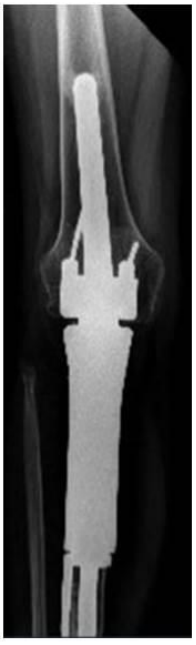
UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

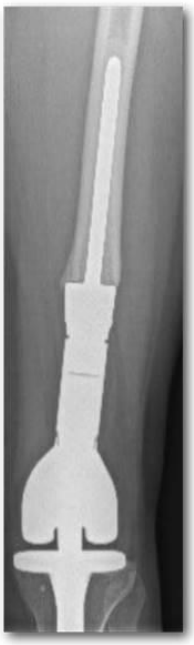
PhD Thesis

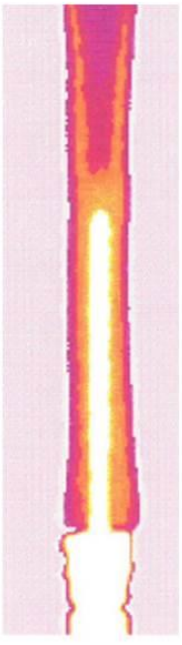
Christina Holm

Surgical treatment of bone sarcomas and aggressive benign bone tumors.

Implant survival, limb survival and adaptive periprosthetic bone remodeling after limb sparing surgery and reconstruction with tumor prostheses and development and comparison of one-year survival prediction models in patients with bone sarcomas







Supervisor: Michael Mørk Petersen, MD, DMSc, Professor

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen: February 14, 2020

Institution: Rigshospitalet, University of Copenhagen, Denmark

Department: Department of Orthopaedic surgery
Musculoskeletal Tumor Section,
Rigshospitalet, University of Copenhagen, Denmark

Author Christina Holm, MD

Title and subtitle: Surgical treatment of bone sarcomas and aggressive benign bone tumors. Implant survival, limb survival and adaptive periprosthetic bone remodeling after limb sparing surgery and reconstruction with tumor prostheses and development and comparison of one-year survival prediction models in patients with bone sarcomas.

Principal supervisor Michael Mørk Petersen, MD, DMSc, Professor,
Department of Orthopedic surgery, Rigshospitalet,
University of Copenhagen, Denmark

Principal Co-supervisor Anders Odgaard, MD, DMSc, Professor
Department of Orthopedic surgery, Gentofte Hospital
University of Copenhagen, Denmark

Co-Supervisor Michala Skovlund Sørensen, MD, PhD

Assessment committee

Chair
Jes Bruun Laurtizen, MD, DMSc, Professor
Department of Orthopedic surgery, Bispebjerg Hospital
University of Copenhagen, Denmark

Danish assessor:
Peter Holmberg, MD, DMSc,
Department of Orthopedic surgery, Aarhus University Hospital, Denmark

International assessor:
Minna Laitinen,
Associate professor, PhD
Helsinki University Hospital
Bone Tumour Unit, Department of Orthopedics and Traumatology
Finland

Table of contents

AKNOWLEDGEMENTS

First and foremost, I would like to express my genuine gratitude to my principal supervisor, Michael Mørk Petersen for introducing me to the field of research in orthopedic oncology, and for making this PhD thesis possible. Your patient academic guidance, always constructive criticism and your great ability to provide me with freedom to act have taught me how to conduct research in the best manner I could imagine, thank you.
I would like to express my appreciation to my co-supervisors, Anders Odgaard and Michala Skovlund Sørensen for their contribution to the thesis. A special thanks to Michala for your thoroughness and integrity to the project. It has been educational to work with you and this project would not have been the same without your support.

To all the patients who participated in this PhD project. I would like to express my deepest gratitude and respect to all of you who even at challenging times of your life, took your time to participate with the intention to provide better treatment for future patients.

A special thanks to everyone at the Musculoskeletal Tumor section at Rigshospitalet: Lise, Iben, Gülsen, Mette, Dorrit, Claus, Allan, Thomas, Gunnar, Sammer, Kolja, Werner. You have all created a very welcoming workplace, it has been a privilege to call this section my workplace. A special thanks to Lise for your patient assistance in synchronizing patient visits and scans to the outpatient clinic. To Chief surgeon Claus Munk Jensen for providing excellent work facilities for research.

To the research nurses, Marina Golemac and Sofie Kofoed Larsen for your great personalities and for your hard work and thoroughness with countless DXA scans and much more. Thank you for patience with me and not at least for creating a very welcoming environment to all patients. A special thanks to Marina, for your great support and effort with the prospective part of this PhD, making it possible for me to have maternity leave during the PhD-study period, thank you.

To Jonathan Forsberg, for a unique opportunity to be a part of your research team at Walter Reed, even for a short time. Thank you for allowing me to conduct research with you and your team. Your dedication to orthopedic oncology is inspiring. To Clare for many hours of support, patience and at the same time good company, I look forward to follow your work. To Jose for making me feel welcome from day one. To Yesse for providing us with the best tips on what to do and where to go in DC. And to everyone else at the office for making me feel welcome.

To my fellow PhD students at Rigshospitalet, with a special thanks to Karen Dyreborg for introducing me from the very beginning and for being good company all the way. A special thanks for taking-over my project when I was traveling to Colombia, thank you.

To Frederik Dessau for your help with the last but certainly not least step of this thesis, the layout, thank you.

Finally, I would like to thank all those I hold the dearest. My parents, for your love and support. My parents-in-law for your invaluable help and care. A Myriam Enciso por cariño y por estar en mi vida. A mis queridos hermanos por mucho cariño. To those I call my second family, my friends. Thank you for continuously encouragement and friendship.
To my husband Michael Holm, for your invaluable love, encouragement and not at least patience, especially through the last months of pace. To my three biggest achievements: Thora, Agnes and little Edwin. Thank you for lots of patience and not at least for enjoying Washington DC as much as me, I dedicate this work to you.

PREFACE

The studies comprising the present thesis have all been conducted at The Musculoskeletal Tumor Section, Department of Orthopedic Surgery, Rigshospitalet in Copenhagen from November 2016 to November 2019 where I was employed and matriculated as a Ph.D. student at the Faculty of Health Science, University of Copenhagen. External validation and development of the prediction model was conducted in collaboration with Orthopedics, USU-Walter Reed Department of Surgery, Bethesda, MD, USA where I was fortunate to have my stay abroad in June and July 2019.

LIST OF PAPERS

Study I:
Holm CE, Bardram C, Riecke AF, Horstmann PF, Petersen MM.
Implant and limb survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses. 50 patients followed for a mean of 14 years.
Int Orthop. 2018 May;42(5):1175-1181.

Study II:
Holm CE, Sørensen MS, Yilmaz M, Petersen MM.
Improvement in failure rate after resection of primary bone tumors and reconstruction with second-generation mega-prostheses?
Manuscript submitted for publication.

Study III:
Holm CE, Horstmann PF, Sørensen MS, Dyreborg K, Petersen MM.
Quantitative measurements of adaptive bone remodeling around the cemented Zimmer® Segmental stem after tumor resection arthroplasty using dual-energy x-ray absorptiometry.
Manuscript submitted for publication.

Study IV:
Holm CE, Grazal CF, Raedkjaer M, Baad-Hansen T, Forsberg JA, Nandra R, Grimer R, Petersen MM, Sørensen MS.
Development and comparison of one-year survival models in patients with primary bone sarcomas.
External validation of a Bayesian Belief Network model and creation and external validation of a new Gradient Boosting Machine model.
Manuscript submitted for publication

SELECTED ABBREVIATIONS

LSS	Limb sparing surgery
BMD	Bone mineral density
MRI	Magnetic Resonance Imaging
PET/CT	Positron Emission Tomography/Computed Tomography
GBM	Gradient Boosting Machine
BBN	Bayesian Belief Network
DXA	Dual energy x-ray absorptiometry
ROI	Region of interest
ROC	Receiver operating characteristic
DCA	Decision curve analysis
TM	Trabecular metal
MBD	Metastatic bone disease
CV	Coefficient of variation
MSTS	The Musculoskeletal Tumour Society Score
DSR	Danish sarcoma registry

ENGLISH SUMMARY

Background: Bone sarcomas are a rare type of malignancies representing less than 0.2 % of all malignancies and approximately 5% of all childhood cancers. The majority of bone sarcomas arise in the lower extremities and pelvis. From the time in the 1970ies where amputation was the only available treatment, great advances in the treatment of bone sarcomas have been achieved and limb sparing surgery (LSS) with bone tumor resection and reconstruction with tumor prostheses is currently offered to more than 90% of patients with bone sarcoma in the long bones. LSS provide immediate fixation with early weight bearing and preservation of function. The incidence of revisions and complications has nonetheless since the introduction of LSS been reported higher compared to primary arthroplasties, along with prolonged rehabilitation. The decision of treatment course often relies on estimated life expectancy.

Purpose: The overall purpose of the present thesis was to investigate the incidence of revisions and complications after LSS and reconstruction with tumor prostheses of the lower extremities over time and examine the local adaptive bone remodeling around the stem of a tumor prosthesis. Furthermore, was an aim of the thesis development and validation of a prediction model for 1-year survival in effort to provide evidence as a basis for further optimization of the surgical treatment of bone sarcomas.

Methods: The thesis comprised four cohorts; two retrospective cohorts of patients having LSS with resection and reconstruction with tumor prostheses in the lower extremities at a tertiary referral center for orthopedic oncology from 1985-2016. One prospective single Center population having LSS with insertion of the Zimmer® Segmental tumor prostheses with trabecular metal (TM) collar and cemented stem, and lastly a retrospective national cohort comprising patients aged > 15 and newly diagnosed with bone sarcoma.

Main results: In study I (1985-2005) we found a 10-year implant survival of 24% (95%CI: 9%-41%) and a 10-year limb survival of 83% (95%CI: 65%-96%). In study II (2006-2016) we found a 10-year cumulative incidence of major revision of 18% (95%CI: 9%-28%) and a 10-year cumulative incidence of amputation of 11% (95%: 3%-18%). In study III, we found a decrease in periprosthetic bone mineral density (BMD) (8-15%) compared to baseline in all four regions of interest after 1-year of follow up, and a decrease in BMD of the affected ankles after 1-year follow up. In study IV external validation of the previously published Bayesian Belief Network (BBN) model yielded poor discriminatory ability with receiver operating characteristic (ROC) analysis and area under the curve (AUC) of 68% (95%CI: 62%-73%). Internal validation of the

developed gradient boosting machine (GBM) prediction model yielded discriminatory ability of 75% (95%CI: 70%-80%). Overall model performance assessed by Brier score was 0.09 (95%CI: 0.077-0.11). Decision curve analysis demonstrated positive net-benefit for probability thresholds above 0.5. External validation of the developed GBM model yielded poor discriminatory ability with an AUC of the ROC curve of 63% (95%CI: 57%-68%). The Brier score was 0.14 (CI95%: 0.12-0.16).

Conclusion: Our long-term results with tumor prostheses support the use of LSS and prosthetic reconstruction with regards to implant and limb survival and also functional outcome. However, long-term survivors must expect repeated revisions. Despite lower incidence of major revisions after the introduction of rotating-hinge prostheses, we did not find statistically significantly improved implant- or limb-survival over time, possibly explained by type II errors, as indicated by wide confidence intervals. The evaluation of adaptive bone remodeling around the cemented Zimmer®Segmental stem and the ankle of the affected limb after 1-year follow up, indicate a slow progressive decrease in BMD, most likely caused by a combination of immobilization and stress shielding. We successfully developed a 1-year survival prediction model using Machine-Learning. By external validation and comparison of the two present prediction models, none of the models yielded satisfactory discriminatory ability. The authors are obligated to continue the ongoing work with modernization and revision of a model suitable for clinical use.

DANISH SUMMARY

Baggrund: Knoglesarkomer er en sjælden kræfttype, der udgør mindre end 0.2% af alle kræftformer og ca. 5% af cancertilfælde hos børn. Størstedelen af knoglesarkomer opstår i underekstremiteterne og bækkenet. Fra 1970'erne hvor amputation var eneste mulige behandling af knoglesarkomer, er der gjort betydelige fremskridt og ekstremitetsbevarende kirurgi med tumorresektion og rekonstruktion med tumorproteser tilbydes nu til mere end 90% af patienter med knoglesarkom i de lange knogler. Ekstremitetsbevarende kirurgi med tumorproteser tillader tidlig vægtbæring og bevarelse af funktion. Udover langvarig genoptræning, er forekomsten af revisioner og komplikationer efter ekstremitetsbevarende kirurgi dog rapporteret højere sammenlignet med almindelig primær alloplastik. Beslutningen om hvilken behandling patienter tilbydes, er ofte afhængig af estimeret rest levetid.

Formål: Det overordnede formål med nærværende afhandling var at identificere og evaluere forekomsten af komplikationer og amputationer efter ekstremitetsbevarende kirurgi med tumorproteser hos patienter med knoglesarkom i underekstremiteterne samt at måle knogleremodelleringen af den protesenære knogle efter indsættelse af tumorprotese. Endvidere var formålet at udvikle og validere en model der kan forudsige patientoverlevelsen 1 år efter man får stillet diagnosen knoglesarkom med det formål, at behandlingen i højere grad kan individualiseres.

Metoder: Afhandlingen omfatter fire kohorter. To retrospektive kohorter med patienter, der fik foretaget ekstremitetsbevarende kirurgi med resektion og rekonstruktion med tumorproteser i perioden 1985-2016. Endvidere en prospektiv kohorte af patienter der fik foretaget ekstremitetsbevarende kirurgi med indsættelse af Zimmer® Segmental tumorproteser med cementeret stem og trabekulær metal (TM) krave. Endeligt, en retrospektiv national kohorte af patienter over 15 år med ny-diagnosticeret knoglesarkom.

Hovedresultater: I studie I (1985-2005) fandt vi en sandsynlighed for overlevelse af protesen efter 10 år på 24% (95% CI: 9%-41%) og en sandsynlighed for overlevelse af ekstremiteten efter 10 år på 83% (95% CI: 65%-96%). I studie II (2006-2016) fandt vi en risiko for fjernelse af protesen på 18% (95% CI: 9%-28%) efter 10 år og en risiko for amputation efter 10 år på 11% (95% CI: 3%-18%). I studie III fandt vi efter 1 år et signifikant fald i knogletætheden på 8-15% i alle fire udvalgte regioner omkring protesen, men vi fandt også fald i knogletæthed i anklen på de opererede ben. I studie IV viste ekstern validering af en tidligere fremført BBN-model dårlig evne til at diskriminere ved evaluering med ROC/AUC: 68% (95% CI: 62% -73%). Intern

validering af den udviklede GBM-model viste god evne til at diskriminere: 75% (95% CI: 70% - 80%) og den præsterede godt ved evaluering med Brier-score: 0,09 (95% CI: 0,077-0.11). Ekstern validering af den udviklede GBM-model viste dog dårlig diskriminations evne ved evaluering med AUC/ROC: 63% (95% Cl: 57% -68%). Brier score var 0.14 (CI95%: 0.12-0.16).

Konklusion: Vores langtidsresultater med tumorproteser understøtter brugen af ekstremitetsbevarende kirurgi i relation til protese- og ekstremitetsoverlevelse samt funktionsniveau. Langtidsoverlevere må dog forvente gentagne revisioner. På trods af en lavere forekomst af større re-operationer efter indførelsen af proteser med rotation i hængselledet, fandt vi ikke en signifikant forbedret protese- eller ekstremitetsoverlevelse over tid, hvilket muligvis kan forklares med type II-fejl, de vide konfidensintervaller taget i betragtning. Evaluering efter 1 år af den adaptive knogleremodellering omkring det cementerede intramedullære Zimmer®Segmental stem og anklen på den afficerede ekstremitet, indikerer et langsomt progressivt fald i knogletæthed, sandsynligvis forårsaget af en kombination af immobilisering og stress shielding. Endelig udviklede vi en model for 1-års overlevelse ved hjælp af Machine-Learning. Ved ekstern validering og sammenligning af de to nærværende modeller, viste ingen af modellerne tilfredsstillende evne til at diskriminere. Forfatterne vil fortsætte det igangværende arbejde med at modernisere og revidere en overlevelses-model, der vil være egnet til klinisk brug.

BACKGROUND

SARCOMAS

Epidemiology and etiology

Sarcomas are a heterogeneous group of relatively rare malignant tumors in the musculoskeletal system comprising 1% of all adult cancers and 10-15% of childhood cancers (1–4). The reported incidence is approximately six to eight per 100,000 inhabitants corresponding to 300 cases per year in Denmark (150 soft tissue sarcomas, 100 retroperitoneal/abdominal sarcomas, 50 bone sarcomas) (5,6). The incidence of sarcomas in general is increasing with age although a broad variety among histological subtypes exist (7). The male/female ratio is mostly reported as 1.5:1 (8,9). The word sarcoma is derived from the Greek meaning fleshy (sarcos) tumor (oma). Sarcomas arise in the body's connective tissues including bone, muscle, cartilage, fat tissue, blood vessels and peripheral nerve-sheaths and hence arise in all parts of the body although most dominant in the extremities (78%) with 64% located in the lower extremities and 10% localized in the upper extremity (10). Sarcomas are derived from the mesodermal germ layer that develops into the mesenchymal structures in the fetus and hence called mesenchymal tumors (2). Sarcomas are classified as soft tissue sarcomas (STS) or bone sarcomas based on the anatomic site of the tumor and are further often named based on their histological origin. The retroperitoneal and gastrointestinal sarcomas are often described separate although they are a subdivision of soft tissue sarcomas (11). In general about 80% of sarcomas originate from soft tissue and 20% from bone (4). Visceral tumors, soft tissue tumors and bone sarcomas each differ in methods of diagnosis, staging and treatment approaches (6). According to the World Health Organization´s (WHO) most recent update in 2013, there exist more than 50 histologic subtypes and 100 distinct diagnostic entities (8,11) yielding a broad heterogeneous morphology and biological behavior of all sarcoma subtypes.

Sarcomas are primarily compartment based with high grade subtypes mostly metastasizing hematogenously to the lungs, which is also the primary cause of sarcoma-specific death (12). Sarcomas more seldom metastasize to the skeleton and rarely to lymph nodes (1,13–16). Approximately 10-20% of patients presents with metastases at diagnosis (17,18). Although the vast majority of sarcomas are suggested to arise de novo (8) certain exposures associated with specific sarcomas has been described (e.g. previous radiation therapy and certain chemicals) as well as rare genetic syndromes predisposing to sarcoma development has been defined (e.g. Pagets disease, Li-Fraumeni syndrome and Neurofibromatosis) (2,3,8). However, the etiology of sarcomas is yet to be found.

Bone sarcomas

Bone sarcomas, which is the scope of this thesis represents less than 0.2% of all malignancies and approximately 5% of childhood cancers (14,17,19,20). Contrary to STS, the age-specific frequency of bone sarcomas is bimodal with first peak in the second decade of life and a second peak after the age of 60 years(8,20). According to the most recent update from WHO, the general anatomic distribution of bone sarcomas is: around the knee (43%), hip and pelvis (31%), shoulder girdle (11%), lower leg (7%), upper limb (5%) and trunk (3%), although variations among subtypes exist (8).

The three main types of bone sarcomas are: Osteosarcoma (35%), chondrosarcoma (25%) and Ewing sarcoma (16%) (8,17). Osteosarcoma and Ewing sarcoma are dominant in childhood, with peak incidence in the second decade of life, whilst chondrosarcoma is the most frequent subtype in adulthood (8,16,21). Osteosarcomas are characterized by the production of osteoid or immature bone. They mainly arise in the metaphyses of the long bones and the high-grade intramedullary osteosarcoma constitutes the most classic and frequent type of osteosarcoma (80%) (17). Ewing sarcomas primarily arise in the diaphysis of long bones or in the pelvis. All types of Ewing sarcoma are high-grade tumors (14). The growth spurt and rapid growth of bones during puberty is often reported to correlate with the peak incidence within the second decade of life and has been suggested to contribute to the pathogenesis of Ewing sarcoma and osteosarcoma (21–23). Chondrosarcomas are characterized by the production of tumor cartilage and mainly arise in the metaphyseal region of the long bones. The vast majority of primary chondrosarcomas are low-grade tumors (14).

DIAGNOSTIC PROCESS

In Denmark, patients are diagnosed according to national guidelines by a multidisciplinary team of orthopedic surgeons, oncologists, radiologists, clinical physiologists, pediatricians, pathologists and if required by other specialist teams (24–26). The aim of the diagnostic workup is to stage the local and distant extent of the tumor to determine the most appropriate course of treatment for each patient.

Clinical presentation

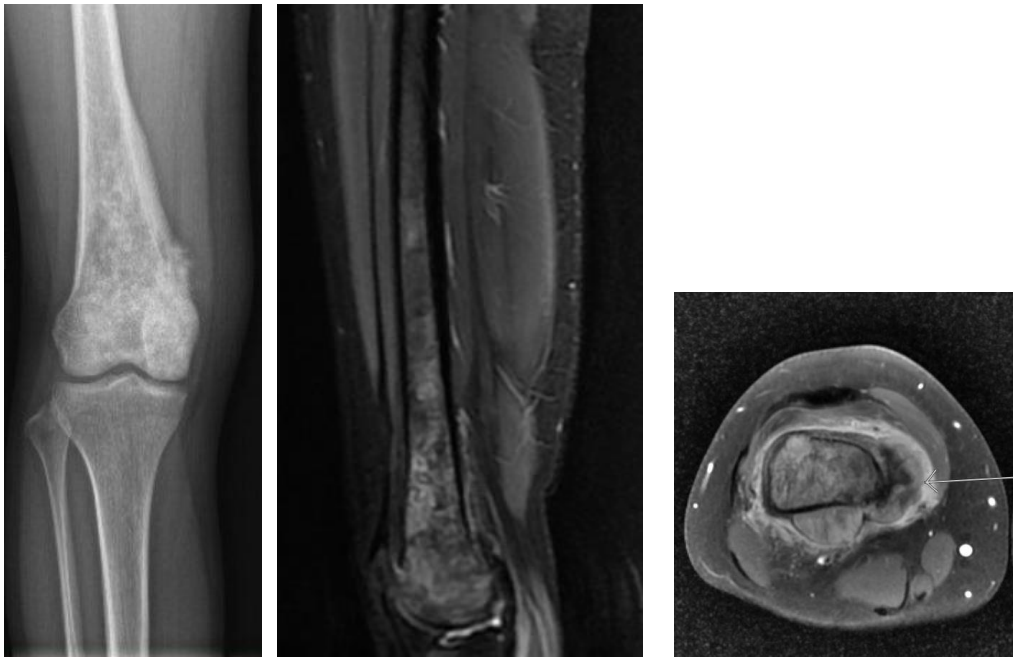
The initial symptom of a bone sarcoma is mainly an unexplained non-mechanical ache in the affected area, often developing to become a persistent pain especially at night and in some cases followed by swelling. Ewing sarcoma may be associated with "B-symptoms": fever, weight loss and malaise, which rarely occur in other bone sarcoma subtypes. For approximately 10% of patients a pathological fracture in the affected area is the first presentation of disease (19,27–29). Due to the non-specific subtle disease manifestations in often, young healthy individuals, symptoms are often misinterpreted as caused by traumas or other benign conditions. In addition, as a result of the rareness and low public and professional awareness, sarcomas are often diagnosed at a late stage (28,30).

Diagnostic imaging

Plain radiographs (**Fig. 1-3**) of the affected area often suggests the diagnosis of bone malignancy in the primary care setting and is a mandatory baseline examination, followed by Magnetic Resonance Imaging (MRI). MRI of the whole bone/compartment (**Fig. 1-3**) is the preferred modality for local staging due to the high accuracy in determining the potential intra- and extra-osseous extension of the tumor into the bone marrow or adjacent soft tissue, including neurological and vascular structures (14,17,24,31).

Fig. 1

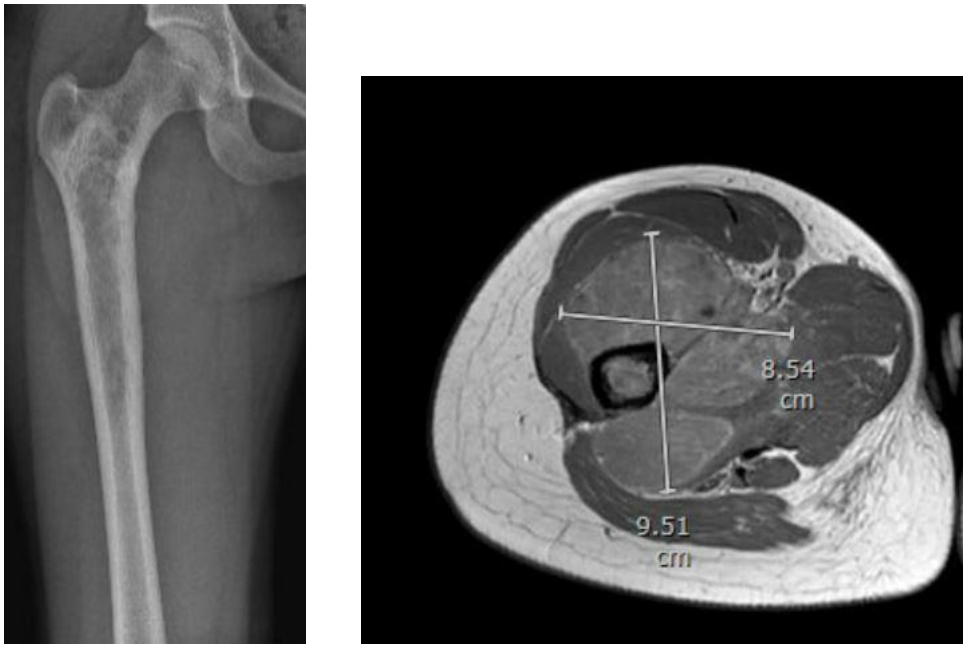
Radiograph and MRI of distal femur suggesting bone malignancy (osteosarcoma)



Anterior-posterior radiograph (left) and the corresponding sagittal MRI image (middle) and axial (right) of distal femur of a 16-year old female, with symptoms during 6 months. The patient was later diagnosed with high-grade osteosarcoma.

Fig.2

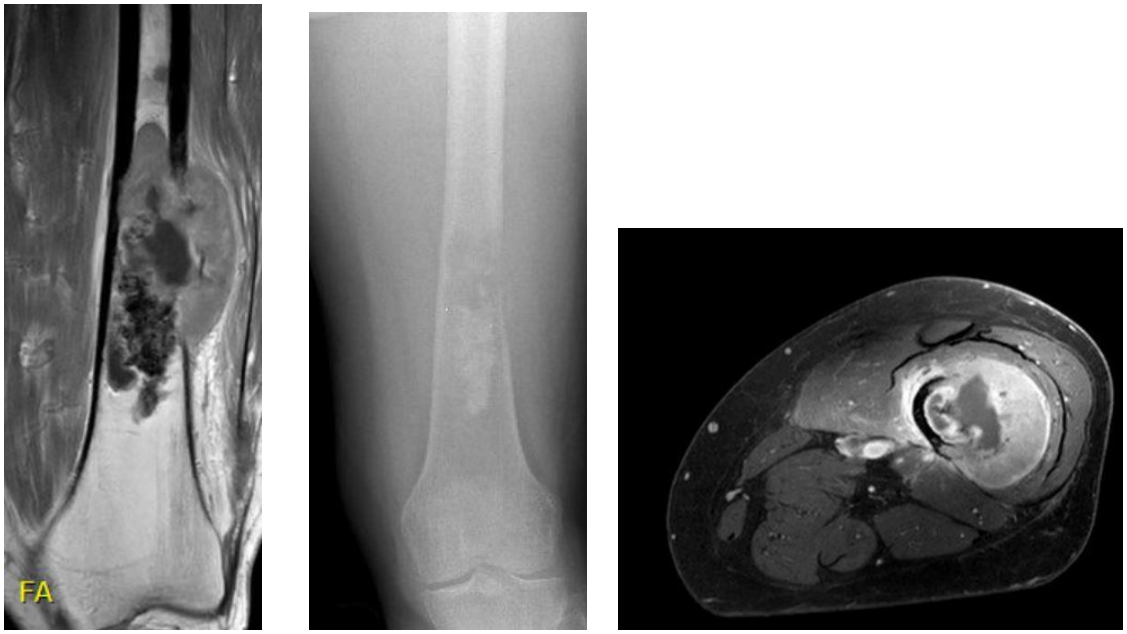
Radiograph and MRI of proximal femur suggesting bone malignancy (Ewing sarcoma)



Radiograph (left) and the corresponding axial MRI image (right) of proximal femur of a 25-year old female with symptoms for a year. The patient was later diagnosed with Ewing sarcoma.

Fig. 3

Radiograph and MRI of distal femur suggesting bone malignancy (chondrosarcoma).

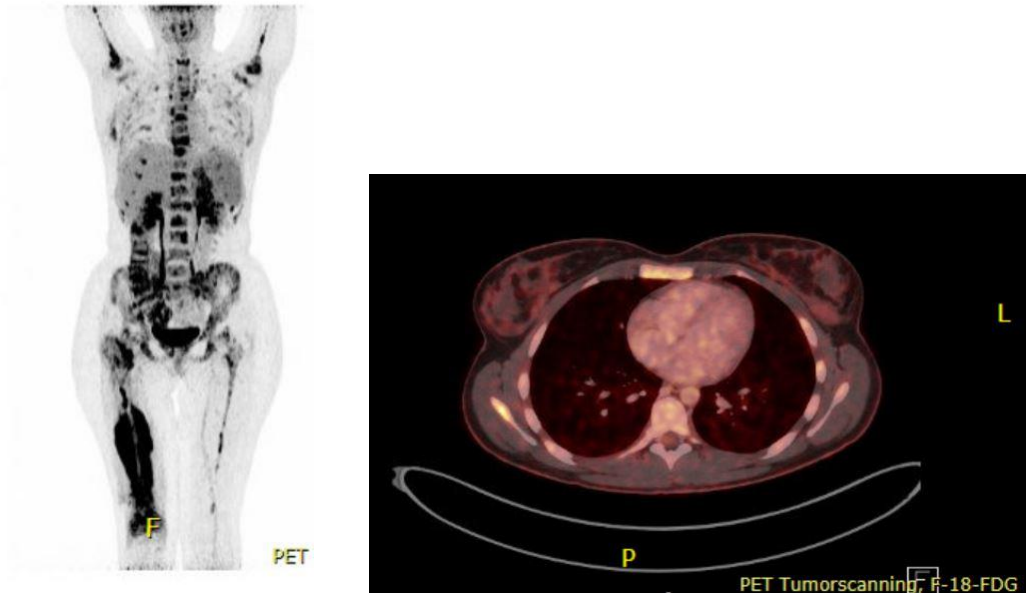


Radiograph (left) and corresponding coronal MRI image (middle) and axial MRI image (right) of a 68-year old female with symptoms for 2 months. The patient was later diagnosed with chondrosarcoma.

If the suspicion of bone malignancy is maintained by MRI, supplementary procedures including biopsy are indicated. According to the Danish guidelines supplementary imaging to determine the distant extent of a tumor includes Computed Tomography scan CT and/or radiograph of the lungs to screen for pulmonary metastases as well as bone scintigraphy to assess bone metastasis. Positron Emission Tomography/Computed Tomography (PET/CT) (**Fig. 4**) is increasingly used for staging by determining potential lung and bone metastases or skip lesions (13,16,17,32–34). According to the Danish national guidelines PET/CT is utilized in patients with high grade bone sarcoma and certain sarcomas with potential atypical pathways of metastasizing (24). Computed Tomography scans (CT) provides superior bone details in terms of better visualization of the periosteal bone formation or cortical destruction and can be used as supplement to radiograph and MRI (16).

Fig.4

PET/CT of a patient with newly diagnosed Ewing sarcoma



PET/CT of an 18-year-old female with a tumor at the right femur. PET/CT demonstrates multiple bone metastases (left) and multiple pulmonal metastases (right). The patient was later diagnosed with Ewing sarcoma.

Tumor grading

The definitive diagnosis is based on histopathological interpretation of a biopsy or resection of tumor (14,16,24). Definitive histological grading of chondrosarcomas has however been a subject for discussion in the literature (35,36) due to the challenging distinction between malignant and benign lesions. Laitinen et al. (37) recently demonstrated a concordance in histological grading of only 43% when comparing preoperative biopsy and resection specimen. Furthermore they found that biopsy downgraded final histology from resection specimen in 50% of cases (37), thus indicating that histological grading based on biopsy is not reliable. They concluded that final histological grading of chondrosarcomas should be based on the highest grade seen in the resection specimen.

According to WHO no generally accepted grading system for bone sarcoma exist (8) although several histologic grading systems have been proposed (38). The two most commonly used histologic grading systems worldwide are the National Cancer Institute grading (NCI) (39) and the grading system by Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) (40). In Denmark bone sarcomas are often graded according to the FNCLCC grading system although it has never been validated for bone sarcomas (8). FNCLCC is based on three parameters: tumor differentiation, mitotic index and tumor necrosis (5,40).

TREATMENT

Along with chemotherapy in some histological subtypes, the current main treatment of bone sarcomas is surgery. Curative intended surgery aims to resect the tumor with wide excision (13,17,41) according to the Enneking classifications (42). The Enneking Classification uses four defined categories of margins:

Intralesional: Dissections within the sarcoma, leaving macroscopic and microscopic tumor.

Marginal: Dissections along the pseudocapsule or within the reactive zone, often leaving microscopic disease at the margin. An adequate dissection for some benign tumors.

Wide excision: Not well defined. Dissections within normal tissue in not defined distance from the pseudocapsule or reactive zone.

Radical: Resect the entire compartment (amputation).

The decision of which surgical principle to choose is an evaluation of risks and benefits of all available options and expected outcomes. Taken into consideration is: patient demographics, tumor size and involvement of anatomically important structures as well as possible response to neoadjuvant chemotherapy and also patient preference.

Surgical options are amputation or limb sparring surgery (LSS) with resection and reconstruction with e.g. rotationplasty, arthrodesis, allografts, autografts and finally the most commonly used reconstruction: tumor prostheses, which is the scope of this thesis.

From amputation to limb sparing surgery

LSS with resection and reconstruction with tumor prostheses has evolved dramatically since 1943 when Austin Moore inserted the first tumor arthroplasty in the proximal femur (43) and decades ahead in 1977 when Marcove et al. (44) reported of the first experimental reconstructions with fixed hinge tumor prostheses in the femoral bone. LSS with resection and reconstruction with tumor prostheses was originally considered experimental due the absence of clinical trials with long-term follow up, and hence only offered to patients with short residual life expectancy (45). Surgery with amputation was the only available treatment, often supplemented with preoperative high-dose radiation (46). Chemotherapy was not a preferred treatment modality and early chemotherapy studies primarily comprised patients with disseminated disease (47).

Despite surgical resection by amputation, early studies demonstrated 5-year survival rates of only 10-20% in patients treated for osteosarcoma (46,47). The introduction of Adriamycin (combined with Methotrexate) in the early 1970's was the first chemotherapeutic treatment to demonstrate improved survival rates of 20%-40% (48,49).

The parallel improvement of imagine modalities with the introduction of CT and bone scintigraphy in the 1970's and also MRI in the 1980's improved the surgical possibilities for LSS dramatically (50). The continuous innovations in imaging modalities until present has allowed accurate staging and hence improved the decision-making and planning of LSS with resection and reconstruction.

The advances in orthopedic implant possibilities, adjuvant chemotherapy and imaging techniques have substantially increased the 5- and 10-year patient survival to currently 60%-90%, depending on subtype (49,51,52). LSS with tumor prostheses has over the last decades therefore gradually replaced amputation as the surgical method of choice (45,53) and LSS is currently offered to more than 90% of all patients with bone sarcoma in the long bones (51,54), combined with chemotherapy in osteosarcomas and Ewing sarcomas. Chondrosarcomas are, however, mainly considered chemo-resistant (55).

Limb sparring surgery

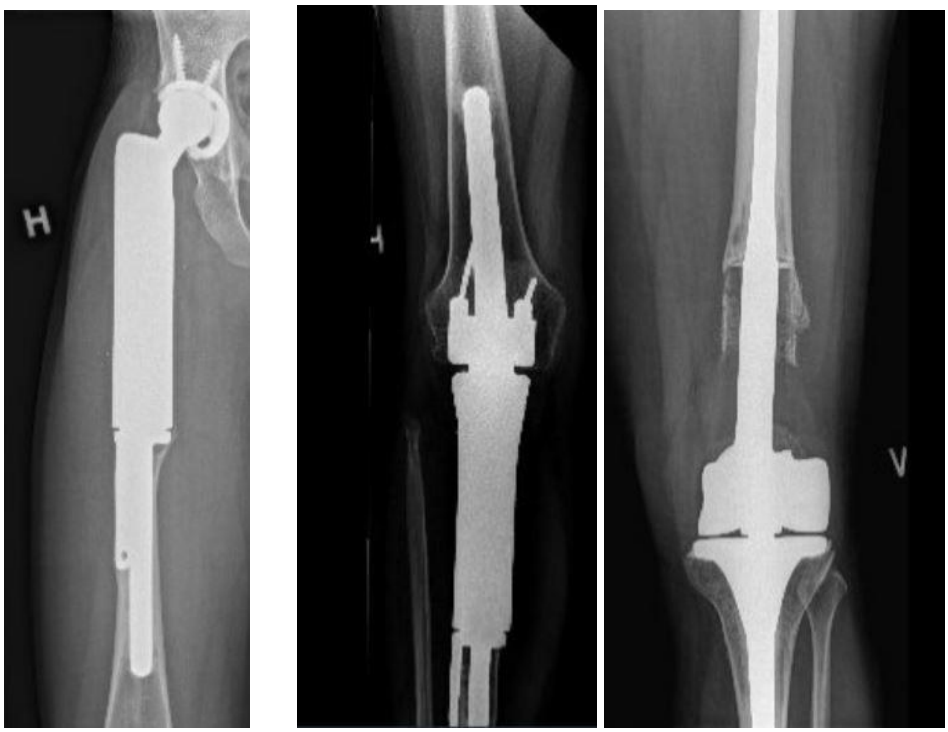
LSS with tumor prostheses is the preferred surgical principle for several reasons: they provide immediate fixation, allowing early weight bearing and also maintenance of function (14,56). The incidence of revisions and complications has nevertheless, since the introduction of tumor prostheses, been higher compared to primary arthroplasty (56–60). Aseptic loosening and deep infection are the most common reported causes for revision and implant failure (51,61–63). The high incidence of complications is affected by several factors. Patients with sarcomas are often subjected to chemotherapy and hence immune compromised combined with a general decline in health due to their malignancy. Also, large surgical exposures, prolonged surgery time combined with extensive loss of tissue and at times limited soft-tissue coverage makes sarcoma patients prone to infections (51,64). Furthermore, Holzer et al. (65) found that osteosarcoma patients subjected to intensive chemotherapy demonstrated significant lower bone mineral density (BMD) when compared to healthy age-matched reference population, and Jeys et al. (59) found a significantly higher risk of infection among patients with bone sarcoma undergoing radiation therapy. In addition, it is well known that implantation of a stem causes bone remodeling and resorption due to the trauma to the bone (66). This is caused by stress shielding, postoperative immobilization and also the operative trauma (67,68). Stress shielding is caused by alterations in loading pattern due to the higher stiffness of the implant. This is according to Wolffs law. The adaptive changes in bone remodeling and loss of BMD predisposes to a higher risk of periprosthetic fractures as breaking strength is directly associated with BMD (69). Furthermore, Andersen et al. (70) demonstrated that low preoperative BMD prior to primary total knee

arthroplasty is related to increased implant migration, hence inferring a higher risk of aseptic loosening due to bone loss. This is supported by Hansen et al. (71) in a prospective study evaluating patients treated with osseointegrated implants after transfemoral amputation. Hansen et al.(71) demonstrated higher risk of implant removal among patients with progressive loss of BMD after surgery. Lastly, in distal femur resections and reconstructions the quadriceps excision and passive extension gait provided by the prosthesis with no support from quadriceps contraction (72–74) has been suggested to cause higher stress to the femur stem and hence a higher risk of aseptic loosening (57,75–77).

