflammatory response may be helpful in identifying patients with very short life expectancies. Models containing SF-36, cytokine and/or chemokine information will be compared to existing models using ROC and DCA to determine which approach is better suited for clinical use. This study is currently underway at Memorial Sloan-Kettering Cancer Center and the Murtha Cancer Center in Maryland (www.ClinicalTrials.gov identifier: NCT01470105). In addition, we plan to extend this project to include all patients with skeletal metastases, not simply those undergoing orthopaedic surgery. However, funding to support this work is pending.

We believe it is important to better characterize the surgeon's estimate. Clearly, a certain amount of experience is necessary to produce accurate decisions; however, the requisite level of experience has not been elucidated. Current work focuses on assessing the relative contributions to model accuracy for differing levels of training in order to determine the effect of clinical experience on this important, but subjective determination.

As mentioned above, this methodology may be applied to any clinical question for which there is prior knowledge in the form of existing, quality data. Other studies are underway to determine whether it is feasible to estimate the likelihood of other untoward outcomes in orthopaedic surgery, such as prosthetic joint infection, reoperations after hip arthroplasty, and mortality after hip fracture surgery.

Acknowledgements

These collaborations, and this thesis, would not have been possible without the unwavering support from my wife **Dawn**, our two children, **Linnea** and **Sofia** och min svärmor **Gullan**.

In addition, I thank **Dr. Rikard Wedin**, my friend and main tutor, who provided guidance, and facilitated access to the most well curated metastatic disease registry data in the world. This work would not have been possible without his logistical support and insight into Swedish healthcare, culture, and politics. Rikard's wife **Lena** and children **Linda**, **Jenny** and **Björn** took me in, and gave me a much-needed home away from home during my time in Sweden.

Dr. John Healey, my mentor, motivated by example and gave me the time, direction and latitude to pursue these efforts.

Professor Henrik Bauer was the voice of reason at times, and taught me the finer points of orthopaedic surgery—and bread making.

Professor Lars Weidenhielm, who first suggested that I should have a Ph.D.—to which I laughed out loud—and provided guidance along the way. I look forward to many rewarding future collaborations, and maybe a regatta or two.

Dr. Benjamin Kyle Potter, my friend, partner, and sounding board. You always had my back. Thank you for helping me understand the value of this thesis. It would not have been possible without your support.

Dr. Karl-Åke Jansson, a big supporter of this work, and for future collaborations, introduced me to my future home in Solsidan (or Saltsjöbaden).

Dr. Mathias Mosfeldt, the luckiest man in the world. I'm sure we'll meet up on an island somewhere, or in the Alps…

Dr. Rüdiger Weiss, who loves football and could solve the world's problems over a pint—but only with written permission from the boss.

Dr. Andrea Piccioli led the Società Italiana di Ortopedia e Traumatologia and **Dr. Silvia Spinelli,** who guided me through the streets of Rome, fed me espresso, and organized a successful external validation study in thirteen Italian centers.

Chris Armao and **Liz Silvius**, the analytics team that helped develop my understanding of Bayesian statistics, **Benjamin Peterson, Tosh Keune** and **Hanna Hensman** who put PATHFx on the web, in English and Swedish; and **Todd Radano**, who introduced me to the field and helped me realize the potential of clinical decision support.

Dr. Matthew Wagner, my sounding board for all things, bioinformatics. Someday my code will look as good as his…

Ann-Britt Wikström, the famous television star, kept track of my credits (the system is really quite complicated), helped me along the way, and taught me that in Sweden, mandatory courses are *really* mandatory.

Dr. Stephen Ahlers, who always has the big picture in mind. Thank you for supporting me along the way.

The staff of the **Orthopaedics Department at the Karolinska University Hospital,** who accepted me as one of their own—they know how to throw a good party!

The success of this thesis depended on the staff and surgeons that collected and curated the excellent data contained within the Scandinavian Skeletal Metastasis Registry.

We obtained thorough plagiarism and grammatical analysis using iThenticate® and Grammarly, prior to publication.

Finally, I thank the Maynard Limb Preservation Fund, and the U.S. National Cancer Institute Soft Tissue Sarcoma Program Project (grant P01 CA047179).

