

From Department of Molecular Medicine and Surgery
Karolinska Institutet, Stockholm, Sweden

PREDICTION OF MORTALITY IN HIP FRACTURE PATIENTS

Mathias Mosfeldt



**Karolinska
Institutet**

Stockholm 2024

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by Universitetsservice US-AB, 2024

© Mathias Mosfeldt, 2024

ISBN 978-91-8017-369-8

Cover illustration: - Optional - Click to enter a clarification regarding the illustration on the front page. To remove, select entire row and press Delete.

Prediction of mortality in hip fracture patients

Thesis for Doctoral Degree (Ph.D.)

By

Mathias Mosfeldt

The thesis will be defended in public at Clarence Crafoord, A5:04, Eugeniav. 3
Karolinska Universitetssjukhuset, Solna Universitetssjukhuset, Solna,
24-05-31, 09:00

Principal Supervisor:

Associate Professor Karl-Åke Jansson
Karolinska Institutet
Department of Department of Molecular
Medicine and Surgery
Division of Orthopaedics

Co-supervisor(s):

Clinical Professor Henrik Jørgensen
University of Copenhagen
Department of Department of Clinical Medicine
Division of Clinical Biochemistry

Opponent:

Associate Professor Max Gordon
Karolinska Institutet
Department of Clinical Sciences Danderyd
Hospital
Division of Orthopedic Surgery

Examination Board:

Associate Professor Cecilia Mellstrand Navarro
Karolinska Institutet
Department of Department of Clinical Science
and Education, Södersjukhuset
Division of Orthopedic Surgery

Professor Kristina Åkesson
Lund University
Department of Clinical Sciences, Malmö
Division of Clinical and Molecular Osteoporosis
Research

Associate Professor Rikard Wedin
Karolinska Institutet
Department of Department of Department of
Molecular Medicine and Surgery
Division of Orthopaedics

Popular science summary of the thesis

For the elderly population, breaking a hip can be a serious injury and one that needs a lot of medical attention. This type of injury is even linked to a higher chance of dying in the years following the incident. In Scandinavia we have some of the highest rates of this fracture globally and in recent years in Sweden, this group of patients has accounted for nearly 3% of all hospital beds. The total amount of fractures globally is projected to rise in the coming years as populations continue to grow older.

With more and more people living longer, the total number of these fractures is increasing. In fact, there is even an increasing number of people who live to be over 100 years old, and this number is expected to keep growing in the coming years. As a result, the age range of people who could potentially suffer a hip fracture now span nearly 50 years. Besides the age range, the only thing these patients have in common is where the bone is fractured, making the group very diverse.

As a thought experiment, imagine a group of people aged between 60 and 110 and try to guess the approximately 30% of them that would not survive more than a year if they suffered a fractured hip. Now imagine guessing the outcome if you just saw one of them, in pain in a hospital bed. Let's try to make it easier; high age, male sex, certain elevated blood sample values, living in an institution, using heart medication, having several other illnesses, and needing support when walking might all increase the risk. Sometimes having even more things to keep in mind at the same time is not all that helpful!

The main goal of this thesis was to create computer programs that could use similar information to accurately predict how likely someone is to die after breaking their hip, to make the example above easier in real life situations. In preparation for this, we first looked at circumstances concerning hip fractures in patients that are 100 years old or older, as this group of patients had not been studied much before. We also investigated if the regular blood tests done when someone is admitted to a hospital with a fractured hip could help identifying the patients at a higher risk of dying, for patients of any age over 60.

We then used information from a database of Danish hip fracture patients to build computer programs to estimate individualized predictions of mortality by using so

called machine learning algorithms that can “learn” and draw conclusions from calculations on large data sets. We also tested the programs we created on a group of Swedish patients to evaluate how well they worked when applied to a new population.

After investigating all hip fracture patients in Denmark over a 17-year period we found that for patients more than 100 years old, having a hip fracture resulted in a much higher risk of dying. But also somewhat surprisingly, we found that the patients that had reached such a high age had fewer illnesses in their medical history than the comparison group. The patients that had reached ages of a 100 years or older had been healthier during the course of their life than the comparison group. Furthermore, we found that some of the blood samples collected upon admission for a hip fracture had value in prediction of three-month mortality. Finally, we were able to develop accurate computer programs to predict the likelihood of dying 1-, 3-, 6-, and 12-months after a hip fracture and we found that these models could also be used for a Swedish population. An online version of these models is currently being developed so that they can be used to identify the patients that are the most at risk. For researchers it might be beneficiary to investigate these patients specifically and in turn, this might lead to a more rational use of resources as subgroups at high risk might benefit from an intervention that has no effect on another. Last but not least, it could be a tool to aid caregivers when providing information to patients and their relatives about the severity of the injury and the medical situation.

Abstract

Background:

Hip fractures are associated with elevated mortality and require extensive health care resources. As populations grow increasingly older the total amount of hip fractures has also escalated. An increasing amount of people live to be more than 100 years old and this group is expected to reach more than 50 times current estimated levels during the course of this century. As a result, the age span of geriatric hip fracture patients is now almost 50 years, and as the defining characteristic of these patients aside from their age is merely the anatomical region of the fracture, it is a very heterogeneous group. Our main aim was to develop and externally validate machine learning models for prediction of mortality after hip fracture to identify which patients are the most at risk. In preparation for this we wanted to explore the value of routine blood samples for prediction of mortality after hip fracture. Furthermore, we wanted to investigate temporal trends of incidence and mortality in the emerging group of centenarian hip fracture patients, and finally train machine learning models for estimations of mortality 1, 3, 6 and 12 months after hip fracture and externally validate these models so they ultimately could be used clinically.

Methods:

Studies 1 and 3 of the thesis are based on a database of consecutive hip fracture patients from Bispebjerg University Hospital in Copenhagen, Denmark, queried at different time points for different data. Study 1 consisted of 792 patients and their age, sex and admission blood samples as well as follow up data on mortality. The predictive value of preoperative blood samples was compared using receiver operating characteristics curves (ROC) and univariate and multivariate regression was performed to determine which blood samples were associated with 3- month mortality.

Study 2 was a nationwide survey of all hip fractures in Denmark over a 17-year period. A total of 517 centenarian patients had suffered hip fractures and temporal trends of incidence and mortality were analyzed. Furthermore, comparisons to hip fracture patients from the same cohort with the age interval 70 to 99 (n = 124,007) were done concerning incidence, mortality and comorbidities.