The implant evolution

The development of orthopedic implant designs and possibilities has continuously aimed to optimize function and prevent implant complications and revisions. The evolution regarding the knees has moved from early fully constrained or fixed-hinge custom-made designs (**fig. 5**) to modern modular rotating hinge designs (**Fig. 6**) with cemented or uncemented fixation, and various coatings have been suggested to improve bone ingrowth.

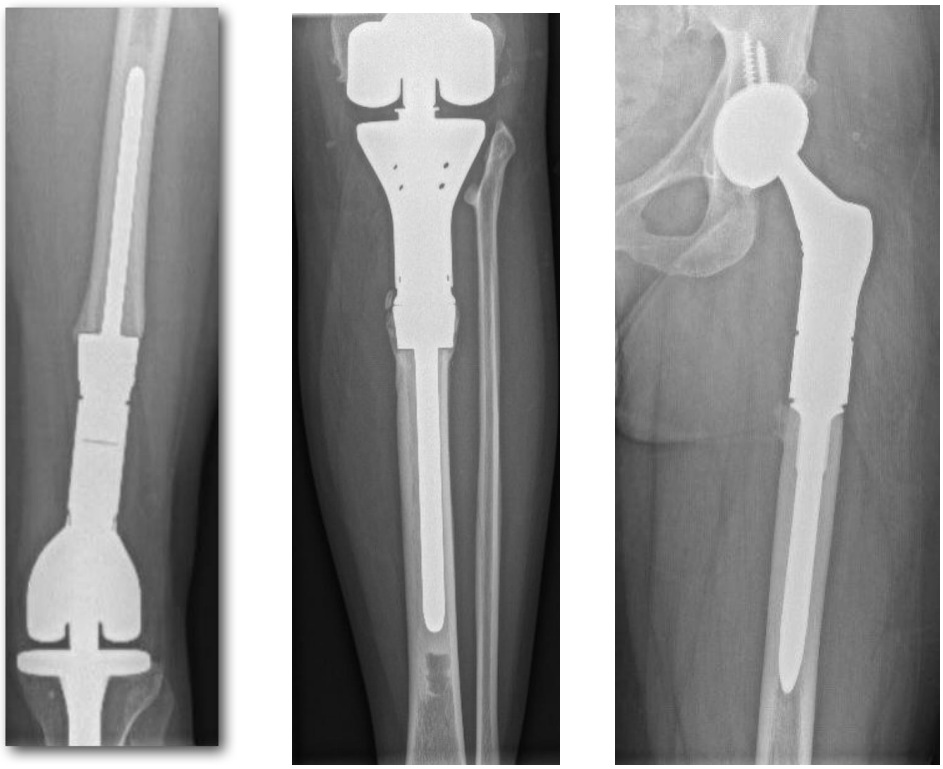
Fig.5



Examples of first-generation implants used. Proximal femur resection and reconstruction with a tumor prosthesis (Howmedica Modular Resection System, Stryker-Howmedica inc. UK) (left). Proximal tibia resection and reconstruction with a fixed-hinge tumor prosthesis Howmedica Modular Resection System, Stryker-Howmedica inc. UK) (middle). Distal femur resection and reconstruction with a rotation-hinge tumor prosthesis (Waldemar Link, Hamburg, Germany) (right).

The early custom-made prostheses were most commonly reported to fail due to stem fractures caused by sub-optimal design and use of alloys unable to withstand the high stress (50). Although the introduction of modular fixed-hinge knee systems in the 1980's provided more patients-specific prostheses (78,79), the fixed-hinge design caused high stress in the bone implant interface (57,61,64). The introduction of rotating-hinge prostheses has therefore been one of the most central improvements to modern treatment with LSS due to the reduction in mechanical stress and bone resorption in the bone implant interface (77).

Fig. 6



Examples of second-generation surgical implants used. Distal femur resection and reconstruction with a rotating-hinge tumor prosthesis (Segmental System, Zimmer, Warsaw, IN, USA) (left). Proximal tibia resection and reconstruction with a rotating-hinge tumor prosthesis (Segmental System, Zimmer, Warsaw, IN, USA) (middle). Proximal femur resection and reconstruction with a tumor prosthesis (Global Modular Replacement System, Stryker, UK.) (right).

In addition, the introduction of trabecular metal or hydroxyapatite coated collars at the prosthesis-bone junction has been suggested to facilitate bony ingrowth and decrease aseptic loosening when compared to implants without collar (61,80,81). Use of cemented and uncemented stem fixations has demonstrated various results of bony ingrowth and rates of aseptic loosening when compared. The results remains inconclusive (57,82–84), and the choice of fixation method mostly relies on surgeon preferences.

Limb salvage

The two most commonly reported causes for amputation after LSS and reconstruction with a tumor prosthesis, are deep infection and local recurrence (61,84). Other causes for amputation such as mechanical complications, severe functional limitations or vascular complications are rarely reported (51). While risk of recurrence in some long-term studies have demonstrated to decrease with time, the risk of deep infection has been demonstrated to persist or increase with time (51). Thus, deep infection is to be considered the current greatest threat for limb survival. Although local recurrence has been reported with lover incidence, it additionally represents a risk of reduced overall survival.

Moving towards personalized evidence-based treatment

Due to the variety and heterogeneity of bone sarcomas, management is based upon individual decision making. The decision to do surgery and which surgical intervention to choose, often relies on predicted survival. Metastasis, tumor size and age are the most commonly suggested prognostic factors for survival at diagnosis (85–87). A prognostic factor is defined as a factor with proven independent impact of a given outcome (e.g. death) regardless of any given treatment. As such, independent prognostic factors are able to identify subgroups with differing risks (e.g. tumor size) and hence they guide decision making (88). Prognostic factors are, however, not powerful enough to guide choice of treatment on an individual level as oppose to validated predictive factors (89). A predictive factor is a factor that identifies differential benefit from a certain treatment depending on the status of the predictive factor (89,90). A prognostic factor can also be a predictive factor but not necessarily; most prognostic factors are not predictive factors (90–92).

Accurate survival prediction for patients with newly diagnosed bone sarcoma would therefore be a considerable aid for the clinician in the decision-making of determining the most appropriate treatment management. A few attempts to create evidence based prediction models for survival in bone sarcoma patients using machine learning techniques has been conducted (93–95). To the best of our knowledge only Bongers et al. (95) has successfully externally validated a 5-year survival prediction model for patients with chondrosarcoma using machine-learning. However, since bone sarcomas often are diagnosed at a late stage, knowledge of predicted short-term survival may mitigate side effects from chemotherapy, extensive rehabilitation or complications after LSS in patients with short residual life expectancy and hence the requirement of proper validated predictive tools of short-term survival is pending.

Machine learning in a clinical setting

Machine learning aims to make the most accurate prediction possible. The technique is based on algorithms that find patterns in preferably large irregular complex sample sizes with nonlinear interactions collected in uncontrolled settings (96). Machine learning makes minimal assumptions about data and the models are solely evaluated by their ability of accurate prediction. Classic statistics is based on well-known methods (e.g. regression) and assumptions aiming to determine inference between variables. In that manner e.g. independent prognostic factors are adjusted for confounding and identified (97). In general, machine learning models do not require any understanding of the underlying mechanisms and are, as stated by Bzdok et al. (96), to a greater extent based on reliability to the empirical capabilities of a given model. Machine learning models are evaluated on a test set, corresponding to evaluating a given regression model by significance test and analysis of confidence intervals.

Due to the substantial differences to classic statistics, the use of machine learning in a clinical setting is subject for discussion (96,98). However, bearing in mind the diverse methodologies and limitations, qualified usage and knowledge of prediction as well as inference are valuable complementary insights in complex clinical cases as with bone sarcomas.

Gradient boosting machines (GBM) is a group of machine learning techniques used to generate non-parametric regression or classification models (99). Gradient boosting uses the ensemble technique, which gradually and sequentially converts weak prediction models (base learners) into stronger. By each boost every new model is subsequently being correlated to the negative gradient of the customized loss function from the previous model. The boosting technique has previously proven to outperform other machine learning models in accuracy and generalizability (100,101) and hence potentially produces a model with consistently higher accuracy compared to conventional single strong machine learning models (101).

OBJECTIVES

The overall objective of the present thesis was to evaluate various aspects of the surgical treatment of bone sarcomas and similar diseases with LSS and reconstruction with tumor prostheses in the lower extremities and furthermore to develop and validate a prediction model for 1-year survival in patients with newly diagnosed bone sarcoma.

Aims and hypotheses

The studies comprising this thesis where based on the following aims:

Study I:

To make a long-term single center evaluation in an early retrospective patient cohort (1985 to 2005) having had LSS and reconstruction with tumor prostheses in the lower extremities.

Hypothesis: As present study was descriptive no hypotheses were tested. Data was presented as descriptive statistics. Postoperative survival, implant survival, limb survival and functional outcome were compared to previously reported results in the peer-reviewed literature.

Study II:

The primary aim was to evaluate if the use of second-generation tumor prostheses for reconstruction after bone resection of primary malignancies in the lower extremities in a retrospective cohort (2006 to 2016) resulted in improved implant survival, limb survival and also functional outcome compared to first generation prostheses (examined in study I).

Hypothesis: We hypothesized that the innovations in e.g. implant possibilities and diagnostic imaging techniques would have resulted in improved implant survival, limb survival as well as functional outcome compared to first generation tumor prostheses (1985 to 2005).

Study III:

To quantitatively measure the adaptive bone remodeling around the intramedullary 130mm Zimmer®Segmental Straight Fluted Cemented Stem in patients with malignant bone tumors receiving tumor prostheses for reconstruction.

Hypothesis: We hypothesized that the use of TM collars together with the intramedullary 130mm cemented Segmental Straight Stem will secure an optimal stem fixation, thus leading only to a limited degree of stress shielding of the peri-prosthetic bone compared to literature.

Study IV:

First, to externally validate the 1-year Bayesian Belief Network (BBN) prediction model for survival of patients with bone sarcomas suggested by Nandra et al. [87] in 2016. Second, to develop and evaluate if a 1-year GBM prediction model outperform the prior BBN model when externally validated.

Hypothesis: We hypothesized that the GBM model would outperform the prior BBN model by measures of discrimination, overall performance and clinical utilization when externally validated.

MATERIAL AND METHODS

Overall study design

The Musculoskeletal Tumor Section, Department of Orthopedic Surgery, Rigshospitalet in Copenhagen (MTS) is a tertiary referral center for orthopedic oncology surgery in the eastern part of Denmark serving approximately 3.9 million citizens. The MTS specializes in bone and soft tissue sarcomas as well as metastatic bone disease (MBD) and aggressive benign bone and soft tissue tumors. Patients with substantiated suspicion of sarcoma or aggressive benign tumor are referred from local hospitals to one of two sarcoma centers in Denmark for further diagnosing, biopsy and treatment. The second center is located at Aarhus University Hospital serving patients in the western part of Denmark.

All studies were non-intervention, observational studies with no estimation of causality. Hence, Study I and II were conducted as retrospective observational studies of patients referred to, the MTS. Study III was performed as a prospective observational study without control. Study IV was performed as a retrospective study comprising a consecutive national historical cohort based on registry data from both sarcoma centers in Denmark.

Due to the Danish social health care system all patients diagnosed with sarcoma, aggressive benign bone or soft tissue tumors as well as MBD are entitled to free government paid treatment in a public medical care system. Hence, all patients diagnosed with sarcoma will be treated at one of the two referral centers for orthopedic oncology.

Data sources

Civil Registration System (CRS)

CRS was established April 2, 1968 and contains information on all Danish Citizens regarding name, date of birth, place of birth, current address, vital status, date of death. All information is linked to the unique 10-digit Civil Personal Register number, and data is updated daily (102). In all 4 cohorts CRS was used for survival status or date of death at follow-up.

The Danish Sarcoma Registry (DSR)

The Danish Sarcoma Registry (103) is a population based prospectively maintained national database since January 1, 2009. Patients from 2000-2008 were later included to DSR by a validation through the Danish Cancer Registry and the Danish National Pathology Registry (104,105). The registry contains information about patient characteristics, tumor characteristics, diagnostic, details on treatment, local and distant recurrence, comorbidity as well as death (study IV).

Patient inclusion

The present thesis is based upon four study populations:

Study I

Method

A consecutive retrospective cohort of patients having LSS and reconstruction with tumor prostheses at MTS for bone sarcoma or aggressive benign tumor of the lower extremities from January 1, 1985 to December 31, 2005 was identified from our local pathology database. Patients with other types of surgical treatment were excluded as well as all surgeries performed as revisions of prior implants were excluded. Patients were followed until death or end of study (January 1, 2015) resulting in a minimum follow-up of 9 years.

Subjects

Fifty patients with a mean age of 34 (6-74) years (F/M 24/26) diagnosed with bone sarcoma (n=44), aggressive benign tumor (n=6). Histological diagnoses were osteosarcoma (n=30), chondrosarcoma (n=9), giant cell tumor (n=6), Ewing sarcoma (n=4), angiosarcoma of bone (n=1). The anatomical locations were distal femur (n=29), proximal femur (n=29), proximal tibia (n=9) and entire femur (n=3) (**Table 1**). The following types of implants were used for reconstruction: HMRS® (Stryker-Howmedica Inc.) (n=32), KMFTR® (Stryker-Howmedica Inc.) (n=2), Link Endo-rotational (Waldemar Link) (n=5), Capanna/MegaC (Waldemar Link) (n=2), Bimetric revision (Biomet) (n=2), Kent hip (Biomet) (n=1), GMRS® (Stryker) (n=1), custom-made (Waldemar Link) (n=1), unknown (n=4).

Study II

Method

A consecutive retrospective cohort (from here referred to as *late cohort)* of all patients who underwent LSS and reconstruction with tumor prostheses at MTS for bone sarcoma, soft tissue sarcoma adjacent to bone or joint or aggressive benign tumor, between January 1, 2006 and December 31, 2016 was identified by manually screening our institutional surgical planning system. Patients with other types of surgical treatment or revision of prior implants were excluded. Patients were followed until death or end of study (January 1, 2018) resulting in a minimum of 2 year follow up. The consecutive cohort from study I is from here referred to as *early cohort*.

Subjects

Seventy-two patients (F/M=30/42), mean age 44 (range 7-84) years diagnosed with bone sarcoma (n=60), soft tissue sarcoma adjacent to the bone (n=9) or an aggressive benign bone tumor (n=3) were identified. The anatomic sites of reconstruction were distal femur (n=33) proximal femur (n=24), proximal tibia (n=12) and entire femur (n=3) (**Table 1**). The following types of implants were used for reconstruction: GMRS (Stryker®) (n=37), Segmental (Zimmer®Biomet) (n=27), Mega C (Link®) (n=5), Link custom made growing prostheses (Link®) (n=3).

Table 1.

Patient characteristics from study I and II

	All patients	**Study I**	**Study II**	**p-value**
No of patients	*n=122*	*n=50*	*n=72*	
Female/male	*54/68*	*24/26*	*30/42*	*p =0.58**
Patients alive at end of study	*75*	*28*	*47*	*p =0.35**
Mean age at surgery (range)	*39 (6-84)*	*34 (6-74)*	*44 (7-84)*	*p = 0.02#*
Location				
Hip	*33 (27%)*	*9 (18%)*	*24 (33%)*	*p =0.06**
Knee	*83 (68%)*	*38 (76%)*	*45 (63%)*	*p =0.17**
Total femur	*6 (5%)*	*3 (6%)*	*3 (4%)*	*p =0.68**
Total number of revisions	*137*	*78*	*59*	*p =0.27**
Angiosarcoma	*3 (3%)*	*1 (2%)*	*2 (3%)*	*N/A*
Ewing sarcoma	*8 (7%)*	*4 (8%)*	*4 (6%)*	*N/A*
Giant cell	*9 (7%)*	*6 (12%)*	*3 (4%)*	*N/A*
Myofibrosarcoma	*5 (4%)*	-	*5 (8%)*	N/A
Synovial cell sarcoma	*1 (1%)*	*1(2%)*	*1 (1%)*	N/A
Rhabdomyosarcoma	*1 (1%)*	-	*1 (1%)*	N/A
Leiomyosarcoma	*1 (1%)*	-	*1 (1%)*	N/A
Sarcoma NOS	*5 (4%)*	-	*5 (7%)*	N/A

#student unpaired t-test, *Fishers exact test

Study III

Method

A prospective cohort comprising patients, who underwent bone tumor resection with LSS, and reconstruction with a cemented Zimmer® Segmental System tumor prosthesis (Zimmer Biomet) in the lower extremities between January 1, 2015 and July 1, 2018 was identified. Inclusion criteria were LSS due to sarcoma, MBD or aggressive benign bone or soft tissue tumor. Inclusion criteria were patients with age > 15 years, patients with no diseases severely affecting the bone metabolism and patients with expected survival more than 1 year (estimated by the surgeon and the investigators).

Subjects

Thirty-three patients were assessed for eligibility (**Fig. 7**). Twenty-eight patients were included, and 21 patients (F/M=12/9, mean age 55 years) diagnosed with a primary bone tumor (n=6), aggressive benign tumor (n=4), myelomatosis (n=2) or MBD (n=9), completed 1-year follow-up (**Table 2**).

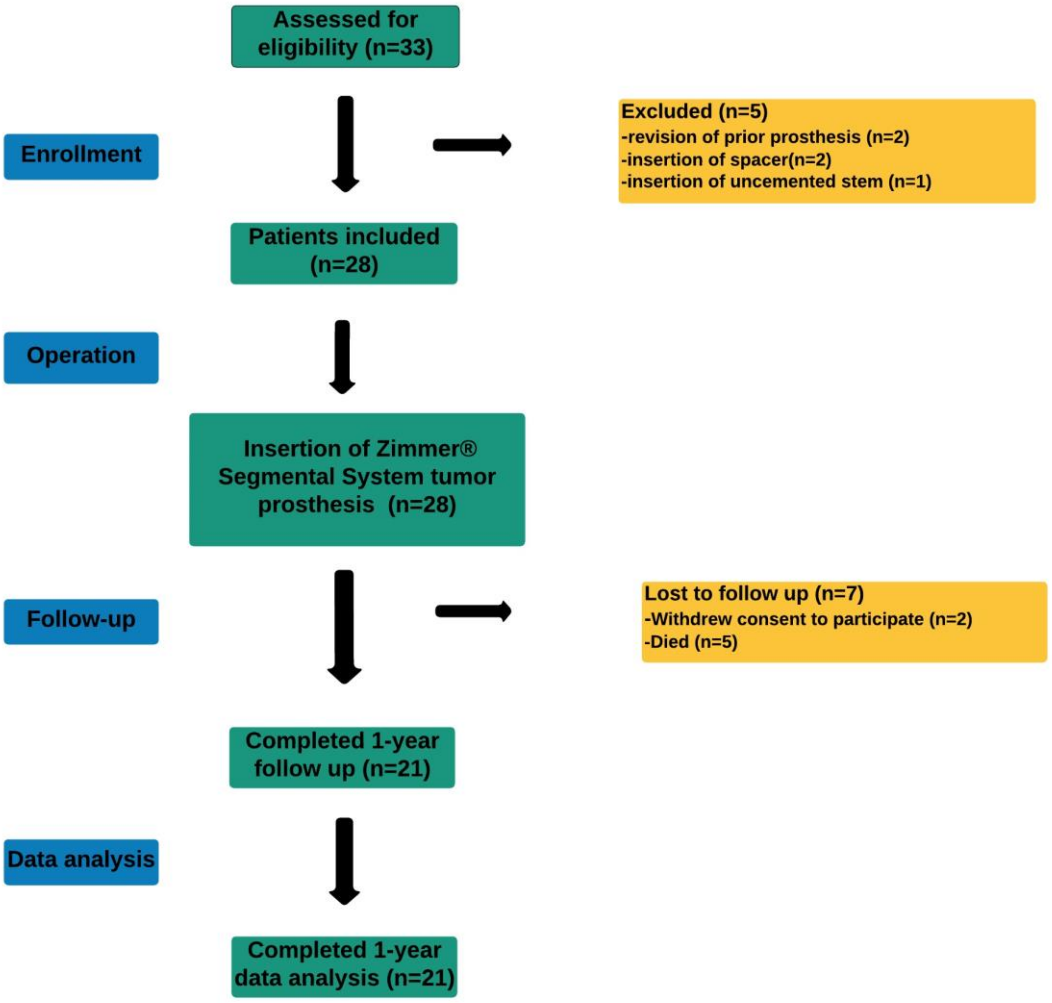
Table 2.

Baseline data of patients (n=21) that completed 1-year follow-up in Study III.

Variable	Level	Total (%)
Gender		
	Female	12 (57%)
	Male	9 (43%)
Age (years)		
	Mean (range)	55 (15-81)
Resection (cm)		
	Mean (range)	15 (10-24)
Resection site		
	Proximal femur	10 (48%)
	Distal femur	9 (43%)
	Proximal tibia	2 (10%)
Pathology		
	Metastasis	9 (43%)
	Giant Cell	4 (19%)
	Chondrosarcoma	2 (10%)
	Myelomatosis	2 (10%)
	Osteosarcoma	2 (10%)
	Myxoid liposarcoma	1 (5%)
	Desmoplastic fibroma	1 (5%)

Fig. 7

Study III flow chart. Enrollment, follow-up, and data analysis.



Study IV

Method

A retrospective national historical cohort comprising patients with age > 15, newly diagnosed with bone sarcoma identified from DSR (5) between January 1, 2000 and June 22, 2016 served as the validation cohort (n=771). Patients were included from the only two tertiary referral Centers for orthopedic oncology in Denmark. All patients were accounted for a minimum of 2-year follow-up until end of study or death.

The training cohort is previously described by Nandra et al. (94). In short, 3493 patients diagnosed and treated for bone sarcomas between 1970 and 2012 at The Royal Orthopaedic Hospital, Birmingham UK, were included in their institutional prospectively maintained database.

Subjects

As intended the demographic and clinical features of the training set and validation set differed (**Table 3**). The proportion of missing values varied among features although most notable in the training set were the tumor size (missing in 51%) and in the validation set the tumor grade (missing in 23%) (**Table 3**).

Table 3
Distribution and comparison of baseline variables between training and validation cohort of study IV

Variable	Level	Training cohort 1970-2012 n=3493 (%)	Validation cohort 2000-2012 n=771 (%)	Total n=4264 (%)	*p*-value
Gender					0.22*
	Female	1451 (42)	338 (44)	1789(42)	
	Male	2042 (59)	430 (56)	2472(58)	
	missing	0	3 (0.4)	3 (0.1)	
Age					<.0001#
	Median (IQR)	23 (14-51)	44 (22-62)	26 (15-53)	
	missing	0	3 (0.4)	3 (0.1)	
Tumor size (cm)					<.0001*
	Median (IQR)	10 (7-13)	6(3-10)	8 (2-12)	
	missing	1796 (51)	0	1796 (42)	
Grade					<.0001*
	High	2641(76)	293 (49)	2934 (72)	
	Intermediate	374 (11)	143 (24)	517 (13)	
	Low	478 (14)	158 (27)	636 (16)	
	missing	0	177 (23)	177 (4)	
Histology					<.0001*
	Osteosarcoma	1572 (45)	174 (25)	1746 (41)	
	Chondrosarcoma	793 (23)	326 (46)	1119 (26)	
	Ewing	653 (19)	114 (16)	767 (18)	
	Sarcoma	182 (5)	26 (3)	191 (4)	
	Chordoma	70 (2)	34 (5)	104 (2)	
	Other (19 histologic diagnoses)	223 (6)	36 (5)	259 (6)	
	missing	0	61(8)	61(1)	
Pathologic fracture at diagnosis					<.0001*
	No	3035 (87)	729 (95)	3764 (88)	
	Yes	458 (13)	42 (5)	500 (12)	
	Missing	0	0	0	
Anatomic location					<.0001*
	Head and Neck	20 (1)	50 (7)	70 (2)	
	Lower Extremity	2118 (61)	355 (47)	2473 (58)	
	Pelvic Girdle	642 (18)	117 (16)	759 (18)	
	Spine	0	32 (4)	32 (1)	
	Upper Extremity	471 (14)	103 (14)	574 (14)	
	Upper Trunk	230 (7)	93 (12)	323 (8)	
	missing	12 (0.3)	21 (3)	33 (1)	
Metastasis at diagnosis					0.63*
	No	3010 (86)	651 (87)	3661 (86)	
	Yes	483 (14)	98 (13)	581 (14)	
	missing	0	22 (3)	22 (0.1)	
Status at 1 year after diagnosis					0.009*
	Alive	3099 (89)	655 (85)	3754 (88)	
	Dead	394 (11)	113 (15)	507 (12)	
	missing	0	3 (0.4)	3 (0.1)	
Year of diagnosis					
	missing	222 (6)	3 (0.4)	225 (5)	-

***Mann-Whitney U-test, #Chi square test**

Outcomes and measurements

Study I

The first outcome was overall survival defined as the probability of survival from the day of surgery to death of all causes or end of study.

Second outcome was implant survival defined as the probability of revision free survival from day of surgery to end of study. Revisions were defined as unplanned prosthesis-related surgery of all causes (including wear of polyethylene, replacement of bushings, superficial aseptic wound revisions etc.) from day of surgery until end of study.

Third outcome was limb survival defined as the probability of amputation free survival. Amputation was defined as amputations of all causes.

Fourth, was incidence of complications classified according to the Henderson classification (106).

Study II

The first outcome was overall survival defined as the probability of survival from the day of surgery to death of all causes or end of study.

Second outcome was the cumulative incidence of revision. Revisions were defined as: *Major revisions*: change or removal of bone-anchored parts of the implant or amputation of the extremity for any cause. *Minor revisions*: all implant-related surgeries without removal of bone anchored parts, including local recurrence, brisement force, DAIR (debridement, antibiotics and implant retention), repositions etc. Aseptic superficial wound revision was not defined as revision.

Third outcome, was the cumulative incidence of amputation defined as amputation of the extremity for all causes.

Fourth outcome, incidence of complications classified according to the Henderson classification (106).

Study III

Primary outcomes were the results of the DXA measurements postoperatively and with follow-up after 3 months, 6 months, and 12 months, defined as the mean BMD (g/cm^2) changes in ROI 1, ROI 2, ROI 3, ROI4.

Secondary outcome, was the BMD (g/cm^2) measurements in both ankles postoperatively and with follow-up after 3 months, 6 months, and 12 months after surgery, defined as the mean BMD in a single ROI (**Fig.8b**). These measurements were used to assess if the periprosthetic BMD

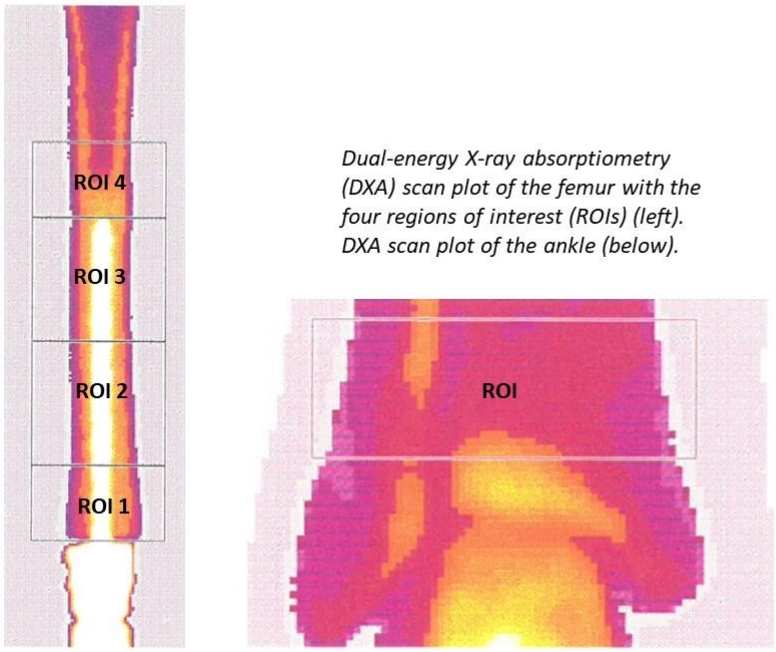
changes were caused by a general decrease in BMD, a response to the surgical trauma, immobilization, or because of stress shielding.

Dual energy x-ray absorptiometry

Bone mineral measurements using dual energy x-ray absorptiometry (DXA) of the periprosthetic bone of the femur or tibia around the stem and adjacent to the TM collar was measured by DXA using a Norland XR-46 scanner (scan resolution 0.5 x 0.5 mm, scan speed 45mm/s). All measurements were performed at the Bone Laboratory, Department of Orthopedics, Rigshospitalet, Denmark. All patients were placed supine with the femur in neutral rotation during scanning. On the computerized scan-plots we selected three regions of interest (ROI) around the stem in the femoral or tibial bone and a ROI adjacent to the TM collar for measurements of local changes in BMD (**Fig. 8a**). A custom-made metal exclusion software facility, which allows a variable threshold for metal exclusion, was used for scan analysis. The threshold ranged from: 4.0 g/cm^2 – 6.0 g/cm^2. The precision of the BMD measurements was calculated from double measurements of 6 patients.

By using the same DXA technique (scan resolution: 1.0 x 1.0 mm; scan speed: 45 mm/s) we performed scans of the ankle of the operated side and the contralateral non-operated side. BMD was measured in a 2-cm long ROI located 1 cm proximal from the ankle joint (**Fig. 8b**) (Study III).

Fig. 8



Dual-energy X-ray absorptiometry (DXA) scan plot of the femur with the four regions of interest (ROIs) (left). DXA scan plot of the ankle (below).

Study IV

Primary outcome was external validation of the published 1-year survival BBN model by Nandra et al. (94), with survival defined as survival 1-year after day of diagnosis. Secondary outcome was the development of a 1-year survival GBM prediction model and third outcome was externally validation of the GBM model.

Clinical outcome

The Musculoskeletal Tumor Society Score (MSTS) (study I, II, III)

The Musculoskeletal Tumor Society Score (MSTS) was introduced in 1983 by Enneking et al. and modified in 1993 (107) for evaluation of the functional outcome after treatment of sarcomas.

The system estimates from bad to very good with parallel assigned values from 0-5 in six categories in the lower extremities:

- Pain
- Function
- Emotional acceptance
- Supports
- Walking
- Gait.

Subsequently the six values are added and divided by the maximum value of 30 and a percent rating is calculated.

Statistical considerations

Study I

Kaplan-Meier survival analysis

Probability of 5- and 10-year overall survival was estimated by Kaplan-Meier analysis. Kaplan-Meier curves also were performed for evaluation of over-all survival (calculated from time of surgery), implant survival (time to 1st revision) and limb survival (time to amputation). 95% confidence intervals (CI) were calculated using Microsoft Excel using Greenwood's formula for calculation of standard error.

Kaplan-Meier survival analysis is based on three assumptions: All censored individuals have the same expected survival as those who continues to be followed, probability of survival is not related to time of inclusion and lastly that a given event happens at a certain time point.

Considerations: Kaplan-Meier method was designed to assume identical risk in censored and uncensored patients without considering other competing risks (such as death) that may alter the probability of the outcome. By censoring patients who die before a potential revision or amputation, Kaplan Meier assumes the risk of revision to be independent of death and hence potentially overestimates the probability of revision- or amputation free survival.

Study II

Demographic data

All data were tested for distribution of normality. Patient demographics were analyzed descriptively and tested for significance by Chi^2-test (categorical variables) and Mann-Whitney test (continuous variables). Confidence intervals (CI) are reported as 95% CI and p-values <0.05 were considered statistically significant.

Kaplan-Meier survival analysis

Overall patient survival was estimated by Kaplan-Meier survival analysis with right censoring. Log-rank test was used for comparison of survival between groups.

Competing risk analysis

Cumulative incidence function including competing risk was used in study II. The Aalen-Johansson estimator was used to assess the cumulated incidence of major and minor implant revisions calculated by a competing risk model with death and amputation as competing risks, and with death

as competing risk when calculating the cumulative incidence of amputation. Cumulative incidence of failures according to the Henderson Classifications was calculated using the competing risk analysis. Grays test was used to assess differences between groups.

Consideration: Since all patients available were included, we did not perform power calculation. This does not exclude an underpowered sample size which is also reflected in the resulting wide confidence intervals. However, by this method we performed the true comparison between cohorts.

Study III

Demographic data

Patient and surgery characteristics were reported as proportion (%) means with standard deviation (SD).

Repeated measures analysis of variance (ANOVA)

All changes in BMD over time were analyzed using repeated measures ANOVA, and students paired t-test for comparison of the stepwise BMD changes over time compared to the first postoperative scanning. P-values below 0.05 were considered significant. Precision of the BMD measurements was evaluated by calculation of the coefficient of variation (CV = SD / mean) x 100%). All data are presented as mean (SD or range). The mean MSTS scores were compared using the students t-test.

Considerations: Repeated measures ANOVA requires complete data since one missing value leads to exclusion of all data of a given patient. Also, repeated measures ANOVA can be biased by outside factors during follow-up. An alternative would be a linear mixed model since it uses all data despite missing values. We consulted a biostatistician and were advised to conduct the repeated measure ANOVA due the more feasible interpretation in a small sample size with only a few missing values. In addition, repeated measures ANOVA are well suited for small sample sizes and we only had few missing data of those who completed 1-year data analysis follow-up. All available data was used when performing students paired t-test. Doing multiple comparisons creates risk of Type I errors (mass-significance effect). Since the study was not a confirmatory study and no hypothesis was statistically tested, correction for multiplicity effect is not required (108,109).

Study IV

Demographic data

All data were tested for distribution of normality. Baseline distributions were reported as proportion (%) medians with interquartile ranges (IQR) or means with standard deviation (SD). Categorical data were compared using Chi^2-test and Mann–Whitney U test was performed for comparison of continuous variables.

Measures of discrimination

The ability of accuracy and discrimination was determined by receiver operating characteristic (ROC) analysis and area under the curve (AUC) (110). A value of 1 is perfect discrimination and a value of 0.5 represents chance. Validation was considered suitable for clinical usage if the AUC under the ROC curve was greater than 0.7 as the lowest acceptable threshold and was determined a priori. Discrimination is a measure of how well the model can separate those who do and do not experience the outcome.

Brier score

Overall predictive model performance was evaluated with the Brier score (111). The Brier score quantifies the compliance between the predicted probability and observed outcome. The reported value between 0 and 1 is the average squared differences between all the predicted and actual outcomes in the cohort, with 0 indicating perfect agreement and 1 indicating perfect disagreement. A score of 0.25 reflects a 50% incidence of outcome and hence scores above 0.25 is also to be considered non-informative (112).

Decision curve analysis

Decision curve analysis (DCA) evaluates the clinical usage of a model by quantifying the consequences of over- or undertreatment (112,113). Prediction models generates a survival probability at a given time point after diagnosis. When a probability of survival is between 0 and 1, decision-making might be more difficult for the clinician. A threshold probability is the point where the expected benefit of surgery is equal to the expected benefit of not treating and where surgeons may become indecisive (114). Threshold possibilities between 0 and 1 are plotted against net benefit on a decision curve. We compared the net-benefit of all thresholds and hence determined the clinical use of the GBM model. A model is considered to be clinical usable if it

demonstrates net benefit across the range of thresholds i.e. is superior to assuming that all patients or no patients would live longer than 1 year.

Considerations: Although several other analyses for evaluation of model performance exist, we suggest that present analyses are well suited for the purpose of study IV.

APPROVALS

Study I:

The study I was approved by the Danish Data Protection Agency (no. 2013-41-2591) and the Danish Health and Medicines Authority (no. 3-3013-894/1).

Study II:

Study II was approved by the Danish Data Protection Agency (j.nr: 2012-58- 0004) and the Danish Health and Medicine Authority (3-3013-2578/1). Informed consent was obtained from all individual participants still alive at inclusion in the study.

Study III:

Study III was approved by the Scientific Ethical Committee of the Capital Region of Denmark (J. No. H-2-2014-105) and the Danish Data Protection Agency (J. No.:2012-58-00004, J.nr.: 2013-41-2591). The study was conducted in accordance with the ethical standards of the national ethical committee and with the 1964 Declaration of Helsinki. Prior to inclusion informed consent was obtained from all participants after written and oral information.

Study IV:

Approval was obtained from The Danish Data Protection Agency (P-2019-54), the Danish Patient Safety Authority (no. 3-3013-2866/1) and The Danish Clinical Quality Program–National Clinical Registries (RKKP) (J.nr.: DSD-2019-06-20).

From The Royal Orthopaedic Hospital, Birmingham UK, all patients consent to their data being stored in their institutional oncology database for research use. The study was authorized by their local ethical board as a service evaluation and did not require any additional procedures.

SUMMARY OF STUDY RESULTS

Study I

Overall survival

The probability of over-all 5-year, 10-year, and 15-year survival was 68% (CI: 55%-81%), 60% (CI: 46%-74%), 55% (CI: 40%-69%) respectively (**Fig. 9**).

Implant survival

The probability of 5-year, 10-year, and 15-year implant revision-free survival was 43% (CI: 27%-61%), 24% (CI: 9%-41%), and 16% (CI:0%-28%) respectively (**Fig. 10**).

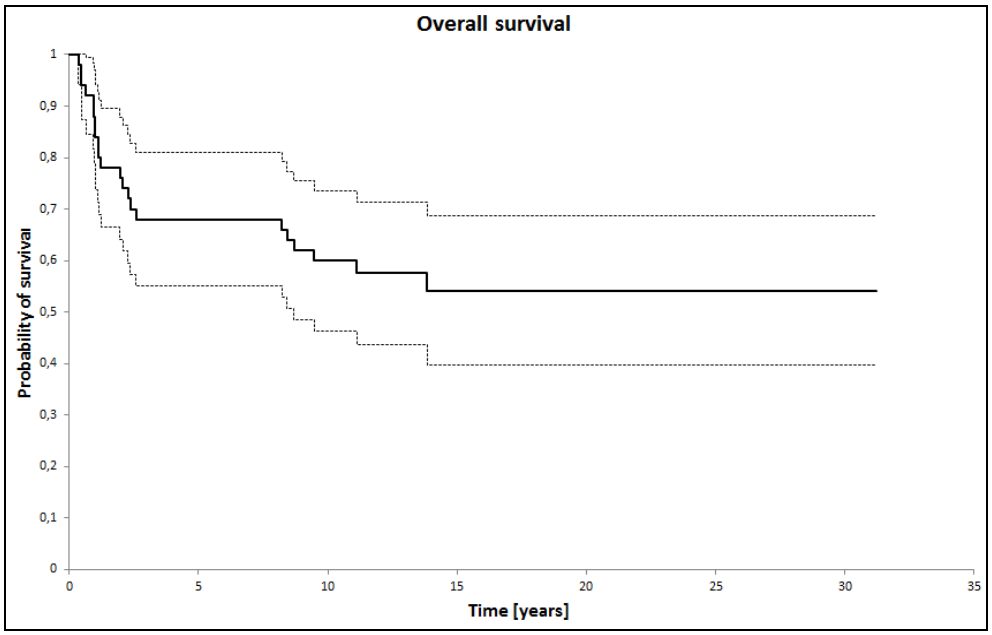
Limb survival

The probability of 5-year, 10-year, and 15-year limb survival was 89% (CI: 80%-100%), 83% (CI: 65%-96%), and 83% (CI: 65%-96%) respectively (**Fig. 11**)

Clinical outcome

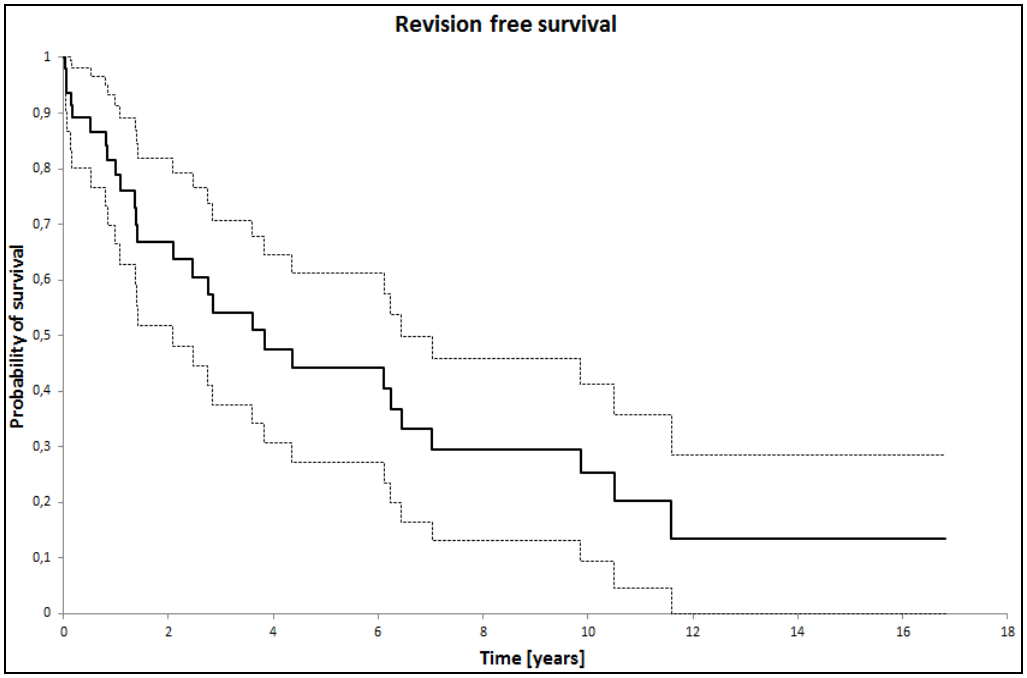
The functional outcome was evaluated in 24 patients (out of 28 patients alive at clinical follow-up) with their limbs still spared after an average of 17 (9-30) years postoperatively. The individual parameters mean scores were: pain 3.9 (0-5), function 2.1 (0-5), emotional acceptance 4.1 (1-5), supports 4.0 (0-5), walking ability 3.8 (0-5), and gait 3.3 (0-5). The mean MSTS score was 21.2 (range 6-30), representing a score of 71%.