References

1. R Core Team. R: A language and environment for statistical computing [Internet]. Available from: http://www.R-project.org/

2. Glare P, Virik K, Jones M, et al. A systematic review of physicians' survival predictions in terminally ill cancer patients. BMJ 2003;327(7408):195–8.

3. Hartsell WF, Desilvio M, Bruner DW, et al. Can physicians accurately predict survival time in patients with metastatic cancer? Analysis of RTOG 97-14. Journal of Palliative Medicine 2008;11(5):723–8.

4. Lamont EB, Christakis NA. Prognostic disclosure to patients with cancer near the end of life. Ann Intern Med 2001;134(12):1096–105.

5. Madan TN. Dying with dignity. Social Science & Medicine 1992;35(4):425–32.

6. Chochinov HM, Hack T, Hassard T, Kristjanson LJ, McClement S, Harlos M. Dignity in the terminally ill: a cross-sectional, cohort study. The Lancet 2002;360(9350):2026–30.

7. Weeks JC, Cook EF, O'Day SJ, et al. Relationship between cancer patients' predictions of prognosis and their treatment preferences. JAMA 1998;279(21):1709–14.

8. Clohisy DR, Le CT, Cheng EY, Dykes DC, Thompson RC. Evaluation of the feasibility of and results of measuring health-status changes in patients undergoing surgical treatment for skeletal metastases. J Orthop Res 2000;18(1):1–9.

9. Harrington KD, Sim FH, Enis JE, Johnston JO, Dick HM, Gristina AG. Methylmethacrylate as an adjunct in internal fixation of pathological fractures. J Bone Joint Surg Am 1976;58(8):1047–55.

10. Wedin R, Bauer HCF. Surgical treatment of skeletal metastatic lesions of the proximal femur: endoprosthesis or reconstruction nail? J Bone Joint Surg Br 2005;87(12):1653–7.

11. Yong M, Jensen AÖ, Jacobsen JB, Nørgaard M, Fryzek JP, Sørensen HT. Survival in breast cancer patients with bone metastases and skeletal-related events: a population-based cohort study in Denmark (1999-2007). Breast Cancer Res Treat 2011;129(2):495–503.

12. Nørgaard M, Jensen AÖ, Jacobsen JB, Cetin K, Fryzek JP, Sørensen HT. Skeletal related events, bone metastasis and survival of prostate cancer: a population based cohort study in Denmark (1999 to 2007). The Journal of Urology 2010;184(1):162–7.

13. Saad F, Lipton A, Cook R, Chen Y-M, Smith M, Coleman R. Pathologic fractures correlate with reduced survival in patients with malignant bone disease. Cancer 2007;110(8):1860–7.

14. Weinfurt KP, Li Y, Castel LD, et al. The significance of skeletal-related events for the health-related quality of life of patients with metastatic prostate cancer. Ann Oncol 2005;16(4):579–84.

15. Coleman RE. Clinical features of metastatic bone disease and risk of skeletal morbidity. Clin Cancer Res 2006;12(20 Suppl):6243s–6249s.

16. Coleman RE. Skeletal complications of malignancy. Cancer 1997;80(8 Suppl):1588–94.

17. Manoso M, Healey JH. Metastatic Cancer To the Bone. In: DeVita VT, Rosenberg S, Hellman S, editors. Cancer. Philadelphia, PA: Lippincott Williams & Wilkins; 2008. p. 2368–71.

18. Sathiakumar N, Delzell E, Morrisey MA, et al. Mortality following bone metastasis and skeletal-related events among women with breast cancer: a population-based analysis of U.S. Medicare beneficiaries, 1999–2006. Breast Cancer Res Treat 2011;131(1):231–8.

19. Coleman RE. Bisphosphonates: clinical experience. Oncologist 2004;9 Suppl 4:14–27.

20. Lipton A, Theriault RL, Hortobagyi GN, et al. Pamidronate prevents skeletal complications and is effective palliative treatment in women with breast carcinoma and osteolytic bone metastases. Cancer 2000;88(5):1082–90.