Study 3 consisted of 1186 patients from the same database as study 1 but from a longer time interval and with a wide range of biochemical and anamnestic data as potential predictors, as well as follow up data on mortality. Three different types of machine learning models, a Random Forest (RF), an extreme gradient boosting (XGB) and a generalized linear model (GLM) were developed for 1-, 3-, 6-, and 12- month mortality using the 10 to 13 most important features selected by the Boruta algorithm. The data was partitioned so that 70% was used for training and 30% was used as a holdout test set, results were compared using the area under the curve (AUC) for receiver operating characteristics (ROC) curves, calibration slope and intercept and decision curve analysis (DCA).

Study 4 was based on 5055 hip fracture patients from Karolinska Solna and Huddinge from a 10-year period, data was collected from RIKSHÖFT and from Karolinska Data (KARDA). The models developed in study 3 were deployed and comparisons using AUC and calibration curves and DCA were performed. The best performing models were recalibrated as event rates were lower in the external validation data than in the development data.

Results:

Elevated creatinine on admission blood samples was associated with an almost threefold increased 3-month mortality and had an AUC of 0.69 (0.64–0.74). The other blood samples under study had lower AUC values and on multivariate regression only age, creatinine, potassium and albumin remained as risk factors.

For the centenarian hip fracture patients, incidence had declined slightly but mortality had remained stable during the study period. The centenarians under study had less registered comorbidities as measured by Charleston comorbidity index (CCI) than the comparison group of younger hip fracture patients. 68 % of centenarian patients had a CCI of 0 versus 46 % in the comparison group, but the mortality was higher.

The developed machine learning models had AUC values of approximately 0.8 for all timepoints but the XGB model was the overall best performer as it was more calibrated and had better DCA. An online tool based on the XGB models has been developed for evaluation and educational purposes (<https://hipfx.shinyapps.io/hipfx/>).

For the external validation study, mortality was lower in the Swedish data set than in the data that the models were developed on. The AUC values especially

for the longer timeframes was acceptable but lower than in the development study and again, it was the XGB model that performed best overall with values of 0.72, 0.74, 0.75 and 0.77 for 1-, 3-, 6-, and 12-month mortality respectively. In line with the observed difference in event rates models were not well calibrated, so the XGB models were recalibrated using bootstrapped isotonic regression with much improved calibration.

Conclusions:

Of the blood samples under study, creatinine was the best predictor of mortality after hip fracture with an AUC of 0.69 which is similar to results in studies evaluating the use of ASA classification, CCI index and Possum score, so admission blood samples could be an interesting addition to prediction models for mortality after hip fracture.

Our findings on centenarian hip fracture patients suggest the compression of morbidity theory proposed in many other studies on subjects older than 100 years which could be an important finding as this age group is expected to increase drastically in the coming century. Based on our results, healthcare needs do not seem to increase proportionally with age for the group under study. Furthermore, as at least some of the known predictors for mortality does not seem to be linearly separable, prediction models for mortality in hip fracture populations might benefit from machine learning techniques to model complex interactions between factors.

The XGB models developed in the third study of the thesis had the best results for prediction of 1-, 3-, 6- and 12-months mortality after hip fracture. The models performed acceptably in the external validation, especially for the later timepoints, but needed recalibration as the mortality rates were different in the development data and the external validation. The previously developed online tool will be updated with the recalibrated models so that they can be used clinically.

List of scientific papers

- I. Value of routine blood tests for prediction of mortality risk in hip fracture patients.

Mathias Mosfeldt, Ole B Pedersen, Troels Riis, Henning O Worm, Susanne van der Mark, Henrik L Jørgensen, Benn R Duus, Jes B Lauritzen.

Acta Orthop. 2012 Feb;83(1):31-5.

- II. Centenarian hip fracture patients: a nationwide population-based cohort study of 507 patients.

Mathias Mosfeldt, Christian M Madsen, Jes B Lauritzen, Henrik L Jørgensen.

Acta Orthop. 2019 Aug;90(4):342-347.

- III. Development and internal validation of a multivariable prediction model for mortality after hip fracture with machine learning techniques.

Mathias Mosfeldt; Henrik Løvendahl Jørgensen; Jes Bruun Lauritzen; Karl-Åke Jansson.

Calcif Tissue Int. 2024 Apr 16. Online ahead of print

- IV. External validation of machine learning models for estimation of mortality 1, 3, 6 and 12 months after hip fracture.

Mathias Mosfeldt; Henrik Løvendahl Jørgensen; Jes Bruun Lauritzen; Karl-Åke Jansson

Draft

Contents

1	Literature review	1
2	Research aims.....	13
3	Materials and methods	15
4	Results	29
5	Discussion.....	35
6	Conclusions	43
7	Points of perspective	45
8	Acknowledgements.....	47
9	References.....	49