Fig. 9



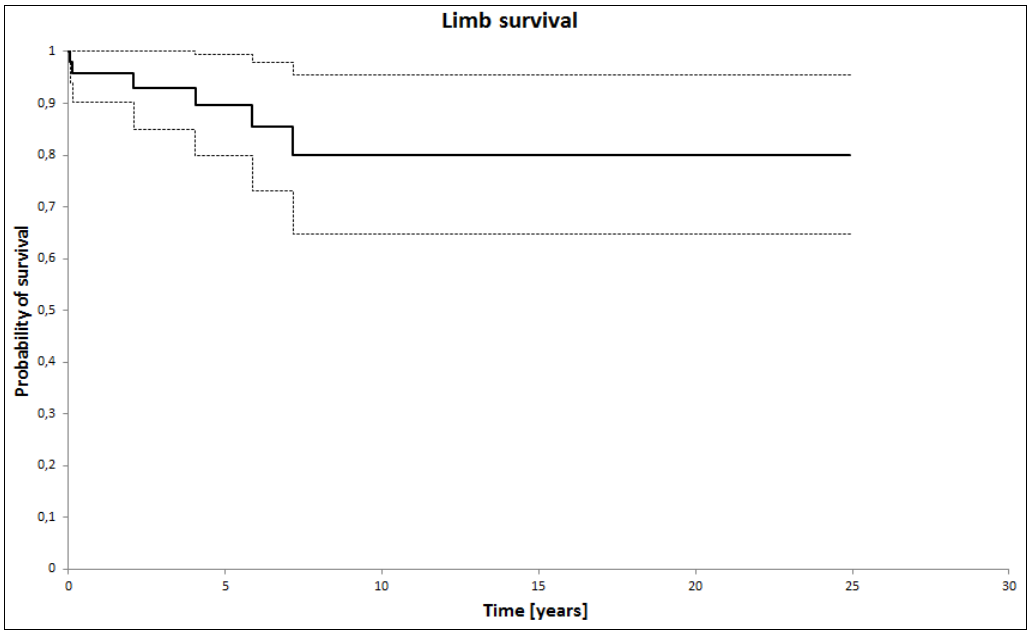
Kaplan-Meier survival curve with 95%CI demonstrating the probability of overall survival after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 1985-2005 (n=50).

Fig. 10



Kaplan-Meier survival curve with 95%CI showing the probability of implant revision-free survival after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 1985-2005 (n =50)

Fig. 11



Kaplan-Meier survival curve with 95%CI showing the probability of limb survival after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 1985-2005 (n =50)

Study II

Overall survival

The probability of over-all 5- and 10-year survival for the present late cohort was 68% (95%CI 57%-79%) and 61% (CI95% 48%-74%) respectively (**Fig. 12**).

Our study showed no difference in overall survival between the early and late cohort (p =0.93) (**Fig. 13**)

Implant survival

The 2, 5, and 10-year cumulative incidence of major revision was 11% (95%CI: 4%-18%), 16% (95%CI: 7%-25%), and 18% (95%CI: 9%-28%) respectively (**Fig. 14**). The 2, 5, and 10- year cumulative incidence of minor revision was 15% (95%CI: 7%-24%), 20% (95%CI: 11%-30%), and 25% (95%CI: 14%-36%) respectively. When comparing cumulative incidence for minor and major revision between early and late cohort, we found no statistically significant difference (p =0.9 and p=0.2 respectively) (**Fig. 15**).

Limb survival

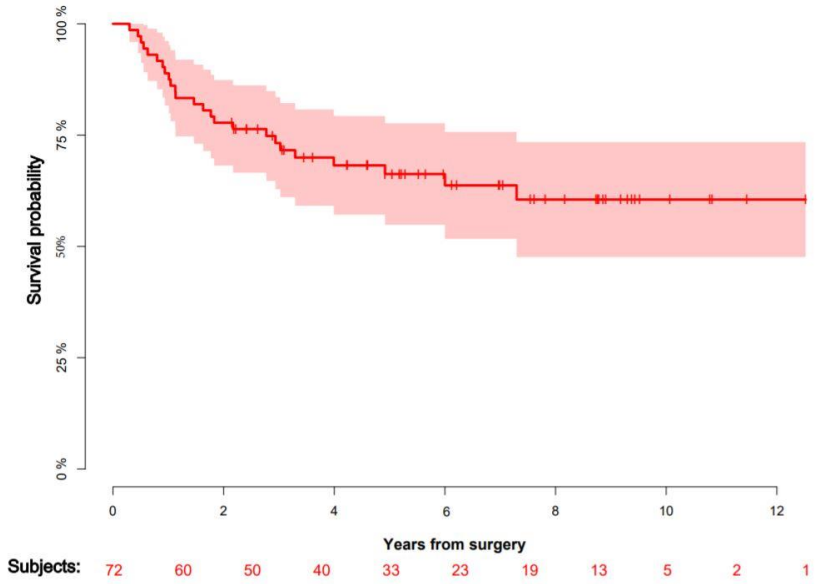
Eight patients in the late cohort were amputated (11%) (resection site: knee (n=5); hip (n=3)). Seven patients (10%) were amputated due to recurrence of tumor, and one patient (1%) due to acute ischemia. The 2, 5, and 10-year cumulative incidence of amputation was 8% (95%CI: 2%-15%), 8% (95%CI: 2%-15%), and 11% (95%CI: 3%-18%), respectively (**Fig. 16**). We found no difference comparing cumulative incidence for amputation between the early and late cohort (p = 0.9) (**Fig. 17**).

Clinical outcome

47 patients were alive during the entire study period, and functional outcome was evaluated in 40 patients) with their limbs still spared after an average of 6 (1.7-12) years postoperatively. Seven of the 47 patients did not have a functional evaluation because they had been amputated (n=3) or were lost to follow-up for various reasons (n=4). The mean (range) individual MSTS score parameters were: pain 3.5 (0-5), function 2.6 (0-5), emotional acceptance 3.5 (0-5), supports 3.8 (0-5), walking ability 3.7 (0-5), and gait 3 (0-5). Mean MSTS score was 20.2 (range 6-30), representing a mean score of 67%.

Fig. 12

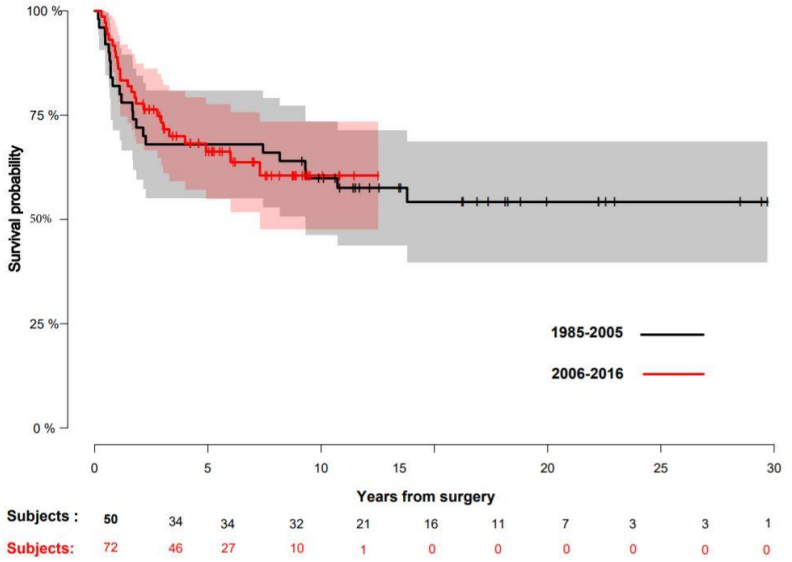
Probability overall survival in late cohort



Kaplan-Meier survival curve with 95%CI demonstrating the probability of overall survival after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 2006-2016 (n=72).

Fig. 13

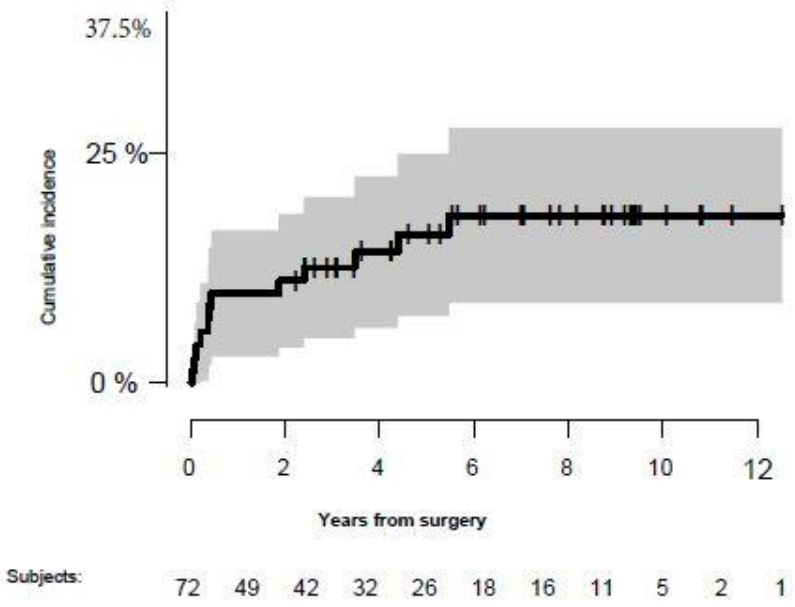
Probability overall survival in early and late cohort



Kaplan-Meier analysis of overall survival in the early and late cohort with 95%CI Log-rank test demonstrated no statistically significant difference (p=0.93) in probability of overall survival between early (black line) and late (red line) cohort.

Fig.14

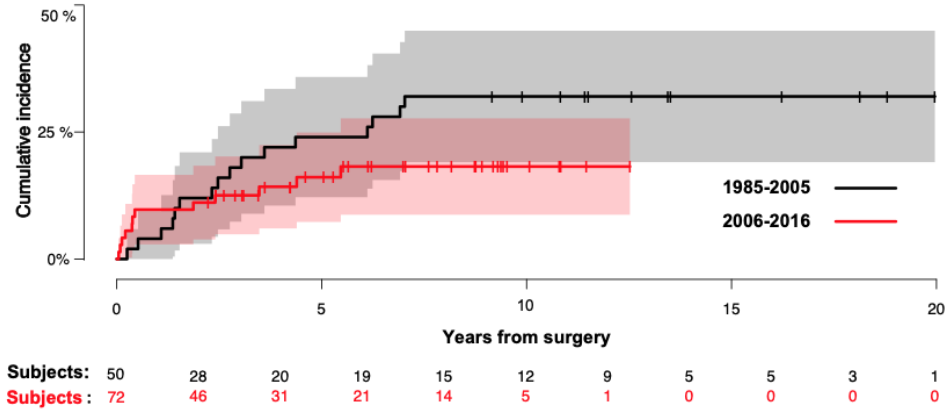
Risk of major revision in late cohort



Cumulative incidence with 95%CI of major revisions resulting in implant failure (amputation, removal of implant or removal of bone anchored parts) after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 2006-2016 (n=72).

Fig. 15

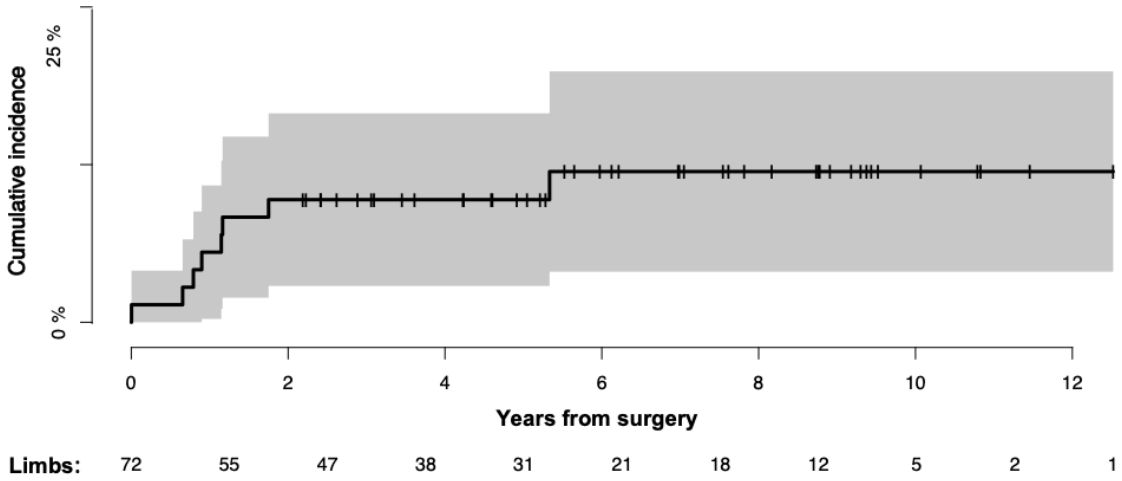
Risk of major revision in early and late cohort



Cumulative incidence with 95%CI of major revisions resulting in implant failure (amputation, removal of implant or removal of bone anchored parts) after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses in the early (n=50) (black line) and late (n=72)(red line) cohort. Grays test demonstrated no statistically significant difference in risk of revision between early and late cohort (p=0.2).

Fig. 16

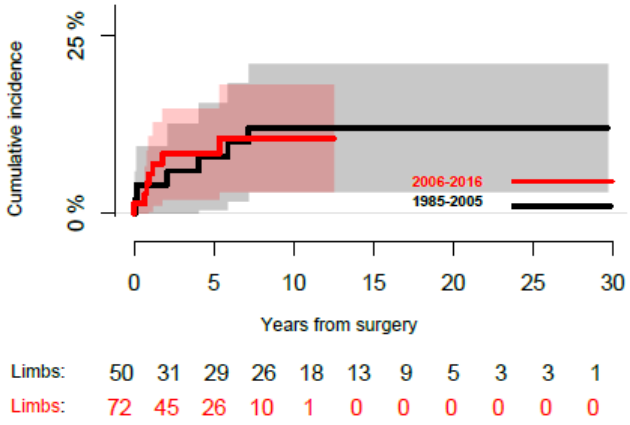
Risk of limb amputation in late cohort



Cumulative incidence with 95%CI of amputation after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses 2006-2016 (n=72).

Fig. 17

Risk of limb amputation in early and late cohort



Cumulative incidence with 95%CI of amputation after resection of primary bone tumors of the lower extremities and reconstruction with tumor prostheses. Grays test demonstrated no statistically significant difference in risk of amputation between early (black line) and late (red line) cohort (p=0.9).

Study III

Clinical outcome

The mean MSTS score was 17 (5-29) after 3 months. The score did not change during the follow-up, and it was 18 (4-30) after 12 months representing a mean score of 59%. After 3 and 6 months the highest score was in the emotional acceptance category (mean score: 3.8) and lowest in the function category (mean score 1.9). One year after surgery patients scored highest in the walking category (3.6) and lowest in function (2.0).

Periprosthetic adaptive bone remodeling

During the first year after surgery bone loss was seen in all four ROI around the tumor prosthesis ending with a statistically significant (using paired t-test) decrease in BMD after 1 year of 8-15% (**Fig. 18**) (**Table 4**). The bone loss after 1 year was most pronounced (14-15%) in the 2 ROI closest to the TM collar and lowest (8%) adjacent to the tip of the stem (ROI 4) (**Table 4**).

Bone remodeling around the ankles

After 3 months the BMD decreased by 6% (p=0.008, paired t-test) in the operated ankle followed by a temporary plateau after 6 months and finally at 1 year of follow up the BMD loss in the operated ankle reached 9% below baseline ($p<0.001$, paired t-test). We found an initial minor decrease of 2% (p=0.12, paired t-test) in BMD in the non-operated ankle after 3 months and it stayed approximately at that level throughout the study period (**Fig. 19**) (**Table 4**).

Table 4.

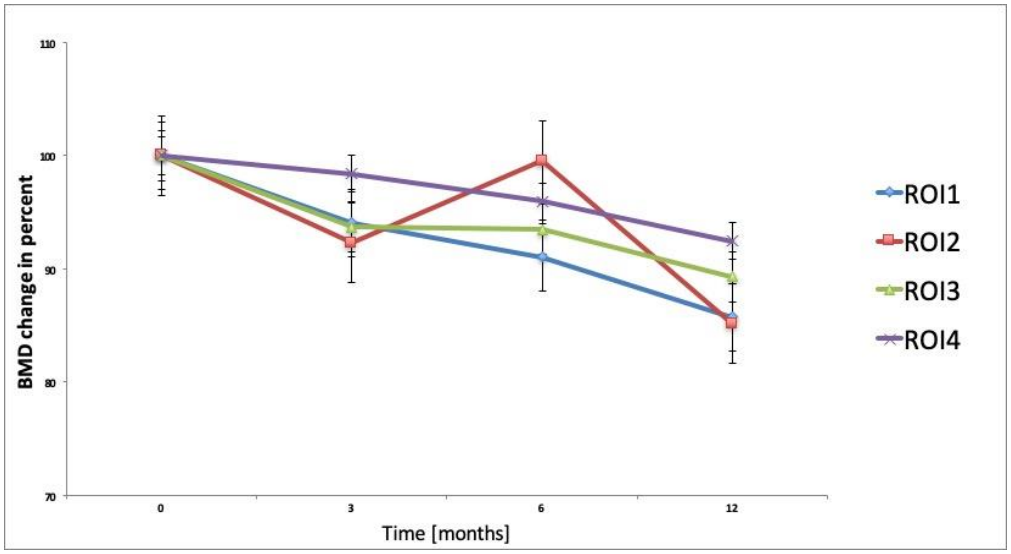
Mean (SD) BMD (g/cm2) and mean percentage changes in the 4 ROIs around the stem and in both ankles (operated and non-operated contralateral legs).

Follow-up	Postoperative (n=21)	3 months (n=18)	6 months (n=21)	12 months (n=21)	p-value[#] 0-12 months (n=18)
ROI1, BMD	2.186 (0.38)	2.056 (0.48)	1.990 (0.46)	1.874(0.27)	0.037
ΔBMD%		-6%	-9%	-14%	
p values (stepwise)*		0.285	0.092	0.004	
CI(95%)		(-0.10-0.33)	(-0.04-0.43)	(0.11-0.52)	
ROI2, BMD	2.248 (0.41)	2.075 (0.53)	2.238 (0.57)	1.914 (0.30)	0.071
ΔBMD%		-8%	-0.4%	-15%	
p values (stepwise)*		0.366	0.95	0.003	
CI(95%)		(-0.15-0.37)	(-0.28-0.30)	(0.13-0.54)	
ROI3, BMD	2.215 (0.43)	2.075 (0.49)	2.071 (0.38)	1.978 (0.3)	0.223
ΔBMD%		-6%	- 7%	-11%	
p values (stepwise)*		0.438	0.117	0.005	
CI(95%)		(-0.16-0.35)	(-0.04-0.33)	(0.08-0.39)	
ROI4, BMD	2.080 (0.42)	2.047 (0.44)	1.948 (0.45)	1.923 (0.45)	0.009
ΔBMD%		-2%	-4%	-8%	
p values (stepwise)*		0.356	0.079	<0.0001	
CI(95%)		(-0.04-0.11)	(-0.01-0.18)	(0.09-0.22)	
Ankle operated, BMD	0.751 (0.15)	0.7048 (0.14)	0.7049 (0.17)	0.681 (0.16)	<0.001
ΔBMD%		-6%	-6%	-9%	
p values (stepwise)*		0.008	0.008	<0.001	
CI(95%)		(0.01-0.06)	(0.02-0.09)	(0.05-0.11)	
Ankle contralateral, BMD	0.806 (0.19)	0.793 (0.18)	0.814 (0.25)	0.788 (0.17)	0.322
ΔBMD%		-2%	+1%	-2%	
p values (stepwise)*		0.12	0.90	0.12	
CI(95%)		(-0.01-0.05)	(-0.08-0.07)	(-0.01-0.05)	

*** students paired t-test (compared to postoperative value), [#]repeated measures ANOVA**

Fig. 18

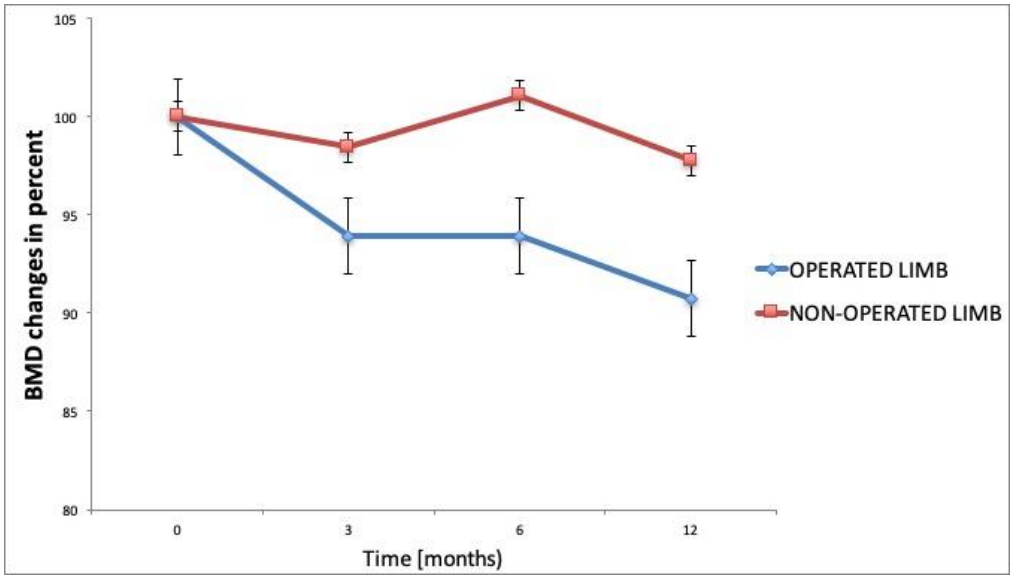
Mean (SE) BMD changes in percent of the 4 ROI.



Mean BMD changes in percent in the four ROI after surgery with 12 months of follow-up.

Fig. 19

Mean (SE) percentage BMD changes in the ankles, operated versus non-operated limb.



Mean BMD changes in percent of the ROI operated- and non-operated ankle after surgery and after 3, 6 and 12 months.

Study IV

External validation of prior BBN model

External validation of the BBN 1-year prediction model yielded poor discriminatory ability with an AUC ROC of 68% (95%CI, 62%-73%) (**Fig.25**), and hence the ability of the model to discriminate between survival and not survival is insufficient when based on this Danish population. Overall model performance evaluated with the Brier score was 0.122 (CI95%: 0.102-0.141).

Development and internal validation of the GBM model

Internal validation by AUC ROC analysis yielded well discriminatory ability with 75% (CI95%: 70%-80%) (**Fig. 26**). The Brier score for overall model performance was 0.09 (CI95%: 0.077-0.11). DCA demonstrated a positive net-benefit for probability thresholds above 0.5. (**Fig. 27**). However, at threshold probabilities below 0.5 the surgeon gains more benefit assuming that all patients are alive. Net benefit was capped at 85% (patients alive after one year).

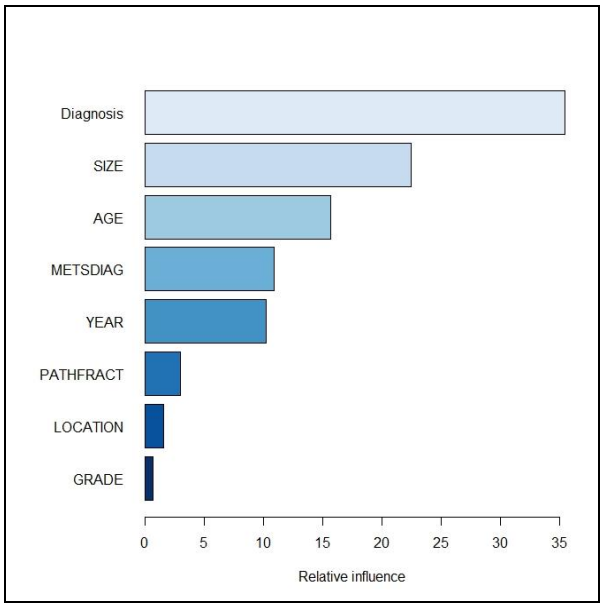
Features ranking highest in variable importance were: Diagnosis, Tumor size and Age (**Fig. 24**).

External validation of the GBM model

External validation of the GBM model yielded poor discriminatory ability with an AUC of the ROC of 63% (CI95%: 57%-68%) (**Fig.28**). The Brier score was 0.14 (CI95%: 0.12-0.16).

Fig. 24

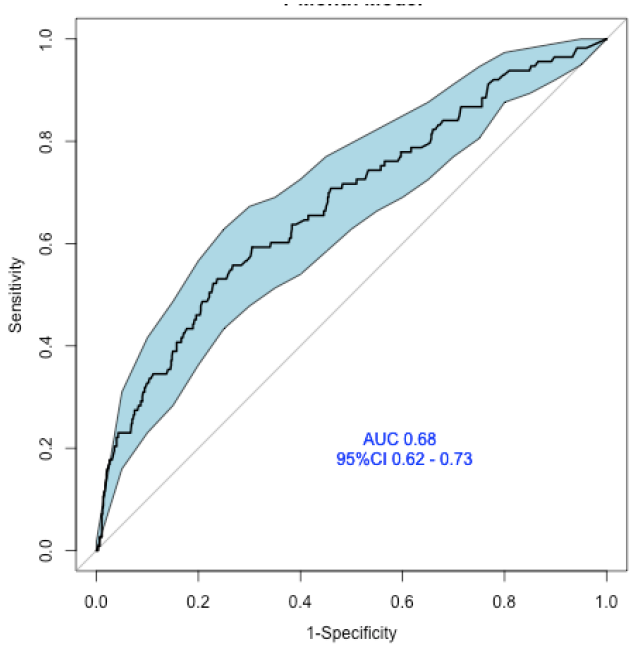
Relative influence of confirmed features



By shuffling copies of all features the chosen Boruta algorithm trains a Random Forest on the overall data. Features are then rejected or confirmed. Confirmed features are ranked with their relative influence in the GBM model as demonstrated.

Fig. 25

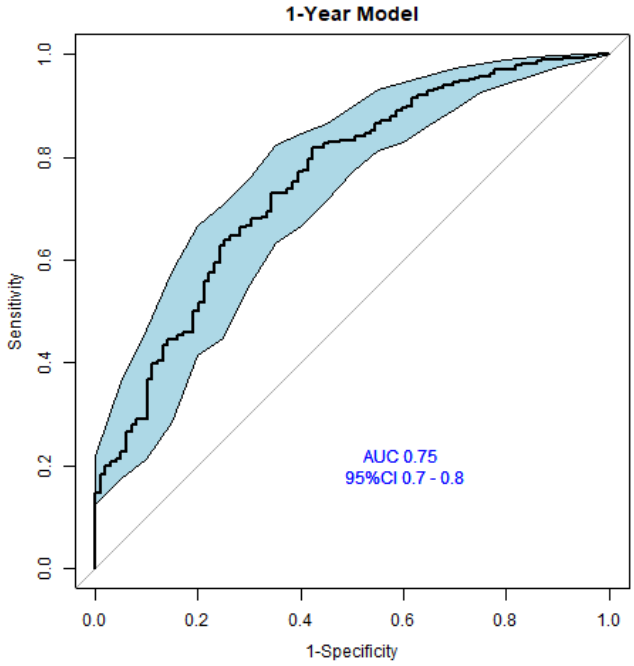
Receiver operating characteristics 1-year survival External validation BBN prediction model



ROC curves of the external validation of the1-year survival BBN model. The discriminatory accurary of the BBN model for survival yielded poor power (0.68).

Fig. 26

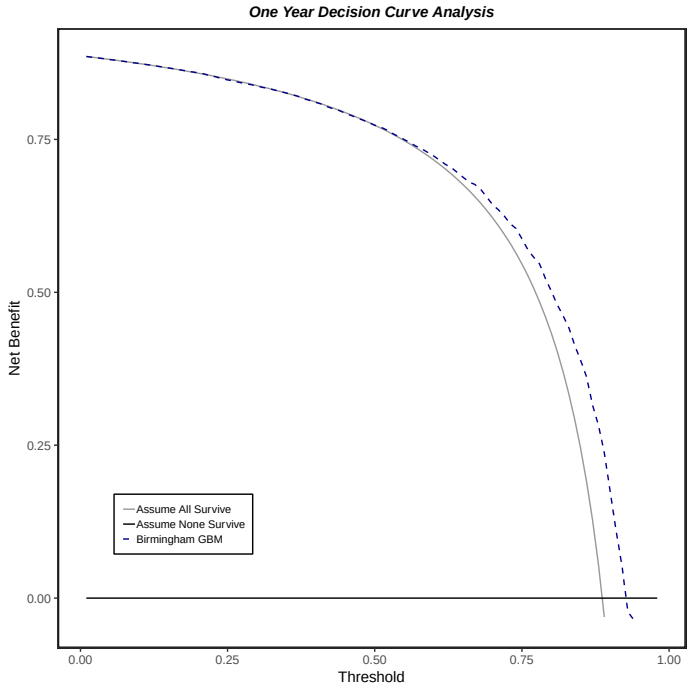
Receiver operating characteristics, 1-year survival
Internal validation GBM prediction model



ROC curves of the internal validation of the 1-year survial GBM model. The discriminatory accuracy of the GBM model for survival was classified as good (0.75).

Fig. 27

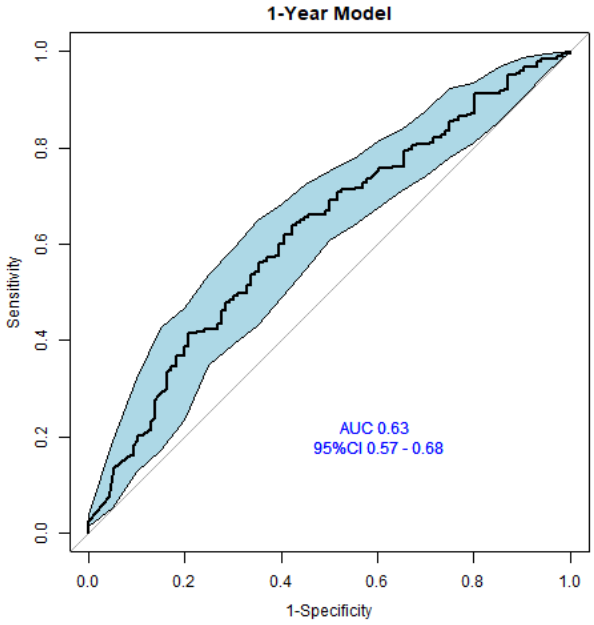
Decision curve analysis graph,
Internal validation GBM prediction model



Net benefit plotted on the decision curve analysis graph against threshold probabilities demonstrating the benefit of intervention based on decision to treat from model output. The curve demonstrates a net-benefit if using the model at thresholds above 0.50 compared to assuming all patients survive. For thresholds below 0.50 the model is no better or no worse that assuming all patients will survive.

Fig. 28

Receiver operating characteristics 1-year survival
External validation GBM prediction model



ROC curves of the external validation of the 1-year survival BBN model.
The discriminatory accurary of the GBM model for survival yielded poor power (AUC: 0.63).

DISCUSSION

Methodological considerations

General considerations

Reviewing the literature, the amount of studies and reports in the field of orthopedic oncology and bone sarcomas is substantial. However, a considerable amount of reported results origins from retrospective observational studies or systematic reviews (51,60,62,84,115). To obtain a genuine association between an intervention and an outcome the preferred method of choice would be a randomized controlled trial (RCT). Since the treatment of bone sarcomas most often is individualized in terms of surgical options, timing and adjuvant treatment modalities, the feasibility of a surgical interventional RCT study is limited. Furthermore, due to the rareness of bone sarcomas it would be challenging to conduct an RCT with a sufficiently powerful sample size within a reasonable time limit. The PARITY investigators (116) recently finished inclusion of patients in a multicenter RCT evaluating the perioperative duration (24 hours versus 5 days) of prophylactic antibiotic treatment after insertion of a tumor prosthesis in the lower extremities. Follow-up has, however, not yet been completed.

For these causes we chose to address the objectives of study I and II by conducting studies of observational retrospective design despite the inherent limitations. By conducting observational retrospective studies, we will not be able to provide sufficient scientific evidence of causality between the intervention and outcomes of interests and hence the provided results in study I and II will be speculations of relationships or associations. However as stated by Song et al. (117), well designed observational studies can provide valid results addressing important clinical research questions as suggested in study I and II. In addition, conducting a retrospective study allows immediate access to all data, and all available patients of interest are included.

Study III is an uncontrolled prospective clinical trial. Depending on study design, prospective trials are often considered superior to observational studies. However, by the absence of control group the results will provide weak evidence of causality (118), which was also not intended. Nevertheless, as stated by Sacca (119) some uncontrolled trials may be regarded as scientifically valid depending on the outcome and given that our aim was to evaluate the adaptive periprosthetic bone remodeling over time without statistical testing of our hypothesis, we believe that the chosen method was suitable (120). Furthermore, the prevalence of uncontrolled trials is

higher among oncology/hematology where feasibility is often a limitation for RCT's (119) as discussed above. We therefore believe that the present design is acceptable for the purpose.

Registry data is an option to obtain large amounts of uniform data when conducting retrospective studies. However, the risk of incomplete, inaccurate and inconsistent data is considerable, as the data collection method is not controlled. Also, registry data may be limited by the predetermined availability of outcomes or variables of interest and choosing variables by their availability may be a limitation. For the purpose of the objectives in study IV, data comprising the validation cohort was collected from DSR. DSR is continuously validated through The Danish Cancer Registry in accordance to applicable rules of RKKP (121), hence overcoming some of these limitations.

Registry data comprising the training cohort was originally described by Nandra et al. (94). Data was obtained from the institutional prospectively maintained database at The Royal Orthopaedic Hospital, Birmingham UK, that dates back over 30 years ago. Although prospectively collected, data has to our knowledge not been validated. Also, since data comprises newly diagnosed patients from a time period (1970-2012) where the diagnostic process has advanced significantly, data may suffer from inaccuracy or may be even missing as indicated by the high percentage (51%) of missing data for the variable tumor size (**table 3**).

A validation process as aimed in study IV requires identical variables in the training and the validation set (122). Besides Alkaline phosphatase, all included features of the prior BBN model by Nandra et al. (94) were accessible from the DSR database for the external validation. Furthermore, for the purpose of comparing the BBN model with the GBM model, it was predefined to use the same variables for the train and external validation process despite the limitation of the one missing variable.

Confounding

A confounding factor correlates with both the exposure and the outcome without being causally related. Furthermore, confounders may be distributed unequally between cohorts subjected to a certain exposure. Observational studies (study I and II) are often influenced by unpredictable and also unknown confounding factors. Potential sources of confounders for revision could be stage at time of surgery, differences in complexity of operation, comorbidity as well as level of function prior to surgery, however, these data were not collected for the purpose of study I and II. To limit confounding an RCT is well suited, as chance will balance possible confounders. The inclusion of patients with soft tissue sarcomas adjacent to the bone in study II as well as patients with MBD in study III may further introduce confounding when comparing to the cohort in study I.

The chosen statistical analysis used in study I and II did not permit accounting for confounders, and even if adjusting for confounding the resulting estimates will be with some degree of uncertainty due to residual confounding. Hence, unmeasured and unknown confounders should be considered when interpreting the results.

Moreover, as oppose to conventional statistical approaches with thoroughly controlled data to mitigate bias, emerging data sources for machine learning are often less structured, without taking confounding into account when assuming causal inference of the used features in the training set (98).

Furthermore, when evaluating prediction models by measures of discrimination, confounding may deviate the location of the ROC curve and hence over- or underestimate the accuracy of discrimination by the model. A way to adjust for confounding could be to stratify the ROC curve (e.g. 60<age>60) (110). We did not collect data on measured or unmeasured data or test for cofounding due to the chosen methods and hence we chose to accept the limitations of the interpretations.

Bias

Study I and II are based on retrospective data and hence susceptible to selection bias.
However, data in both studies comprised consecutive included cohorts from a single Center with complete follow-up of all patients due to the Danish Civil Registration System (102) and should thus not be subjected to selection bias in this matter.
In study III, intervention is not randomly assigned and since the cohort is not consecutive, selection bias is present. Also, attrition bias (loss to follow up) may occur in long-term prospective follow-up studies. However, in patients with malignancies, attrition may occur sooner. The inclusion of patients suffering from MBD in study III may introduce a greater risk of loss to follow up caused by death or attrition. To mitigate attrition bias, a follow-up period of only 1 year was used in study III.

Registry data is prone to information bias due to misclassification. The validation cohort in study IV comprised several objective variables (tumor size, grade, histologic diagnosis), which may introduce this bias. Further, by converting the histologic grade variable from the DSR for the purpose of equality between the train and validation set, we might have added further information bias to the final model. However, due to the completeness of DSR and since data is collected prospectively, information bias is to be considered low.

Limitations

Due to the chosen design we did not perform power calculation in any of our studies. In study I, II and IV we included all patients available in all our studies and our cohorts should therefore in principle be representative to the population of interest. This does however not exclude the risk of underpowered studies and risk of type II errors when comparing outcomes from study I and II. We did not find significant differences between the cohorts in study I/II, which could be caused by a type II error as indicated by the wide confidence intervals.
Also, given that patients in study I-II alone comprised patients considered eligible to LSS with tumor prostheses, and that some of these patients may have passed away during neo-adjuvant treatment, the estimated survival probabilities in study I and II are less comparable to studies reporting general probability of overall survival in patients with suffering from bone sarcoma.
Furthermore, when estimating overall survival in study I and II we did not explore in cause of death and hence may have underestimated overall survival since patients may have died of other causes than their malignancies.

No standards of practice exist in the design of machine learning models as illuminated by Balki et al. (123). Large sample sizes has previously been recognized to be the single biggest influence on design and performance of models together with the rule of thumb with ten events per predictor parameter of interest as also recognized for classic statistic prediction models, although the latter has been subject for discussion (123–127). Chen et al. suggest that a small modern cohort is superior to historic larger sample sizes when using machine learning (98). This is supported by Park et al. (128) suggesting that robust validation of a model depends on an adequately target population, preferably prospective. Few attempts have been made to overcome the lack of knowledge in identifying an adequate sample size for machine-learning prediction models (125) and due to our results it is speculative if the Danish population cohort used for validation of the BBN and GBM model was adequate for validation in terms of sample size and events per variable although the cohort was modern with limited missing data.
Also, in Machine-learning a variable is often valuable if it improves the prediction ability of the model although it may not have proven to be an independent prognostic or predictive factor.

Finally, the cohort used to train the GBM model and the prior BBN model comprised data ranging from 1970-2012 and the significant improvements in overall survival during those years may have skewed model outcome into poorer overall survival as well as yielding poorer generalizability when validated on a modern cohort.