21. Rosen LS, Gordon D, Kaminski M, et al. Long-term efficacy and safety of zoledronic acid compared with pamidronate disodium in the treatment of skeletal complications in patients with advanced multiple myeloma or breast carcinoma. Cancer 2003;98(8):1735–44.

22. Sathiakumar N, Delzell E, Morrisey MA, et al. Mortality following bone metastasis and skeletal-related events among men with prostate cancer: a population-based analysis of US Medicare beneficiaries, 1999–2006. Prostate Cancer Prostatic Dis 2011;14(2):177–83.

23. Wedin R, Bauer HC, Wersäll P. Failures after operation for skeletal metastatic lesions of long bones. Clin Orthop Relat Res 1999;(358):128–39.

24. Pockett RD, Castellano D, Mcewan P, Oglesby A, Barber BL, Chung K. The hospital burden of disease associated with bone metastases and skeletal-related events in patients with breast cancer, lung cancer, or prostate cancer in Spain. Eur J Cancer Care (Engl) 2010;19(6):755–60.

25. Schulman KL, Kohles J. Economic burden of metastatic bone disease in the U.S. Cancer 2007;109(11):2334–42.

26. Arneson TJ, Shuling Li, Yi Peng, et al. Estimated number of prevalent cases of metastatic bone disease in the US adult population. CLEP 2012;(4):87–93.

27. Sugiura H, Yamada K, Sugiura T, Hida T, Mitsudomi T. Predictors of survival in patients with bone metastasis of lung cancer. Clin Orthop Relat Res 2008;466(3):729–36.

28. Lynch AC, Sullivan CR, Dahlin DC. Pathologic fractures of femur treated at Mayo Clinic (1950-1959). Arch Surg 1965;90:127–32.

29. Perez CA, Bradfield JS, Morgan HC. Management of pathologic fractures. Cancer 1972;29(3):684–93.

30. Behr J, Dobozi W, Badrinath K. The treatment of pathologic and impending pathologic fractures of the proximal femur in the elderly. Clin Orthop Relat Res 1985;198:173–8.

31. Lote K, Wallöe A, Bjersand A. Bone metastasis. Prognosis, diagnosis and treatment. Acta Radiol Oncol 1986;25(4-6):227–32.

32. Nielsen OS, Munro AJ, Tannock IF. Bone metastases: pathophysiology and management policy. J Clin Oncol 1991;9(3):509–24.

33. Borel Rinkes I, Wiggers T, Bouma W, van Geel A, Boxma H. Treatment of manifest and impending pathologic fractures of the femoral neck by cemented hemiarthroplasty. Clin Orthop Relat Res 1990;260:220–3.

34. Yazawa Y, Frassica FJ, Chao EY, Pritchard DJ, Sim FH, Shives TC. Metastatic bone disease. A study of the surgical treatment of 166 pathologic humeral and femoral fractures. Clin Orthop Relat Res 1990;(251):213–9.

35. van der Hulst RRWJ, van den Wildenberg FAJM, Vroemen JPAM, Greve J-WM. Intramedullary nailing of (impending) pathologic fractures. Journal of Trauma and Acute Care Surgery 1994;36(2):211.

36. Algan S, Horowitz S. Surgical treatment of pathologic hip lesions in patients with metastatic disease. Clin Orthop Relat Res 1996;332:223–31.

37. Kunisada T, Choong PF. Major reconstruction for periacetabular metastasis: early complications and outcome following surgical treatment in 40 hips. Acta Orthop Scand 2000;71(6):585–90.

38. Schürmann M, Gradl G, Andress HJ, Kauschke T, Hertlein H, Lob G. Metastatic lesions of the humerus treated with the isoelastic diaphysis prosthesis. Clin Orthop Relat Res 2000;380:204–14.

39. Wedin R, Bauer HC, Rutqvist LE. Surgical treatment for skeletal breast cancer metastases: a population-based study of 641 patients. Cancer 2001;92(2):257–62.

40. Steensma M, Boland PJ, Morris CD, Athanasian E, Healey JH. Endoprosthetic treatment is more durable for pathologic proximal femur fractures. Clin Orthop Relat Res 2012;470(3):920–6.