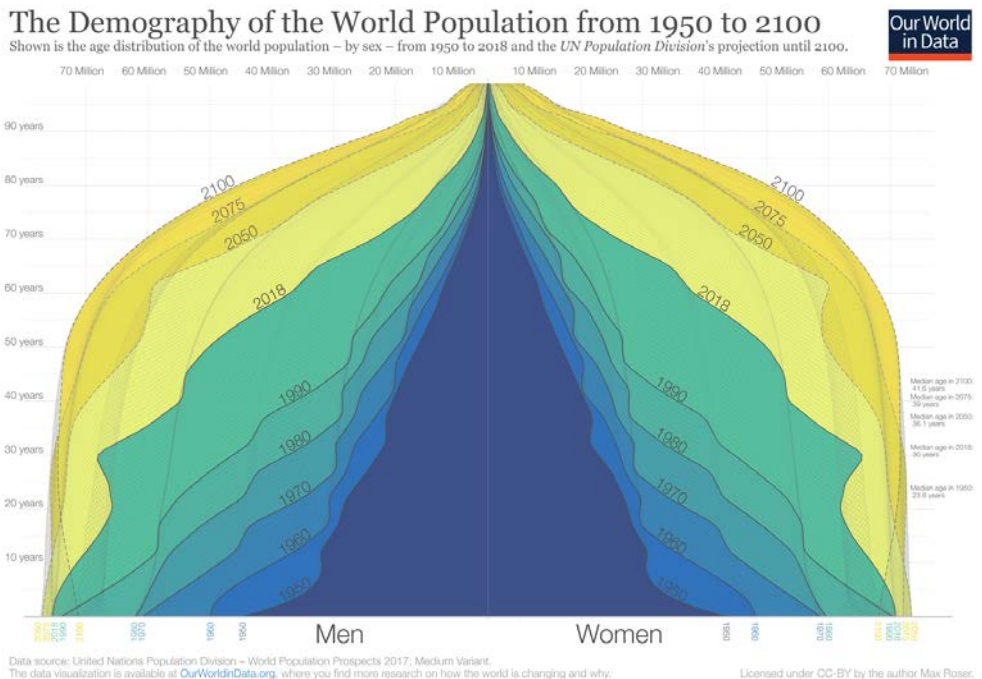
List of abbreviations

ASA score	American Society of Anaesthesiologists Physical Status Classification System. A score from 1 that denotes a healthy patient with no acute or chronic disease to 6 that denotes a patient that is declared brain dead.
AUC	Area under the Curve refers to the total area under a "Receiver Operating Characteristics" curve and is used a measure of a binary classifiers ability to discriminate between to outcomes.
CCI	Charlson Comorbidity Index is a weighted index that was initially developed to predict short term mortality but has been used more as a proxy for burden of comorbidities.
CRS	Civil Registration System in Denmark, virtually complete and contains information on all citizens residing legally in Denmark.
CRN	10-digit civil registration number uniquely identifying individuals in Denmark.
DNPD	Danish National Prescription Database, containing information on all prescribed drugs dispensed at pharmacies in Denmark since 1995.
DNPR	Danish National Patient Registry, a registry containing information on all hospital admissions in Denmark since 1977.
GLM	Generalized Linear Model, a flexible generalization of ordinary linear regression.
HD	Helsinki Declaration, a statement of ethics concerning medical research on humans initiated in 1964 and considered one of the most important documents in this field.

ICD10 codes	International Statistical Classification of Diseases and Related Health Problems, 10th Revision codes, used for medical classification.
LOOCV	Leave-One-Out Cross-Validation, a method for estimating the performance of a predictive model by training on all but one data point and testing on the one left out, repeating this process for each data point.
OR	Odds Ratio, a measure of association between an exposure and an outcome.
POSSUM	"Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity," a scoring system used to predict the risk of morbidity and mortality following surgery.
RF	Random Forest, a tree based machine learning algorithm for classification and regression tasks.
ROC curve	Receiver Operating Characteristics curves are a graphical representation of a binary classifiers performance at different thresholds. For each threshold setting the true positive rate is plotted against the false positive rate.
XGB	XGBoost, a machine learning algorithm based on gradient boosting for classification and regression tasks.

Introduction

Hip fractures represent a significant health concern, particularly within the context of an aging global population. While life expectancy continues to rise and the proportion of elderly individuals in populations is growing globally [1], incidence rates have remained stable so that the total amount of hip fractures increased by 93% from 1990 to 2019 [2].



[1]

This demographic shift poses considerable challenges for healthcare systems worldwide. In addition to the clinical implications, the economic burden associated with hip fractures is substantial, encompassing direct medical costs, rehabilitation expenses, and indirect costs for caregivers [3]. Given the finite resources of healthcare systems and the escalating demand for services driven by population aging, exploration of modern machine learning strategies to help guide clinical management is tempting, as the advent of AI and Big Data is already integrated in many aspects of our daily life to great effect. In this context, the application of machine learning holds promise as a tool to predict outcomes on an individual level by leveraging data in new ways. Considering the

vast amount of high-level data potentially available from daily healthcare to researchers in medicine, exploration of these avenues to increase effectiveness seems obvious.

Increased life expectancy also unveils a subgroup of subjects that has not yet been studied much. There are approximately half a million people worldwide that are over 100 years of age (centenarians) today, but these numbers are estimated to increase 10-fold by the year 2100 [4]. While there are few studies concerning hip fractures in centenarian patients, previous studies on centenarians have showed several characteristics of the group that differ from younger geriatric patients [5]. As such, investigating emerging populations, such as centenarian patients intersect with a looser definition of prediction as it delves into the realm of uncovering patterns and deciphering trends, that in turn might give opportunities to extrapolate insights from a cohort that defies conventional norms.

This dissertation endeavors to contribute to the advancement of knowledge by providing insight about the emerging group of centenarian hip fracture patients by exploring temporal trends of mortality, incidence and associated morbidity in a nationwide study from Denmark spanning a 17-year period.

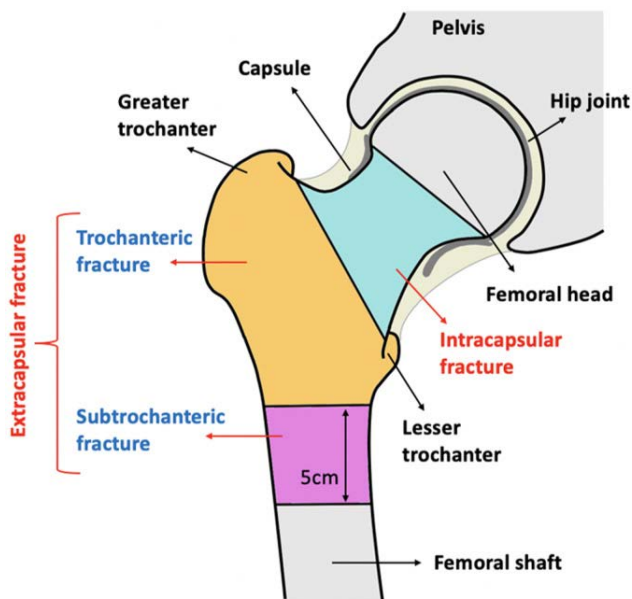
Furthermore, by exploring the value of routine blood samples for prediction of mortality after hip fracture. And finally, by development and subsequent external validation of prediction models for mortality in geriatric hip fracture patients that are accurate, easy to use and rely on data that is already routinely collected upon admission to ultimately improve care and outcomes.

1 Literature review

1.1 Definition & classification

Hip fracture is a general term used to describe the fractures in the proximal part of the femur, pertaining to the neck of femur (also known as cervical, medial or intracapsular), through the trochanters (also known as pertrochanteric, lateral or extracapsular) or immediately below the trochanters (also known as subtrochanteric).

In Sweden, based on numbers from the National registry of hip fractures for the year 2023 (RIKSHÖFT), the proportions of the different fracture types were 51% cervical/ neck of the femur, 3% basocervical/ base of the neck of femur, 38% trochanteric and 8% subtrochanteric [6].



Picture from "Important perioperative factors, guidelines and outcomes in the management of hip fracture." Kyriacou and Khan, 2021. [7]

Several classification systems for fracture types have been suggested and among these the "Garden classification" is one of the most implemented [8]. Originally introduced in 1961, it classifies fractures of the neck of femur into 4 categories depending on the anteroposterior radiograph as follows: incomplete, complete without displacement, complete with partial displacement and complete with full displacement. While later studies have shown that other ways of classifying these fractures might be more beneficial in regards to

development of non-union [9], it is still in daily use across many orthopedic departments across the globe. Similarly for trochanteric fractures, the Evans classification originally introduced in 1949 [10] is still in use, most often in the modified Evans-Jensen version, and classifies trochanteric fractures as follows: type 1 with two non-displaced fragments, type 2 with two displaced fragments, type 3 with three fragments and loss of posterolateral support, type 4 with three fragments and loss of medial support, type 5 has loss of posterolateral and medial support. In addition there is the more pragmatic AO/OTA classification [11].