COMPARISON OF MAIN FINDINGS

Overall survival

In study I and II we found similar probabilities of 5- and 10-year overall patient survival and hence no statistically significant differences between the cohorts were identified. Our results were comparable to previous findings of similar studies, with reported survival rates of 50%-70% (10,15,49,51,57,129–131). Stratifying for age, Mirabello et al. (23) demonstrated that children with osteosarcoma had a 5-10% better overall survival than patients up to 50 year. Passing 60 years survival rate decreased to 24% and further to 11% for patients older than 80 years (10,86,132). Due to our sample size we did not stratify for age.

Good chemotherapy response, defined as tumor necrosis > 90%, has been the most consistently reported prognostic factor for overall survival in patients with osteosarcoma and Ewing sarcoma (15,31,49,85,87,133–136). However, as illuminated in several studies, the continuous development in chemotherapy and improved overall survival seems to have reached a plateau since the 1980's (10,49,131). Anninga et al. (10) remained inconclusive when attempting to address whether or not intensification or change of neoadjuvant chemotherapy improves survival in poor responders with osteosarcoma. In a review Ferguson et al. (15) did the same findings and due to the reported efficiency of neoadjuvant chemotherapy Jaffe et al. (129) initiated a trial treating sarcomas exclusively with chemotherapy. The study did however not justify the exclusively utilization of chemotherapy.

Prediction of overall survival

For patients suffering from osteosarcoma, several prognostic factors for survival present at the time of diagnosis has been suggested (85). Viewing literature there is a broad consensus that the presence of metastases at diagnosis is the factor with greatest impact on prognosis (49,85–87). In study IV we found that the five features ranking highest of relative influence on bone sarcoma survival according to present GBM model were: histologic diagnosis, tumor size, age, metastasis at diagnosis, year of diagnosis.

Slade et al. (137) reported that 25% of patients with osteosarcoma under the age of 21 years had metastases at diagnosis with 50% of the metastases being pulmonary and furthermore, that patients with extra-pulmonary metastatic disease had significantly poorer survival compared to patients with metastasis limited to the lungs (137). Other suggested factors such as tumor size, alkaline phosphatase, tumor site, histologic subtype and gender has consistently been reported as prognostic factors for survival (85–87,133–136,138). However, as stated in a systematic review

by Bramer et al. (85), strong unsuspected prognostic factors may not become significant when attempting to evaluate prognostic factors in small underpowered sample sizes.

Hence, uncertainty about independent prognostic factors still exists and prognostication of individual patients is a remaining clinical concern. The effect of chemotherapy has been reported to be one of the most reliable prognostic factors (87) although not available at time of diagnosis. A survival prediction model would therefore be a considerable aid for the clinician, when it is time to plan the appropriate individual treatment at the time of diagnosis, and this was the reason for us to do study IV. The overall purpose of prediction is to allocate patients to be optimal and duly treated in order to achieve the best outcome and ultimately improve survival.

Only a few attempts to create prediction models for survival in bone sarcoma patients using machine learning techniques has been suggested (93–95). As oppose to Bongers et al. (95) who successfully externally evaluated a 5-year survival prediction model for chondrosarcomas, we did not succeed in externally validating the BBN or the GBM model on the Danish population cohort and hence none of the models are suitable for clinical use based on present externally validations. Since both BBN and GBM prediction models previously has demonstrated to be suitable for clinical use (101,139,140), we speculate that the use of only one histologic subgroup of bone sarcomas as performed by Quirina et al. (93) may have led to a greater extent of homogeneity in the training cohort and hence sufficient generalizability for successful external validation.

As proposed by Moons et al. (141), a model that does not perform well when externally validated should first be updated or revised. Revision includes modification of predictor-outcome associations and/or adding feature candidates for the train process, a process that is warranted for the present models in this thesis, bearing in mind the statement by Moons: Just because a model is good to predict does not mean it is useful clinically (142).

Towards target therapy treatment?

Recent studies evaluating cytogenetic analysis of sarcomas, has identified certain molecular aberrations associated with specific histological subtypes (6,20,143). The advances into the molecular and cytogenetic characterizations of sarcomas, has been suggested to improve diagnostic possibilities as well as identification of potential targeting therapy markers (2,3,6,20,22). Target therapy implies that a marker is expected to be a proven predictive factor hence also suitable as a variable for prediction models where most factors currently are prognostic (89). The identification of predictive factors is not straightforward. Although a lack in consensus on how to carry out clinical trials for identification of predictive factors it is commonly considered, that it is not possible to assess predictive significance of a potential predictive factor without a clinical trial including a control group (88,90–92,97), a challenging task due to the above mentioned limitations and methodological considerations in the field of orthopedic oncology and well cited by Hingorani et al. (91): patients testing negative for a given predictive factor may still benefit from the given treatment and should not automatically be excluded. Adding solid predictive factors in the revision or creation of future prediction models is warranted.

Limb survival

Secondary amputation after LSS is a devastating event for any patient and especially for patients with a severe course of illness and perhaps sparse residual life expectancy.

Various rates from 6% to 30% of secondary amputation after LSS has been reported (51,52,84). We found our results in study I (12%) and II (11%) comparable to literature. Interestingly, the causes of amputation in study I and II were different. In study I, various causes including deep infection led to amputation with only a slight predominance of local recurrences. In study II the main part of amputations were caused by local recurrence, and none by deep infection. Comparing to literature, the most common reported causes to amputation are deep infection and local recurrence (41,61,63,144–146) although Zeegen et al. (60) found equal distribution of amputation between infection and recurrence in their study. Likewise our findings in study I and II, other causes to amputation such as mechanical complications, severe functional limitations or vascular compromise are rarely reported (51,146,147). In a very long-term evaluation (mean follow-up 29 years), Grimer et al. (51) reported a decreased risk of local recurrence with time while the risk of deep infection was manifest. However, Jeys et al. (146) reported that 10% of all amputations were performed after 5 years, and were caused equal by local recurrence or infection. Since none of these studies consider death as a competing risk, the reversal findings

could be influenced by the possible lower survival of patients with local recurrence in long-term follow-up studies. By using competing risk analysis, we found an increased 5- and 10-year risk of amputation in study II (8% and 11% respectively), which is comparable to the sparse studies evaluating limb survival after LSS with tumor prostheses by using competing risk analysis (148,149). In study I we did not adjust for death as a competing risk but also found a decreased 5- and 10-year probability of limb survival.

Local recurrence

Risk of local recurrence causing amputation is associated with poor margins. However, there exist conflicting reports whether local recurrence in osteosarcoma patients is independently associated with overall survival separately from chemotherapy response (49). Since the patient survival in general has plateaued parallel with significant improvements in imaging and surgical techniques, the surgical improvements may not affect overall survival. As stated by Anderson (49) there has come to be an acceptance of higher rates of local recurrence in LSS compared to amputation. To what extent this affects overall survival is controversial. Bacci et al. (150) demonstrated that free margins and poor chemotherapy response yielded better overall survival compared to poor margins and good chemotherapy response. Bertrand et al. (130) reported inferior survival in patients with poor margins in a review comprising 241 patients treated between 1999 and 2001, and Ferguson et al. (15) stated that failure to achieve complete gross resection with free margins leads to high risk of local recurrence and poor overall survival. This is supported by Stevenson et al. (151) who demonstrated significant reduced risk of local recurrence with wider margins in chondrosarcoma grade I-III and furthermore that local recurrence in chondrosarcoma grade II and III is significantly associated with disease specific overall survival. In addition, Tsuda et al. (152) also demonstrated higher recurrence rate and mortality in patients with poor surgical margins of surgically treated peripheral chondrosarcomas of the pelvis. Since chondrosarcomas are considered chemoresistant, these findings support that local recurrence due to sufficient surgical margins, is independently associated with overall survival, although biological heterogeneity of the subtypes limits general considerations.

Nonetheless, Grimer et al. (153) reported that increased rates of local recurrence did not affect overall survival and thus advocated to choose LSS despite poor chemotherapy response. The same findings with increased local recurrence rates, without compromising overall survival was found by Myers et al. (61). In a recent systematic review Thornley et al. (62) reported that incidence of local recurrence was only reported in 5% of all patients and hence the previous reverse findings may be caused by varied and inconsistent reporting.

Deep infection

Various infection rates from 2% to 20% of primary tumor prostheses has been reported (41,59,115,154–156). We found comparable results in study I and II, with rates of 16% and 11% respectively.

Since rotating-hinge prosthesis has been taken into use, many studies now report deep infection as the most common cause for revision compared to previous reports of mechanical complications causing most revisions (61,75,77). The incidence of infection has been reported to continue in long-term follow-up studies (51,61). In study II we found that the risk of deep infection after 10 years plateaued which, however, might be caused by the mid-term follow-up in the study. Proximal tibia is the predominant location for deep infection and hence also the most common place for amputations caused by deep infection (146). This has been suggested to be associated to insufficient wound coverage after extensor mechanism reconstruction (57,59,64,75,82,146). The higher incidence of infections in the proximal tibia is confirmed by Puchner et al. (149) who demonstrated a 5- and 10-year cumulated incidence of infection of 17% and 22% respectively when evaluating LSS and reconstruction with tumor prostheses of the proximal tibia. By using competing risk analysis these results are fully comparable to present results of study II although demonstrating a fairly higher incidence of infections. However, the cohort by Puchner et al. (149) as well comprised patients suffering from MBD who are more prone to infections due to immune-suppression and a poor general health condition.

It is well-known that complete muscular coverage of the tumor prosthesis reduces the risk of infections that are related to wound dehiscence (157). The introduction of extensor mechanism reconstruction with medial gastrocnemius flap, has been reported to reduce infection rate as it allows adequate coverage of the prostheses as well as wound coverage (57,78,158,159). Furthermore, silver coated implants and iodine surface coatings has been suggested (57,160,161) and in some studies also showed promising results reducing infection rate (57,162–164). However, long-term follow-up will be needed for evaluation of metallic-coated prostheses in terms of documentation of adverse systemic side effects. Improved monitoring and early detection may as well reduce the amount of deep infections. Given that infection is one of the most common complications and one of the most common causes to amputation, deep infection is to be considered the current greatest threat for limb survival.

Implant survival

Aseptic loosening

Incidence of aseptic loosening has been reported with various rates from 2% to 11% (41,77). Rotating hinge prostheses has been reported to reduce the torsional stress in the bone-implant interface hence reducing aseptic loosening and stem breakage (57,106). Furthermore, by allowing multiple degrees of movement, rotating-hinge prostheses intends to reduce wear by dispersing stress throughout the condylar surfaces (165). Our findings suggest the same mechanical improvements of implants: In study I aseptic loosening constituted the vast majority of all revisions (23%). Although we did not identify any statistically significant difference, the demonstrated incidence of aseptic loosening in study II was markedly reduced (14%). Previous studies comparing rotating hinge with fixed hinge prostheses reports likewise decreased rates in aseptic loosening when comparing rotating hinge prostheses with fixed hinge prostheses (57,61,106,157,166,167). Although slightly lower, we found our results from study II fully comparable to Puchner et al. (149) who by competing risk analysis demonstrated the true incidence of aseptic loosening of 6% and 16%. By comparison between rotating-hinge and fixed hinge-prostheses Puchner et al. (149) did not demonstrate significant differences, which could be caused by type II error due to the low sample size as in present study II.

It is well known that aseptic loosening is predominant in distal femur prostheses (57,75–77) which is suggested to be due to the quadriceps excision and the passive extension gait provided by the prosthesis with no support from quadriceps contraction (72–74). Likewise, in both our first and second-generation prostheses distal femur was the most common site for aseptic loosening followed by proximal tibia. Since bone sarcomas most often arise in the distal femur this represents a substantial clinical concern.

Use of cemented stems has previous been associated to aseptic loosening (157). The use of cementless stems has therefor intended to mitigate aseptic loosening by fascilitating bony ingrowth combined with the use of anatomically curved stems. Curved stems are known to have higher torque resistance than straight cemented stems and hence possibly reduce loosening (64,82,156,168–170). The role of fixation with regards to aseptic loosening remains however unknown (82,170,171). Pala et al. (172) compared cemented fixations with uncemented fixation in the lower extremitites and found no difference in incidence of aseptic loosning with a mean follow-up of 28 months . This is supported in a recent systematic review by Haijie et al. (84) who concluded that method of fixation is not a determinant with regards to aseptic loosening. As such, we consider our results from study I and II where we used various types of fixation, comparable with literature (82,156,173,174).

Adaptive periprosthetic bone remodeling

In study III we demonstrated that adaptive periprosthetic bone remodeling due to stress shielding and immobilization is associated with progressive bone loss and hence represents an increased risk of periprosthetic fractures, and potentially more complex surgery in case of later revision surgery. Although a few attempts has been made to investigate adaptive periprosthetic bone remodeling after insertion of tumor prostheses (175,176), we are to the best of our knowledge, the first to quantitatively measure the adaptive bone remodeling in a prospective population of this sample size with 1-year follow up, and there exist no previously reported longitudinal results of the periprosthetic bone remodeling after resection and reconstruction with the cemented Zimmer® Segmental tumor prosthesis.

Likewise our findings, Lan et al. (177) and Andersen et al. (176) found increased reduction in BMD with lesser distance to the part of the stem connecting to the reconstruction with extension pieces and the joint, corresponding to the Gruen Zones 1, 2, 6, and 7, when measuring bone remodeling around primary hip stems (178,179). This pattern in bone remodeling along the stem is also well-described after both cemented and uncemented primary hip arthroplasty (178,180–182). However, the evaluation of BMD changes over time by Lan et al.(177) was based upon measurements in one selected ROI using the contralateral leg as reference which strongly limits comparison. Vennesma et al. (179) showed that the affected limb should always be used as reference and that patients must be followed prospectively in order to obtain exact BMD changes after surgery. In addition, Kröger et al. (183) demonstrated local differences in BMD between the affected and the non-affected limb and therefore stated that BMD measurements years after surgery are invalid, with regard to true bone loss, when comparison is made with the contralateral limb.

Several studies have reported significant periprosthetic loss in BMD around the cemented and uncemented femur stem within the first 3 months after primary hip arthroplasty followed by an increase or plateau in BMD after 6 months (179,183). Reverse to these findings, our results indicate a progressive remodeling and loss in BMD after one year. Even though we used cemented fixation which allows immediate weight bearing for all prostheses, we suggest that the demonstrated progressive bone remodeling after 1 year partly is a reflection, of the well-known prolonged rehabilitation after insertion of a tumor prostheses. Furthermore, loss of bone stock in relation to chemotherapy is well-described (65) as well as the age-related decay in BMD (184) which are factors that further could have affected the demonstrated results in study III. Nonetheless, we found a relatively slow decrease in BMD during the first year after surgery in all our ROI, and we speculate that this partly could be explained by the intended fixation of the

TM collar to bone with less load transfer directly to the tip of the stem and hence reduced stress shielding adjacent to the joint.

The precision error of BMD measurement in study III expressed as the CV was 2%-5%, and thus comparable to previous findings when measuring BMD around cemented hip and knee arthroplasties (180,183,185). The demonstrated CV by Andersen et al. (176) was slightly lower when evaluating the uncemented proximally hydroxyapatite-coated femur stem. This is likely explained by the bone-cement interface included in our measurements. Lan et al. (177) found CV comparable to ours although they evaluated uncemented stems. However, their measures are based upon smaller ROI, and Gehrchen et al. (186) demonstrated that small ROI-size is associated with poorer precision.

The decrease in BMD of the affected ankle after 1-year was 9% and the non-affected ankle was close to baseline (2%). These findings indicate that the periprosthetic BMD changes during follow-up are caused by stress shielding combined with immobilization and to a lesser extend a general decrease in BMD.

Functional outcome

Although hypothesized, we did not find improved MSTS score when comparing the cohort in study I and II. Study I comprised the time period when LSS with tumor prostheses gradually substituted amputation. Hence, LSS with tumor prostheses was mainly offered to patients with long residual life expectancy and high functional status. Also, given that functional outcome after LSS is directly related to the amount of functional muscle preserved (157), we speculate that the lack of improvement in functional outcome with second-generation prostheses partly reflect that LSS was offered to the vast majority of bone sarcoma patients including those with severe stages and possibly need of greater magnitude of resection to achieve wide resection. Puchner et al. (149) demonstrated a MSTS of 82% with a mean follow-up of 1-359 month. The vast majority of their cohort comprised patients under the age of 30, which could explain the relatively high functional outcome score. They too did not find any differences in functional outcome when comparing fixed- and rotating hinge prostheses.

In study III the MSTS score yielded poorer outcome when compared to previous studies (77,187,188). Harvey et al. (187) evaluated functional outcome in patients with MBD, who underwent LSS with resection and reconstruction with a tumor prosthesis or intramedullary fixation. With a mean follow-up of 16 months they found a mean MSTS score of 24 for patients treated with tumor prostheses. However, as bone metastases most often arise in the proximal

femur, the study only included proximal femoral reconstructions. It is well known that primary arthroplasty around the knee requires prolonged rehabilitation compared to hip arthroplasty. Also, the quadriceps incision when performing distal femur tumor arthroplasty further prolongs rehabilitation. Given that 48% (n=10) of the implants in study III were inserted in the distal femur (n=8) or proximal tibia (n=2), we speculate that the site of reconstruction partly reflects the present moderate functional outcome level. In addition, resection of MBD can be accomplished by marginal resection as oppose to resection of sarcomas with intended wide excision and free margins. Hence bone sarcoma patients undergoing LSS with tumor prostheses are more exposed to tissue loss and impaired functional outcome. We therefor suggest that the unchanged poor outcome in study III is mainly caused by the required prolonged rehabilitation throughout the first year after surgery and to a lesser extent reflects the expected functional outcome for long-term survivors.

PERSPECTIVES AND FUTURE RESEARCH

As illuminated throughout this thesis the rareness as well as the biological heterogeneity of bone sarcomas is a considerable limitation to achieve sound evidence-based conclusions for improvements in management and ultimately overall survival.
Methodologically high-quality studies with uniform study designs and consistent uniform reporting of implant failures, preferably in international collaborative databases or multicenter studies to achieve sufficient sample sizes overcoming type II errors, is pending. The use of competing risk analysis is advisable to achieve the true risk of recurrence, revision or amputation.

Improvement of overall patient survival must be of highest priority for further research in the field of orthopedic oncology. Moving towards individualized evidence-based treatment, the development of validated prediction models is suggested to ultimately improve overall survival. Perhaps prediction models for each main subtype of bone sarcoma will be more feasible, to obtain generalizability? Nevertheless, the potential improvement of any prediction model is to be evaluated in an impact study (141). Also, detection of evidence-based predictive factors for target therapy should aim to restrict therapy, perhaps also for LSS?

With the continuous advances in imaging and surgical techniques including vascular reconstruction and use of free flaps, patients previously considered unfit for LSS due to unresectable tumors can now be offered LSS, thus an increased number of patients having LSS with tumorprostheses is to be expected. Numerous studies including this thesis has unveiled causes associated with revisions, implant failures and amputations. Future studies will need to aim on how to prevent the adverse events after LSS.

CONCLUSION

We identified the true risk of implant and limb survival associated with LSS and reconstruction with first-generation fixed-hinge and second-generation rotating-hinge tumor prostheses. By consistent reporting and follow-up, our results from a highly specialized center, justifies the use of LSS and reconstruction with tumor prostheses, and acceptable functional outcome, although long-term survivors must expect repeated revisions. Comparison of first- and second-generation tumor prostheses demonstrated no statistically differences in terms of implant- and limb-survival, presumably due to type II error as indicated by wide confidence intervals and patient group heterogeneity. Furthermore, identification of overall survival in a population-based cohort of sarcoma patients, demonstrated no changes in overall patient survival over time, as also reported in literature.

Quantitative measurements of the adaptive periprosthetic bone remodeling around the cemented Zimmer® Segmental stem demonstrated significant decrease (8%-15%) in BMD in all four ROI's after 1–year, most likely caused by a combination of immobilization and stress shielding as indicated by the decrease in BMD of 9% in the affected ankle 1-year after surgery.

Lastly, we can conclude that the 1-year Bayesian belief network survival model proposed by Nandra et al. is not recommendable for clinical usage based on a Danish population cohort validation. The successfully created GBM 1-year prediction model did not outperform the prior BBN model and revision is pending. The authors are obligated to continue the ongoing work with modernization of a model advisable for clinical use.

REFERENCES

1. Kubo T, Furuta T, Johan MP, Ochi M. Prognostic significance of 18 F-FDG PET at diagnosis in patients with soft tissue sarcoma and bone sarcoma; Systematic review and meta-analysis. Eur J Cancer. 2016;58:104–11.
2. Helman LJ, Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer. 2003;3(9):685–94.
3. Lahat G, Lazar A, Lev D. Sarcoma Epidemiology and Etiology: Potential Environmental and Genetic Factors. Surg Clin North Am. 2008;88(3):451–81.
4. Lewin J, Puri A, Quek R, Ngan R, Alcasabas AP, Wood D, et al. Management of sarcoma in the asia-pacific region: Resource-stratified guidelines. Lancet Oncol. 2013;14(12):e562–70.
5. Jørgensen, Peter Holmberg, Lausten Gunnar Schwarz PAB. The Danish Sarcoma Database. Clin Epidemiol. 2016;8:685–90.
6. Ducimetière F, Lurkin A, Ranchère-Vince D, Decouvelaere AV, Péoc'h M, Istier L, et al. Incidence of sarcoma histotypes and molecular subtypes in a prospective epidemiological study with central pathology review and molecular testing. PLoS One. 2011;6(8):e20294.
7. Dorfman HD, Czerniak B. Bone Cancers. CANCER Suppl. 1995;75(1):203–10.
8. World Health Organization WHO, Fletcher C, Bridge JA, Hogendoorn PCW, Mertens F. WHO Classification of Tumours of Soft Tissue and Bone : WHO Classification of Tumours, vol. 5. 2013.
9. SEER. Cancer Epidemiology in Older Adolescents and Young Adults 15 to 29 Years of Age. Cancer. 2000;1975–2000.
10. Anninga JK, Gelderblom H, Fiocco M, Kroep JR, Taminiau AHM, Hogendoorn PCW, et al. Chemotherapeutic adjuvant treatment for osteosarcoma: Where do we stand? Eur J Cancer. 2011;47(16):2431–45.
11. Jo VY, Fletcher CDM. WHO classification of soft tissue tumours: An update based on the 2013 (4th) edition. Pathology. 2014;46(2):95–104.
12. Raut CP. Tumor biological aspects of epithelial versus mesenchymal tumors of the gastrointestinal tract. Visc Med. 2018;34(5):342–6.
13. Rajiv Rajani GPC. Treatment of Bone tumors. Surg Pathol Clin. 2012;5:301–18.
14. Grimer R, Athanasou N, Gerrand C, Judson I, Lewis I, Morland B, et al. UK guidelines for the management of bone sarcomas. Sarcoma. 2010;2010.
15. Ferguson WS, Goorin AM. Current treatment of osteosarcoma. Cancer Invest. 2001;19(3):292–315.
16. Casali PG, Bielack S, Abecassis N, Aro HT, Bauer S, Biagini R, et al. Bone sarcomas: ESMO-PaedCan-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29(August):iv79–95.
17. Sybil Biermann J, Adkins DR, Agulnik M, Benjamin RS, Brigman B, Butrynski JE, et al. Bone Cancer: Clinical practice guidelines in oncology. JNCCN J Natl Compr Cancer Netw. 2013;11(6):688–723.
18. Jeys L, Morris G, Evans S, Stevenson J, Parry M, Gregory J. Surgical Innovation in Sarcoma Surgery. Clin Oncol. 2017;29(8):489–99.
19. Rosenberg AE. Bone Sarcoma Pathology: Diagnostic Approach for Optimal Therapy. Am Soc Clin Oncol Educ B. 2017;37:794–8.
20. Franchi A. Epidemiology and classification of bone tumors. Clin Cases Miner Bone Metab. 2012;9(2):92–5.
21. Eyre R, Feltbower RG, James PW, Blakey K, Mubwandarikwa E, Forman D, et al. The epidemiology of bone cancer in 0 - 39 year olds in northern England, 1981 - 2002. BMC Cancer. 2010;10(357).
22. Xiao W, Mohseny AB, Hogendoorn PCW, Cleton-Jansen A-M. Mesenchymal stem cell transformation and sarcoma genesis. Clin Sarcoma Res. 2013;3(1):10.
23. Mirabello L, Troisi RJ, Savage SA. International osteosarcoma incidence patterns in children and adolescents, middle ages and elderly persons. Int J Cancer. 2009;125(1):229–34.
24. Sundhedsstyrelsen. Pakkeforløb for sarkomer i knogle og bløddele. 2012. 51 p.
25. Keller J. Sarkomer og aggressive benigne tumorer Referenceprogram udarbejdet under Dansk Sarkomgruppe (DSG). 2013; Available from: https://www.ortopaedi.dk/fileadmin/sarkom/2013-referenceprogram-godkendt-i-2023.pdf
26. Behandling af sarkomer og aggressive , benigne tumorer Behandlingsvejledning. 2007;(August).
27. Coleman R, Body JJ, Aapro M, Hadji P, Herrstedt J, Group W. clinical practice guidelines Bone health in cancer patients : ESMO Clinical Practice clinical practice guidelines. Ann Oncol. 2014;25(March):124–37.
28. Grimer RJ, Sneath RS. Diagnosing malignant bone tumours. J Bone Jt Surg - Br Vol. 1990;72(5):754–6.
29. Grimer RJ, Briggs TWR. Earlier diagnosis of bone and soft-tissue tumours. J Bone Jt Surg - Br Vol. 2010;92-B(11):1489–92.
30. Maguire M. Early symptoms of bone and soft tissue sarcomas: Could they be diagnosed earlier? Ann R Coll

Surg Engl. 2012;94(6):451.
31. Gibbs CP, Weber K, Scarborough MT. Malignant Bone Tumors. J Bone Jt Surg - Am Vol. 2001;83(11):1728–45.
32. Eary JF, Conrad EU. Imaging in sarcoma. J Nucl Med. 2011;52(12):1903–13.
33. Benz MR, Tchekmedyian N, Eilber FC, Federman N, Czernin J, Tap WD. Utilization of positron emission tomography in the management of patients with sarcoma. Curr Opin Oncol. 2009;21(4):345–51.
34. Becher S, Oskouei S. PET Imaging in Sarcoma. Orthop Clin North Am. 2015;46(3):409–15.
35. Roitman PD, Farfalli GL, Ayerza MA, Múscolo DL, Milano FE, Aponte-Tinao LA. Is Needle Biopsy Clinically Useful in Preoperative Grading of Central Chondrosarcoma of the Pelvis and Long Bones? Clin Orthop Relat Res. 2017;475(3):808–14.
36. Jennings R, Riley N, Rose B, Rossi R, Skinner JA, Cannon SR, et al. An Evaluation of the Diagnostic Accuracy of the Grade of Preoperative Biopsy Compared to Surgical Excision in Chondrosarcoma of the Long Bones. Int J Surg Oncol. 2010;2010:1–4.
37. Laitinen MK, Stevenson JD, Parry MC, Sumathi V, Grimer RJ, Jeys LM. The role of grade in local recurrence and the disease-specific survival in chondrosarcomas. J Bone Jt Surg - Br Vol. 2018;100B(5):662–6.
38. Oliveira AM, Nascimento AG. Grading in soft tissue tumors: Principles and problems. Skeletal Radiol. 2001;30(10):543–59.
39. Costa J, Wesley RA, Glatstein E, Rosenberg SA. The grading of soft tissue sarcomas results of a clinicohistopathologic correlation in a series of 163 cases. Cancer. 1984;53(3):530–41.
40. Coindre J. Grading of Soft Tissue Sarcomas. Arch Pathol Lab Med. 2006;130(October):1448–53.
41. Ahlmann ER, Menendez LR, Kermani C, Gotha H, Surgeon Kermani OC, Fellow Gotha RH. Survivorship and clinical outcome of modular endoprosthetic reconstruction for neoplastic disease of the lower limb. J Bone Jt Surg - Br Vol. 2006;88:790–5.
42. Enneking William F., Spanier S. Suzanne GMA. A system for the surgical staging of musculoskeletal sarcoma. Clin Orthop Relat Res. 1980;153.
43. Moore AT, Bohlman HR. Metal hip joint: a case report. 1942. Clin Orthop Relat Res. 2006;453:22–4.
44. Marcove RC, Lewis MM, Rosen G, Huvos AG. Total Femur and Total Knee Replacement. Clin Orthop Relat Res. 1977;126(Jul-Aug):147–52.
45. Rosen G., Murphy L., Huvos A. G., Gutierrez M. MRC. Chemotherapy, en bloc resection, and prosthetic bone replacement in the treatment of osteogenic sarcoma. Cancer. 1976;37:1–11.
46. Friedman MA, Carter SK. The therapy of osteogenic sarcoma: Current status and thoughts for the future. J Surg Oncol. 1972;4(5–6):482–510.
47. Sutow WW, Sullivan MP, Fernbach DJ. Adjuvant chemotherapy in primary treatment of osteogenic sarcoma. Proc Am Assoc Cancer Res. 1974;Vol. 15:77.
48. O'Bryan RM, Luce JK, Talley RW, Gottlieb JA, Baker LH, Bonadonna G. Phase II evaluation of adriamycin in human neoplasia. Cancer. 1973;32(1):1–8.
49. Anderson ME. Update on Survival in Osteosarcoma. Orthop Clin North Am. 2016;47(1):283–92.
50. Hwang JS, Mehta AD, Yoon RS, Beebe KS. From amputation to limb salvage reconstruction: Evolution and role of the endoprosthesis in musculoskeletal oncology. J Orthop Traumatol. 2014;15(2):81–6.
51. Grimer RJ, Aydin BK, Wafa H, Carter SR, Jeys L, Abudu A, et al. Very long-term outcomes after endoprosthetic replacement for malignant tumours of bone. J Bone Jt Surg - Br Vol. 2016;98(6):857–64.
52. Holm CE, Bardram C, Riecke AF, Horstmann P, Petersen MM. Implant and limb survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses fifty patients followed for a mean of forteen years. Int Orthop. 2018;42(5):1175–81.
53. Rosen G. Preoperative (neoadjuvant) chemotherapy for osteogenic sarcoma: A ten year experience. Orthopedics. 1985;8(5):659–64.
54. Kotz RI. Progress in musculoskeletal oncology from 1922 - 2012. Int Orthop. 2014;38(5):1113–22.
55. Angelini A, Guerra G, Mavrogenis AF, Pala E, Picci P, Ruggieri P. Clinical outcome of central conventional chondrosarcoma. J Surg Oncol. 2012;106(8):929–37.
56. Pala E, Angelini A, Trovarelli G, Berizzi A, Ruggieri P. Current megaprostheses options. Curr Orthop Pract. 2017;28(6):520–3.
57. Goulding KA, Gaston CL, Grimer RJ. Outcomes and Options for Prosthetic Reconstruction After Tumour Resection About the Knee. Curr Surg Reports. 2014;2(42).
58. Ruggieri P, Bosco G, Pala E, Errani C, Mercuri M. Local recurrence, survival and function after total femur resection and megaprosthetic reconstruction for bone sarcomas. Clin Orthop Relat Res. 2010;468(11):2860–6.
59. Jeys LM, Grimer RJ, Carter SR, Tillman RM. Periprosthetic infection in patients treated for an orthopaedic oncological condition. J Bone Jt Surg - Am Vol. 2005;87:842–9.
60. Zeegen EN, Aponte-Tinao LA, Hornicek FJ, Gebhardt MC, Mankin HJ. Survivorship analysis of 141 modular metallic endoprostheses at early followup. Clin Orthop Relat Res. 2004;90025(420):239–50.

61. Myers GJC, Abudu AT, Carter SR, Tillman RM, Grimer RJ. Endoprosthetic replacement of the distal femur for bone tumours. Long-term results. J Bone Jt Surg - Br Vol. 2007;89-B(4):521–6.
62. Thornley P, Vicente M, Macdonald A, Evaniew N, Ghert M, Velez R. Causes and Frequencies of Reoperations After Endoprosthetic Reconstructions for Extremity Tumor Surgery : A Systematic Review. 2019;2118:894–902.
63. Houdek MT, Wagner ER, Wilke BK, Wyles CC, Taunton MJ, Sim FH. Long term outcomes of cemented endoprosthetic reconstruction for periarticular tumors of the distal femur. Knee. 2016;23(1):167–72.
64. Pala E, Trovarelli G, Angelini A, Maraldi M, Berizzi A, Ruggieri P. Megaprosthesis of the knee in tumor and revision surgery. Acta Biomed. 2017;88:129–38.
65. Holzer G, Krepler P, Koschat MA, Grampp S, Dominkus M, Kotz R. Bone mineral density in long-term survivors of highly malignant osteosarcoma. J Bone Jt Surg - Br Vol. 2003;85(2):231–7.
66. Petersen MM, Gehrchen PM, Nielsen PK, Lund B. Loss of bone mineral of the hip assessed by DEXA following tibial shaft fractures. Bone. 1997;20(5):491–5.
67. Bohr HH, Lund B. Bone mineral density of the proximal tibia following uncemented arthroplasty. J Arthroplasty. 1987;2(4):309–12.
68. Soininvaara TA, Miettinen HJA, Jurvelin JS, Alhava EM, Kröger HPJ. Bone mineral density in the proximal femur and contralateral knee after total knee arthroplasty. J Clin Densitom. 2004;7(4):424–31.
69. Petersen MM, Jensen NC, Gehrchen PM, Nielsen PK, Nielsen PT. The relation between trabecular bone strength and bone mineral density assessed by dual photon and dual energy X-ray absorptiometry in the proximal tibia. Calcif Tissue Int. 1996;59(4):311–4.
70. Andersen MR, Winther NS, Lind T, Schrøder HM, Flivik G, Petersen MM. Low Preoperative BMD Is Related to High Migration of Tibia Components in Uncemented TKA—92 Patients in a Combined DEXA and RSA Study With 2-Year Follow-Up. J Arthroplasty. 2017;32(7):2141–6.
71. Hansen RL, Langdahl BL, Jørgensen PH, Petersen KK, Søballe K, Stilling M. Changes in periprosthetic bone mineral density and bone turnover markers after osseointegrated implant surgery: A cohort study of 20 transfemoral amputees with 30-month follow-up. Prosthet Orthot Int. 2019;43(5):508–18.
72. Capanna R, Morris HG, Campanacci D, Del Ben M, Campanacci M. Modular uncemented prosthetic reconstruction after resection of tumours of the distal femur. J Bone Jt Surg - Br Vol. 1994;76(2):178–86.
73. van Krieken FM, den Heeten GJ, Pedersen DR, Brand RA, Crowninshield RD. Prediction of muscle and joint loads after segmental femur replacement for osteosarcoma. Clin Orthop Relat Res. 1985;(198):273–83.
74. Ruggieri P, Mavrogenis AF, Pala E, Abdel-Mota'al M, Mercuri M. Long term results of fixed-hinge megaprostheses in limb salvage for malignancy. Knee. 2012;19(5):543–9.
75. Gosheger G, Gebert C, Ahrens H, Streitbuerger A, Winkelmann W, Hardes J. Endoprosthetic reconstruction in 250 patients with sarcoma. Clin Orthop Relat Res. 2006;
76. Mittermayer F, Krepler P, Dominkus M, Schwameis E, Sluga M, Heinzl H, et al. Long-term followup of uncemented tumor endoprostheses for the lower extremity. Clin Orthop Relat Res. 2001;(388):167–77.
77. Pala E, Henderson ER, Calabrò T, Angelini A, Abati CN, Trovarelli G, et al. Survival of current production tumor endoprostheses: Complications, functional results, and a comparative statistical analysis. J Surg Oncol. 2013;108(6):403–8.
78. Mavrogenis AF, Pala E, Angelini A, Ferraro A, Ruggieri P. Proximal tibial resections and reconstructions: Clinical outcome of 225 patients. J Surg Oncol. 2013;107(4):335–42.
79. Meyers PA, Heller G, Healey J, Huvos A, Lane J, Marcove R, et al. Chemotherapy for nonmetastatic osteogenic sarcoma: the Memorial Sloan-Kettering experience. J Clin Oncol. 1992;10(1):5.
80. Chao EYS, Fuchs B, Rowland CM, Ilstrup DM, Pritchard DJ, Sim FH. Long-Term Results of Segmental Prosthesis Fixation by Extracortical Bone-Bridging and Ingrowth. J Bone Jt Surg - Am Vol. 2004;86A(5):948–55.
81. Saran N, Zhang R, Turcotte RE. Osteogenic protein-1 delivered by hydroxyapatite-coated implants improves bone ingrowth in extracortical bone bridging. Clin Orthop Relat Res. 2011;469(5):1470–8.
82. Flint MN, Griffin AM, Bell RS, Ferguson PC, Wunder JS. Aseptic loosening is uncommon with uncemented proximal tibia tumor prostheses. Clin Orthop Relat Res. 2006;450:52–9.
83. Bhangu AA, Kramer MJ, Grimer RJ, O'Donnell RJ. Early distal femoral endoprosthetic survival: Cemented stems versus the Compress® implant. Int Orthop. 2006;30(6):465–72.
84. Haijie L, Dasen L, Tao J, Yi Y, Xiaodong T, Wei G. Implant Survival and Complication Profiles of Endoprostheses for Treating Tumor Around the Knee in Adults: A Systematic Review of the Literature Over the Past 30 Years. J Arthroplasty. 2018;33(4):1275–87.
85. Bramer JAM, Van Linge JH, Grimer RJ, Scholten RJPM. Prognostic Factors in localized Extremity Osteosarcoma; a Systematic Review. Eur J Surg Oncol. 2009;35:1030–6.
86. Duchman KR, Gao Y, Miller BJ. Prognostic factors for survival in patients with high-grade osteosarcoma using the Surveillance, Epidemiology, and End Results (SEER) Program database. Cancer Epidemiol. 2015;39(4):593–9.
87. Bielack SS, Kempf-Bielack B, Delling G, Exner GU, Flege S, Helmke K, et al. Prognostic factors in high-

grade osteosarcoma of the extremities or trunk: An analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. J Clin Oncol. 2002;20(3):776–90.
88. Sargent DJ, Conley BA, Allegra C, Collette L. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol. 2005;23(9):2020–7.
89. M.paesmans. Prognostic and predictive factors for lung cancer. Breathe. 2012;9(2):113–22.
90. Clark GM. Prognostic factors versus predictive factors: Examples from a clinical trial of erlotinib. Mol Oncol. 2008;1(4):406–12.
91. Hingorani AD, Van Der Windt DA, Riley RD, Abrams K, Moons KGM, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. BMJ. 2013;346(February):1–9.
92. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ. 2013;346(2):1–11.
93. Quirina C.B.S, Bernstein KDA, Calderon SAL, Schwab JH. Can Machine-learning Techniques Be Used for 5-year Survival Prediction of Patients With Chondrosarcoma ? Clin Orthop Relat Res. 2018;476:2040–8.
94. Nandra R, Parry M, Forsberg J, Grimer R. Can a Bayesian Belief Network Be Used to Estimate 1-year Survival in Patients With Bone Sarcomas? Clin Orthop Relat Res. 2017;475:1681–9.
95. Bongers MER, Thio QCBS, Karhade AV, Stor ML, Raskin KA, Lozano Calderon SA, DeLaney TF, Ferrone ML SJ. Does the SORG Algorithm Predict 5-year Survival in Patients with Does the SORG Algorithm Predict 5-year Survival in Patients with Chondrosarcoma ? An External Validation. Clin Orthop Relat Res. 2019;00(May):1–8.
96. Bzdok D, Altman N, Krzywinski M. Points of Significance: Statistics versus machine learning. Nat Methods. 2018;15(4):233–4.
97. Simms L, Barraclough H, Govindan R. Biostatistics primer: What a clinician ought to know-prognostic and predictive factors. J Thorac Oncol. 2013;8(6):808–13.
98. Chen JH, Asch SM, Alto P. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. N Engl J MED. 2018;376(26):2507–9.
99. Schapire RE, Avenue P, Room A. The Boosting Approach to Machine Learning An Overview. MSRI Work Nonlinear Estim Classif. 2002;1–23.
100. Chen Y, Jia Z, Mercola D, Xie X. A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index. Comput Math Methods Med. 2013;2013:8 pages.
101. Natekin A, Knoll A. Gradient boosting machines , a tutorial. Front Neurorobot. 2013;7(December).
102. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014;29(8):541–9.
103. Jørgensen PH, Lausten GS, Pedersen AB. The Danish sarcoma database. Clin Epidemiol. 2016;8:685–90.
104. Safwat A, Baerentzen S. Population-based Aarhus Sarcoma Registry : validity , completeness of registration , and incidence of bone and soft tissue sarcomas in western Denmark. Clin Epidemiol. 2013;5(March):45–56.
105. Bjerregaard B, Larsen OB. The Danish pathology register. Scand J Public Health. 2011;39(7):72–4.
106. Henderson ER, Groundland JS, Pala E, Dennis JA, Wooten R, Cheong D, et al. Failure mode classification for tumor endoprostheses: retrospective review of five institutions and a literature review. J Bone Joint Surg Am. 2011;93(5):418–29.
107. Enneking WF, Dunham W, Gebhardt M, Malawer M PD. A System for the Functional Evaluation of Reconstructive Procedures After Surgical Treatment of Tumors of the Musculoskeletal System. Clin Orthop Relat Res. 1991;(286):241–6.
108. Bender R, Lange S. Adjusting for multiple testing - When and how? J Clin Epidemiol. 2001;54(4):343–9.
109. Ranstam J, Kärrholm J, Pulkkinen P, Mäkelä K, Espehaug B, Pedersen AB, et al. Statistical analysis of arthroplasty data. II. Guidelines. Acta Orthop. 2011;
110. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Casp J Intern Med. 2013;4(2):627–35.
111. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol. 2010;63(8):938–9.
112. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. Epidemiology. 2010;21(1):128–38.
113. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic Progn Res. 2019;3(1):1–8.
114. Forsberg JA, Ma DS, Chen Q, Vickers A. Treating Metastatic Disease Which Survival Model Is Best Suited for the Clinic ? Clin Orthop Relat Res. 2013;471:843–50.
115. Pala E, Henderson ER, Calabrò T, Angelini A, Abati CN, Trovarelli G, et al. Survival of current production tumor endoprostheses: Complications, functional results, and a comparative statistical analysis. J Surg Oncol. 2013;
116. Ghert M. Prophylactic antibiotic regimens in tumour surgery (PARITY): A pilot multicentre randomised controlled trial. Bone Jt Res. 2015;4(9):154–62.
117. Song W.J CCK. Plast Reconstr Surg. 2010;126(6):2234–42.