41. Ward WG, Holsenbeck S, Dorey FJ, Spang J, Howe D. Metastatic disease of the femur: surgical treatment. Clin Orthop Relat Res 2003;(415 Suppl):S230–44.

42. Nilsson J, Gustafson P. Surgery for metastatic lesions of the femur: Good outcome after 245 operations in 216 patients. Injury 2008;39(4):404–10.

43. Sarahrudi K, Greitbauer M, Platzer P, Hausmann J-T, Heinz T, Vécsei V. Surgical treatment of metastatic fractures of the femur: A retrospective analysis of 142 patients. The Journal of Trauma: Injury, Infection, and Critical Care 2009;66(4):1158–63.

44. Talbot M, Turcotte RE, Isler M, Normandin D, Iannuzzi D, Downer P. Function and health status in surgically treated bone metastases. Clin Orthop Relat Res 2005;438:215–20.

45. Schneiderbauer MM, Knoch von M, Schleck CD, Harmsen WS, Sim FH, Scully SP. Patient survival after hip arthroplasty for metastatic disease of the hip. J Bone Joint Surg Am 2004;86(8):1684–9.

46. Parrish FF, Murray JA. Surgical treatment for secondary neoplastic fractures. J Bone Joint Surg Am 1970;52(4):665–86.

47. Wedin R. Surgical treatment for pathologic fracture. Acta Orthop Scand Suppl 2001;72(302):1–29.

48. Lin PP, Mirza AN, Lewis VO, et al. Patient survival after surgery for osseous metastases from renal cell carcinoma. J Bone Joint Surg Am 2007;89(8):1794–801.

49. Narazaki D, Alverga Neto C, Baptista A, Camargo M. Prognostic factors in pathologic fractures secondary to metastatic tumors. Clinics 2006;61:313–20.

50. Nathan S, Healey J, Mellano D, et al. Survival in patients operated on for pathologic fracture: implications for end-of-life orthopedic care. Journal of Clinical Oncology 2005;23(25):6072–82.

51. Hansen BH, Keller J, Laitinen M, et al. The Scandinavian Sarcoma Group Skeletal Metastasis Register. Survival after surgery for bone metastases in the pelvis and extremities. Acta Orthop Scand Suppl 2004;75(311):11–5.

52. Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. American Journal of Clinical Oncology 1982;5(6):649.

53. Bauer H, Wedin R. Survival after surgery for spinal and extremity metastases: prognostication in 241 patients. Acta Orthop 1995;66(2):143–6.

54. Tokuhashi Y, Matsuzaki H, Toriyama S, Kawano H, Ohsaka S. Scoring system for the preoperative evaluation of metastatic spine tumor prognosis. Spine 1990;15(11):1110–3.

55. Karnofsky DA. The Clinical Evaluation of Chemotherapeutic Agents in Cancer. Columbia University Press; 1949.

56. Tokuhashi Y, Matsuzaki H, Oda H, Oshima M, Ryu J. A revised scoring system for preoperative evaluation of metastatic spine tumor prognosis. Spine 2005;30(19):2186–91.

57. Yamashita T, Siemionow KB, Mroz TE, Podichetty V, Lieberman IH. A prospective analysis of prognostic factors in patients with spinal metastases:

use of the revised Tokuhashi score. Spine 2011;36(11):910–7.

58. Nathan SS, Chan L, Tan WL, et al. The need for a system of prognostication in skeletal metastasis to decide best end-of-life care - a call to arms. Ann Acad Med Singap 2010;39(6):476–81.

59. Adamina M, Tomlinson G, Guller U. Bayesian statistics in oncology: a guide for the clinical investigator. Cancer 2009;115(23):5371–81.

60. Cheng SH, Horng C-F, Clarke JL, et al. Prognostic index score and clinical prediction model of local regional recurrence after mastectomy in breast cancer patients. Int J Radiation Oncology Biol Phys 2006;64(5):1401–9.

61. Damato B, Eleuteri A, Fisher AC, Coupland SE, Taktak AFG. Artificial Neural Networks estimating survival probability after treatment of choroidal melanoma. Ophthalmology 2008;115(9):1598–607.