1.2 Treatment

Historically, hip fracture patients have required a lot of resources and long hospital stays, often in traction, for several months. Today, practically all patients are treated surgically and along with several other medical advances in perioperative care, both hospital time, morbidity and mortality have been significantly improved. Interestingly, Evans noted perhaps ahead of his time, in his article from 1949, that *"...it seems clear that a case can be made for operative fixation of these fractures as a routine treatment. If any exception is to be made it is not in the older age groups, for in these patients operation is a life-saving measure;.."* [10]. Length of stay in the non-operated group in that article was 15 weeks (6 - 36) and in the operated group 7 weeks (3 - 16). In comparison the average length of stay according to the Swedish registry "RIKSHÖFT" for 2022 patients was 6,7 days [6].

Surgery is performed with screws or hip replacement for the cervical fractures and intramedullary nailing or dynamic hip screws for the trochanteric fractures. Methods of fixation are chosen depending on the degree of dislocation and in some cases surgeons' preference and according to the 2023 annual report from RIKSHÖFT, varies quite substantially across Sweden for some fracture types. As of January 2024 there are national guidelines with recommendations for many aspects of the care including choice of surgical treatment [12].

1.3 Epidemiology

Hip fractures in geriatric populations are considered to be fragility fractures and the average age of patients in Sweden 2022 was 82 years. 65% of the patients were female. Approximately 16 000 hip fractures occur annually in Sweden and estimated direct and indirect costs amount to 14 billions SEK annually [6]. In a

study on the “5 largest countries in the EU plus Sweden” costs were estimated to account for €37.5 billion annually [13].

Worldwide annual incidence of hip fracture was estimated to 2.7 millions in an article by Oden et al. in 2013 [14].

Age-standardized incidence rates of hip fractures varies tenfold between countries for both men and women, and the Scandinavian countries have some of the highest incidence rates in the world. According to Kanis et al.,[15] Swedish women have an incidence rate of 539 per 100,000 in stark contrast to 20/100,000 in South Africa and 2/100,000 in Nigeria.

1.4 Risk factors for mortality after hip fracture

1.4.1 Age

Increasing age has perhaps unsurprisingly been demonstrated in several articles to be one of the most important factors associated with an increased risk of mortality after hip fracture [16, 17].

1.4.2 Sex

Male gender has been identified in many studies of risk factors to be associated with an increased risk of mortality after hip fracture. In a large metaanalysis of thirty-three studies with a total of 462,699 patients they found male gender to be predictive of 30-day mortality with an Odds Ratio (OR) of 2.0 (95% CI: 1.85–2.18)[16]. In a meta-analysis from 2010 excess mortality after a hip fracture was elevated even after 10 years of follow-up and was higher for men than for women at all timepoints measured [18]. Several studies have pointed out that mortality after hip fracture among men is almost double that of women [19, 20].

1.4.3 Comorbidities

Several studies have found associations between classification systems for comorbidities such as ASA grade and Charlson Comorbidity Index (CCI) and mortality after hip fracture [16, 17, 21, 22].

ASA score	
5-category classification system used to assess the physical status of patients	<ol style="list-style-type: none"> 1. A normal, healthy patient 2. A patient with mild systemic disease 3. A patient with severe systemic disease 4. A patient with severe systemic disease that is a constant threat to life 5. A moribund patient who is not expected to survive without the operation
CCI score	
Assesses a variety of co-morbidities	1 each: myocardial infarct, congestive heart failure, peripheral vascular disease, dementia, cerebrovascular disease, chronic lung disease, connective tissue disease, ulcer, chronic liver disease
Each condition, depending on the associated risk of dying, is assigned with a score of 1, 2, 3 or 6	2 each: hemiplegia, moderate or severe kidney disease, diabetes, diabetes with complication, tumor, leukaemia, lymphoma
	3 each: moderate or severe liver disease
	6 each: malignant tumour, metastasis, AIDS

From "Metaanalysis of risk factors for mortality in patients with hip fracture", Norring-Agerskov et al., 2013.[22]

The predictive value of ASA and Charlson comorbidity Index on mortality after hip fracture has also been evaluated several times in the literature [23, 24], with relatively poor discriminatory ability and area under the curve (AUC) values for receiver operating characteristic curves of between 0.59 and 0.67 for 3- and 12-month mortality. The ROC is a graphical representation of a binary classifiers performance at different threshold values, and the AUC or c-statistic as it is sometimes called, is the probability that a positive outcome has a higher estimated probability than a negative outcome for two randomly chosen samples in the population examined. An AUC of 0.5 represents the predictive value of flipping a coin or randomly guessing a binary outcome and a value of 1.0 would be perfect predictions. Values in between have sometimes been considered directly comparable for the predictive value of different models but as explained in section 3.2.1, more parameters are needed to evaluate the predictive performance of a model. Several thresholds have been suggested labeling intervals such as poor (0.6-0.7), fair (0.7-0.8), good (0.8-0.9) and excellent (0.9-1) but thresholds are somewhat arbitrary and could be misleading as pointed out by Alba et al [25]. Despite this, the AUC is often the only metric of performance in studies of predictive models and while it is very informative it does have limitations in this regard and should preferably be supplemented with measures of calibration and clinical utility [26].

Living arrangements, dementia and walking ability.

It has been shown in several studies that living in a nursing home or similar facility is associated with an increased risk of mortality after hip fracture [16]. Many of the conditions that necessitates such living arrangements have also been found to be risk factors for mortality after hip fracture, so rather than the actual living arrangements posing a risk it is likely that it is the compound risk of poor walking ability, dementia and overall frailty.

A metanalysis of the effect of dementia on the outcome after hip fracture based on 30 studies with a total of 1,037,049 patients found negative effects on regaining ability to walk, mortality and several medical complications [27]. In that study, dementia increased the risk of one year mortality with an OR of 2.10 (95% CI 1.74–2.54). Other studies have found similar associations [28, 29].

Walking ability is associated with excess mortality after hip fracture [30, 31] and could also be one of the factors contributing to elderly patients needing institutional living. Several previously published models for estimating mortality after hip fracture also use this a predictor [32, 33].

While it might be a crude measurement of its constituents, using institutional or aided living arrangements as risk factor might have an advantage as it is an easily available parameter for retrospective studies, also avoiding the problems associated with diagnosing dementia and quantifying the ability to walk preoperatively for a patient upon presentation with a fracture.