118. Eccles M, Grimshaw J, Campbell M, Ramsay C. Research designs for studies evaluating the effectiveness of change an improvement strategies. Qual Saf Heal Care. 2003;12(1):47–52.
119. Saccà L. The uncontrolled clinical trial: Scientific, ethical, and practical reasons for being. Intern Emerg Med. 2010;5(3):201–4.
120. Ranstam J. Hypothesis-generating and confirmatory studies, Bonferroni correction, and pre-specification of trial endpoints. Acta Orthop. 2019;90(4):297.
121. Dansk Sarkom Database (DSD) Årsrapport 2016. 2016; Available from: http://dsg.ortopaedi.dk/wp-content/uploads/2017/02/Arsrapport-2015-for-Sarkom.pdf
122. Forsberg JA, Wedin R, Bauer H, Hansen BH, Laitinen M, ST Clement, Keller J, Boland PJ HJ. External validation of the Bayesian Estimated Tools for Survival (BETS) models in patients with surgically treated skeletal metastases. BMC Cancer. 2012;12(493).
123. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. Can Assoc Radiol J. 2019;70(4):344–53.
124. Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. J Clin Epidemiol. 2019;76(2016):175–82.
125. Riley RD, Jr FEH, Ie K, Joie S, Burke DL, Gm K, et al. Minimum sample size for developing a multivariable prediction model : PART II - binary and time-to-event outcomes. Stat Med. 2019;38(February 2018):1276–96.
126. Raudys SJ JA. Small sample size effects in statistical pattern recongnition: Recommendations for practitioners. IEEE Trans Pattern Anal Mach Intell. 1991;13(3):252–64.
127. Courvoisier DS, Combescure C, Agoritsas T, Gayet-Ageron A, Perneger T V. Performance of logistic regression modeling: Beyond the number of events per variable, the role of data structure. J Clin Epidemiol. 2011;64(9):993–1000.
128. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology. 2018;286(3):800–9.
129. Jaffe N, Carrasco H, Raymond K, Ayala A, Eftekhari F. Can cure in patients with osteosarcoma be achieved exclusively with chemotherapy and abrogation of surgery? Cancer. 2002;95(10):2202–10.
130. Bertrand TE, Cruz A, Binitie O, Cheong D, Letson GD. Do Surgical Margins Affect Local Recurrence and Survival in Extremity, Nonmetastatic, High-grade Osteosarcoma? Clin Orthop Relat Res. 2016;474(3):677–83.
131. Allison DC, Carney SC, Ahlmann ER, Hendifar A, Chawla S, Fedenko A, et al. A meta-analysis of osteosarcoma outcomes in the modern medical era. Sarcoma. 2012;2012:10 pages.
132. Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: Data from the surveillance, epidemiology, and end results program. Cancer. 2009;115(7):1531–43.
133. Bacci G, Longhi A, Versari M, Mercuri M, Briccoli A, Picci P. Prognostic factors for osteosarcoma of the extremity trerated with neoadjuvant chemotherapy: 15-Year experience in 789 patients treated at a single institution. Cancer. 2006;106(5):1154–61.
134. Smeland S, Müller C, Alvegard TA, Wiklund T, Wiebe T, Björk O, et al. Scandinavian Sarcoma Group Osteosarcoma Study SSG VIII: prognostic factors for outcome and the role of replacement salvage chemotherapy for poor histological responders. Eur J Cancer. 2003;39(4):488–94.
135. Whelan JS, Jinks RC, McTiernan A, Sydes MR, Hook JM, Trani L, et al. Survival from high-grade localised extremity osteosarcoma: Combined results and prognostic factors from three European osteosarcoma intergroup randomised controlled trials. Ann Oncol. 2012;23(6):1607–16.
136. Sæter G, Elomaa I, Wahlqvist Y, Alvegård TA, Wiebe T, Monge O, et al. Prognostic factors in bone sarcomas Prognostic factors in bone sarcomas. Acta Orthop Scand Suppl. 1997;68(273):156–60.
137. Slade AD, Warneke CL, Hughes DP, Lally PA, Lally KP, Hayes-Jordan AA, et al. Effect of concurrent metastatic disease on survival in children and adolescents undergoing lung resection for metastatic osteosarcoma. J Pediatr Surg. 2015;50(1):157–60.
138. Glasser DB, Lane JM, Huvos AG, Marcove RC, Rosen G. Survival, prognosis, and therapeutic response in osteogenic sarcoma. The memorial hospital experience. Cancer. 1992;69(3):698–708.
139. Stojadinovic A, Peoples GE, Libutti SK, Henry LR, Eberhardt J, Howard RS, et al. Development of a clinical decision model for thyroid nodules. BMC Surg. 2009;9(1):1–11.
140. Forsberg JA, Wedin R, Boland PJ, Healey JH. Can We Estimate Short- and Intermediate-term Survival in Patients Undergoing Surgery for Metastatic Bone Disease? Clin Orthop Relat Res. 2017;475(4):1252–61.
141. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: Application and impact of prognostic models in clinical practice. BMJ. 2009;338(7709):1487–90.
142. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: Validating a prognostic model. BMJ. 2009;338(7708):1432–5.
143. Thorn AP, Daugaard S, Christensen LH, Jarle IB, Petersen MM. YKL-40 protein in osteosarcoma tumor tissue. APMIS. 2016;(17):453–61.

144. Malawer MM, Chou LB. Prosthetic survival and clinical results with use of large-segment replacements in the treatment of high-grade bone sarcomas. J Bone Jt Surg - Am Vol. 1995 Aug 1;77(8):1154–65.
145. Grimer RJ, Carter SR, Tillman RM, Sneath RS, Walker PS, Unwin PS, et al. Endoprosthetic replacement of the proximal tibia. J Bone Jt Surgery-British Vol. 1999;81B(3):488–94.
146. Jeys LM, Grimer RJ, Carter SR, Tillman RM. Risk of amputation following limb salvage surgery with endoprosthetic replacement, in a consecutive series of 1261 patients. Int Orthop. 2003;27(3):160–3.
147. Janssen SJ, Langerhuizen DWG, Schwab JH, Bramer JAM. Outcome after reconstruction of proximal femoral tumors: A systematic review. J Surg Oncol. 2019;119(1):120–9.
148. Yilmaz M, Sørensen MS, Saebye C, Baad-Hansen T PM. Long - term results of the Global Modular Replacement System tumor prosthesis for reconstruction after limb - sparing bone resections in orthopedic oncologic conditions : Results from a national cohort. J Surg Oncol. 2019;120(July 2018):183–92.
149. Puchner SE, Kutscha-Lissberg P, Kaider A, Panotopoulos J, Puchner R, Böhler C, et al. Outcome after reconstruction of the proximal tibia - Complications and competing risk analysis. PLoS One. 2015;10(8):1–14.
150. Bacci G, Ferrari S, Mercuri M, Bertoni F, Picci P, Manfrini M, et al. Predictive factors for local recurrence in osteosarcoma. 540 Patients with extremity tumors followed for minimum 2.5 years after neoadjuvant chemotherapy. Acta Orthop Scand. 1998;69(3):230–6.
151. Stevenson JD, Laitinen MK, Parry MC, Sumathi V, Grimer RJ, Jeys LM. The role of surgical margins in chondrosarcoma. Eur J Surg Oncol. 2018;44(9):1412–8.
152. Tsuda Y, Evans S, Stevenson JD, Parry M, Fujiwara T, Laitinen M, et al. Is the Width of a Surgical Margin Associated with the Outcome of Disease in Patients with Peripheral Chondrosarcoma of the Pelvis? A Multicenter Study. Clin Orthop Relat Res. 2019;477(11):2432–40.
153. Grimer RJ, Taminiau AM, Cannon SR. Surgical outcomes in osteosarcoma. J Bone Jt Surg - Br Vol. 2002;84:395–400.
154. Lee SH, Oh JH, Lee KS, Yoo KH, Kim HS. Infection after prosthetic reconstruction in limb salvage surgery. Int Orthop. 2002;26(3):179–84.
155. Hardes J, Gebert C, Schwappach A, Ahrens H, Streitburger A, Winkelmann W, et al. Characteristics and outcome of infections associated with tumor endoprostheses. Arch Orthop Trauma Surg. 2006;126(5):289–96.
156. Gosheger G, Gebert C, Ahrens H, Streitbuerger A, Winkelmann W, Hardes J. Endoprosthetic reconstruction in 250 patients with sarcoma. Clin Orthop Relat Res. 2006;450(450):164–71.
157. Henshaw R, Malawer M. Review of Endoprosthetic Reconstruction in Limb-sparing Surgery. Musculoskelet Cancer Surg. 2001;381–402.
158. Wang TY, Dormans JP, Chang B. Soft-tissue optimization of limb salvage with knee endoprosthesis: The 10-year experience at the children's hospital of philadelphia. Ann Plast Surg. 2012;69(5):560–4.
159. Trieb K, Göggel M, Dürr HR. Proximaler Tibiaersatz mit Transfer des M. Gastrocnemius. Oper Orthop Traumatol. 2012;24(3):263–71.
160. Healey JH. CORR Insights™: High Infection Rate Outcomes in Long-bone Tumor Surgery with Endoprosthetic Reconstruction in Adults: A Systematic Review. Clin Orthop Relat Res. 2013;471(6):2028–9.
161. Dhanoa A, Ajit Singh V, Elbahri H. Deep Infections after Endoprosthetic Replacement Operations in Orthopedic Oncology Patients. Surg Infect (Larchmt). 2015;16(3):323–32.
162. Hardes J, Von Eiff C, Streitbuerger A, Balke M, Budny T, Henrichs MP, et al. Reduction of periprosthetic infection with silver-coated megaprostheses in patients with bone sarcoma. J Surg Oncol. 2010;101(5):389–95.
163. Tsuchiya H, Shirai T, Nishida H, Murakami H, Kabata T, Yamamoto N, et al. Innovative antimicrobial coating of titanium implants with iodine. J Orthop Sci. 2012;17(5):595–604.
164. Donati F, Di Giacomo G, D'Adamio S, Ziranu A, Careri S, Rosa MA, et al. Silver-Coated Hip Megaprosthesis in Oncological Limb Savage Surgery. Biomed Res Int. 2016;2016:6 pages.
165. Carty CP, Dickinson IC, Watts MC, Crawford RW, Steadman P. Impairment and disability following limb salvage procedures for bone sarcoma. Knee. 2009;16(5):405–8.
166. Choong PFM, Sim FH, Pritchard DJ, Rock MG, Chao EYS. Megaprostheses after resection of distal femoral tumors. A rotating hinge design in 30 patients followed for 2-7 years. Acta Orthop Scand. 1996;67(4):345–51.
167. Kawai A, Healey JH, Boland PJ, Athanasian EA, Jeon DG. A rotating-hinge knee replacement for malignant tumors of the femur and tibia. J Arthroplasty. 1999;14(2):187–96.
168. Mavrogenis AF, Pala E, Angelini A, Calabro T, Romagnoli C, Romantini M, et al. Infected Prostheses after Lower-Extremity Bone Tumor Resection: Clinical Outcomes of 100 Patients. Surg Infect (Larchmt). 2015;16(3):267–75.
169. Mittermayer F, Windhager R, Dominkus M, Krepler P, Schwameis E, Sluga M, et al. Revision of the Kotz type of tumour endoprosthesis for the lower limb. J Bone Jt Surg - Br Vol. 2002;84(3):401–6.

170. Unwin PS, Cannon SR, Grimer RJ, Kemp HBS, Sneath RS, Walker PS. Aseptic loosening in cemented custommade prosthetic replacements for bone tumours of the lower limb. J Bone Jt Surg - Br Vol. 1996;7878(1):5–13.
171. Plotz W, Rechl H, Burgkart R, Messmer C, Schelter R, Hipp E, et al. Limb salvage with tumor endoprostheses for malignant tumors of the knee. Clin Orthop Relat Res. 2002;(405):207–15.
172. Pala E, Mavrogenis AF, Angelini A, Henderson ER, Letson GD, Ruggieri P. Cemented versus cementless endoprostheses for lower limb salvage surgery. J BUON. 2013;18(2):496–503.
173. Pala E, Trovarelli G, Calabr?? T, Angelini A, Abati CN, Ruggieri P. Survival of Modern Knee Tumor Megaprostheses: Failures, Functional Results, and a Comparative Statistical Analysis. Clin Orthop Relat Res. 2015;473(3):891–9.
174. Schwartz AJ, Kabo JM, Eilber FC, Eilber FR, Eckardt JJ. Cemented distal femoral endoprostheses for musculoskeletal tumor: Improved survival of modular versus custom implants. Clin Orthop Relat Res. 2010;468(8):2198–210.
175. Davis AM, Devlin M, Griffin AM, Wunder JS, Bell RS. Functional outcome in amputation versus limb sparing of patients with lower extremity sarcoma: A matched case-control study. Arch Phys Med Rehabil. 1999;80(6):615–8.
176. Andersen MR, Petersen MM. Adaptive Bone Remodeling of the Femoral Bone After Tumor Resection Arthroplasty With an Uncemented Proximally Hydroxyapatite-Coated Stem. J Clin Densitom. 2016;19(2):202–7.
177. Lan F, Wunder JS, Griffin a M, Davis a M, Bell RS, White LM, et al. Periprosthetic bone remodelling around a prosthesis for distal femoral tumours. Measurement by dual-energy X-ray absorptiometry (DEXA). J Bone Jt Surg - Br Vol. 2000;82(1):120–5.
178. Bodén HSG, Sköldenberg OG, Salemyr MOF, Lundberg HJ, Adolphson PY. Continuous bone loss around a tapered uncemented femoral stem: A long-term evaluation with DEXA. Acta Orthop. 2006;77(6):877–85.
179. Venesmaa PK, Kröger HPJ, Jurvelin JS, Miettinen HJA, Suomalainen OT, Alhava EM. Periprosthetic bone loss after cemented total hip arthroplasty: A prospective 5-year dual energy radiographic absorptiometry study of 15 patients. Acta Orthop Scand. 2003;74(1):31–6.
180. Nygaard M, Zerahn B, Bruce C, Søballe K, Borgwardt A. Early periprosthetic femoral bone remodelling using different bearing material combinations in total hip arthroplasties: a prospective randomised study. Eur Cell Mater. 2004;8(5):815–9.
181. Abu-Rajab RB, Watson WS, Walker B, Roberts J, Gallacher SJ, Meek RMD. Peri-prosthetic bone mineral density after total knee arthroplasty. J Bone Jt Surg - Br Vol. 2006;88(5):606–13.
182. Tanzer M, Kantor S, Rosenthall L, Bobyn JD. Femoral remodeling after porous-coated total hip arthroplasty with and without hydroxyapatite-tricalcium phosphate coating: A prospective randomized trial. J Arthroplasty. 2001;16(5):552–8.
183. Kröger H, Miettinen H, Arnala I, Koski E, Rushton N, Suomalainen O. Evaluation of periprosthetic bone using dual-energy X-ray absorptiometry: Precision of the method and effect of operation on bone mineral density. J Bone Miner Res. 2009;11(10):1526–30.
184. Brockstedt H, Kassem M, Eriksen EF, Mosekilde L, Melsen F. Age- and sex-related changes in iliac cortical bone mass and remodeling. Bone. 1993;14(4):681–91.
185. Seki T, Omori G, Koga Y, Suzuki Y, Ishii Y, Takahashi HE. Is bone density in the distal femur affected by use of cement and by femoral component design in total knee arthroplasty? J Orthop Sci. 1999;4(3):180–6.
186. Gehrchen PM, Petersen MM, Nielsen PK, Lund B. Influence of region size on bone mineral measurements along femoral stems in THA. HIP Int. 2000;10(4):204–8.
187. Harvey N, Ahlmann ER, Allison DC, Wang L, Menendez LR. Endoprostheses last longer than intramedullary devices in proximal femur metastases. Clin Orthop Relat Res. 2012;470(3):684–91.
188. Tunn PU, Pomraenke D, Goerling U, Hohenberger P. Functional outcome after endoprosthetic limb-salvage therapy of primary bone tumours - A comparative analysis using the MSTS score, the TESS and the RNL index. Int Orthop. 2008;32(5):619–25.

International Orthopaedics (2018) 42:1175–1181
https://doi.org/10.1007/s00264-018-3861-7

ORIGINAL PAPER

Implant and limb survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses fifty patients followed for a mean of forteen years

Christina Enciso Holm[1] · Christian Bardram[1] · Anja Falk Riecke[1] · Peter Horstmann[1] · Michael Mørk Petersen[1]

Received: 18 October 2017 / Accepted: 21 February 2018 / Published online: 12 March 2018

Abstract

Purpose Previous studies reported variable outcome and failure rates after mega-prosthetic reconstructions in the lower extremities. The purpose of this study was to make a long-term single-center evaluation of patients treated with limb-sparing surgery and reconstruction with mega-prostheses in the lower extremities.

Methods We identified 50 patients (osteosarcoma ($n = 30$), chondrosarcoma ($n = 9$), osteoclastoma ($n = 6$), Ewing sarcoma ($n = 4$), angiosarcoma ($n = 1$)), who underwent limb-sparing reconstruction of the lower extremities (proximal femur ($n = 9$), distal femur ($n = 29$), proximal tibia ($n = 9$), and the entire femur ($n = 3$)) between 1985 and 2005. Surviving patients not lost to follow-up were evaluated using the MSTS score. Causes of failure were classified according to the Henderson classification. Kaplan-Meier survival analysis was used for evaluation of patient, prosthesis, and limb survival.

Results Twenty-eight patients were alive at follow-up. Fifty-four percent had revision surgery ($n = 27$). The ten year patient survival was 60% (95%CI 46–74%); the ten year implant survival was 24% (95%CI 9–41%), and the ten year limb survival rate was 83% (95%CI 65–96%). Type 1 failure occurred in 9%, type 2 in 16%, type 3 in 28%, type 4 in 18%, and type 5 in 3%. Mean MSTS score was 21 (range, 6–30), representing a median score of 71%.

Conclusions Our long-term results with mega-prostheses justify the use of limb-salvage surgery and prosthetic reconstruction. Our results are fully comparable with other findings, with regard to limb and prosthesis survival, but also with regard to functional outcome.

Keywords Mega-endoprostheses · Bone tumors · Limb-salvage surgery · Tumour endoprosthesis infection · Amputation · MSTS

Introduction

Since the late 1980s, a change was made at our clinic, which gradually substituted amputation as the method of choice for treating bone sarcomas towards limb-sparing surgery. The change was a result of progress in orthopaedic implant possibilities, surgical advances, and the introduction of chemotherapy as a part of the treatment for many of these patients [1]. Previous studies have shown that a similar time-related change in treatment strategy took place in other hospitals, where 73% of patients were treated with amputation between 1982 and 1989 [2, 3], 42% during 1990–1997 [4], and only 6% in 1997–2000 [5].

The overall survival of patients that have had limb-sparing surgery has been evaluated continuously and has been found to be comparable to, or even better than, patients treated with amputation [6]. When performing limb-sparing bone tumour resection of the lower extremities, such as the proximal femur, distal femur, or proximal tibia resections, reconstructions are most often performed using mega-prostheses. The implant survival of mega-prostheses are for various reasons lower compared to ordinary primary hip or knee arthroplasty in patients suffering from degenerative joint diseases [7]. However, the five year implant survival of mega-prostheses has increased since the 1980s from approximately 20 to 85% [8, 9]. The purpose of this study was to make a long-term evaluation of patients treated in our department between 1985 and 2005 with limb-sparing surgery and reconstruction with

✉ Christina Enciso Holm
christina.holm@dadlnet.dk

[1] The Musculoskeletal Tumor Section, The Department of Orthopedic Surgery, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark

mega-prostheses after resection of bone sarcomas or giant cell tumours of bone in the lower extremities. Our aim was to assess the treatment regimen in a relatively small orthopaedic oncology centre, and evaluate if it resulted in an acceptable functional outcome, avoiding amputation.

Patients and methods

A retrospective search in our local pathology database was conducted. We identified 50 patients, with a mean age (at the time of surgery) of 34 (6–74) years ($F/M = 24/26$), who underwent limb-sparing surgery and reconstruction with a mega-prosthesis for a bone sarcoma ($n = 44$) or a giant cell tumour of bone ($n = 6$) of the lower extremities at our department from 1985 to 2005 (Table 1). The histological diagnoses were osteosarcomas ($n = 30$), chondrosarcomas ($n = 9$), giant cell tumours ($n = 6$), Ewing sarcomas ($n = 4$), and angiosarcoma of bone ($n = 1$). The anatomical location of the bone tumours resulted in the following limb-sparing resections: distal femur ($n = 29$), proximal femur ($n = 9$), proximal tibia ($n = 9$), and entire femur ($n = 3$).

The types of tumour prostheses used for the primary reconstruction exhibited some variation. We used 35 prostheses from Stryker or Stryker-Howmedica Inc. (HMRS® prostheses ($n = 32$), KMFTR® prostheses ($n = 2$), GMRS® prosthesis ($n = 1$)), eight prostheses from Valdemar Link (Endorotational system ($n = 5$), Campana ($n = 2$), custom-made ($n = 1$)), and three prostheses from Biomet (Bimetric revision ($n = 2$), Kent hip ($n = 1$)). In four patients, the prosthesis used could not be determined with certainty. In general, uncemented stems were preferred when using the HMRS® prostheses.

Overall survival time data were collected from the Danish Civil Register on January 1, 2015 giving complete survival data follow-up for all patients living in Denmark [10]. Twenty-eight patients were alive at the time of follow-up, and patient follow-up ranged from seven days to 29 years with a mean of 14.3 years. The short follow-up was one patient living in the Faroe Islands who returned to the island immediately after surgery.

Patient records were used to determine the date of diagnosis, age at diagnosis, date of surgery, type of used implant, tumour histology, tumour site, date and course of revision surgery, or amputation. Among the patients who were still alive, 24 of 28 were available for clinical evaluation using the Enneking score (MSTS) [11]. Revision surgery was defined as all kinds of prosthesis-related surgery, hence also including minor surgical procedures performed because of, e.g., polyethylene wear, and all failures were classified according to Henderson et al. [12].

Statistical analyses were performed with the use of the IBM SPSS Statistics Version 22 statistical package. All data are presented as mean values together with total range. Kaplan-Meier survival analyses were performed for evaluation of overall survival (calculated from time of surgery), implant survival (time to first revision) and limb survival (time to amputation). 95% confidence intervals (CI) were calculated using Microsoft Excel using Greenwood's formula for calculation of standard error.

Table 1 Patient demographics and implant revision data

Patient demographics	
Number of patients	50
Female/male	24/26
Age at surgery (years)	
Mean (range)	34 (6–74)
Follow-up (years)	
Mean (range)	14 (0–29)
Patient survival	
Deaths during follow-up	22/50
Tumour histology	
Giant cell tumour/bone sarcoma	6/44
Time to death after surgery (years)	
Mean (range)	3.7 (0–14)
Implant revisions	
Total number of revisions	78
Patients with 1 revision	10
Patients with 2 revisions	3
Patients with 3 or more revisions	14

Results

Twenty-two patients died after an average of 3.7 (0.4–13.9) years following surgery, and causes of death have not been explored further (Table 1). The probability of overall 5-, 10-, and 15-year survival was 68% (CI 55–81%), 60% (CI 46–74%), and 55% (CI 40–69%) respectively (Fig. 1).

In total, 27 patients underwent revision surgery (54%). Average time from primary surgery to the first revision was three years and ranged from 14 days to 10.5 years. The primary causes for the first revision were as follows: polyethylene wear ($n = 6$, 12%) and deep infection ($n = 6$, 12%). Other causes for the first revision were the following: aseptic loosening ($n = 5$, 10%), chronic instability ($n = 3$, 6%), stem fracture ($n = 2$, 4%), periprosthetic fracture ($n = 1$, 2%), protrusion of a bipolar head through acetabulum ($n = 1$, 2%), positive margin after tumour resection ($n = 1$, 2%), and compartment syndrome ($n = 1$, 2%). One patient had no files from the first revision explaining what lead to an amputation. In total, 27 patients underwent 78 revision surgeries (Table 2). Ten patients (37%) experienced only one revision, while three

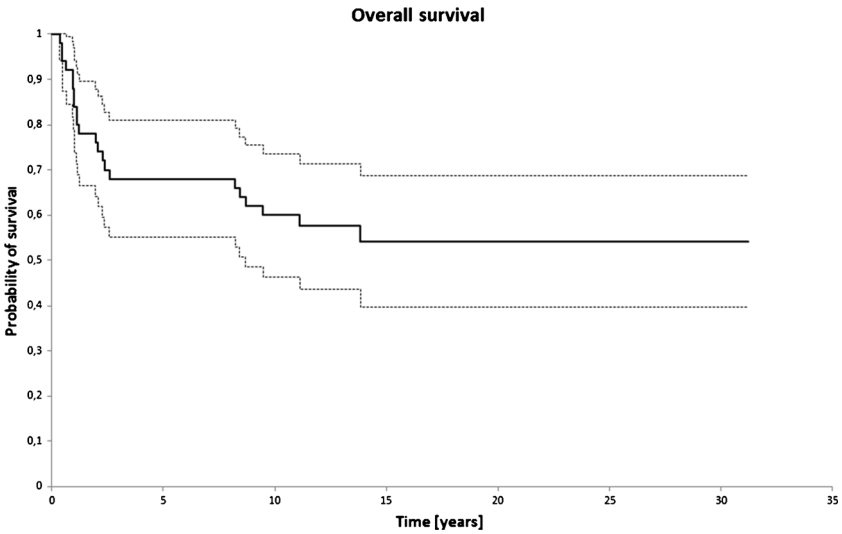


Fig. 1 Kaplan-Meier survival curve with 95% confidence limits showing the probability of overall survival after resection of primary bone tumours of the lower extremities and reconstruction with mega-prostheses ($n = 50$)

patients (11%) experienced two revisions, and 14 patients (52%) had three or more revisions (Table 1). The most frequent cause of revision in general was deep infection ($n = 19$, 24%), aseptic loosening ($n = 18$, 23%), and polyethylene wear ($n = 16$, 21%) (Table 2). In general, eight (16%) patients were infected, all with deep infection, and infections caused a total of 19 revisions (24%); 14 of those (18%) were two-stage revision surgeries, which all counted for two revisions. One patient was treated only with a DAIR (debridement, antibiotics, and implant retention) procedure, three patients were treated with both DAIR and two-stage. Seven patients (14%) experienced resolution of infection after revision. Only one patient (2%) with early infection, initially treated with two-stage revision, was subsequently amputated due to recurrence of infection. Deep infection occurred in six primary implants (12%), after a mean of 1.8 years (range 60 days to 7 years). Four patients (8%) had deep infection in their revision implants, three of them (6%) had revision because of aseptic loosening with change of stem, and one patient (2%) had a stem extension. Infection in revision implants occurred after a mean of 3.0 years (range 1–6.7 years). Two patients (4%) with infection in revision implant also had deep infection in their primary implant.

Table 2 The causes of all ($n = 78$) revisions performed in 50 patients treated with resection of a primary bone tumor of the lower extremities and reconstruction with a mega-prosthesis

Causes of performed revisions in general	
Patients having revision	27
Number of revisions	78
Deep infection	19
Polyethylene wear	16
Aseptic loosening	18
Instability	10
Fractured stem	5
Periprosthetic fracture	4
Recurrence or progression of tumor	2
Severe symptoms from earlier revision	1
Compartment	1
Ulceration through acetabulum	1
Unknown	1

The probability of 5-, 10-, and 15-year implant revision-free survival was 43% (CI 27–61%), 24% (CI 9–41%), and 16% (CI 0–28%) respectively. (Fig. 2).

Out of the 78 revisions registered, only 71 revisions could be classified according to the Henderson classification of mode of failure. One patient ($n = 1$) had no description for the cause of revision, and other non-classifiable revisions were compartment syndrome ($n = 1$), and synovectomia/brisement, without removal of the endoprosthesis ($n = 3$). One patient ($n = 1$) was amputated as a consequence of severe symptoms from an earlier revision, and one patient ($n = 1$) had exchange of the polyethylene due to infection, without removal of the whole prosthesis. According to the Henderson classification, the predominant type of the endoprosthetic failure was mechanical failure ($n = 54$, 76%), and the most frequent mode of failure was type 3 (=structural failure) ($n = 26$, 37%) followed by type 2 (=aseptic loosening) ($n = 18$, 25%) and type 1 (=soft tissue failure) ($n = 10$, 14%). Seventeen (24%) revisions were performed because of non-mechanical types of failure, and the most frequent mode of failure was type 4 (=infection) ($n = 15$, 21%) followed by type 5 (tumour progression) ($n = 2$; 3%) (Fig. 4).

Six patients (12%) underwent amputation of the affected limb. The primary cause was recurrence or progression of tumour ($n = 2$). Other causes were infection ($n = 1$), ischemia of the limb ($n = 1$), and severe discomfort after earlier revision ($n = 1$). One patient had no adequate files explaining the cause of amputation. Out of the six patients, three were alive at the

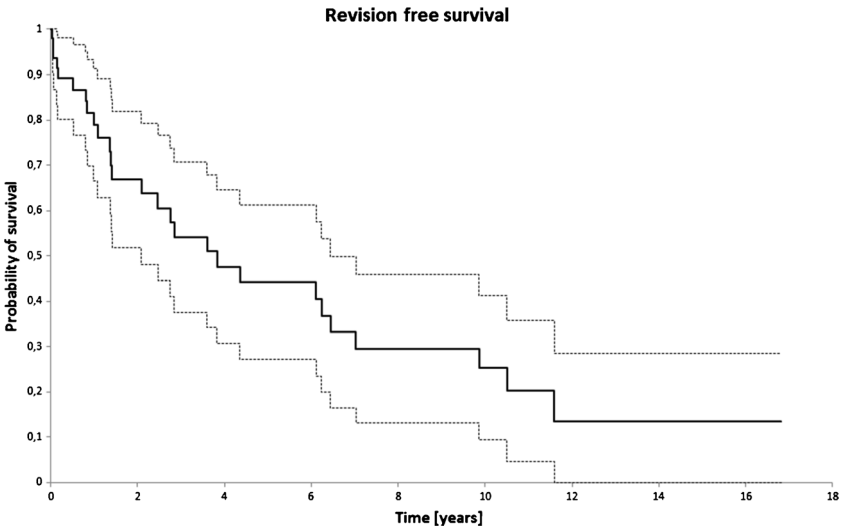


Fig. 2 Kaplan-Meier survival curve with 95% confidence limits showing the probability of implant revision-free survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses ($n = 50$)

time of follow-up. The probability of five, ten and 15-year limb survival was 89% (CI 80–100%), 83% (CI 65–96%), and 83% (CI 65–96%) respectively (Fig. 3).

The functional outcome was evaluated in 24 patients with their limbs still spared after an average of 17 (9–30) years post-operatively. The individual parameter mean scores were as follows: pain 3.9 (0–5), function 2.1 (0–5), emotional acceptance 4.1 (1–5), supports 4.0 (0–5), walking ability 3.8 (0–5), and gait 3.3 (0–5). The mean MSTS score was 21.2 (range 6–30), representing a median score of 71%.

Discussion

We found that for mega-prostheses as reconstruction in connection with limb-sparing surgery, the probability of five, ten and 15-year implant revision-free survival was 43, 24, and 16%, while the probability of five, ten and 15-year limb survival was 89, 83, and 83%. Aseptic loosening, wear of polyethylene, and infection caused most revisions. According to the Henderson classification [13], mechanical endoprosthetic failures were the most frequent type of failure and with type 3 failures (=structural failures) being the predominant failure mode (Fig. 4).

Limb-salvage surgery is now accepted as fully comparable with amputation, without decreasing long-term survival and also with better functional outcome [1, 14, 15]. Bone sarcomas are the fourth most common cancer in individuals under the age of 25 [16], thereby placing great demands to longevity and function of the reconstructions. Furthermore, cancer patients are more prone to complications, due to the impaired immune system, longer surgery time, and greater loss of tissue and structures [7]. Several studies have explored factors affecting endoprosthetic survival, and our aim was to complement and, hopefully, extend the understanding of prosthesis survival, especially regarding the long-term implant and extremity survival and the long-term level of function that can be obtained with the use of mega-prostheses as reconstruction in connection with limb-sparing surgery.

Earlier and newer studies show that infection and aseptic loosening are the main complications after tumour prosthetic reconstruction [13, 17–22].

We found that deep infection was responsible for 12% of the first revisions and 24% of all revisions. Of the patients, 2% was amputated because of infection, and the remaining experienced limb-salvage resolution, without recurrence of infection at the final follow-up. Our results correlate with earlier findings. Jeys et al. [23] and Rubio et al. [21] reported infection rates of 11 and 13%, respectively. Pala et al. [24] and Mavrogenis et al. [25] reported infection rates of 9.3 and 8.6% respectively. Jeys et al. [23] also reported a high incidence of infection in the first two years, which also correlates with our results. These findings are possibly explained by the ongoing oncologic treatment. Although only one of our patients with deep infection underwent amputation, our results emphasizes that infection is a severe risk factor in limb-salvage surgery on cancer patients.

In our long-term follow up, we also found aseptic loosening as one of the primary modes of endoprosthetic failure. Five patients (10%) had aseptic loosening as the cause of the first revision, and in total, 18 revisions (23%) in eight patients were performed because of aseptic loosening. [18] described higher risk for aseptic loosening in patients with greater bone length resected [18]. They also concluded that aseptic loosening is a midterm complication, i.e., not seen in short-term follow-up studies. In our study, aseptic loosening was seen after an average time of 3.5 years (range 1.1–6.9). Zeegen et al. did a short-term follow-up and did not find a higher risk of implant failure after long resections [22]. Also, several short-term follow-up studies found aseptic loosening as one of the most common failure modes [19, 21]. Our results are comparable with other studies, but too underpowered to

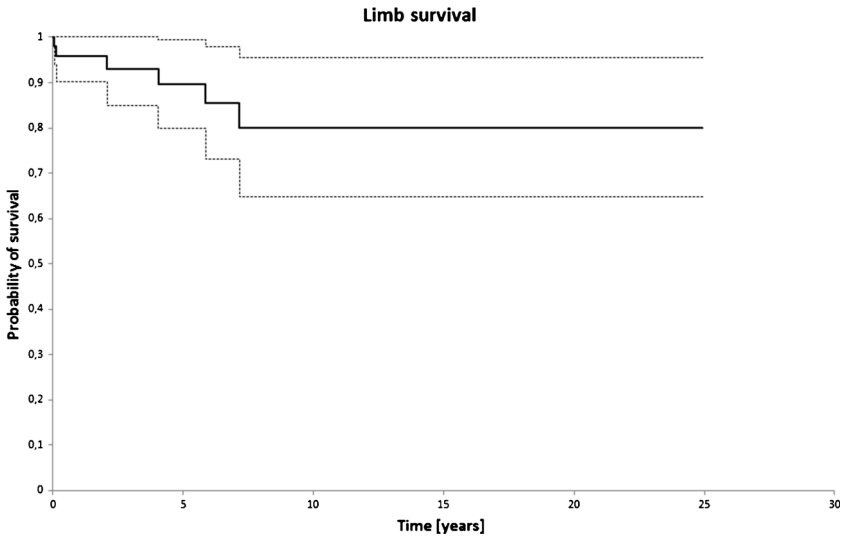


Fig. 3 Kaplan-Meier survival curve with 95% confidence limits showing the probability of limb survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses ($n = 50$)

clarify potential differences. Some studies have found lower risk of aseptic loosening in uncemented prostheses [7, 25, 27–29], while Houdek et al. did not observe that difference [7]. We primarily used uncemented prostheses, but also found aseptic loosening as one of the most common causes of implant failure.

Six patients (12%) had exchange of polyethylene as the first revision, and a total of 16 (32%) patients were revised because of polyethylene wear, and exchange of polyethylene amounted to 21% of all revisions. Revision with change of polyethylene is considered to be minor surgery, but is also known to significantly reduce implant failure [30]. We found no infections related to revision with change of polyethylene, which is emphasized in our fully comparable life and limb five and ten year surviving rates.

Classified according to the Henderson classification [13], our primary failure modes were mechanical ($n = 54$), versus non-mechanical ($n = 17$). Structural failures (type 3) ranked the highest ($n = 26$). We defined polyethylene wear as a structural failure, according to Henderson. Our results correlate with those in literature. Most studies find soft tissue failures as the least common failure [13]. Although not comparable, we found tumour progression (type 5) as the least common failure ($n = 2$), next to soft tissue failure ($n = 10$). Other studies found recurrence of tumor to be the main cause of amputation [31].

Several authors found that survival of endoprosthetic reconstruction is dependent on anatomic site [14, 22, 32, 33]. Henderson et al. [13] found a significant difference in failure mode, depending on location [13]. It is well known that proximal femur reconstructions have better survival than distal femur and proximal tibia reconstructions [14, 20, 22, 24]. Although our cohort is of too limited size for such comparison, which also was not intended, we note that 29 (58%) of our prostheses were distal femur reconstructions, nine (18%) were proximal tibia, nine (18%) were proximal femur, and three reconstructions (6%) were whole femur resections. This could

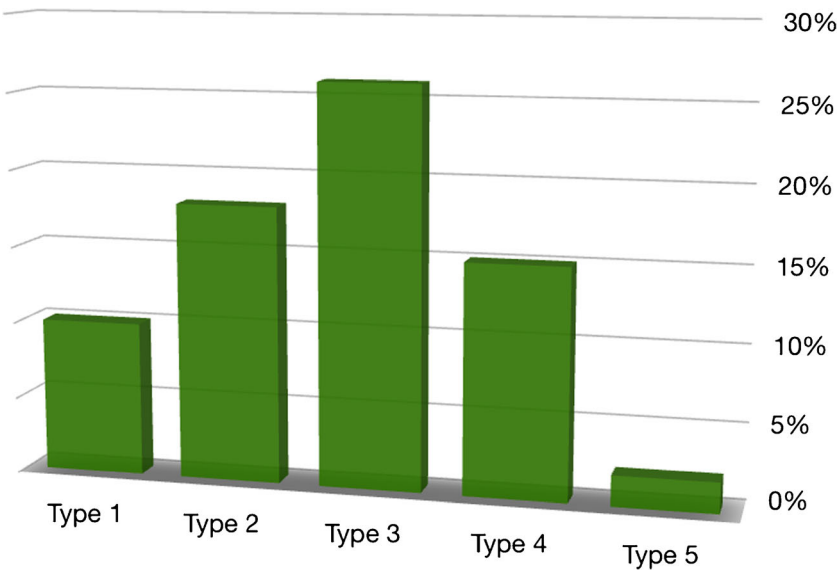


Fig. 4 Chart showing the overall incidence (%) of endoprosthetic failure according to Henderson's five failure modes for all anatomic sites. Type 1 = soft tissue failure (14%), type 2 = aseptic loosening (25%), type 3 = structural failure (37%), type 4 = infection (21%), type 5 = tumour progression (3%). Seventy-six percent mechanical failures vs. 24% non-mechanical failures

be one explanation of our moderate cumulative five and ten year prosthesis survival rate (respectively 43 and 24%). Also, we registered any kind of revision as an event in our Kaplan-Meier survival analysis; thus, also revisions not involving the prosthesis itself and exchange of polyethylene because of wear in well-fixed prostheses were included in the cumulative prosthesis survival rate.