62. Heriot AG, Tekkis PP, Smith JJ, et al. Prediction of postoperative mortality in elderly patients with colorectal cancer. Diseases of the Colon & Rectum 2006;49(6):816–24.

63. Jerez-Aragonés JM, Gómez-Ruiz JA, Ramos-Jiménez G, Muñoz-Pérez J, Alba-Conejo E. A combined neural network and decision trees model for prognosis of breast cancer relapse. Artificial intelligence in medicine 2003;27(1):45–63.

64. Lisboa P, Taktak A. The use of artificial neural networks in decision support in cancer: a systematic review. Neural Networks 2006;19(4):408–15.

65. Taktak AFG, Fisher AC, Damato BE. Modelling survival after treatment of intraocular melanoma using artificial neural networks and Bayes theorem. Phys Med Biol 2003;49(1):87–98.

66. Khan J, Wei J, Ringnér M, et al. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine 2001;7(6):673–9.

67. Chen Q, Vansant G, Oades K, et al. Diagnosis of the small round blue cell tumors using multiplex polymerase chain reaction. Journal of Molecular Diagnostics 2007;9(1):80.

68. Kattan MW, Heller G, Brennan MF. A competing-risks nomogram for sarcoma-specific death following local recurrence. Stat Med 2003;(22):3515–25.

69. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. Circulation 2007;115(7):928–35.

70. Vickers AJ, Elkin EB. Decision Curve Analysis: A novel method for evaluating prediction models. Med Decis Making 2006;26(6):565–74.

71. Hunink M, Glasziou P, Siegel J, et al. Decision Making in Health and Medicine with CD-ROM. Cambridge University Press; 2001.

72. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications.

Ann Surg 2004;240(2):205–13.

73. Katagiri H, Takahashi M, Wakai K, Sugiura H, Kataoka T, Nakanishi K. Prognostic factors and a scoring system for patients with skeletal metastasis. J Bone Joint Surg Br 2005;87(5):698–703.

74. Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN. A Bayesian derived network of breast pathology co-occurrence. Journal of Biomedical Informatics 2008;41(2):242–50.

75. Peterson LT, Ford EW, Eberhardt J, Huerta TR, Menachemi N. Assessing differences between physicians' realized and anticipated gains from electronic health record adoption. Journal of medical systems 2011;35(2):151–61.

76. Online artificial neural network analysis tool [Internet]. Available from: http://home.ccr.cancer.gov/ oncology/oncogenomics/tool.htm

77. Buccheri G, Ferrigno D, Tamburini M. Karnofsky and ECOG performance status scoring in lung cancer: a prospective, longitudinal study of 536 patients from a single institution. Eur J Cancer 1996;32A(7):1135–41.

78. Pearl J. Introduction to Probabilities, Graphs and Causal Models. In: Causality. Cambridge England: Cambridge University Press; 2009. p. 1–38.

79. Pearl J. Bayesian Inference. In: Morgan G, editor. Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: 1988. p. 29–73.

80. Nisbet R, Elder J IV, Miner G. Model Evaluation and Enhancement. In: Handbook of Statistical Analysis Data Mining Applications. Academic Press; 2009. p. 285–311.

81. Peirce CS. The numerical measure of the success of predictions. Science 1884;4(93):453–4.

82. Chernew ME, Hirth RA, Sonnad SS, Ermann R, Fendrick AM. Managed care, medical technology, and health care cost growth: A review of the evidence. Med Care Res Rev 1998;55(3):259–88.

83. van den Hout A, Matthews FE. Estimating dementia-free life expectancy for Parkinson's patients using Bayesian inference and microsimulation. Biostatistics 2009;10(4):729–43.

84. Mackillop WJ, Quirt CF. Measuring the accuracy of prognostic judgments in oncology. Journal of Clinical Epidemiology 1997;50(1):21–9.

85. Muers MF, Shevlin P, Brown J. Prognosis in lung cancer: physicians' opinions compared with outcome and a predictive model. Thorax 1996;51(9):894–902.

86. Quinn R, Drenga J. Perioperative morbidity and mortality after reconstruction for metastatic tumors of the proximal femur and acetabulum. The Journal of arthroplasty 2006;21(2):227–32.