1.4.4 Waiting time to surgery

The issue of waiting time to surgery and perioperative mortality has been somewhat controversial and is currently a quality indicator for hip fracture care in both Sweden and Denmark. Several other countries such as Norway, Germany, the UK, Australia and the US have guidelines or recommendations concerning the timing of surgery but with slightly different time frames, such as “the day of admission or the day after” in the UK and 48 hours in the USA [12].

As pointed out in the national Swedish guidelines for hip fracture care, while time frames and recommendations might vary slightly between countries, time to surgery should preferably be preparation time and not waiting time only [12]. In a study by Daugaard et al. of 38 O2O hip fracture patients, they found risk of death

to increase with an odds ratio of 1.3 for every 24 hours of surgical delay [34]. Of note from that study was also that neither admission during weekends/ holidays or time of day influenced mortality rates. A similar study from Sweden based on data from 59,675 over a five-year period found limited association to mortality four months after surgery. After stratifying for ASA grade, they only found an association between mortality and waiting time of more than 24 hours for patients with ASA 3 and 4 [35]. The same group also found an increased risk for adverse events for patients with waiting time for surgery of more than 24 hours, again only for patients with ASA 3 or 4 [36].

1.5 Mortality

Increased mortality after hip fractures is firmly established in the literature. In a meta-analysis from 2010 they found a 5 to 8-fold increased risk of death the first three months after a fracture when comparing to age and sex matched control groups [37]. The risk decreased over time but never returned to the levels seen in control groups even after 10 years according to that study.

Mortality in a large study consisting of all 122,923 hip fractures that occurred in Denmark between the years of 1999 and 2012 remained stable over the study period and was 9.9% at 1-month and 28.0% at one year [19].

In a similar study on Swedish patients based on the entire population over the age of 60 between the years 1997 and 2017, results were reported separately for men and women. Females had a 30-day mortality of approximately 6% and a one-year mortality of 22%, while the same numbers for males were 11% and 33% respectively.

Mortality rates remained stable during the study period as well but it was noted that a change was seen in the composition of the populations so that the amount of comorbidities as measured by CCI scores increased [20].

1.6 Centenarians and hip fractures

Studies suggest that centenarian populations are going to increase in the coming years which has led to interest about how hip fractures affect individuals in this age group.

Many articles on centenarian hip fractures have been characterized by the difficulty in finding sufficient amounts of patients and due to the small size of this population most studies use retrospective cohorts to describe

characteristics, however with smaller study sizes it is difficult to draw conclusions that can be applied on a larger scale as descriptions of the group as a whole.

In a meta-analysis on the subject, 23 studies were included with a total of 6970 patients, of which only 6 studies had more than 100 patients and 13 studies had less than 50 patients [38].

One month mortality was 22% and one-year mortality 54% so significantly higher than what is usually reported in hip fracture populations in general.

1.7 Prediction models for mortality after hip fracture

Several attempts have been made to predict mortality in this patient group, both using models developed for other purposes or patient groups, and specific models for hip fracture patients either using more traditional “frequentist” type statistics or machine learning approaches, some of the latter with a wide array of predictors making them impractical for clinical use as discussed in the following sections.

1.7.1 Models not using machine learning techniques

The Nottingham Hip Fracture Score (NHFS) originally published by Maxwell, Moran et al. seems to be the model that has found the most widespread use of the previously published models [39]. In contrast to most other models, it has been externally validated several times [40].

The scoring system incorporates age (66–85 yr, ≥ 86 yr), sex (male), number of co-morbidities (≥ 2), mini-mental test score (≤ 6 out of 10), admission hemoglobin concentration (≤ 10 g/dl), living in an institution, and presence of malignant disease to predict 30-day mortality. In the original study the area under the curve (AUC) for receiver operating characteristics (ROC) for mortality was 0.72.

The NHFS has also been published in a slightly altered version with improved results and termed The Almelo Hip Fracture Score (AHFS)[32]. AUC in the AHFS version was reported to be 0.82 for 30-day mortality when patients are divided into a high, medium and a low-risk group. The authors argue that a prediction model is more accurate on a group level than on an individual level and that these three risk groups enough to correctly inform patients about the prognosis and to make clinical decisions.

In the article by Pugely et al. age >80 years, male gender, partially dependent functional status, totally dependent, malignancy, and ASA class were independently predictive of mortality and were chosen for the risk score model [41]. 30-day mortality was predicted with an AUC of 0.702. Both the original NHFS and the system by Pugely et al. were based on rather large groups of patients, 4967 and 4331 respectively.

In a study by Jiang et al. 3891 patients with hip fractures were identified in a Canadian health region from 1994 to 2000 [42]. This study was designed to only include baseline preoperative characteristics so comparisons and risk adjustments between different hospitals would be possible. The parameters included are all pre-existing co-morbidities as well as age, sex and admission from long-term care. In addition to the demographic data the parameters that remained in the final model with the with greatest independent risk were malnutrition, renal failure, pneumonia, pre-existing malignancy, and previous myocardial infarction in decreasing order of significance. Outcome measurements were in-hospital and one-year mortality with AUC values of 0.82 and 0.74 respectively.

The "Physiological and Operative Severity Score for the enUmeration of mortality and morbidity"(POSSUM) score was originally developed and introduced in 1991 for evaluation of perioperative risk with a mixture pre- and perioperative factors in general surgery patients [43]. It is still in use today although recent evaluations have found it to be too imprecise to be recommended [44]. A modified version for orthopedic patients has been proposed [45], but it has had varied results in hip fracture populations. In a meta-analysis that evaluated the observed to expected ratios for prediction of mortality after hip fracture for two versions, the O-Poosum and the P-Poosum, they found the ratios to be 0.61 (0.16–2.38) and 0.68 (0.49–0.95) respectively [46]. There were pooled results of nine studies including 3649 patients for O-Poosum and four studies with 1794 patients for P-Poosum with both systems severely overestimating mortality.

In a study by Burgos et al. from 2008 they evaluated the predictive value for mortality after hip fracture of six risk scoring systems that were nonspecific for this population [23]. ASA classification could predict 90-day mortality with an AUC on ROC curves of 0.60. The other scoring systems used in the study were RISK-VAS, Barthel, Goldman, POSSUM and Charlson and the AUC for these on 90-day mortality was 0.68, 0.69 0.43, 0.64, 0.59 respectively.

In a Swedish study from 2009 by Söderquist et al., [47] a prediction model that was a combination of age, gender, ASA and a questionnaire for diagnosing cognitive dysfunction (Short Portable Mental Status Questionnaire, SPMSQ) was created. In this study an AUC of 0.74 for 24-month mortality was attained. The relatively long time period makes comparisons to other studies difficult, however it could be argued that this makes it easier to obtain accurate results in an ageing population. As observed by Haentjens et al., the increased risk of mortality is most acute in the months after the fracture and substantially decreases over the course of the following years. This trend is also evident in the Kaplan Meier plots in the article by Söderquist et al.