Although we did no comparison between implant sites and amputees, our MSTS median score (71%) is fully comparable with other findings in literature [34–36]. Mean MSTS function for patients with distal femur mega-prosthetic reconstructions ranges in the literature from 78 to 86% [24]. With few exceptions, we evaluated all patients alive ($n = 24$), making our result representable for our cohort. Of those patients evaluated with MSTS, functional score in general was the lowest with a median score of 2.1 (range 0–4). Tun et al. and McGoveran et al. reported the same results [34, 37]. A possible explanation could be that the mega-prosthesis examined are mainly early generation, but also by the necessary soft tissue and bone loss during tumour removal could be of importance. We notice that the moderate functional score did not cause low scores in pain, walking ability, or emotional acceptance. Davis et al. found significantly lower disability after amputation compared to limb-sparing surgery [38], emphasizing the importance of limb-salvage surgery.

Our study has certain limitations that need to be discussed. It is a retrospective study, non-randomized, with a small number of patients, exposing it to selection biases. Differences in tumor histology, stage, and adjuvant treatment are factors that all could affect the results. Further, different manufacturers, implant design, and anatomical sites are heterogenous in our cohort. Also, we had MSTS scores on only 48% of patients. Our study is too underpowered to detect differences between the abovementioned factors, which also was not our aim. Although underpowered, our study is a single-centre study. Different manufactures are used, but only few surgeons were responsible for insertion of all prostheses. Our aim was to address the outcome of our reconstructions, and we consider our results to be useful knowledge in spite of these limitations.

Conclusion

Our long-term evaluation of the treatment regimen, limb-sparing surgery and reconstruction with tumour prostheses in the lower extremities performed in a relatively small orthopaedic oncology center because of primary malignant bone tumor of giant cell tumours of bone, resulted in a low risk of amputation (15-year limb survival probability of 83%) and a good functional outcome (average MSTS score of 21, 9–30 years post-operatively). However, long-term survivors must expect to undergo implant revision surgery, since the 15-year probability of implant revision-free survival was only 16%.

Acknowledgements We thank consultant pathologist Søren Daugaard.

Compliance with ethical standards

Ethical approval The study has been approved by the Danish Data Protection Agency (no. 2013-41-2591) and the Danish Health and Medicines Authority (no. 3-3013-894/1).

Informed consent Informed consent was obtained from all individual participants alive included in the study.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Rougraff BT, Simon MA, Kneisl JS, Greenberg DB, Mankin HJ (1994) Limb salvage compared with amputation for osteosarcoma of the distal end of the femur. A long-term oncological, functional, and quality-of-life study. J Bone Joint Surg Am 76(5):649–656
2. Saeter G, Alvegard TA, Elomaa I et al (1991) Treatment of osteosarcoma of the extremities with the T-10 protocol, with emphasis on the effects of preoperative chemotherapy with single-agent high-dose methotrexate: a Scandinavian Sarcoma Group study. J Clin Oncol 9:1766–1775
3. Rydholm A, Gustafson P, Rooser B et al (1991) Limb-sparing surgery without radiotherapy based on anatomic location of soft tissue sarcoma. J Clin Oncol 9:1757–1765
4. Smeland S, Müller C, Alvegard TA et al Scandinavian Sarcoma Group Osteosarcoma Study SSG VIII: prognostic factors for outcome and the role of replacement salvage chemotherapy for poor histological responders. Eur J Cancer 39:488–494
5. Ferrari S, Smeland S, Mercuri M et al (2005) Neoadjuvant chemotherapy with high-dose ifosfamide, high-dose methotrexate, cisplatin, and doxorubicin for patients with localized osteosarcoma of the extremity: a joint study by the Italian and Scandinavian Sarcoma Groups. J Clin Oncol 23:8845–8852. https://doi.org/10.1200/JCO.2004.00.5785
6. Grimer RJ, Taminiau AM, Cannon SR (2002) Surgical subcommitte of the European osteosarcoma intergroup; Surgical outcomes in osteosarcoma. J Bone Joint Surg Br 84(3):395–400
7. Houdek MT, Wagner ER, Wilke BK et al (2016) Long term outcomes of cemented endoprosthetic reconstruction for periarticular tumors of the distal femur. Knee 23:167–172. https://doi.org/10.1016/j.knee.2015.08.010
8. Hardes J, Gebert C, Schwappach A et al (2006) Characteristics and outcome of infections associated with tumor endoprostheses. Arch Orthop Trauma Surg 126:289–296. https://doi.org/10.1007/s00402-005-0009-1
9. Qadir I, Umer M, Baloch N (2012) Functional outcome of limb salvage surgery with mega-endoprosthetic reconstruction for bone tumors. Arch Orthop Trauma Surg 132:1227–1232. https://doi.org/10.1007/s00402-012-1542-3
10. Schmidt M, Pedersen L, Sørensen HT (2014) The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 29:541–549. https://doi.org/10.1007/s10654-014-9930-3
11. Wada T, Kawai a IK et al (2007) Construct validity of the Enneking score for measuring function in patients with malignant or aggressive benign tumours of the upper limb. J Bone Joint Surg Br 89:659–663. https://doi.org/10.1302/0301-620X.89B5.18498
12. Pala E, Henderson ER, Calabrò T et al (2013) Survival of current production tumor endoprostheses: complications, functional results,

and a comparative statistical analysis. J Surg Oncol 108:403–408. https://doi.org/10.1002/jso.23414
13. Henderson ER, Groundland JS, Pala E et al (2011) Failure mode classification for tumor endoprostheses: retrospective review of five institutions and a literature review. J Bone Joint Surg Am 93:418–429. https://doi.org/10.2106/JBJS.J.00834
14. Torbert JT, Fox EJ, Hosalkar HS et al (2005) Endoprosthetic reconstructions. Clin Orthop Relat Res NA:51–59. https://doi.org/10.1097/01.blo.0000179735.37089.c2
15. Simon MA, Aschliman MA, Thomas N, Mankin HJ (1986) Limb-salvage treatment versus amputation for osteosarcoma of the distal end of the femur. J Bone Joint Surg Am 68:1331–1337
16. Carty CP, Dickinson IC, Watts MC et al (2009) Impairment and disability following limb salvage procedures for bone sarcoma. Knee 16:405–408. https://doi.org/10.1016/j.knee.2009.02.006
17. Mittermayer F, Krepler P, Dominkus M et al (2001) Long-term followup of uncemented tumor endoprostheses for the lower extremity. Clin Orthop Relat Res:167–177
18. Kemp HBS, Sneath RS, Walker PS (1996) Custom-made prosthetic replacements for bone tumours of the lower limb. Bone Joint J 78–B:5–13
19. Morgan HD, Cizik AM, Leopold SS et al (2006) Survival of tumor megaprostheses replacements about the knee. Clin Orthop Relat Res 450:39–45. https://doi.org/10.1097/01.blo.0000229330.14029.0d
20. Gosheger G, Gebert C, Ahrens H et al (2006) Endoprosthetic reconstruction in 250 patients with sarcoma. Clin Orthop Relat Res 450:164–171. https://doi.org/10.1016/S0276-1092(08)70278-5
21. Rubio D, Serrano M, Wang E (2013) Tumour endoprosthetic reconstruction for primary aggressive and malignant bone tumours of the distal femur. Malays Orthop J 7(3):1–5. https://doi.org/10.5704/MOJ.1311.007
22. Zeegen EN, Aponte-Tinao LA, Hornicek FJ et al (2004) Survivorship analysis of 141 modular metallic endoprostheses at early followup. Clin Orthop Relat Res 90025:239–250. https://doi.org/10.1097/01.blo.0000126224.97525.5b
23. Jeys LM, Grimer RJ, Carter SR, Tillman RM (2005) Periprosthetic infection in patients treated for an orthopaedic oncological condition. J Bone Joint Surg Am 87:842–849. https://doi.org/10.2106/JBJS.C.01222
24. Pala E, Trovarelli G, Calabrò T et al (2015) Survival of modern knee tumor megaprostheses: failures, functional results, and a comparative statistical analysis. Clin Orthop Relat Res 473:891–899. https://doi.org/10.1007/s11999-014-3699-2
25. Mavrogenis AF, Pala E, Angelini A et al (2015) Infected prostheses after lower-extremity bone tumor resection: clinical outcomes of 100 patients. Surg Infect 16:267–275. https://doi.org/10.1089/sur.2014.085
26. Unwin PS, Cannon SR, Grimer RJ, Kemp HB, Sneath RS, Walker PS (1996) Aseptic loosening in cemented custom-made prosthetic replacements for bone tumours of the lower limb. J Bone Joint Surg Br 78(1):5–13
27. Flint MN, Griffin AM, Bell RS et al (2006) Aseptic loosening is uncommon with uncemented proximal tibia tumor prostheses. Clin Orthop Relat Res 450:52–59. https://doi.org/10.1016/S0276-1092(08)70276-1
28. Capanna R, Scoccianti G, Frenos F et al (2014) What was the survival of megaprostheses in lower limb reconstructions after tumor resections? Clin Orthop Relat Res 473:820–830. https://doi.org/10.1007/s11999-014-3736-1
29. Schwartz AJ, Kabo JM, Eilber FC et al (2010) Cemented distal femoral endoprostheses for musculoskeletal tumor: improved survival of modular versus custom implants. Clin Orthop Relat Res 468:2198–2210. https://doi.org/10.1007/s11999-009-1197-8
30. Mittermayer F, Windhager R, Dominkus M et al (2002) Revision of the Kotz type of tumour endoprosthesis for the lower limb. J Bone Joint Surg Br 84:401–406
31. Jeys LM, Grimer RJ, Carter SR, Tillman RM (2003) Risk of amputation following limb salvage surgery with endoprosthetic replacement, in a consecutive series of 1261 patients. Int Orthop 27:160–163. https://doi.org/10.1007/s00264-003-0429-x
32. Malawer MM, Chou LB (1995) Prosthetic survival and clinical results with use of large-segment replacements in the treatment of high-grade bone sarcomas. J Bone Joint Surg Am 77:1154–1165
33. Jeys LM, Kulkarni a GRJ et al (2008) Endoprosthetic reconstruction for the treatment of musculoskeletal tumors of the appendicular skeleton and pelvis. J Bone Joint Surg Am 90:1265–1271. https://doi.org/10.2106/JBJS.F.01324
34. Tunn PU, Pomraenke D, Goerling U, Hohenberger P (2008) Functional outcome after endoprosthetic limb-salvage therapy of primary bone tumours—a comparative analysis using the MSTS score, the TESS and the RNL index. Int Orthop 32:619–625. https://doi.org/10.1007/s00264-007-0388-8
35. Aksnes LH, Bauer HCF, Jebsen NL et al (2008) Limb-sparing surgery preserves more function than amputation: a Scandinavian sarcoma group study of 118 patients. J Bone Joint Surg Br 90:786–794. https://doi.org/10.1302/0301-620X.90B6.19805
36. Sharil AR, Usm MO, Nawaz AH et al (2013) Early functional outcome of resection and endoprosthesis replacement for primary tumor around. Knee 7:30–35
37. McGoveran BM, Davis AM, Gross AE, Bell RS (1999) Evaluation of the allograft-prosthesis composite technique for proximal femoral reconstruction after resection of a primary bone tumour. Can J Surg 42:37–45
38. Davis AM, Devlin M, Griffin AM et al (1999) Functional outcome in amputation versus limb sparing of patients with lower extremity sarcoma: a matched case-control study. Arch Phys Med Rehabil 80:615–618. https://doi.org/10.1016/S0003-9993(99)90161-2

Improvement in failure rate after resection of primary bone tumors and reconstruction with second-generation megaprostheses?

Christina Holm[1]

Michala Skovlund Sørensen[1]

Müjgan Yilmaz[1]

Michael Mørk Petersen[1]

[1]The Musculoskeletal Tumor Section

The Department of Orthopedic Surgery

Rigshospitalet,

University of Copenhagen,

Denmark

Correspondence to:
Christina Holm, M.D.
The Department of Orthopedic Surgery,
Rigshospitalet, University of Copenhagen, Denmark
Phone: +45 35452365
Email: christina.holm@dadlnet.dk

Funding: There is no funding source

Keywords: Bone sarcoma, Global Modular Replacement System, Zimmer® Segmental prostheses, survival, revision, tumor prosthesis

Abstract

Objective: Previous multicenter studies reports variable outcome and failure rates after mega-prosthetic reconstructions for tumors in the lower extremities. Purpose of this study was to evaluate if the use of second-generation mega-prosthesis for reconstruction after resection of primary malignancy result in lower incidence of implant revision and limb amputation incidence than first generation prostheses.

Methods: A retrospective single center study of 72 consecutive patients (F/M=30/42), mean age 44 (range 7-84) years with bone or soft tissue sarcomas (n=67) or aggressive benign bone tumors (n=5) having surgery between 2006 and 2016 with bone resection and reconstruction with mega-prostheses. Causes of failure were classified according to Henderson Classification. Kaplan-Meier survival analysis was used for evaluation of overall patient survival. Fine and Gray competing risk analysis was used for assessing cumulative incidence for implant revision and limb amputation incidence. Functional outcome was evaluated with MSTS score.

Results: Forty-seven patients were alive at follow-up. Mean follow-up was 6 years (range 2 – 13 years). Twenty-eight patients (39%) underwent revision for all causes. Overall 10-year patient survival was 61% (95%CI 48-74%); 10-year cumulative risk of major revision was 18% (95%CI 9%-28%). Ten-year cumulative incidence of limb amputation was 11% (95%CI: 3%-18%). The overall predominant failure mode causing revision was non-mechanical (51%). Mean MSTS score was 20 (67%) (range 0-30).

Conclusion: We found no difference in risk of major revision, minor revision or amputation as well as functional outcome when comparing to our first-generation prostheses. For future evaluations of tumor prostheses, we advocate using competing risk analyzes.

Introduction

Advances in treatment of bone sarcomas have over the past three decades led to a gradually increased patient survival from 15% to 60-70% [1]. Although with various follow-up time, some studies even report survival rates of 90% [2]. This improvement is generally a result of the introduction of neo-adjuvant and adjuvant chemotherapy [3]. The possibility to downstage tumors before surgery facilitated the development of limb-sparing tumor resection and reconstruction instead of amputation. This, combined with advances in diagnostic and imaging techniques allowing more accurate pre-surgical planning has resulted in that limb-sparing surgery now is the method of choice offered to 90%-95% of all patients [4].

Over the years there has been a simultaneous development of orthopedic implant designs and possibilities. Due to the mechanical stress in the bone implant interface in first generation fixed-hinge prostheses, the introduction of rotating hinge prostheses has been one of the most central improvements to modern treatment of limb sparing surgery with reconstruction around the knee [5].

As described by Holm et al.[6], a change was made at our clinic in the late 1980's, which gradually substituted amputation as the method of choice for treating bone sarcomas towards limb sparing surgery. As a result of the concurrent improved implant possibilities we at our Center introduced the GMRS® prostheses in 2005 and the Zimmer® Segmental prostheses in 2011 as second-generation prostheses. Both prostheses are modular with cemented and cementless stems for press-fit fixations.

The aim of the current study was to evaluate implant revision, amputation incidence, and functional outcome after the use of second-generation mega-prostheses and compare to our previously used first-generation prostheses. We hypothesized that the improvement in orthopedic implant possibilities had resulted in a lower implant revision incidence, a lower amputation incidence and as well as an improved functional outcome compared to our first-generation prostheses.

Material and Methods

Participants

In order to compare outcomes between two generations of mega-prostheses the current cohort is compared to a previously published cohort [6] from now on called the early cohort (due to an earlier surgery period but from the same center and treated by the same surgons as in current study).

Late cohort

We conducted a retrospective search by manually screening our institutional surgical planning system identifying all patients having limb-sparing surgery with an implant and diagnosed with a primary bone sarcoma, a soft tissue sarcoma located very close to a joint/bone or an aggressive benign bone tumor between January 1, 2006 and December 31, 2016. All patients who underwent reconstruction with mega-prostheses combined with a bone resection for a primary bone sarcoma, a soft tissue sarcoma located very close to a joint/bone or an aggressive benign bone tumor in the lower extremities in that period were included. All patients had complete minimum two-year follow-up with regards to overall survival. Patients who received other surgical treatment than bone resection and reconstruction with mega-prostheses, due to sarcomas or aggressive benign bone tumors were not included in the study, and we included patients of all ages.
We identified 72 patients (F/M=30/42, mean age 44 (range 7-84) years) who underwent limb-sparing surgery and reconstruction with a mega-prosthesis for a bone sarcoma (n=60), a soft tissue sarcoma (n=9) or an aggressive benign bone tumor (n=3). Patient characteristics and tumor histology are summarized in **Table 1.** The anatomic sites of reconstruction were: distal femur (n=33) proximal femur (n=24), proximal tibia (n=12) and entire femur (n=3). The following types of implants were used for reconstruction: GMRS (Stryker®) (n=37), Segmental (Zimmer®Biomet) (n=27), Mega C (Link®) (n=5), Link custom made growing prostheses (Link®) (n=3).

Early cohort

Fifty patients who underwent resection and reconstruction with mainly first-generation mega-prostheses in the lower extremities at our musculoskeletal tumor center between 1985 and 2005 due to a primary bone sarcoma (n=44) or an aggressive benign bone

tumor (n=6), has previously been described in detail by Holm et al. [6], and patient characteristics and tumor histology are summarized in **Table 1**.

Ethics

The study was approved by the Danish Data Protection Agency (j.nr: 2012-58- 0004) and the Danish Health and Medicine Authority (3-3013-2578/1). Informed consent was obtained from all individual participants still alive at inclusion in the study.

Variables

From patient records we obtained: gender, age, date of diagnosis, date of surgery, anatomical tumor site and type of implant. Also, from patient records we registered all subsequent types of surgeries related to the prosthesis. Due to the Danish Civil Registration System no patients were lost to follow up and exact date of death was known for all patients [7]. From the Danish National Pathology Registry (DNPR) [8] we found histopathological diagnosis and date for debut of cancer. For some patients, histopathological diagnosis was defined post-surgically after biopsy was obtained during the surgical procedure where the mega-prosthesis was implanted.

For patients alive in the study period, postoperative functional outcome was evaluated with the Enneking score (MSTS-score) [9].

Implant follow-up

Major revisions were defined as all unplanned implant related surgeries with removal or exchange of bone anchored components in a primary implant, or amputation of the extremity for any cause. Minor revisions were defined as all implant-related surgeries without removal of bone-anchored parts, such as change of polyethylene, repositions with or without insertion of a constrained liner, brisement forcé or local recurrence without contamination of the prosthesis. Both closed and open hip repositions due to dislocation were defined as revisions. Also, all DAIR (debridement, antibiotics and implant retention) surgeries were registered as minor revisions and two-stage surgeries due to deep infection were considered as two revisions. All planned extensions of growing prostheses were not defined as revisions. Minor superficial and aseptic wound revisions were not registered. All revisions were registered until death or end of the study period December 31, 2018.

Furthermore, endoprosthetic complications and failures were classified according to the Hendersons Failure mode Classification [10].

Statistical Analysis

Overall patient survival was calculated based on day of surgery until death or end of study December 31, 2018. Kaplan-Meier survival analysis with right censoring was used for evaluation of the probability of overall patient survival. Log-rank test was used for comparison of survival between groups.

Since the Kaplan-Meier method assumes identical risk in censored and uncensored patients the Aalen-Johansson estimator was used to assess the cumulated incidence of major and minor implant revisions calculated by a competing risk model with death and amputation as competing risks, and with death as competing risk when calculating the cumulative incidence of amputation. Cumulative incidence of failures according to the Henderson Classifications was calculated using the competing risk analysis. Grays test, log-rank test and Chi square test were used to assess differences between groups. Confidence intervals are reported as 95% confidence intervals (95% CI) and p-values <0.05 are considered statistically significant.

Statistical analysis was performed using software R (R Foundation, Vienna, Austria).

Results

Overall patient survival

In the late cohort 25 patients (35%) had died at the end of study, after an average of 2.1 years (109 days-7 years) following surgery. Forty-seven patients (65%) were alive with a mean follow-up of 6 years (2-13 years). Causes of death have not been explored further.

The probability of over-all 5- and 10-year survival for the present late cohort was 68% (CI95% 57%-79%) and 61% (CI95% 48%-74%) respectively **(Fig. 1).**

Our study showed no difference in overall survival between the early and late cohort (p =0.93) **(Fig. 1)**.

Limb Survival

Eight patients in the late cohort were amputated (11%) (resection site: knee (n=5); hip (n=3)). Seven patients (10%) were amputated due to recurrence of tumor, and one

patient (1%) due to acute ischemia. The 2, 5, and 10-year cumulative incidence of amputation was 8% (CI: 2%-15%), 8% (CI: 2%-15%) and 11% (CI: 3%-18%), respectively **(Fig. 2)**. We found no difference comparing cumulative incidence for amputation between the early and late cohort (p = 0.9) (**Fig. 2).**

Incidence of revisions

In the late cohort 28 patients (39%) underwent revision surgery for all causes. Average time from primary surgery to first revision was 1.2 years (1 day – 7.4 years). Major revision by anatomic site included knee (n=10), hip (n=2) and for minor revision, knee (n=7), hip (n=6) and total femur (n=1). A total of 50 revisions were conducted in the late cohort. Distribution of all revisions for both cohorts, are described in **Table 2**. Main causes for first revision in general in the present late cohort were deep infection (n=6; 8%) followed by wear of polyethylene (n = 5; 7%) and aseptic loosening (n=5; 7%).

The 2, 5 and 10-year cumulative incidence of major revision was 11% (CI: 4%-18%), 16% (CI: 7%-25%) and 18% (CI: 9%-28%) respectively (**Fig. 3**). The 2, 5 and 10- year cumulative incidence of minor revision was 15% (CI: 7%-24%), 20% (CI: 11%-30%) and 25% (CI: 14%-36%) respectively **(Fig. 4)**.

When comparing cumulative incidence for minor and major revision between early and late cohort, we found no difference (p =0.9 and p=0.2 respectively) **(Fig. 3,4)**.

Average time from surgery to major revision was 4.1 years (range: 17 days- 12.5 years).

Deep infections

In the late cohort 8 patients (11%) had a total of 17 major (n=15) and minor (n=2) revisions due to deep infection. The 5- and 10-year cumulative incidence for deep infection was 11% (CI: 4%-19%) and 11% (CI: 4%-19%) respectively (**Fig.5**). We found no difference between the two cohorts (p=0.9) (**Fig.5**). Seven patients (10%) had deep infection in a primary implant; three patients (4%) had re-infection and their secondary implant and one patient (1%) had infection in a secondary implant inserted for other causes than infection. The 5- and 10-year cumulative incidence for deep infection after revision was 17% (CI: 2%-32%) and 17% (CI:2%-32%) respectively (**Fig.6**). We found no difference for deep infection after revision between the two cohorts (p=0.81) (**Fig.6**).

Henderson Classification

Complications according to Henderson et al. [10] are summarized in Table 3. Out of a total of 50 major and minor revisions according to the Henderson classification 39 could be classified.

The overall predominant failure mode was non-mechanical (n=20, 51%), whereas mechanical failures constituted 49% (n=19) (**Table 3**). We found no differences in failure mode type 1-5 between the early and late cohort (**Table 3**).

Functional outcome

In the late cohort 47 patients were alive during the entire study period, and functional outcome was evaluated in 40 after an average of 6 (1.7-12) years postoperatively. Seven of the 47 patients did not have a functional evaluation because they had been amputated (n=3) or were lost to follow-up for various reasons (n=4). The mean (range) individual MSTS score parameters were: pain 3.5 (0-5), function 2.6 (0-5), emotional acceptance 3.5 (0-5), supports 3.8 (0-5), walking ability 3.7 (0-5), and gait 3 (0-5). Mean MSTS score was 20.2 (range 6-30), representing a mean score of 67%.

Discussion

The probability of over-all 5- and 10-year survival of our late cohort was 68% and 61% respectively. The 2-, 5- and 10-year cumulative incidence of major revisions was 11%, 16% and 18% respectively. The 2-, 5- and 10-year cumulative incidence for amputation was 8%, 8% and 11% respectively. Deep infection and recurrence of tumor caused most revisions. According to the Henderson classification non-mechanical endoprosthetic failures were the most frequent type of failure.

Our study showed no improved overall patient survival in the late cohort, although our results are similar to other studies investigating patient survival [5]. Also, although not significant, we found decreased incidences of deep infections (36% vs. 10%) and aseptic loosening (38% vs. 24%) in present cohort compared to our first-generation prostheses. We speculate that these findings are partly due to the improvements in implant and imaging technology. Furthermore, as opposed to the previous decades 95% of all patients with bone sarcomas are now offered limb-sparing surgery despite poor life

expectancy [4], and hence the heterogeneity in terms of diagnosis, staging and adjuvant oncological treatment undoubtedly has increased the comorbidity and risk for post-surgical complications. We therefore hypothesize that our negative statistical findings could be a result of selection bias over time between the two cohorts or loss of power in analysis due to small sample size. Also, the difference in revision rates reported from previous studies could in part be explained by the lack of using a competing risk model for implant revision [11]. We believe that this is particularly important when estimating risks for patients with sarcoma because of their high mortality rates and increased risk of revision over time due to implant wear-out, especially in long-term follow up studies.
In a systematic review Thornley et al. [12] recently described low quality of reporting and inconsistency with regards to follow up and surveillance among studies, hence limiting interstudy comparison due to high heterogenicity. Henderson et al. [10] has suggested five classifications of failure mode in order to obtain consistency. However, any comparison across studies between five subgroups with competing risk analysis will be limited due to the often, small sample sizes, and also inconsistence reporting. Previous findings with chi-square test, does not take revisions over time into consideration. To detect any potential differences across studies, we suggest competing risk, and furthermore broader consistency with fewer categories and thus larger sample sizes.

The reported literature on the two prostheses used in the current study is sparse. Pala et al. [13] and Yilmaz et al. [14] evaluated the rotating-hinge Global Modular Reconstruction System (GMRS) prosthesis. They report higher implant survival than us; however, Pala et al. excluded soft tissue failures and revisions caused by local recurrence. Although our results are not directly comparable to Pala et al. [13], both studies reveal reduced revision rates compared to fixed hinge prostheses.

It is well-known that it is difficult to achieve free margins with soft tissue sarcomas since they are often poor circumscribed [15]. Recurrence of soft tissue sarcomas or highly malignant bone sarcomas with significant soft tissue components was the main cause for amputation in the late cohort of the present study (n=6), where 8 patients had amputation (10%); thus this is comparable to previous studies reporting various rates from 6% to 23% [16]. In contrast to present late cohort, patients in the early cohort, having soft tissue sarcomas or highly malignant osteosarcomas with large soft tissue

components were not offered limb sparing surgery [6], and this is probably an explanation for the initially higher cumulative incidence for amputation and poorer overall survival after amputation in the late cohort, although not statistically significant (**Fig.2 c**).

Incidence of aseptic loosening has been reported with decreasing rates from 2% to 11% in recent studies evaluating rotating-hinge prostheses [5,17,18]. In our study aseptic loosening constituted fewest major revisions in contrast to the early cohort [6]. Also, we found a difference between Henderson type 3 failures i.e. wear of polyethylene and stem-fractures. We consider our results to support the hypothesis that rotating hinge prostheses reduces the torsional stress in the bone-implant interface hence reducing aseptic loosening and stem breakage, although our result undoubtedly also reflects sparse follow-up time.

The most common cause for revision in the late cohort of our study was deep infection with distal femur and proximal tibia as the most common sites. We found no difference in the risk of deep infection between the two cohorts. All patients had salvage after revision for deep infection with no need for subsequent amputation at final follow up. However, the 5- and 10-year incidence for deep infection after revision was 17%, with deep infection constituting the majority of the prior revisions. We found no difference in overall survival between patients with or without infection. Although all patients had salvage after deep infection our results emphasizes that infection remains a severe complication after reconstruction despite improved adjuvant oncological treatment and implant possibilities. Silver coated implants and iodine surface coating has been suggested [19] and in some studies also showed promising results reducing infection rate [2].

We found no improvement in mean MSTS score in the late cohort compared to the early cohort. Reviewing literature mean MSTS score ranges from 66%-82% [20,21]. We speculate that the lack of improvement with second-generation prostheses partly reflects that only patients with high functional status in the early cohort were offered limb sparing surgery.

Limitations

Some limitations to our study need to be discussed. Our study is a retrospective, nonrandomized study prone to selection bias. However, to the best of our knowledge randomized controlled trials to evaluate and determine optimal treatment with reconstruction for these patients is not an option. Furthermore, comorbidity is without doubt also of great importance. Although comparisons were drawn between subgroups, present cohort was heterogeneous in terms of diagnosis, staging and adjuvant oncological treatment, which undoubtedly affected outcome and hence the interpretation and comorbidity of the patients was not evaluated due to study design. Due to the varying variables, our study admittedly is too underpowered to detect all intended potential statistical differences. Nonetheless, we found markedly lower incidence rates of major revisions between the cohorts. Furthermore, our study is a single center study with consistent follow up and reporting of outcomes. Also, all patients having limb-sparing surgery with reconstruction in the lower extremities were included with none lost to follow up.

Conclusion

We found that the 2-, 5- and 10-year cumulative incidence of major revision was 11%, 16% and 18% respectively. The 2-, 5- and 10-year cumulative incidence for amputation was 8%, 8% and 11% respectively. Deep infection and recurrence of tumor caused most revisions. We found no significant difference in risk of major revision despite lower incidences. We found no significant difference in minor revision or amputation as well as regarding functional outcome when comparing to our first-generation prostheses. For future evaluations of tumor prostheses, we advocate using competing risk analyzes in order to achieve valid estimate of implant revision and limb survival, in order to conduct direct insterstudy comparison.

Acknowledgement:

Conflict of interest: The authors declare that they have no conflict of interest

References

[1] J.A.M. Bramer, J.H. Van Linge, R.J. Grimer, R.J.P.M. Scholten, Prognostic Factors in localized Extremity Osteosarcoma; a Systematic Review, 2009. https://pure.uva.nl/ws/files/1478580/78023_05.pdf (accessed September 13, 2018).

[2] K.A. Goulding, C.L. Gaston, R.J. Grimer, Outcomes and Options for Prosthetic Reconstruction After Tumour Resection About the Knee, Curr. Surg. Reports. 2 (2014) 42. doi:10.1007/s40137-013-0042-x.

[3] G. Bacci, S. Ferrari, F. Bertoni, P. Ruggieri, P. Picci, A. Longhi, R. Casadei, N. Fabbri, C. Forni, M. Versari, M. Campanacci, -term outcome for patients with nonmetastatic osteosarcoma of the extremity treated at the istituto ortopedico rizzoli according to the istituto oLongrtopedico rizzoli/osteosarcoma-2 protocol: an updated report, J. Clin. Oncol. 18 (2000) 4016. doi:10.1200/JCO.2000.18.24.4016.

[4] L.H. Aksnes, H.C.F. Bauer, N.L. Jebsen, G. Follerås, C. Allert, G.S. Haugen, K.S. Hall, Limb-sparing surgery preserves more function than amputation: a Scandinavian sarcoma group study of 118 patients., J. Bone Joint Surg. Br. 90 (2008) 786–794. doi:10.1302/0301-620X.90B6.19805.

[5] E. Pala, E.R. Henderson, T. Calabrò, A. Angelini, C.N. Abati, G. Trovarelli, P. Ruggieri, Survival of current production tumor endoprostheses: Complications, functional results, and a comparative statistical analysis, J. Surg. Oncol. (2013). doi:10.1002/jso.23414.

[6] C.E. Holm, C. Bardram, A.F. Riecke, P. Horstmann, M.M. Petersen, Implant and limb survival after resection of primary bone tumors of the lower extremities and reconstruction with mega-prostheses fifty patients followed for a mean of forteen years, Int. Orthop. 42 (2018) 1175–1181. doi:10.1007/s00264-018-3861-7.

[7] M. Schmidt, L. Pedersen, H.T. Sørensen, The Danish Civil Registration System as a tool in epidemiology, Eur. J. Epidemiol. 29 (2014) 541–549. doi:10.1007/s10654-014-9930-3.

[8] B. Bjerregaard, O.B. Larsen, The Danish pathology register, Scand. J. Public Health. (2011). doi:10.1177/1403494810393563.

[9] T. Wada, A. Kawai, K. Ihara, M. Sasaki, T. Sonoda, T. Imaeda, T. Yamashita, O. Surgeon, Construct validity of the Enneking score for measuring function in patients with malignant or aggressive benign tumours of the upper limb, J Bone Jt. Surg [Br]. 89 (2007) 659–63. doi:10.1302/0301-620X.89B5.

[10] E.R. Henderson, J.S. Groundland, E. Pala, J.A. Dennis, R. Wooten, D. Cheong, R. Windhager, R.I. Kotz, M. Mercuri, P.T. Funovics, F.J. Hornicek, H.T. Temple, P. Ruggieri, G.D. Letson, Failure mode classification for tumor endoprostheses: retrospective review of five institutions and a literature review., J. Bone Joint Surg. Am. 93 (2011) 418–29. doi:10.2106/JBJS.J.00834.

[11] E. Pala, G. Trovarelli, A. Angelini, M. Maraldi, A. Berizzi, P. Ruggieri, Megaprosthesis of the knee in tumor and revision surgery, Acta Biomed. 88 (2017) 129–138. doi:10.23750/abm.v88i2.

[12] P. Thornley, M. Vicente, A. Macdonald, N. Evaniew, M. Ghert, R. Velez, Causes and Frequencies of Reoperations After Endoprosthetic Reconstructions for Extremity Tumor Surgery : A Systematic Review, 2118 (2019) 894–902. doi:10.1097/CORR.0000000000000630.

[13] E. Pala, G. Trovarelli, T. Calabr??, A. Angelini, C.N. Abati, P. Ruggieri, Survival of Modern Knee Tumor Megaprostheses: Failures, Functional Results, and a Comparative Statistical Analysis, Clin. Orthop. Relat. Res. 473 (2015) 891–899. doi:10.1007/s11999-014-3699-2.

[14] M.M. Petersen, Long - term results of the Global Modular Replacement System tumor prosthesis for reconstruction after limb - sparing bone resections in orthopedic oncologic conditions : Results from a national cohort, (2019) 1–10. doi:10.1002/jso.25490.

[15] M.A.N.H. Shiu, E.L.B. Castro, S. Hajdu, J.G. Fortner, Surgical Treatment of 297 Soft Tissue Sarcomas of the Lower Extremity, (1975) 597–602.

[16] M.T. Houdek, E.R. Wagner, B.K. Wilke, C.C. Wyles, M.J. Taunton, F.H. Sim, Long term outcomes of cemented endoprosthetic reconstruction for periarticular tumors of the distal femur, Knee. 23 (2016) 167–172. doi:10.1016/j.knee.2015.08.010.

[17] F. Mittermayer, P. Krepler, M. Dominkus, E. Schwameis, M. Sluga, H. Heinzl, R. Kotz, Long-term followup of uncemented tumor endoprostheses for the lower extremity., Clin. Orthop. Relat. Res. (2001) 167–177.

[18] G.J.C. Myers, A.T. Abudu, S.R. Carter, R.M. Tillman, R.J. Grimer,

Endoprosthetic replacement of the distal femur for bone tumours: LONG-TERM RESULTS, J. Bone Jt. Surg. - Br. Vol. 89-B (2007) 521–526. doi:10.1302/0301-620X.89B4.18631.

[19] J.H. Healey, CORR Insights™: High Infection Rate Outcomes in Long-bone Tumor Surgery with Endoprosthetic Reconstruction in Adults: A Systematic Review, Clin. Orthop. Relat. Res. 471 (2013) 2028–2029. doi:10.1007/s11999-013-2893-y.

[20] P.U. Tunn, D. Pomraenke, U. Goerling, P. Hohenberger, Functional outcome after endoprosthetic limb-salvage therapy of primary bone tumours - A comparative analysis using the MSTS score, the TESS and the RNL index, Int. Orthop. (2008). doi:10.1007/s00264-007-0388-8.

[21] P. Ruggieri, G. Bosco, E. Pala, C. Errani, M. Mercuri, Local recurrence, survival and function after total femur resection and megaprosthetic reconstruction for bone sarcomas, Clin. Orthop. Relat. Res. 468 (2010) 2860–2866. doi:10.1007/s11999-010-1476-4.