87. Bollen L, van der Linden YM, Pondaag W, et al. Prognostic factors associated with survival in patients with symptomatic spinal bone metastases: a

retrospective cohort study of 1,043 patients. Neuro-oncology 2014;16(7):991–8.

88. Nicolini A, Carpi A, Rossi G. Cytokines in breast cancer. Cytokine Growth Factor Rev 2006;17(5):325–37.

89. Guida M, Casamassima A, Monticelli G, Quaranta M, Colucci G. Basal cytokines profile in metastatic renal cell carcinoma patients treated with subcutaneous IL-2-based therapy compared with that of healthy donors. J Transl Med 2007;5:51.

90. Marjon PL, Bobrovnikova-Marjon EV, Abcouwer SF. Expression of the pro-angiogenic factors vascular endothelial growth factor and interleukin-8/CXCL8 by human breast carcinomas is responsive to nutrient deprivation and endoplasmic reticulum stress. Mol Cancer 2004;3:4.

91. Moretti S, Chiarugi A, Semplici F, et al. Serum imbalance of cytokines in melanoma patients. Melanoma Res 2001;11(4):395–9.

92. Negrier S, Perol D, Menetrier-Caux C, et al. Interleukin-6, Interleukin-10, and Vascular Endothelial Growth Factor in Metastatic Renal Cell Carcinoma: Prognostic Value of Interleukin-6--From the Groupe Francais d'Immunotherapie. Journal of Clinical Oncology 2004;22(12):2371.

93. Bozcuk H, Uslu G, Samur M, et al. Tumour necrosis factor-alpha, interleukin-6, and fasting serum insulin correlate with clinical outcome in metastatic breast cancer patients treated with chemotherapy. Cytokine 2004;27(2-3):58–65.

94. De Vita F, Orditura M, Galizia G, et al. Serum interleukin-10 levels as a prognostic factor in advanced non-small cell lung cancer patients. Chest 2000;117(2):365–73.

95. Matzaraki V, Alexandraki KI, Venetsanou K, et al. Evaluation of serum procalcitonin and interleukin-6 levels as markers of liver metastasis. Clin Biochem 2007;40(5-6):336–42.

96. Lin M-T, Lin B-R, Chang C-C, et al. IL-6 induces AGS gastric cancer cell invasion via activation of the c-Src/RhoA/ROCK signaling pathway. International Journal of Cancer 2007;120(12):2600–8.

97. Shariat SF, Andrews B, Kattan MW, Kim J, Wheeler TM, Slawin KM. Plasma levels of interleukin-6 and its soluble receptor are associated with prostate cancer progression and metastasis. Urology 2001;58(6):1008–15.

98. Soria J-C, Moon C, Kemp BL, et al. Lack of interleukin-10 expression could predict poor outcome in patients with stage I non-small cell lung cancer. Clin Cancer Res 2003;9(5):1785–91.

99. Parsons HM, Habermann EB, Tuttle TM, Al-Refaie WB. Conditional survival of extremity soft-tissue sarcoma: results beyond the staging system. Cancer 2011;117(5):1055–60.

100. Davis FG, McCarthy BJ, Freels S, Kupelian V, Bondy ML. The conditional probability of survival of patients with primary malignant brain tumors:

surveillance, epidemiology, and end results (SEER) data. Cancer 1999;85(2):485–91.

101. Mas N, Seinfeld J. Is managed care restraining the adoption of technology by hospitals? J Health Econ 2008;27(4):1026–45.

102. Pearl J. Causal inference in statistics: An overview. Statistics Surveys 2009;3:96–146.

103. Dreyer NA, Garner S. Registries for robust evidence. JAMA 2009;302(7):790–1.

104. Beeley C. Web application development with R using Shiny. Birmingham, England: Packt Publishing Ltd; 2013.

105. Freedman AN, Seminara D, Gail MH, et al. Cancer risk prediction models: a workshop on development, evaluation, and application. J Natl Cancer Inst 2005;97(10):715–23.

106. Sox HC. Diagnostic Decision: Probability theory in the use of diagnostic tests. Ann Intern Med 1986;104(1):60–6.

Paper Reprints