1.7.2 External validation of models not using ML techniques

There is large discrepancy between articles describing the development of models and articles to perform external validation and few of the studies above have sought to externally validate these systems after initial publication. Furthermore, few provide metrics of calibration or clinical usefulness.

In a study by Kau and Kwek for validation of the NHFS on an Asian population, it was noted that accuracy of prediction was good for 1-month and poorer for 3- and 12-month mortality and it was noted that different factors might be at play for different time points which could be important to keep in mind for future models [48]. The NHFS has been externally validated many times. In a meta-analysis featuring both ML models and models not using machine learning techniques, the pooled AUC of external validation of the NHFS yielded a c-statistic of 0.70 (95% CI: 0.68 - 0.72)[49]. The NHFS has also been externally validated specifically for a Swedish population and had a c-statistic of 0.67 (95% CI: 0.59-0.74) for 1-month mortality [50]. In the same study the POSSUM score was also evaluated with an AUC of 0.66 (95% CI: 0.59-0.72). Calibration was measured with observed:expected (O:E) ratios, and the authors mention that the none of these models seemed to be well calibrated. There is also a table where the estimated risk is divided in tentiles of riskbands effectively functioning as a written version of flexible calibration curves. While both algorithms seem poorly calibrated is also worth noticing that only 62 patients were deceased which will make this type of rendition of calibration unstable.

As previously mentioned, most of the published prediction models have not undergone subsequent external validation. In a single study that featured several of the models mentioned in part 2.5.1, performance was evaluated for several

systems on the same population [51]. They found that none of the models showed excellent discriminative power (in this article defined as AUC > 0.80), but the models created specifically for a hip fracture population performed best. The models evaluated in this study were the Charlson Comorbidity Index (CCI), Orthopedic POSSUM, E-PASS, a risk model by Jiang et al., a model by Holt et al. and the NHFS. The hip fracture specific models (Jiang, Holt, NHFS) all performed well with AUC values of 0.76–0.78.

1.7.3 Models using machine learning techniques.

With the onset of big data and machine learning models, several attempts have been made to estimate mortality after hip fractures using these techniques. Many of the articles in the field describe the development of models with wide arrays of predictors and few possibilities for readers to use the models in practice, such as the model by DeBaun et al [52]. This model uses a cohort of 19 835 patients with 47 different predictors to estimate 30-day mortality and achieved excellent results with c-statistics of between 0.83 to 0.92 depending on the model type used. While these might be excellent results, they are hard to interpret as there is no mention of calibration, clinical utility or how the results are meant to be utilized. As noted in the introduction of article 3 in this thesis, many of the previously published articles on prediction of mortality after hip fracture are either of a descriptive nature with no means for the reader to use or externally validate the results [53–55], focus on a smaller subgroup of patients such as critically ill, fractures only of the neck of femur, very high short term risk of mortality [56–58] or finally as in the article from DeBaun et al. using a wide variety of pre- and postoperative parameters making them impractical for assessment of patients upon admission to the hospital [52, 55]. An interesting point is made in this respect in an editorial in *Clinical Orthopedics and Related Research (CORR)* that states the following: *“Papers that simply show that a prediction or diagnosis can be made using machine learning, but do not give readers the ability to use the tool for themselves, are of little interest and we don’t expect to publish many of these.”*[59].

Meta-analyses to compare results of ML models in predictions of mortality and other clinical outcomes after hip fracture have been published [49, 60]. This type of study poses several challenges and elucidates the need for structured reporting concerning development of models as suggested by the TRIPOD guidelines [61], and a methodology that considers necessary performance metrics for a thorough evaluation as suggested by Steyerberg and Vergouwe

[26]. Also of note in the meta-analysis by Lex et al, none of the 39 studies included had undergone external validation.

2 Research aims

The overall main aim was to develop and externally validate prediction models for mortality after hip fractures to provide accurate individualized estimates based on preoperative parameters. In preparation for this we explored the associations between preoperative blood samples and mortality, as well as temporal trends of incidence and mortality in the oldest patients.

2.1 Specific aims

Study 1

To study the predictive value for mortality after hip fracture of routine blood samples obtained at admission.

Study 2

To investigate temporal trends of incidence and mortality in centenarian hip fracture patients in a nationwide study spanning almost two decades and compare this group to younger geriatric hip fracture patients with regards to mortality and comorbidities.

Study 3

To develop accurate predictions models for 1-, 3-, 6- and 12-month mortality by training and comparing several different machine learning methods using easily available preoperative parameters.

Study 4

To externally validate the models developed in study 3 on a Swedish population and ultimately make the models available to use clinical use.

3 Materials and methods

3.1 Study populations & study design

3.1.1 Study 1

For the first study of the thesis, "*Value of routine blood tests for prediction of mortality risk in hip fracture patients*", we used a prospectively collected database of consecutive hip fracture patients from a university hospital in Copenhagen, Denmark (Bispebjerg University Hospital). The data was collected by a study nurse and the attending physician for the purpose of hip fracture research and consisted of 792 consecutive patients that presented with any type of hip fracture between October 2008 and July 2009. Patients below the age of 60 and with known malignant diseases were excluded. Data consisted of age, sex, plasma creatinine, plasma potassium, blood hemoglobin, plasma albumin, blood platelet count, plasma CRP, blood glucose, blood leukocyte count and plasma sodium. The outcome was mortality and vital status was collected from the civil registry system.

A retrospective analysis of the prospectively collected data was conducted to evaluate the predictive value for mortality of routine blood samples taken on admission in hip fracture patients. Univariate and multivariate logistic regression analyses were used to examine the relationship between mortality and the risk factors in the study. Receiver operation characteristic (ROC) curves and the area under the curve (AUC) were calculated for the blood samples in the study to determine their predictive value.

3.1.2 Study 2

The study population in the second study of the thesis, "*Centenarian hip fracture patients: a nationwide population-based cohort study of 507 patients*" consisted of all hip fracture patients in Denmark between January 1, 1996 and December 31, 2012. Patients were identified using ICD-10 codes in the Danish National Patient Registry (DNPR), and the full cohort consisted of 154,047 patients. The main study group consisted of 508 patients aged 100 or more and a comparison group of 124,007 patients aged 70 to 99 was extracted as well. The DNPR contains information on all hospital contacts for the Danish population so data on comorbidities was extracted from the same registry. This data was recorded in the study as Charlson comorbidity index (CCI) if patients had a hospital contact registered with one of the comorbidities in the CCI prior to the

fracture date. Only hospital contacts are registered in the DNPR so a CCI of 0 means that no hospital contacts for any of the comorbidities in the index had taken place. All citizens residing legally in Denmark have a unique 10-digit civil registration number (CRN) that is used in all registries as well as the Civil Registration System (CRS) that contains data on age, sex, emigration and vital status. Registration is mandatory by law and the system has proved to be an excellent tool for research as Denmark also has over 70 national registries to monitor quality of care for specific conditions that use the same CRN to register patients [62].