Table 1. *Summarizing patient characteristics*

	All patients	Early cohort	Late cohort	p-value
No of patients	*n=122*	*n=50*	*n=72*	
Female/male	*54/68*	*24/26*	*30/42*	*p =0.58**
Patients alive at end of study	*75*	*28*	*47*	*p =0.35**
Mean age at surgery (range)	*39 (6-84)*	*34 (6-74)*	*44 (7-84)*	*p = 0.02#*
Location				
Hip	*33 (27%)*	*9 (18%)*	*24 (33%)*	*p =0.06**
Knee	*83 (68%)*	*38 (76%)*	*45 (63%)*	*p =0.17**
Total femur	*6 (5%)*	*3 (6%)*	*3 (4%)*	*p =0.68**
Total number of revisions	*137*	*78*	*59*	*p =0.27**
Angiosarcoma	*3 (3%)*	*1 (2%)*	*2 (3%)*	*N/A*
Ewing sarcoma	*8 (7%)*	*4 (8%)*	*4 (6%)*	*N/A*
Giant cell	*9 (7%)*	*6 (12%)*	*3 (4%)*	*N/A*
Myofibrosarcoma	*5 (4%)*	-	*5 (8%)*	N/A
Synovial cell sarcoma	*1 (1%)*	*1(2%)*	*1 (1%)*	N/A
Rhabdomyosarcoma	*1 (1%)*	-	*1 (1%)*	N/A
Leiomyosarcoma	*1 (1%)*	-	*1 (1%)*	N/A
Sarcoma NOS	*5 (4%)*	-	*5 (7%)*	N/A

#student unpaired t-test, *Fishers exact test

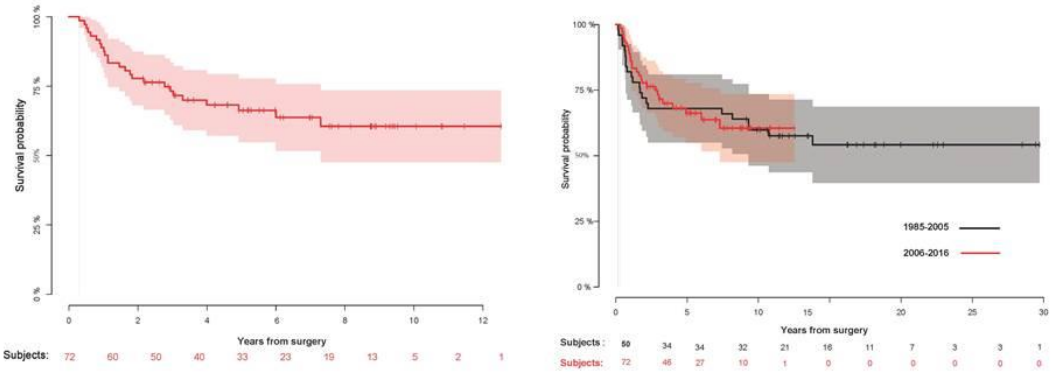
Table 2. *Causes and numbers of performed revisions in general*

	All patients	Early cohort	Late cohort	p-value
Number of revisions	n=128	n=78	n=50	0.31
Aseptic loosening	25	18	7	0.33
Polyethylene wear	20	16	4	0.007
Deep infection	36	19	17	0.6
Instability	15	10	5	1
Fractured stem	7	5	2	0.6
Periprosthetic fracture	4	4	0	0.05
Recurrence or progression of tumor	12	2	10	0.20
Severe symptoms from earlier revision	1	1	0	1
Compartment	1	1	0	0.4
Reduced function (brisement forcé)	3	0	3	1
Ulceration through acetabulum	1	1	0	0.4
Ischemia	1	0	1	1
Unknown	1	1	0	0.8

Table 3. *Showing the Aalen-Johansson estimate of the 5- and 10-year risk of complications to surgery classified according to Henderson et al. [10] of both cohorts, using Grays test to assess the difference between cohorts.*

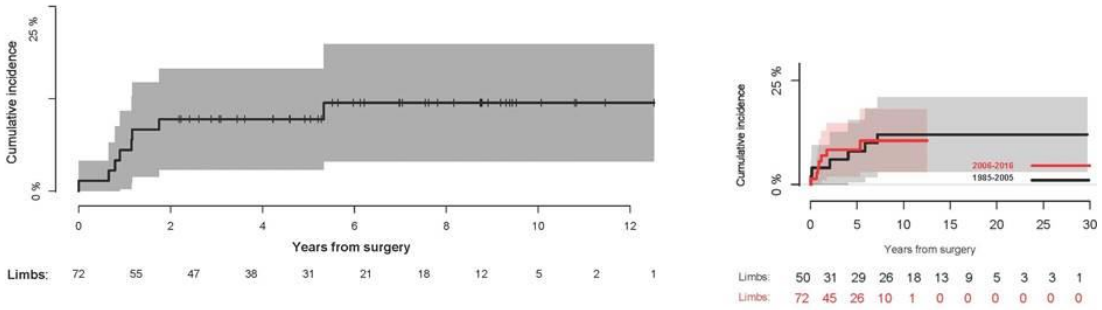
	No. primary implants	***Type 1*** *(Soft-tissue failure)*	***Type 2*** *(Aseptic loosening)*	***Type 3*** *(Structural failure)*	***Type 4*** *(Infection)*	***Type 5*** *(Tumor progression)*	*Total complications*
Old cohort	n=50	n=10 (20%))	n=18 (36%)	n=26 (52%)	n=15 (30%)	n=2 (4%)	n=71
Cum.inc.							
5 year		8.0%	12.0%	10.0%	10.0%	2.0%	
10 year		8,0%	14.0%	24.0%	14.0%	4.0%	
Late cohort	n=72	n=5 (7%)	n=6 (8%)	n=8 (11%)	n=15 (21%)	n=6 (8%)	n=39
Cum.inc.							
5 year		4.7%	4.5%	6.0%	10%	8.3%	
10 year		4.7%	9.0%	11.0%	10%	8.3%	
Grays test		p =0.4	p =0.3	p =0.2	p =0.6	p =0.3	

Fig. 1



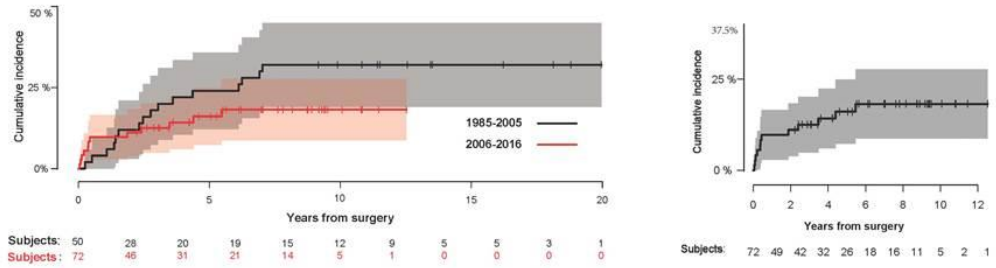
Probability of overall survival for late patient cohort (n=72) (left). Overall survival between patients in the early and late cohort, *p*=0.929 (right)

Fig. 2



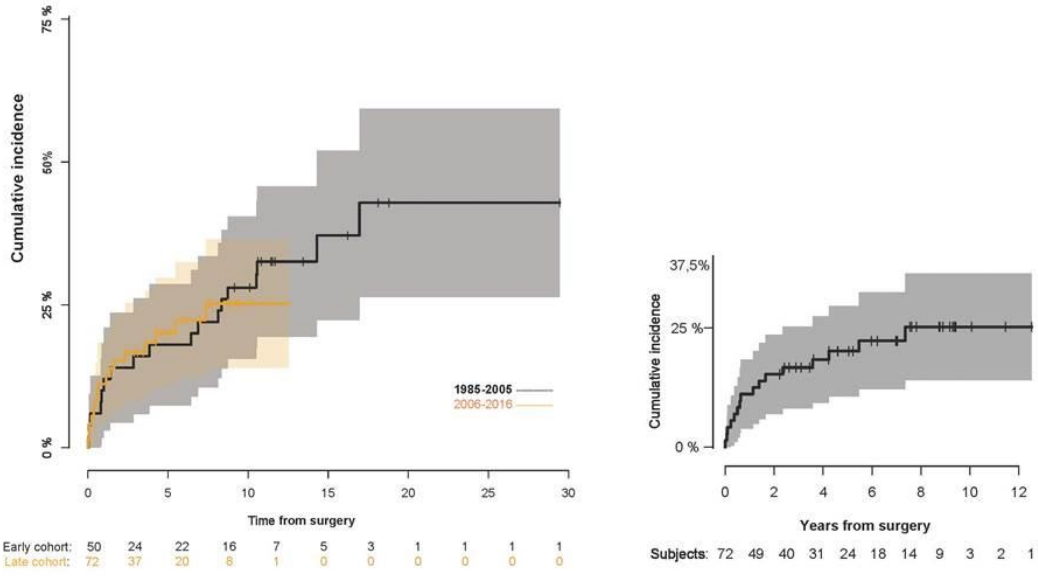
The Aalen-Johansson estimate of the risk of amputation in the late cohort (left) and the risk of amputation between patients in the early and late cohort *p*=0.865 (right).

Fig. 3



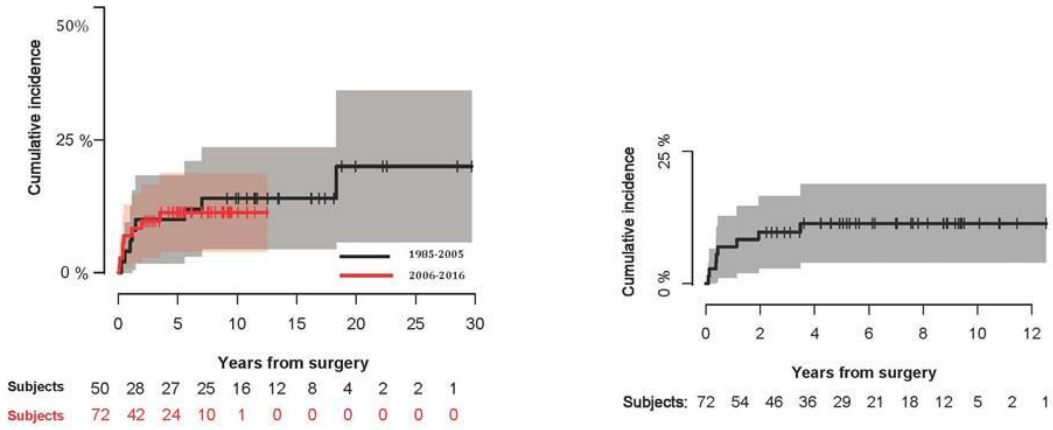
The Aalen-Johansson estimate for risk of major revision in early and late cohort (left), and major revision in the late cohort (right).

Fig. 4



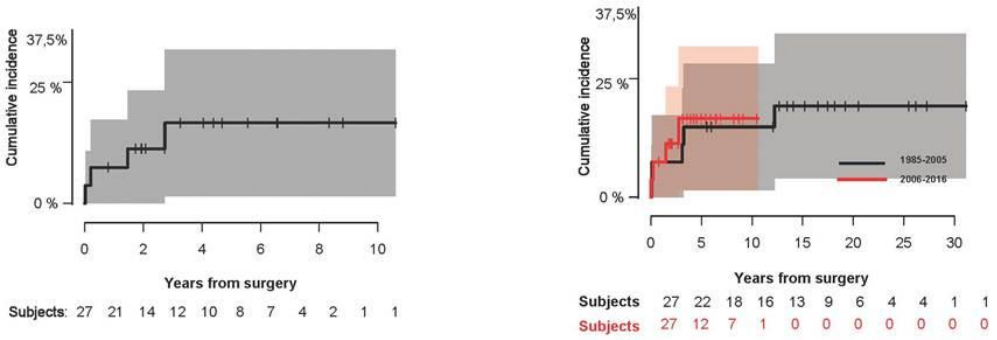
The Aalen-Johansson estimate for risk of minor revision in early and late cohort (left), and minor revision in the late cohort (right).

Fig. 5



The Aalen-Johansson estimate for risk of revision of deep infection in the early and late cohort *p*= 0.831 (left) and the risk of deep infection in late cohort (right).

Fig. 6



The Aalen-Johansson estimate for risk of deep infection after revision in the late cohort (left) and the risk of deep infection after revision between early and late cohort *p*=0.81(right).

Quantitative Measurements of Adaptive Bone Remodeling Around the Cemented Zimmer® Segmental Stem After Tumor Resection Arthroplasty Using DXA.

Christina Holm[1], M.D. Ph.D. stud.

Peter Horstmann[1], M.D. Ph.D.

Michala Skovlund Sørensen[1], M.D. Ph.D

Karen Dyreborg[1], M.D. Ph.D. stud.

Michael Mørk Petersen[1], professor, DMSc

[1]Department of Orthopedic Surgery

Rigshospitalet

University of Copenhagen

Denmark

Corresponding author:

Department of Orthopedic Surgery

[1]Rigshospitalet, University of Copenhagen,

Blegdamsvej 9, DK-2100 Copenhagen Ø

Denmark

Email: christina.holm@dadlnet.dk

Phone: +45 35453545

Abstract

Limb salvage surgery (LSS) is the preferred method for treatment of patients with sarcomas and to a greater extent also to patients with metastatic bone disease. The aim of the present study was to evaluate the adaptive remodeling of the periprosthetic cortical bone after insertion of a tumor prosthesis with cemented stem.

A prospective study of 21 patients (F/M=12/9), mean age 55 years (range 15-81) with metastatic bone disease (n=9), sarcomas (n=8) or aggressive benign tumors (n=4) who underwent bone resection due to a tumor, and reconstruction with a tumor-prosthesis (Zimmer® Segmental 130 mm straight fluted cemented stem with trabecular meta (TM) collars) in the proximal femur (n=10), distal femur (n=9) or proximal tibia (n=2) . Measurements of bone mineral density (BMD) (g/cm^2) were done postoperatively and after 3, 6, and 12 months using dual-energy X-ray absorptiometry. BMD was measured in 4 regions of interest around the cemented stem and in one region of interest 1 cm proximal from the ankle joint of the affected limb and measurement of the contralateral ankle was used as reference. Repeated measures ANOVA and students paired t-test was used to evaluate BMD changes over time.

At 1-year follow-up, BMD decreased compared to baseline in all four regions of interest with a statistically significant bone loss of 8-15%. The bone loss was most pronounced (14-15%) in the 2 regions of interest closest to the trabecular meta (TM) collar and lowest (8%) adjacent to the tip of the stem.

After one year the decrease in bone mineral density of the ankle on the affected limb was 9% and the contralateral ankle was close to baseline, thus suggesting that the periprosthetic bone mineral density changes during follow-up, mainly are caused by stress shielding and immobilization.

Keywords: Tumor prostheses, Stress shielding, bone mass density, dual-energy X-ray absorptiometry

Introduction

Limb sparing surgery (LSS) is today the preferred surgical treatment of bone sarcomas in the lower extremities [12]. The same patient survival is reported if LSS is performed in the majority of cases instead of amputation [9]. Following bone tumor resection, reconstruction of the affected limb is usually done using tumor prostheses in order to save the function of the affected limb. The same technique (LSS and tumor prostheses) are increasingly applied in the treatment of patients suffering from bone destruction because of metastatic bone disease (MBD) [30].

Modern tumor prostheses are attached to the bone using cemented or uncemented intramedullary stems. After implantation of the stem, periprosthetic loss of bone stock in close relation to the stem is to be expected for various reasons such as the bone reaction to the operative trauma, postoperative immobilization, chemotherapy, and stress shielding [7, 10, 13]. Compared to ordinary primary hip or knee arthroplasty, LSS and reconstruction using tumor prostheses causes greater operative trauma and prolonged rehabilitation. Furthermore, patients will often be in need of chemotherapy prior and after surgery. Stress shielding after primary hip arthroplasty is well known and occurs around the well-fixed un-cemented and cemented stems and is characterized by thinning of the compact diaphyseal bone adjacent to the stem [14]. Stress shielding represents a considerable clinical problem after insertion of tumor prostheses due to increased risk of periprosthetic fracture, and also greater complexity in case of revision [18].

Stress shielding after joint replacement surgery has previously been reported using quantitative densitometric techniques in several studies [1, 22, 26]. Dual-energy x-ray absorptiometry (DXA) [24, 35] has been used extensively for quantitative and precise measurements of changes in bone mineral density (BMD) in close relation to both cemented and un-cemented orthopedic implants [17, 19, 25]. Only three studies [2, 7, 18] of partly cross-sectional design (with no immediate postoperative measurements performed) or with a very limited number of patients with inserted tumor prostheses because of malignant bone tumor resection have been published and no real prospective quantitative measurements of the adaptive bone remodeling around the fixation stems of tumor prostheses exist.

In that perspective the aim of the present study was, in a prospective design using DXA, to quantitatively measure the adaptive bone remodeling around the intramedullary 130mm Zimmer® Segmental straight fluted cemented stem in patients with malignant bone tumors receiving tumor prostheses. We hypothesize that the use of trabecular metal (TM) collars together with the

intramedullary 130 mm cemented Segmental stem will secure an optimal stem fixation, thus reducing stress shielding of the periprosthetic cortical bone compared to the sparse previous reports.

Material and Methods

Patient Population

Between January 1, 2015 and July 1, 2018, 33 patients who underwent bone tumor resection with LSS and reconstruction with a Zimmer® Segmental System tumor prosthesis (Zimmer Biomet) in the lower extremities, were evaluated for inclusion in the study (**Fig. 1**). It was predefined to exclude patients with age < 15 years, patients with diseases severely affecting the bone metabolism and patients with expected survival below 1 year (estimated by the surgeon and the investigators). Twelve patients were excluded for various reasons, and 21 patients (F/M=12/9, mean age 55 years) diagnosed with a primary bone tumor (n=6), an aggressive benign tumor (n=4), myelomatosis (n=2) or MBD (n=9) (**Table 1**) completed 1-year follow-up (**Fig. 1).** All surgeries were carried out by, or under supervision of, an experienced tumor joint replacement surgeon at a tertiary referral center for orthopedic oncology. All reconstructions were done using the Zimmer® Segmental tumor prostheses with an intramedullary 130mm straight fluted stem for cementation and a TM collar (**Fig. 2**). Patients were mobilized with full weight-bearing using crutches the day after surgery. Clinical evaluation of the treatment was conducted by using the Enneking score (MSTS-score) [34] after 3, 6 and 12 months.

DXA Evaluation

BMD (g/cm^2) of the periprosthetic bone of the femur or tibia around the stem and adjacent to the TM collar was measured by DXA using a Norland XR-46 scanner (scan resolution 0.5 x 0.5 mm, scan speed 45mm/s) postoperatively and after 3, 6, and 12 months. All patients were placed supine with the femur in neutral rotation during scanning. On the computerized scan-plots, we selected three regions of interest (ROI) around the stem in the femoral or tibial bone and one ROI adjacent to the TM collar for measurements of local changes in BMD over time around the fixation stem: a 2.5-3 cm long area for the bone adjacent to the TM collar (ROI 1), a 5 cm area comprising the middle part of the stem (ROI 2), a 5 cm area comprising the distal part of the stem (ROI 3) and a 3 cm long area comprising the bone adjacent to the tip of the stem (ROI4) (**Fig.3**). A custom-made metal exclusion software facility, which allows a variable threshold for metal exclusion, was used for scan analysis. The threshold (range: 4.0 g/cm^2 – 6.0 g/cm^2) used, was not the same in all patients but in

each individual, it was kept the same. The precision of the BMD measurements was calculated from double measurements of 6 patients, and we found a mean coefficient of variation (CV) of 5% (range 0.8%-16%), 3% (range 0.1%-12%), 2% (range 0.4%-8.5%), and 3% (range 0.7%-6%) for ROI1, ROI2, ROI3, and ROI4 respectively.

Using the same DXA technique (scan resolution: 1.0 x 1.0 mm; scan speed: 45 mm/s), we also performed scans of the ankle of the operated side and the contralateral non-operated side postoperatively and after 3, 6, and 12 months to address a potential decrease in BMD caused by immobility or general decrease. BMD was measured in a 2-cm long ROI located 1 cm proximal from the ankle joint (**Fig. 3**). These scans were performed as previously described and the precision error for measurements of BMD in this ROI is very low [23].

Statistics

The BMD data was considered normally distributed. All changes in BMD over time were analyzed using repeated measures ANOVA and students paired t-test for comparison of the step-wise BMD changes over time compared to the first postoperative scanning. P-values below 0.05 were considered significant. Precision of the BMD measurements was evaluated by calculation the coefficient of variation (CV = (standard deviation (SD) / mean) x 100%). All data is presented as mean (SD or range). The statistical analysis was performed using software R (R foundation, Vienna, Austria).

Results

Clinical Results

The mean MSTS score was 17 (5-29) after 3 months. The score did not change during the follow-up, and it was 18 (4-30) after 12 months representing a mean score of 59%. After 3 and 6 months, the highest score was in the emotional acceptance category (mean score: 3.8) and lowest in the function category (mean score 1.9). One year after surgery, patients scored highest in the walking category (3.6) and lowest in function (2.0).

BMD Changes Around the Stem

We found a significant decrease in periprosthetic BMD during 1-year follow-up in all ROI's, however, in ROI2 and ROI3 statistical significance was only obtained using t-test (0-12 months) and not by the ANOVA analysis (**Table 2).** The greatest reduction in BMD, 1 year after surgery, was in ROI2 (15%). Within the first 3 months, ROI2 showed the highest decrease in BMD of 8% (p=0.366) compared to baseline. From 3 to 6 months, BMD increased close to baseline in ROI2 (-0.4%) followed by a further decrease in BMD after 12 months of 15% below baseline (p =0.003). In ROI1, adjacent to the TM collar, the BMD progressively decreased from 6% within the first 3 months until 14% below baseline after 1 year (p =0.004). In ROI3, closest to the tip of the stem, BMD decreased 6% after 3 months and gradually decreased further to 11% below baseline after 1-year of follow-up (p =0.005). ROI4 adjacent to the tip of the stem showed the lowest decrease in BMD within all follow-up measures although statistically significant after 12 months (8%, p <0.0001).

BMD Changes of the Ankles

After 3 months, the BMD decreased by 6% (p=0.008) in the operated ankle followed by a temporary plateau after 6 months, and finally at 1-year of follow up, the BMD loss in the operated ankle reached 9% below baseline (p=<0.001). We found an initial minor decrease of 2% (p=0.12) in BMD in the non-operated ankle after 3 months and it stayed approximately at that level throughout the study period (**Table 2**).

Discussion

During the first year after surgery, significant BMD changes were seen in all four ROI around the 130mm cemented stem of the Zimmer® Segmental tumor prosthesis ending with a significant bone loss after 1 year of 8-15%. The bone loss was most pronounced (14-15%) in the 2 ROIs closest to the TM collar and lowest (8%) adjacent to the tip of the stem.

To our knowledge, there exist no previous reported longitudinal results of the periprosthetic bone remodeling after resection and reconstruction with the cemented Zimmer® Segmental tumor prosthesis. Only a few studies have investigated the periprosthetic bone remodeling after insertion

of a tumor prosthesis [3, 7, 18]. As in the present study, Lan et al. [18] and Andersen et al. [3] found a further reduction in bone mineral with increased distance from the distal part of the stem towards the extension pieces, or prostheses, corresponding to the Gruen Zones 1, 2, 6 and 7 [4, 33]. The same pattern in BMD changes along the stem, as demonstrated by Lan et al., was found in a cross-sectional study with a mean time of 31.8 months after surgery, using the contralateral leg as reference [18]. However, the evaluation of BMD changes over time by Lan et al. was based upon measurements in one selected ROI which limits comparison. Vennesma et al. [33] demonstrated that to obtain exact measurements of BMD changes after surgery, the operated side should always be reference and patients should be followed prospectively. Likewise, Kröger et al. [17] demonstrated that there are local differences in BMD between limbs and stated that BMD measurements years after surgery compared with contralateral values are invalid. The absolute and relative changes in BMD across all ROI within the present follow up are comparable to the remodeling around stems used in other tumor prostheses as demonstrated by Andersen et al. [5]. Davis et al. [7] evaluated bone remodeling around the Kotz Modular Femur Tibia Reconstruction with a mean of 90.2 months after surgery and their results indicated that BMD reached a plateau. However, their study was cross-sectional using the contralateral limbs as reference and an interstudy comparison is therefore questionable.

The pattern in bone remodeling along the Zimmer® Segmental stem is corresponding to other findings after both cemented and uncemented primary hip arthroplasty [1, 4, 20, 31]. Bone remodeling and bone resorption adjacent to the proximal part of the stem is caused by distal transfer load of the prostheses due to the greater stiffness of the stem. Thus, the periprosthetic bone close to the artificial joint itself is more prone to stress shielding.

Several studies investigating primary hip arthroplasty reported a pronounced periprosthetic loss in BMD around the cemented and uncemented femur stem within the first 3 months after surgery followed by an increase or plateau after 6 month [17, 33]. The adaptive changes in bone remodeling caused by the surgical trauma to the bone after arthroplasty has been suggested to be long lasting despite increased postoperative activity [24, 29]. However, Brodner et al. [6] and Huang et al. [11]found increased BMD in the distal Gruen zones after 5 and 3 year follow up respectively and Korovessis et al.[16] found increased BMD at the greater and minor trochanter after 4 years follow-up. Our results indicate a progressive remodeling and loss in BMD after one year.

Even though we used cemented fixation for all our prostheses with immediate weight bearing, the demonstrated progressive bone remodeling after 1-year could partly be explained by the well

known required prolonged rehabilitation and immobilization after inplantation of tumor prostheses. This is due to prolonged surgery time and extensive loss of tissue. Furthermore, loss of bone stock in relation to chemotherapy is well described [10] and given the mean age in the present cohort, the well known age-related decay [5] in BMD will further affect the risk of progressive bone resorption after surgery.

It is well known from primary hip or knee arthroplasty that lesser stem stiffness, shorter stems and also coating may contribute to retain normal load transfer, and thus enhance bone preservation [19, 21, 31]. The various long-term follow-up results in periprosthetic BMD shows that adaptive bone remodeling after surgery also may contribute to better fixation as opposed to loosening and that it could depend on fixation method of the prostheses due to advantageous distribution and transmission of load. We speculate that the relative slow decrease in BMD until 1-year after surgery in all our ROI partly could be explained by the intended fixation of the TM collar with less load transfer to the tip of the stem and hence reduced stress shielding adjacent to the joint. However, inter study comparison in general is difficult due to differences in measurement of BMD, prostheses, methods of fixation and also patient cohort with regards to age, gender and comorbidity.

The average MSTS score was 22.3 (range: 14-30) 1 year after surgery. The patients scored highest in the walking and gait (average: 4.3) categories and lowest in function and supports (average: 3.3) categories.
The average MSTS score is slightly poorer compared to other studies evaluating tumor prostheses [22, 32]. Due to the need for prolonged rehabilitation after insertion of tumor prostheses, we suggest that the difference is partly caused by the relatively short follow up in our study compared to other studies. Also, we speculate that the MSTS score reflects that our cohort also comprised patients with MBD, which is often a group of patients in poor general health condition. Nevertheless, we find our results comparable to the 1-year evaluation by Andersen et al. [3].

To assess to what extend the periprosthetic changes in BMD were caused by stress shielding, immobilization or a general decrease in BMD for other causes, we performed DXA scans of both ankles. The immobilization of the operated limb is considered to be reflected by the decrease in BMD of the affected ankles. After 1-year, the decrease in BMD of the operated ankle was 9% and the non-operated ankle was close to baseline (2%). These findings indicate that the periprosthetic

BMD changes during follow-up are caused by stress shielding combined with immobilization and to a lesser extend a general decrease in BMD.

We found a precision of BMD measurement of CV 2%-5% which is slightly higher compared to Andersen et al. evaluating the uncemented proximally Hydroxyapatite-Coated femur stem [3]. This could partly be explained by the bone-cement interface in our measurements. Lan et al. evaluated the Kotz Modular Femoral Tibial Reconstruction stems with screw fixation and found CV comparable to ours despite the fact, that they evaluated uncemented stems. However, their measures are based upon smaller ROI and since Gehrchen et al. [8] demonstrated that lesser ROI is associated with poorer precision, the smaller ROI size therefore could be an explanation. Nevertheless, we find our CV comparable to previous findings of cemented hip and knee arthroplasty which has proven to be adequate values to detect small adaptive bone remodeling changes [17, 20, 27].

Some limitations need to be addressed. Our sample size is relatively small and non-randomized. However, to the best of our knowledge randomized controlled trials, to evaluate different implants and methods of fixation for these patients, is not an option. Also, repeated measures can be biased by outside factors including outcome during follow-up. In addition, in case of missing values, repeated measure ANOVA, excludes all data of the participant. Furthermore, repeated measures is well suited for small sample size and despite 7 patients lost to follow up, we have only few missing data of those who completed 1-year data analysis follow-up and all available data was used when performing post-hoc students paired t-test. Nevertheless, to the best of our knowledge present study demonstrates the largest sample size in a prospectively designed study evaluating bone remodeling around a tumor prosthesis with 1-year follow-up.

In conclusion, we successfully evaluated the early adaptive bone remodeling around the cemented Zimmer® Segmental stem and the TM collar, used for reconstruction after tumor resection in the lower extremities. Our results indicated a slow progressive decrease in BMD of 8%-15% after 1-year follow up, and the periprosthetic bone loss is considered (from the results of BMD measurements of the ankles) to be caused by a combination of stress shielding and immobilization.

Acknowledgements

DXA scans were performed by the research-nurses Marina Golemac and Sofie Kofoed Larsen. Financial support for the study was received from Zimmer Biomet.

Compliance with Ethical Standards

Research Involving Human Participants and Informed Consent

The study was approved by the Scientific Ethical Committee of the Capital Region of Denmark (J. No. H-2-2014-105) and the Danish Data Protection Agency (J. No.:2012-58-00004). The study was conducted in accordance with the ethical standards of the national ethical committee and with the 1964 Declaration of Helsinki. Prior to inclusion informed consent was obtained from all participants after written and oral information.

References

1. Abu-Rajab RB, Watson WS, Walker B, Roberts J, Gallacher SJ, Meek RMD. Peri-prosthetic bone mineral density after total knee arthroplasty. *J. Bone Jt. Surg. - Ser. B.* 2006;88:606–613.

2. Andersen MR, Petersen MM. Adaptive Bone Remodeling of the Femoral Bone After Tumor Resection Arthroplasty With an Uncemented Proximally Hydroxyapatite-Coated Stem. *J. Clin. Densitom.* 2016.

3. Andersen MR, Petersen MM. Adaptive Bone Remodeling of the Femoral Bone After Tumor Resection Arthroplasty With an Uncemented Proximally Hydroxyapatite-Coated Stem. *J. Clin. Densitom.* 2016;19:202–207. Available at: http://dx.doi.org/10.1016/j.jocd.2015.02.009.

4. Bodén HSG, Sköldenberg OG, Salemyr MOF, Lundberg HJ, Adolphson PY. Continuous bone loss around a tapered uncemented femoral stem: A long-term evaluation with DEXA. *Acta Orthop.* 2006;77:877–885.

5. Brockstedt H, Kassem M, Eriksen EF, Mosekilde L, Melsen F. Age- and sex-related changes in iliac cortical bone mass and remodeling. *Bone.* 1993;14:681–691.

6. Brodner W, Bitzan P, Lomoschitz F, Krepler P, Jankovsky R, Lehr S, Kainberger F, Gottsauner-Wolf F. Changes in bone mineral density in the proximal femur after cementless total hip arthroplasty. A five-year longitudinal study. *J. Bone Jt. Surg. - Ser. B.* 2004;86:20–26.

7. Davis AM, Damani M, White LM, Wunder JS, Griffin AM, Bell RS. Periprosthetic bone remodeling around a prosthesis for distal femoral tumors: Longitudinal follow-up. *J. Arthroplasty.* 2005;20:219–224.

8. Gehrchen PM, Petersen MM, Nielsen PK, Lund B. Influence of region size on bone mineral measurements along femoral stems in THA. *HIP Int.* 2000;10:204–208.

9. Grimer RJ, Taminiau AM, Cannon SR. Surgical outcomes in osteosarcoma. *J. Bone Jt. Surg.* 2002;84:395–400.

10. Holzer G, Krepler P, Koschat MA, Grampp S, Dominkus M, Kotz R. Bone mineral density in long-term survivors of highly malignant osteosarcoma. *J. Bone Jt. Surg.* 2003;85:231–237. Available at: http://www.bjj.boneandjoint.org.uk/cgi/doi/10.1302/0301-620X.85B2.13257.

11. Huang Q, Shen B, Yang J, Zhou Z ke, Kang P de, Pei F xing. Changes in bone mineral density of the acetabulum and proximal femur after total hip resurfacing arthroplasty. *J. Arthroplasty.* 2013;28:1811–1815. Available at: http://dx.doi.org/10.1016/j.arth.2013.03.020.

12. Hwang JS, Mehta AD, Yoon RS, Beebe KS. From amputation to limb salvage reconstruction: Evolution and role of the endoprosthesis in musculoskeletal oncology. *J. Orthop. Traumatol.*

2014;15:81–86.
13. Jarvinen M, Kannus P. Current Concepts Review-Injury of an Extremity as a Risk Factor for the Development of Osteoporosis. *J Bone Jt. Surg Am*. 1997;79:263–276.
14. Kerner J, Huiskes R, Van Lenthe GH, Weinans H, Van Rietbergen B, Engh CA, Amis AA. Correlation between pre-operative periprosthetic bone density and post-operative bone loss in THA can be explained by strain-adaptive remodelling. *J. Biomech.* 1999;32:695–703.
15. Kishida Y, Sugano N, Nishii T, Miki H, Yamaguchi K, Yoshikawa H. Preservation of the bone mineral density of the femur after surface replacement of the hip. *J. Bone Jt. Surg. - Ser. B.* 2004;86:185–189.
16. Korovessis P, Droutsas P, Piperos G, Michael A, Baikousis A, Stamatakis M. Course of bone mineral content changes around cementless Zweymueller total hip arthroplasty: A 4-year follow-up study. *Arch. Orthop. Trauma Surg.* 1997;116:60–65.
17. Kröger H, Miettinen H, Arnala I, Koski E, Rushton N, Suomalainen O. Evaluation of periprosthetic bone using dual-energy X-ray absorptiometry: Precision of the method and effect of operation on bone mineral density. *J. Bone Miner. Res.* 2009;11:1526–1530.
18. Lan F, Wunder JS, Griffin a M, Davis a M, Bell RS, White LM, Ichise M, Cole W. Periprosthetic bone remodelling around a prosthesis for distal femoral tumours. Measurement by dual-energy X-ray absorptiometry (DEXA). *J. Bone Joint Surg. Br.* 2000;82:120–125.
19. Marchetti ME, Steinberg GG, Greene JM, Jenis LG, Baran DT. A prospective study of proximal femur bone mass following cemented and uncemented hip arthroplasty. *J. Bone Miner. Res.* 2009;11:1033–1039.
20. Nygaard M, Zerahn B, Bruce C, Søballe K, Borgwardt A. Early periprosthetic femoral bone remodelling using different bearing material combinations in total hip arthroplasties: a prospective randomised study. *Eur. Cell. Mater.* 2004;8:815–819.
21. Nygaard M, Zerahn B, Bruce C, Søballe K, Borgwardt A. Early periprosthetic femoral bone remodelling using different bearing material combinations in total hip arthroplasties: a prospective randomised study. *Eur. Cell. Mater.* 2004;8.
22. Pala E, Trovarelli G, Angelini A, Maraldi M, Berizzi A, Ruggieri P. Megaprosthesis of the knee in tumor and revision surgery. *Acta Biomed*. 2017;88:129–138. Available at: www.actabiomedica.it.
23. Petersen MM, Gehrchen PM, Nielsen PK, Lund B. Loss of bone mineral of the hip assessed by DEXA following tibial shaft fractures. *Bone*. 1997;20:491–495.
24. Petersen MM, Nielsen PT, Lauritzen JB, Lund B. Changes in bone mineral density of the

proximal tibia after uncemented total knee arthroplasty: A 3-year follow-up of 25 knees. *Acta Orthop.* 1995;66:513–516.
25. Rathsach Andersen M, Winther N, Lind T, Schrøder HM, Petersen MM. Bone remodeling of the proximal tibia after uncemented total knee arthroplasty: secondary endpoints analyzed from a randomized trial comparing monoblock and modular tibia trays—2 year follow-up of 53 cases. *Acta Orthop.* 2019;3674.
26. Roberston DD, Mintzer CM, Weissman BN, Ewald FC, LeBoff M, Spector M. Distal Loss of Femoral Bone following Total Knee Arthroplasty. *J. Bone Jt. Surg.* 1994;76:66–76.
27. Seki T, Omori G, Koga Y, Suzuki Y, Ishii Y, Takahashi HE. Is bone density in the distal femur affected by use of cement and by femoral component design in total knee arthroplasty? *J. Orthop. Sci.* 1999;4:180–186.
28. Soininvaara T, Kröger H, Jurvelin JS, Miettinen H, Suomalainen O, Alhava E. Measurement of bone density around total knee arthroplasty using fan-beam dual energy x-ray absorptiometry. *Calcif. Tissue Int.* 2000;67:267–272.
29. Soininvaara TA, Miettinen HJA, Jurvelin JS, Alhava EM, Kröger HPJ. Bone mineral density in the proximal femur and contralateral knee after total knee arthroplasty. *J. Clin. Densitom.* 2004;7:424–431.
30. Sorensen MS, Gregersen KG, Grum-Schwensen T, Hovgaard D, Petersen MM. Patient and implant survival following joint replacement because of metastatic bone disease: A cross-sectional study of 130 patients with 140 joint replacements. *Acta Orthop.* 2013;84:301–306.
31. Tanzer M, Kantor S, Rosenthall L, Bobyn JD. Femoral remodeling after porous-coated total hip arthroplasty with and without hydroxyapatite-tricalcium phosphate coating: A prospective randomized trial. *J. Arthroplasty*. 2001;16:552–558.
32. Tunn PU, Pomraenke D, Goerling U, Hohenberger P. Functional outcome after endoprosthetic limb-salvage therapy of primary bone tumours - A comparative analysis using the MSTS score, the TESS and the RNL index. *Int. Orthop.* 2008.
33. Venesmaa PK, Kröger HPJ, Jurvelin JS, Miettinen HJA, Suomalainen OT, Alhava EM. Periprosthetic bone loss after cemented total hip arthroplasty: A prospective 5-year dual energy radiographic absorptiometry study of 15 patients. *Acta Orthop. Scand.* 2003;74:31–36.
34. Wada T, Kawai a, Ihara K, Sasaki M, Sonoda T, Imaeda T, Yamashita T. Construct validity of the Enneking score for measuring function in patients with malignant or aggressive benign tumours of the upper limb. *J. Bone Joint Surg. Br.* 2007;89:659–663.

Table 1.

Baseline data of the patients (n=21) that completed 1-year follow-up

Variable	Level	Total (%)
Gender		
	Female	*12 (57%)*
	Male	*9 (43%)*
Age (years)		
	Mean (range)	*55 (15-81)*
Resection (cm)		
	Mean (range)	*15 (10-24)*
Resection site		
	Proximal femur	*10 (48%)*
	Distal femur	*9 (43%)*
	Proximal tibia	*2 (10%)*
Pathology		
	Metastasis	*9 (43%)*
	Giant Cell	*4 (19%)*
	Chondrosarcoma	*2 (10%)*
	Myelomatosis	*2 (10%)*
	Osteosarcoma	*2 (10%)*
	Myxoid liposarcoma	*1 (5%)*
	Desmoplastic fibroma	*1 (5%)*

Table 2.

Mean (SD) BMD (g/cm2) in the 4 ROIs around the stem and in both ankles (operated and non-operated contralateral legs).

Follow-up	**Postoperative (n=21)**	**3 months (n=18)**	**6 months (n=21)**	**12 months (n=21)**	**p-value[#] 0-12 months (n=18)**
ROI1, BMD	2.186 (0.38)	2.056 (0.48)	1.990 (0.46)	1.874(0.27)	0.037
ΔBMD%		-6%	-9%	-14%	
p values (stepwise)*		0.285	0.092	0.004	
CI(95%)		(-0.10-0.33)	(-0.04-0.43)	(0.11-0.52)	
ROI2, BMD	2.248 (0.41)	2.075 (0.53)	2.238 (0.57)	1.914 (0.30)	0.071
ΔBMD%		-8%	-0.4%	-15%	
p values (stepwise)*		0.366	0.95	0.003	
CI(95%)		(-0.15-0.37)	(-0.28-0.30)	(0.13-0.54)	
ROI3, BMD	2.215 (0.43)	2.075 (0.49)	2.071 (0.38)	1.978 (0.3)	0.223
ΔBMD%		-6%	- 7%	-11%	
p values (stepwise)*		0.438	0.117	0.005	
CI(95%)		(-0.16-0.35)	(-0.04-0.33)	(0.08-0.39)	
ROI4, BMD	2.080 (0.42)	2.047 (0.44)	1.948 (0.45)	1.923 (0.45)	0.009
ΔBMD%		-2%	-4%	-8%	
p values (stepwise)*		0.356	0.079	<0.0001	
CI(95%)		(-0.04-0.11)	(-0.01-0.18)	(0.09-0.22)	
Ankle operated, BMD	0.751 (0.15)	0.7048 (0.14)	0.7049 (0.17)	0.681 (0.16)	<0.001
ΔBMD%		-6%	-6%	-9%	
p values (stepwise)*		0.008	0.008	<0.001	
CI(95%)		(0.01-0.06)	(0.02-0.09)	(0.05-0.11)	
Ankle contralateral, BMD	0.806 (0.19)	0.793 (0.18)	0.814 (0.25)	0.788 (0.17)	0.322
ΔBMD%		-2%	+1%	-2%	
p values (stepwise)*		0.12	0.90	0.12	
CI(95%)		(-0.01-0.05)	(-0.08-0.07)	(-0.01-0.05)	

* students paired t-test, [#]repeated measures ANOVA

Fig.1

Flow chart. Enrollment, follow-up, and data analysis.