We calculated yearly absolute crude incidence rates by dividing the number of centenarian hip fractures each year by the total number of centenarians in the population that year and expressed as events per 1000 persons. This has previously been done in similar publications [57] and is a very close approximation to summing up person time at risk over the year. The reason for this is that although roughly 50% die over the course of the year and is therefore no longer at risk, these individuals are continuously “replaced” during the year by 99-year old subjects that turn 100. Trends in incidence rates over the course of the study period were analyzed using a log-linear regression model assuming a Poisson distribution.

Mortality was described as the proportion of patients dead at different timepoints and compared using univariate analysis with Pearson’s chi-squared test. Mortality was also analyzed using the Kaplan–Meier estimator and compared using the log-rank test. Median survival times were calculated with the Kaplan–Meier estimator. Trends in 30-day and 1-year mortality over time were investigated using a log-binomial regression model to calculate relative risk (RR) estimates. The model used vital status as a function of calendar year or as a function of calendar year, age and sex, to describe changes in mortality over time. Two-sided P -values of less than 0.05 were considered statistically significant.

3.1.3 Study 3

The third study of the thesis *“Development and internal validation of a multivariable prediction model for mortality after hip fracture with machine learning techniques”*, uses a wider array of parameters from the database in study 1 and after a longer period of data collection, from September 2008 to September 2010.

Blood samples were taken on admission for the purpose of the database and anamnestic data was recorded by the attending physician or the study nurse. The dataset contained 1465 patients with no known malignant disease, aged 60 or older. Younger patients were excluded as we wanted to investigate mortality after hip fractures in an aging population where mortality rates are much higher [63, 64]. Patients with ASA 5 or 6 (n=1) were also excluded.

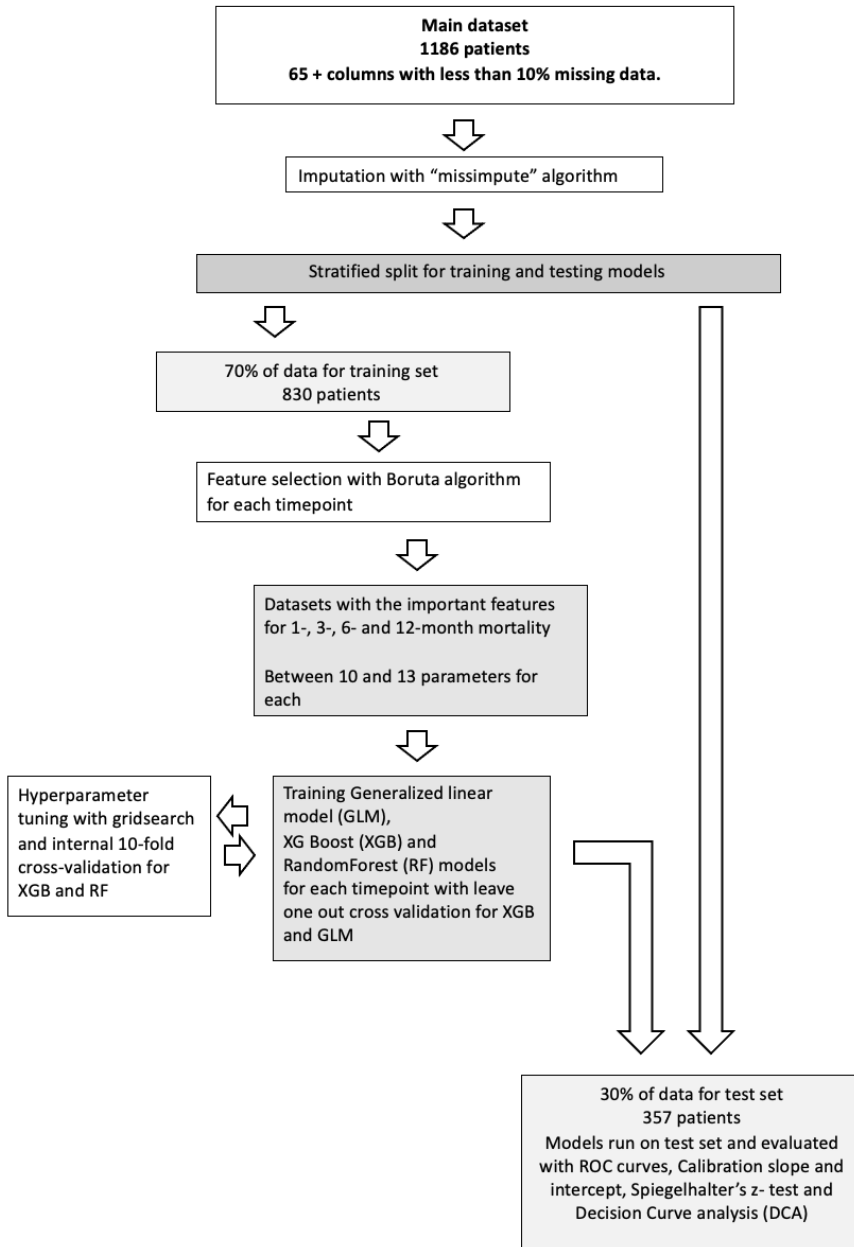
Data on vital status was collected from the CRS in September 2013. The following data was available for further analysis; age, sex, types of medication at admission, anesthesia type, fracture type, type of operation, type of permanent residence (own home, nursing home, homeless) and where patients were admitted from (own home, assisted living, rehab, hospice, hospital, nursing home), New Mobility Score (NMS) [24], American Society of Anesthesiologists physical status (ASA) score, body mass index (BMI), survival (yes/no) at 1, 3, 6 and 12 months after admission, biochemistry (hemoglobin, potassium, sodium, creatinine, calcium, albumin, glucose). During the study period a dedicated orthogeriatric ward was introduced with both geriatricians and orthopedic surgeons attending to patients and this was entered in the database as a potential predictor for mortality (orthogeriatrics (y/n)).

For 109 patients there were no registered blood samples. Exploring the data before imputation, the only pattern that could be found was that 169 patients were missing both albumin and calcium seemingly not at random. No statistical significance could be found between mortality and the cases mentioned above so they were assumed to be missing independently of the outcome and listwise deletion was performed.