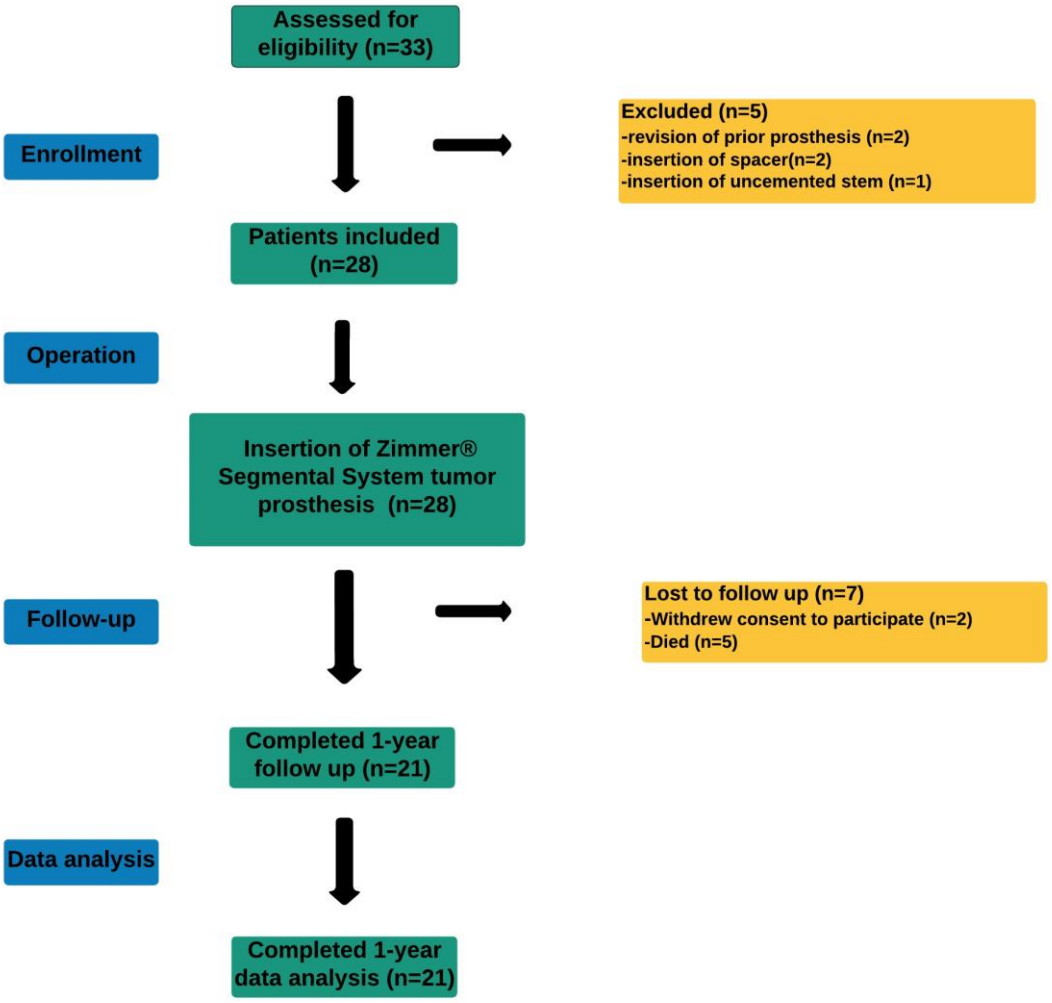


Fig.2

X-ray of proximal femur tumor arthroplasty, Cemented Zimmer® Segmental stem (left). X-ray of distal femur arthroplasty, Cemented Zimmer® Segmental stem (middle). X-ray of proximal tibia arhroplasty, Cemented Zimmer® Segmental stem (right).

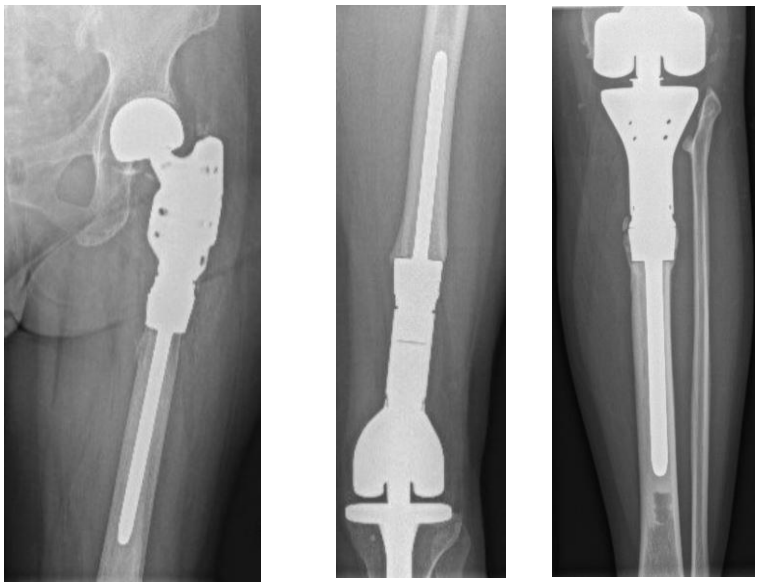
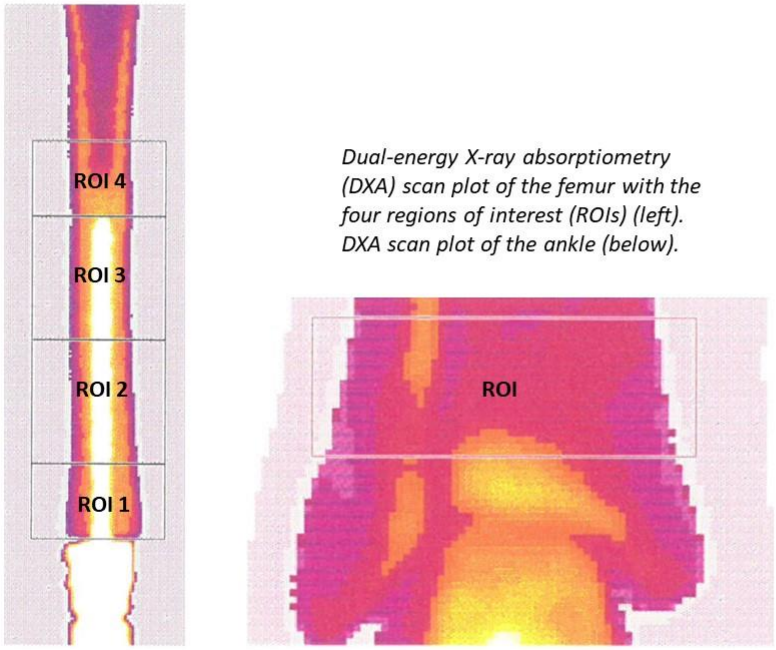


Fig. 3.



Dual-energy X-ray absorptiometry (DXA) scan plot of the femur with the four regions of interest (ROIs) (left). DXA scan plot of the ankle (below).

Development and comparison of one-year survival models in patients with primary bone sarcomas

External validation of a Bayesian belief network model and creation and external validation of a new Gradient Boosting Machine model.

Christina Holm[1],Clare F. Grazal[2],Mathias Raedkjaer[3],Thomas Baad-Hansen[3], Rajpal Nandra[4], Robert Grimer[4], Jonathan Forsberg[2] , Michael Moerk Petersen[1], Michala Skovlund Sørensen[1]

[1]The Musculoskeletal Tumor Section, The Department of Orthopedic Surgery, Rigshospitalet, University of Copenhagen, Denmark, [2]Orthopaedics, USU-Walter Reed Department of Surgery, Bethesda, MD, USA, [3]Department of Orthopaedic Surgery, Tumor Section, Aarhus University Hospital, Aarhus, Denmark, [4] The Royal Orthopaedic Hospital, Birmingham, UK

Correspondence to:

Christina Holm, M.D.

[1]The Department of Orthopedic Surgery,

Musculoskeletal Tumor Section,Rigshospitalet

University of Copenhagen,

Blegdamsvej 9, 2100 Copenhagen Ø

Denmark

Phone: +45 35453545

Email: christina.holm@dadlnet.dk

Abstract

Background

Bone sarcomas often presents late with advanced stage at diagnosis, resulting in varying short-term survival. In 2016 Nandra et al., generated a Bayesian belief network model (BBN) for 1-year survival of patients with bone sarcomas. The purpose of present study is to: 1) External validate the prior 1-year BBN prediction model for survival of patients with bone sarcomas, 2) To develop a Gradient Boosting machine (GBM) model using Nandra et al.'s cohort and evaluate if the GBM model outperform the BBN model suggested by Nandra et al. when externally validated on an independent Danish population cohort.

Material and Methods

The training cohort comprised 3493 patients newly diagnosed with bone sarcoma from the institutional prospectively maintained database at The Royal Orthopaedic Hospital, Birmingham UK. The validation cohort comprised a total of 771 patients with newly diagnosed bone sarcoma included from The Danish Sarcoma Registry between January 1st, 2000 and June 22^{sd}, 2016. We performed area under receiver operator characteristic curve (AUC ROC) analysis, Brier score and decision curve analysis (DCA) to evaluate the predictive performance of the models.

Results

External validation of the BBN 1-year prediction model demonstrated an AUC ROC of 68% (95%CI, 62%-73%). AUC ROC of the GBM model demonstrated: 75% (95%CI: 70%-80%), overall model performance by Brier score was 0.09 (95%CI: 0.077-0.11) and DCA demonstrated a positive net-benefit for threshold probabilities above 0.5. External validation of the developed GBM model demonstrated AUC ROC of 63% (95%CI: 57%-68%) and the Brier score was 0.14 (95%CI: 0.12-0.16).

Conclusion

External validation of the 1-year Bayesian belief network survival model yielded poor outcome and the model is not recommendable for clinical usage based on a Danish population cohort validation. The developed Gradient Boosting Machine 1-year survival model did not outperform the prior Bayesian belief network model and modernization is pending.

Background

Accurate survival prediction for patients with newly diagnosed bone sarcoma would be a considerable aid for the clinician when deciding the most appropriate treatment. Bone sarcomas often presents late with advanced stage at diagnosis, resulting in varying short-term survival (1). In some settings, the decision to perform surgery or more commonly, which surgical treatment to choose, relies in part, on the prediction of estimated survival. Patients with expected short- term survival sometimes may be better served with a minor operative procedure to relief pain and maintain quality residual life or perhaps no surgery- rather than major surgery with amputation or bone resection and insertion of a tumor prosthesis with associated higher risk of complications and prolonged rehabilitation. Prognostic factors for survival in bone sarcomas has been suggested (2) and management guidelines exist (3,4). However, decision of treatment management is a case by case matter due to the broad heterogeneity among bone sarcoma patients. To our knowledge, only a few attempts to create prediction models for survival in bone sarcoma patients using machine learning techniques has been conducted (5,6).

One-year survival is for many cancer groups, as well as bone sarcomas, an indicator of early/late stage at diagnosis (1,7). Nandra et al. (6), generated a Bayesian belief network model (BBN) for 1-year survival of patients with bone sarcomas and demonstrated five factors with conditional dependencies for survival one year after surgery. Bayesian belief network modeling has been used to develop decision support tools in numerous oncologic diagnoses including skeletal metastases and soft tissue sarcomas (6,8,9). To our knowledge no other research group besides Nandra et al. (6) within our field has performed prediction of short time survival for patients with bone sarcomas using this technique. The model has, however; never undergone external validation and thus the clinical use of the tool remains unknown. Nandra et al. (6) developed the BBN 1-year survival prediction model by using commercially available machine learning software (FasterAnalytics™; DecisionQ, Washington, DC, USA), which was originally developed to analyze video cassette sales. As many research communities are transitioning away from proprietary modeling methods in favor of open source software including R (R Foundation, Vienna, Austria) or Python (Python Software Foundation, Wilmington, DE, USA) that are now widely used in the field of machine learning. Open source software is advantageous not only because it is available at low or no cost, but also because it is inherently transparent. Code may be

published as a supplement to peer reviewed manuscripts. This allows for independent validation, as well as continuous development and optimization by the research community in an effort to refine and customize functions (10).

Gradient boosting machines (GBM) is a group of machine learning techniques used to generate non-parametric regression or classification models (11). Gradient boosting uses the ensemble technique, which gradually and sequentially converts weak models into stronger stronger. By each boost every new model is subsequently being correlated to the negative gradient of the customized loss function from the previous model. The boosting technique has previous proven to outperform other machine learning models in accuracy and generalizability (10,12) and hence produces a model with consistently higher accuracy compared to conventional single strong machine learning models (10).

In that perspective the purpose of this study is to: 1) External validate the 1-year BBN prediction model for survival of patients with bone sarcomas by Nandra et al. (6), 2) To develop a GBM model using Nandra et al.'s (6) cohort and evaluate if the GBM model outperform the BBN model suggested by Nandra et al. when externally validated on an independent Danish population cohort.

Material and Methods

The training cohort for this study was originally described by Nandra et al. [20]. Briefly, 3493 patients with newly diagnosed bone sarcomas treated between 1970 and 2012 at The Royal Orthopaedic Hospital, Birmingham UK, were included from their institutional prospectively maintained database. The same cohort was used as training cohort for the creation of the GBM model in present study. From the Danish Sarcoma Registry (13) a cohort of patients (n=771) newly diagnosed with bone sarcomas between 2000-2016 was obtained and was used to comprise the external validation cohort for the BBN model by Nandra et al. (6) as well as for external validation of the GBM model proposed in present study. Ethical approval was obtained from The Danish Data Protection Agency (no. P-2019-54) and the Danish Patient Safety Authority (no. 3-3013-2866/1).

External validation of prior BBN Model

The validation cohort comprised a total of 771 patients with newly diagnosed bone sarcoma included from The Danish Sarcoma Registry (DSR) (13) between January 1 2000 and June 22 2016. The Danish Sarcoma Registry is a prospectively maintained national database since January 1 2009. Patients from 2000-2008 were later included to DSR by a validation through the Danish Cancer Registry and the Danish National Pathology Registry (14). Patients were included from the only two tertiary referral Centers for orthopedic oncology in Denmark. All patients were accounted for a minimum of 1-year follow up and due to the Danish Civil Registration System (15) where exact date of death is known for all Danish patients. Survival was defined as the time from the first contact to a tertiary referral Center to date of death or complete one-year follow-up. Besides three foreign citizens no patients were lost to follow-up. One-hundred-and-thirteen patients out of 771 (15%) died within 1 year follow up (**Fig. 1**).

The BBN model by Nandra et al. (6), included 11 candidate features for final analysis: Age, Gender, Tumor Size at diagnosis, Location, Grade, Alkaline phosphatase, Metastasis at diagnosis, Pathologic fracture at diagnosis, Diagnosis, Tumor Site, Status 1 year after diagnosis, Year of diagnosis. Danish Sarcoma Registry comprises patient characteristics, tumor characteristics, treatment data and death (13) thus most of the required variables for this validation.

Alkaline phosphatase was not available from Danish Sarcoma Registry and hence was not included for validation. In the validation cohort tumor grade was defined by the Myhre-Jensen classification until 2004 (16) and from 2004 and onwards by the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC) (17). Essential for validation is that features used in the training cohort and validation cohort are identic and hence tumor grades were converted as follows: Grade I = 1 = low, Grade II = 2 = intermediate, Grade IIIa and IIIb = 3 = high. No other variables where converted.

Using the Danish validation set we then determined the ability of accuracy and discrimination by receiver operating characteristic (ROC) analysis and area under the curve (AUC) (18). Validation was considered successful if the AUC under the ROC curve was greater than 0.7 as the lowest acceptable threshold and was determined a priori. In essence the area under the curve is

interpreted as the probability that a person who experienced the outcome (death) had a higher predicted probability than the person who did not experience the outcome and by so discrimination is a measure of how well the model can separate those who do and do not experience the outcome. A value of 1 is perfect discrimination and a value of 0.5 represents chance. Overall predictive model performance was evaluated with the Brier score (19). The Brier score quantifies the compliance between the predicted probability and observed outcome. The reported value between 0 and 1 is the average squared differences between all the predicted and actual outcomes in the cohort, with 0 indicating perfect agreement and 1 indicating perfect disagreement. However a score of 0.25 reflects a 50% incidence of outcome and hence scores above 0.25 is also to be considered noninformative (20). The BBN model was used "as-is" by Nandra et al. without prior refitting or optimization and no other imputation of data was used. Validation of the BBN model was performed by using commercially available software (FasterAnalytics^{TM}, DecisionQ Corp., Washington, DC, USA).

Development of Gradient Boosting Machine model

To mitigate overfitting a ten-fold cross validation of the training cohort initially was conducted. By randomization data was split into ten unique test and train set with balanced events per variable. Each test and train set comprised 20% and 80% of data respectively. A GBM model was trained on a training set (n=2794) and subsequently tested on the corresponding test set (n=699).

For correct comparison it was decided not to exclude or include other variables than used by Nandra et al. (6). Due to missing data Alkaline Phosphate was excluded. Tumor sites were subcoded into 5 location categories (6) (Table 1). Decision trees were chosen as base-learners. Since the outcome variable was binary the Bernoulli loss function (10) was chosen. Missing data was imputed by using missForest (21). For feature selection we chose the Boruta train algorithm (22). By shuffling copies of all features the Boruta algorithm trains a Random Forest (23) on the overall data and features are consequently either rejected or confirmed and further ranked with their relative influence in the model. Due to the customizability and efficiency GBM models are prone to overfitting (10); selection and hyper-tuning of parameters is therefore crucial to outcome.

A preliminary baseline model was created with various parameter selections for the hyper-tuning process. Final parameters selected were: *shrinkage=0.01, interaction depth=3, bag fraction=0.8, N.minobsinnode=5*. The optimum number of iterations with minimum loss were n=536. (**Fig.7**). The code is included as supplementary material. We performed internal validation using the test set comprising 699 cases not used for development of the model. We performed external validation on the Danish validation set. For both assessments we used the same metrics as used for external validation of the BBN model: discrimination by ROC analysis and AUC (18), overall performance using the Brier score. Discrimination and Brier score is one aspect of model performance but does not provide information of the utility of the model for clinical use. Decision curve analysis (DCA) overcomes this limitation by quantifying the consequences of over- or undertreatment and is increasingly being used for the assessment of prediction models for clinical use. Prediction models generates a survival probability at a given time point after diagnosis. If the probability is 1 or near 1 the surgeon will presumably not be in doubt weather to treat or if the probability is near 0 the surgeon probably will choose not do surgical intervention. When the probability of survival is between 0 and 1 decision-making might be more difficult for the clinician. The threshold probability is the point where the expected benefit of surgery is equal to the expected benefit of not treating and where surgeons may become indecisive (24) .Assuming that the decision to do surgical intervention is solely based on the outcome of the prediction model a range of threshold possibilities between 0 to 1 are plotted against net benefit on a decision curve. The broad range of thresholds possibilities to evaluate the prediction model is essential since thresholds are patient- or clinician dependent (25). We compared the net-benefit of all thresholds and hence determined the clinical use of the model. A model is considered to be clinical usable if it demonstrates net benefit across the range of thresholds i.e. is superior to assuming that all patients or no patients would live longer than 1 year. As illuminated by Vickers et al. (25) net benefit is defined as a patient that will undergo appropriate treatment (surgery) or opposite will not undergo treatment based on prediction model outcome.

Baseline distributions between the training cohort and the validation cohort where compared using non-parametric tests. Mann–Whitney U test (for unpaired data) was used for continuous variables and chi-square test for categorical variables.

We used Rstudio (R Foundation, Vienna, Austria) for development and external validation of the GBM model and comparison of baseline distributions between the train and validation set.

Results

As intended the demographic and clinical features of the test set and validation set differed (Table 1). Features that differed significantly were age at diagnosis, Tumor size, grade, diagnosis, pathologic fracture at diagnosis, tumor location and status one year after diagnosis. The nonsignificant observations were gender (p=0.63) and metastasis at Diagnosis (p=0.22). The proportion of missing values varied among features although most notable in the train set were the tumor size (missing in 51%) and in the validation set grade (missing in 23%) (**Table 1**).

External validation of the BBN model

External validation of the BBN 1-year prediction model yielded poor discriminatory ability with an AUC ROC of 68% (95%CI, 62%-73%) (**Fig.2**), and hence the ability of the model to discriminate between survival and not survival is insufficient when based on this Danish population. Overall model performance evaluated with the Brier score was 0.122 (95%CI: 0.102-0.141).

Internal validation of the Gradient Boosting Machine model

Internal validation by AUC ROC analysis yielded well discriminatory ability with 75% (95%CI: 70%-80%) (**Fig.3**). The Brier score for overall model performance was 0.09 (95%CI: 0.077-0.11). Decision curve analysis demonstrated a positive net-benefit i.e. above the lines assuming none or all patients are alive one year after diagnosis. Hence supporting that the model is suitable for clinical use for probability thresholds above 0.5 (**Fig.4**). However, at threshold probabilities below 0.5 the surgeon gains more benefit assuming that all patients are alive. Nandra et al. (6) demonstrated similar findings when performing DCA analysis of the BBN model (0.5). Net benefit was capped at 85% (patients alive after one year) given the definition that net benefit is one patient being treated appropriate according to the output of the prediction model. These

findings are also similar to the findings by Nandra et al. (85.5%). Features that ranked highest in variable importance were: Diagnosis, Tumor size and Age (**Fig.1**).

External validation of the GBM model

External validation of the GBM model yielded poor discriminatory ability with an AUC of the ROC curve of 63% (95%CI: 57%-68%) (**Fig.5**), and hence did the GBM model not outperform the BBN when external validated on this Danish cohort. The Brier score was 0.14 (95%CI: 0.12-0.16). Since we cannot recommend the model for clinical use based on this external validation we did not perform DCA.

Discussion

Individually treatment strategy for patients with newly diagnosed bone sarcoma is primarily dependent on estimated short-term survival. No 1-year prediction model for survival using machine learning technique has to our knowledge been successfully externally validated for clinical use and hence the aim of this study was to evaluate two survival prediction models and preferably provide clinicians with a validated decision-tool to support determination of treatment strategy for patients with bone sarcoma.

Limitations

Certain study limitations in the validation cohort needs to be addressed. First, although data was drawn from a prospectively maintained database, all data is to be considered retrospective. Second, the data comprises only patients from a Danish population and origins from two tertiary referral centers with equal treatment strategy and hence may not represent the desired heterogeneity used to test the model for generalizability. However, data was chosen due to the no loss to follow up and limited missing data. Also, patients included were not selected for surgery but comprised all patients with newly diagnosed bone sarcoma. Nevertheless, the selection bias may cause the model to be less robust. Third, we did not explore in cause of death and death for other causes than the cancer diagnosis might have added inaccuracy to the model towards underestimation of

survival. Fourth, the requirement of equal features for validation is essential and although GBM and BBN techniques are particularly feasible with missing data we acknowledge the missing data for Alkaline Phosphatase in the validation cohort although it was excluded in the training cohort as described by Nandra et al. (6). Alkaline Phosphatase has previously proven to be prognostic for patients with osteosarcoma (26) and it is possible that inclusion of Alkaline Phosphatase would have improved prediction accuracy of the GBM model. Also, by converting the histologic grade variable for the purpose of equality we might have added further observation bias to the final model. Fifth, the external validation on both models is performed on a smaller cohort compared to the train set and with significant differences in baseline characteristics besides gender and metastasis at diagnosis (**Table 1**). These differences could partly be explained by the large samplesize where even small differences are detected as well as the different time periods from were patient were included. Obviously the 1-year survival has changed from 1970 to 2012 due to considerably improvements in diagnostics techniques and treatment modalities (2) as also seen by the significant difference in 1 year survival between train and validation cohort (**Table 1**). The improvement in survival over time could explain why the year of diagnosis variable and 1-year survival ranks high in variable importance (**Fig**.**1**). However, year of diagnosis is not a reproducibly variable and consequently a time variable should be used thoughtfully for prediction as it in present study could tend to underestimate survival and hence could risk to under-treat patients.

Discussion

While several prediction models for short term survival has been developed for patients with metastatic bone disease (8,27) and soft tissue tumors (9), only separate prognostic factors for survival has been illuminated for patients with bone sarcoma (1,28) with Nandra et al. (6) being the first research team to develop a prediction model for short term survival. We successfully created a GBM model for 1-year survival that nevertheless yielded poor performance when externally validated on a Danish population-based cohort as well as the BBN model by Nandra et al. (6) which may be caused by both models being trained on the same data set. Besides emphasizing the importance of external validation of prediction models prior to clinical use these results raise some interesting questions to the discussion on the eligibility of prediction models

and data used to train them. Although researchers often seek to obtain as much data as possible to achieve greater power to detect any potential differences, Chen et al. (29) demonstrated that modern data in small sample sizes used to train prediction models, has greater impact for prediction compared to larger historical sample sizes. Also, the predictive power of models is not only a product of a given algorithm but also the variables used train them. One of the main causes to overfitting is too many features compared to the number of observations and hence the demonstrated overfitting of the prior BBN model by Nandra et al. (6) and the present GBM model could partly be explained by the significant improved survival between 1970 and 2016 (2) with subsequently gradually decreased outcome (events). We suggest that the considerably improved treatment for patients with bone sarcoma in general and the resulting better survival during that time period, affects the survival outcome variable and hence also affect the generalizability of both models when being validated with a modern cohort.

We chose to create a GBM model for several reasons. As with BBN, GBM models are capable of handling large non-parametric sample sizes with complex interactions and substantial missing or outlying data (10). However, GBM models has proven to provide higher and more accurate prediction compared to other conventional single machine learning methods and in some studies also when compared to other ensemble methods such as bagging (10,30). Some obvious advantages of the GBM technique, is the customizability and fully transparency. However, another common cause of overfitting is too powerful models and since GBM models tends to continuously mitigate any errors during process they are prone to overfitting if not duly regulated (10). One could be tempted to train the model with a high number of base-learners with many splits and subsequently boost the model with numerous iterations to obtain high accuracy. Nevertheless, beyond any given optimal number of iterations the model will predict the training cohort with consequently increased loss and decreased generalizability; hyper-tuning of parameters is therefore a crucial balance. We speculate that the capability of the GBM model combined with the use of a historical train set partly explain why the present model was not successfully validated on this Danish cohort despite significant differences in patient demographics between the training and validation cohort. It proves the power and potential of the GBM algorithm, but it also obligates the clinician having in mind the purpose of creating prediction models as emphasized by Chen et al. (31). The aim of developing a prediction model

was however not to substitute the clinical prediction and assessment more than to assist the clinical decision making.

The five features with highest rank of relative influence in our GBM model were: Diagnosis, tumor size, Age, metastasis at diagnosis, year at diagnosis (**Fig. 1**). Although not in the same order of relative importance as demonstrated by Nandra et al. (6) tumor size, age and metastasis at diagnosis were also identified in the BBN model to have the largest prognostic effect (6). This emphasizes the use of these features for future prediction models for survival. Further, in order to strengthen the model and circumvent observational bias the use of objective variables such as biochemical markers should be taken in consideration as several biochemical markers has proven to be well suited as features for prediction models in patients with bone metastasis (27). Likewise, Serum Lactate Dehydrogenase and molecular markers as p-53 and p-glycoprotein has been reported to have prognostic value for patients with osteosarcoma (32). Thorn et al. found a positive correlation between high YKL-40 protein expression in tumor tissue and longer overall survival in osteosarcoma patients (33). To our knowledge no biochemical or molecular marker has been used as feature for development of prediction models using machine-learning technique, for patients with bone sarcoma.

Estimating one-year survival in this patient population is challenging. In the knowledge of the complexity and heterogeneity of patients with bone sarcoma the intention should not to replace the clinical assessment and prediction rather than to assist decision-making in the clinical setting. Identification and inclusion of other variables is therefore warranted and the authors are obligated to continue the ongoing work with development and improvement of prediction models for bone sarcoma patients while we also hope to encourage other institutions to validate present GBM model on a non-Scandinavian population.

Conclusion

External validation of the 1-year Bayesian belief network survival model yielded poor outcome and the model is not recommendable for clinical usage based on a Danish population cohort validation. We successfully created a Gradient Boosting Machine 1-year survival model. When external validated the Gradient boosting machine model did not outperform the Bayesian belief network model. Development or modernization of a model to predict mortality in patients with newly diagnosed bone sarcoma is pending.

References

1. Nandra R, Hwang N, Matharu GS, Reddy K, Grimer R. One-year mortality in patients with bone and soft tissue sarcomas as an indicator of delay in presentation. Ann R Coll Surg Engl. 2015;97:425–33.
2. Bramer JAM, Van Linge JH, Grimer RJ, Scholten RJPM. Prognostic Factors in localized Extremity Osteosarcoma; a Systematic Review. Eur J Surg Oncol. 2009;35:1030–6.
3. Benjamin RS, Brigman B, Chow W, Iii EUC, Frassica DA, Frassica FJ, et al. Bone Cancer Clinical Practice Guidelines in Oncology. 2010;8(6):688–712.
4. Federman N, Bernthal N, Eilber FC, Tap WD. The multidisciplinary management of osteosarcoma. Curr Treat Options Oncol. 2009;10(1–2):82–93.
5. Quirina C.B.S, Bernstein KDA, Calderon SAL, Schwab JH. Can Machine-learning Techniques Be Used for 5-year Survival Prediction of Patients With Chondrosarcoma ? Clin Orthop Relat Res. 2018;476:2040–8.
6. Nandra R, Parry M, Forsberg J, Grimer R. Can a Bayesian Belief Network Be Used to Estimate 1-year Survival in Patients With Bone Sarcomas? Clin Orthop Relat Res. 2017;475:1681–9.
7. National Cancer Intelligence Network. One Year Cancer Survival Trends, England , 2000-2004 [Internet]. London: NCIN; 2008. Available from: www.ncin.org.uk
8. Forsberg JA, Eberhardt J, Boland PJ, Wedin R, Healey JH. Estimating survival in patients with operable skeletal metastases: An application of a Bayesian belief network. PLoS One. 2011;6(5).
9. Forsberg JA, Healey JH, Brennan MF. A probabilistic analysis of completely excised high-grade soft tissue sarcomas of the extremity: An application of a Bayesian belief network. Ann Surg Oncol. 2012;19(9):2992–3001.
10. Natekin A, Knoll A. Gradient boosting machines , a tutorial. Front Neurorobot. 2013;7(December).
11. Schapire RE, Avenue P, Room A. The Boosting Approach to Machine Learning An Overview. MSRI Work Nonlinear Estim Classif. 2002;1–23.
12. Chen Y, Jia Z, Mercola D, Xie X. A Gradient Boosting Algorithm for Survival Analysis via Direct Optimization of Concordance Index. Comput Math Methods Med. 2013;2013:8 pages.
13. Jørgensen, Peter Holmberg, Lausten Gunnar Schwarz PAB. The Danish Sarcoma Database. Clin Epidemiol. 2016;8:685–90.
14. Safwat A, Baerentzen S. Population-based Aarhus Sarcoma Registry : validity , completeness of registration , and incidence of bone and soft tissue sarcomas in western Denmark. Clin Epidemiol. 2013;5(March):45–56.
15. Schmidt M, Pedersen L, Sørensen HT. The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol. 2014;29(8):541–9.
16. Jensen OM, Høgh J, Østgaard SE, Nordentoft AM, Sneppen O. Histopathological grading of soft tissue tumours. Prognostic significance in a prospective study of 278 consecutive

cases. J Pathol. 1991;163(1):19–24.

17. Coindre J. Grading of Soft Tissue Sarcomas. Arch Pathol Lab Med. 2006;130(October):1448–53.

18. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Casp J Intern Med. 2013;4(2):627–35.

19. Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol. 2010;63(8):938–9.

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

21. Stekhoven DJ, Bühlmann P. MissForest — non-parametric missing value imputation for mixed-type data. Bioinforma. 2012;28(1):112–8.

22. Kursa MB. Feature Selection with the Boruta Package. J STAT SOFTW. 2010;36(11).

23. Breiman LEO. Random Forests. Mach Learn. 2001;45:5–32.

24. Forsberg JA, Ma DS, Chen Q, Vickers A. Treating Metastatic Disease Which Survival Model Is Best Suited for the Clinic ? Clin Orthop Relat Res. 2013;471:843–50.

25. Vickers AJ, Calster B Van, Steyerberg EW. Net benefit approaches to the evaluation of prediction models , molecular markers , and diagnostic tests. BMJ. 2016;352(6).

26. Bacci G, Picci P, Ferrari S, Orlandi M, Ruggieri P, Casadei R, et al. Prognostic Significance of Serum Alkaline Phosphatase Measurements in Patients with Osteosarcoma Treated with Adjuvant or Neoadjuvant Chemotherapy.

27. Sørensen MS, Silvius EC, Khullar S, Hindsø K, Forsberg JA, Petersen MM. Biochemical Variables are Predictive for Patient Survival after Surgery for Skeletal Metastasis . A Prediction Model Development and. Open Orthop J. 2018;(2018):469–81.

28. Anderson ME. Update on Survival in Osteosarcoma. Orthop Clin North Am. 2016;47(1):283–92.

29. Chen JH, Alagappan M, Goldstein MK, Asch SM, Altman RB. International Journal of Medical Informatics Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets. Int J Med Inform [Internet]. 2017;102:71–9. Available from: http://dx.doi.org/10.1016/j.ijmedinf.2017.03.006

30. Dietterich TG. An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees : Bagging , Boosting , and Randomization. Mach Learn. 2000;40:139–57.

31. Chen JH, Asch SM, Alto P. Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations. N Engl J MED. 2018;376(26):2507–9.

32. Kim YBPHS, Lee JHOSH. The co-expression of p53 protein and P-glycoprotein is correlated to a poor prognosis in osteosarcoma. Int Orthop. 2001;24:307–10.

33. Thorn AP, Daugaard S, Christensen LH, Jarle IB, Petersen MM. YKL-40 protein in

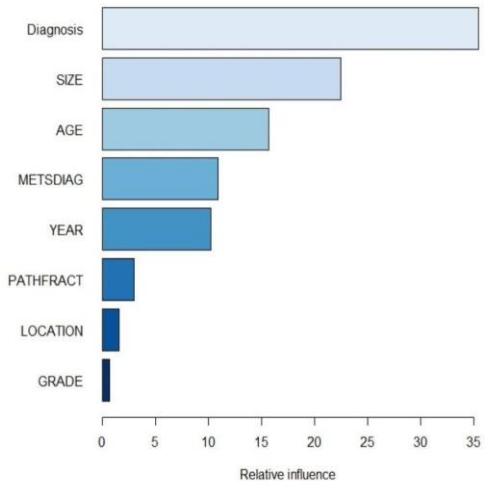
osteosarcoma tumor tissue. APMIS. 2016;(17):453–61.

Table 1
Distribution and comparison of baseline variables between training and validation cohort of study IV

Variable	Level	Training cohort 1970-2012 n=3493 (%)	Validation cohort 2000-2012 n=771 (%)	Total n=4264 (%)	*p*-value
Gender					0.22*
	Female	1451 (42)	338 (44)	1789(42)	
	Male	2042 (59)	430 (56)	2472(58)	
	missing	0	3	3	
Age					<.0001#
	Median (IQR)	23 (14-51)	44 (22-62)	26 (15-53)	
	missing	0	3	3	
Tumor size (cm)					<.0001*
	Median (IQR)	10 (7-13)	6(3-10)	8 (2-12)	
	missing	1796	0	1796	
Grade					<.0001*
	High	2641 (76)	293 (49)	2934 (72)	
	Intermediate	374 (11)	143 (24)	517 (13)	
	Low	478 (14)	158 (27)	636 (16)	
	missing	0	177	177	
Histology					<.0001*
	Osteosarcoma	1572 (45)	174 (25)	1746 (41)	
	Chondrosarcoma	793 (23)	326 (46)	1119 (26)	
	Ewings	653 (19)	114 (16)	767 (18)	
	Sarcoma	182 (5)	26 (3)	191 (4)	
	Chordoma	70 (2)	34 (5)	104 (2)	
	Other (19 histologic diagnoses)	223 (6)	36 (5)	259 (6)	
	missing	0	61	61	
Pathologic fracture at diagnosis					<.0001*
	No	3035 (87)	729 (95)	3764 (88)	
	Yes	458 (13)	42 (5)	500 (12)	
	Missing	0	0	0	
Anatomic location					<.0001*
	Head and Neck	20 (1)	50 (7)	70 (2)	
	Lower Extremity	2118 (61)	355 (47)	2473 (58)	
	Pelvic Girdle	642 (18)	117 (16)	759 (18)	
	Spine	0	32 (4)	32 (1)	
	Upper Extremity	471 (14)	103 (14)	574 (14)	
	Upper Trunk	230 (7)	93 (12)	323 (8)	
	missing	12	21	33	
Metastasis at diagnosis					0.63*
	No	3010 (86)	651 (87)	3661 (86)	
	Yes	483 (14)	98 (13)	581 (14)	
	missing	0	22	22	
Status at 1 year after diagnosis					0.009*
	Alive	3099 (89)	655 (85)	3754 (88)	
	Dead	394 (11)	113 (15)	507 (12)	
	missing	0	3	3	
Year of diagnosis					
	missing	222	3	225	-

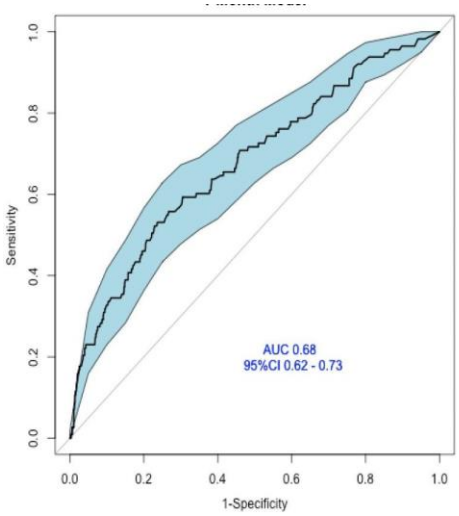
*Mann-Whitney U-test, #Chi square test

Fig. 1



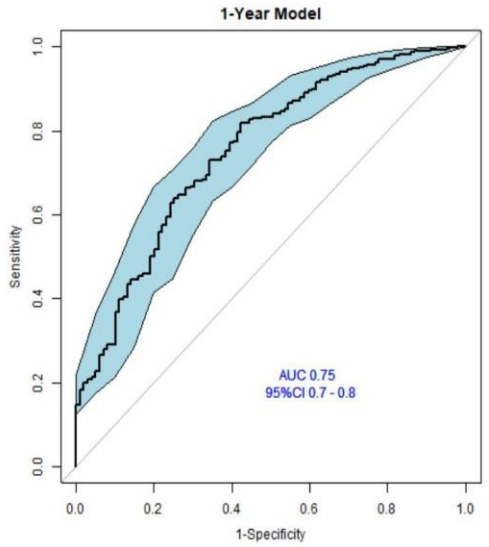
By shuffling copies of all features the chosen Boruta algorithm trains a Random Forest on the overall data. Features are then rejected or confirmed. Confirmed features are ranked with their relative influence in the GBM model as demonstrated

Fig. 2



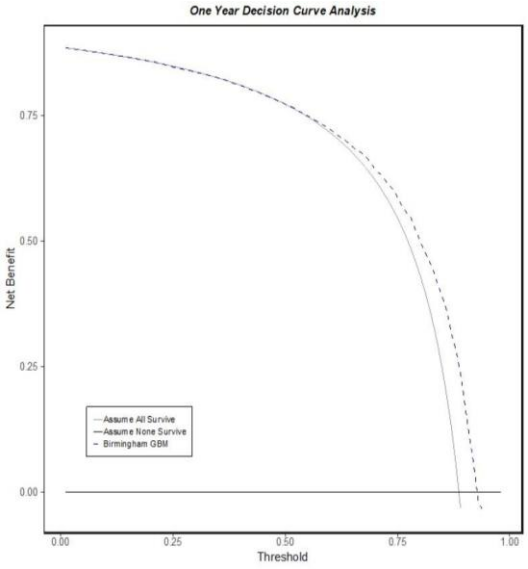
ROC curves of the external validation of the1-year survival BBN model. The discriminatory accurary of the BBN model for survival yielded poor power (0.68).

Fig. 3



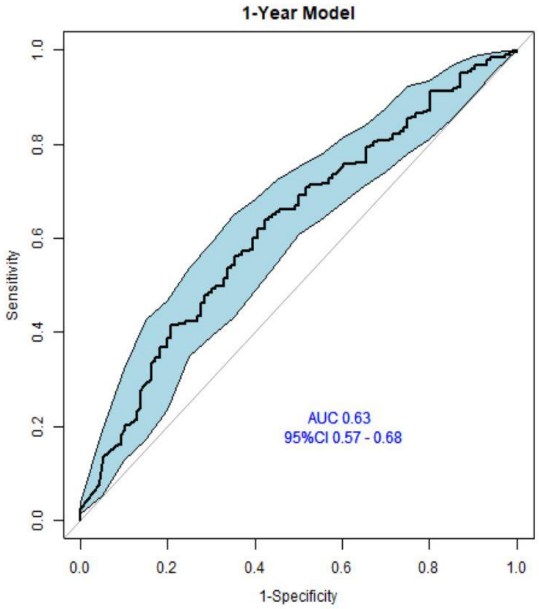
ROC curves of the internal validation of the 1-year survial GBM model. The discriminatory accuracy of the GBM model for survival was classified as good (0.75).

Fig. 4



Net benefit plotted on the decision curve analysis graph against threshold probabilities demonstrating the benefit of intervention based on decision to treat from model output. The curve demonstrates a net-benefit if using the model at thresholds above 0.50 compared to assuming all patients survive. For thresholds below 0.50 the model is no better or no worse that assuming all patients will survive.

Fig. 5



ROC curves of the external validation of the 1-year survival BBN model. The discriminatory accurary of the GBM model for survival yielded poor power (AUC: 0.63.)