For the rest of the dataset no feature had more than 10% missing data. Missing data was handled with the missForest algorithm which is a multiple imputation technique based on random forest modeling that works on both continuous and categorical data [65]. The available data is used to build models predicting the missing values of the dataset and continues for several iterations, each iteration using an improved data set as missing data has been replaced. This algorithm was found to be superior to other commonly used techniques for imputation such as KnnImpute, and MICE (Multiple Imputation by Chained Equations) in the original article, in some cases by over 50%.

A summary of the processes is presented in the figure below which is an excerpt from the article *"Development and internal validation of a multivariable*

prediction model for mortality after hip fracture with machine learning techniques”, published in Calcified Tissues International .



In order to have train/ test sets with similar proportions, a stratified 70/30 split was performed, and the test set of 357 patients was only used in the final evaluation of the models.

As we wanted to create a model that would be easy to use clinically, feature selection was performed using the Boruta algorithm for 1-, 3-, 6- and 12-month mortality to reduce the number of features before model building. This is a wrapper method that uses Random Forest modelling to find the features that have importance for the outcome. Wrapper methods create subsets of candidate features and use the performance of models built on the subsets to evaluate the importance of features. In this case data is duplicated and shuffled at random to create a set of "shadow features". Models are then built on the original data and the shuffled duplicated data so that the importance of features in each set can be compared. Real features that have a higher importance than the random shadow features over several iterations are deemed important. This was performed for each of the timepoints with 1000 iterations for each. Finally smaller training sets with only the chosen parameters for each time point were created and Random Forest (RF), eXtreme Gradient Boosting (XGB) and Generalized Linear Models (GLM) models were trained on the 830 patients of the 70% split for 1-, 3-, 6- and 12-month mortality.

RF models are ensemble tree-based models that are effective for nonlinear data, robust for outliers and can be used on data that contains both numeric and categorical predictors. Simply put it builds a forest of decision trees and uses majority voting to reach a decision on a classification problem. The random part of the name refers to one of the features of this model that makes it different from regular tree-based models. For each node of the tree, the best predictor is chosen from a random subset of predictors ensuring that a wider range of predictors is used and making the model less prone to overfitting to the data set. The random subset is sampled with replacement, so-called "bootstrapping", and the final result of the model is produced by aggregating the decisions of all the separate trees that are built on different parts of the data set. This process is referred to as "bagging" which is an acronym for "bootstrap aggregation" and the result is that the model is less sensitive to overfitting to the training data and is generalizable.

According to the original developers of the RF algorithm, the model effectively performs “leave one out cross validation” (LOOCV) during the training so there is no need for a separate validation set or cross validation [66].

XGB models are also a type of ensemble tree-based models but the development uses “gradient boosting” instead of “bagging”. The term boosting refers to combining several weak models to create a strong model. Gradient boosting refers to developing sequential models and using the error residuals of the previously built model to fit the next one, with the final output being the weighted sum of all the previously built trees. Training was done using LOOCV in order to best compare the results to the results of the random forest model.

GLM is an adaptation of linear regression but using a link function to make the dependent variable linearly related to the predictors. Also, GLM allows for dependent variables to have a non-normal distribution. We wanted to use this model to examine if the classes were linearly separable and if less complicated models could be used for mortality prediction in this population. Training was also done using LOOCV.

Hyperparameters for both the XGB and the RF models were tuned using a grid search and internal 10-fold cross-validation on the training sets.

The final models were employed on the test sets and evaluated using calibration- at- large, calibration slope and intercept, Spiegelhalter's z-test, AUC and DCA. Flexible calibration curves were not included as our test-set did not have more than 200 cases and non-cases which is recommended for flexible calibration curves to be used [67].

3.1.4 Study 4

For the external validation study all patients that were treated for a hip fracture at Karolinska University Hospital Huddinge and Solna over a 10-year period between the 1st of January 2010 and the 31st of December 2020 were included. As this study was intended to externally validate the models from study 3, patients below 60 years of age, with ASA 5 or 6 and patients with known malignant diseases or metastatic fractures were excluded like in the original study. To obtain all the necessary data two different registries were used. The personal registration number was used to match the data after which it was deleted so all further analysis was done with anonymous data.

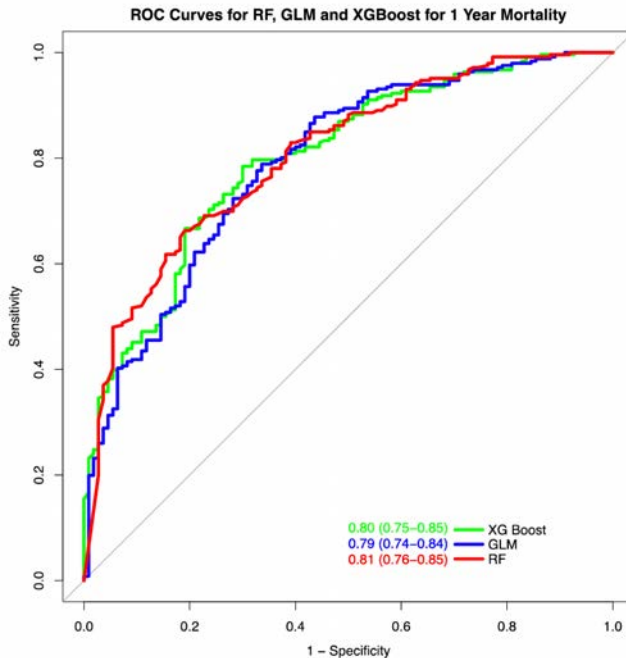
The Swedish National Hip Fracture Registry [6] was used for identifying the patients that had been treated for a hip fracture during the study. This is a national registry that collects a wide range of data related to hip fracture treatment and outcomes with a high grade of completeness. From this registry we were able to collect data on mobilization and walking aids, permanent living arrangements, where patients were admitted from, admission date, Age, BMI, Sex, ASA grade and time of death if applicable.

For blood samples and medication data, the Karolinska database (KARDA) could be used. This is a database that queries the electronic patient charts and was the source of data concerning preoperative levels of Creatinine, Potassium, Albumin, Hemoglobin, Calcium and if patients were under treatment with cardiac medication at the time of admission (any of the following alone or in any combination: diuretics, betablockers, digoxin, vitamin K antagonists, organic nitrates).

3.2 Outcome measures for prediction models

3.2.1 Receiver Operating Characteristics & the c-statistic / Area under the curve

In the field of medical predictions, it is common to use the “Area under the Curve” (often referred to as AUC or the c-statistic) from “Receiver Operator Characteristics” curves (often referred to as ROC curves) to measure accuracy of predictions. The curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity) and is a measure of “discrimination”, or the model’s ability to discriminate between two outcomes. Below is an example of ROC curves comparing three different ML models for prediction of 1 year mortality after hip fracture and their associated AUC values with confidence intervals, from study 3 of this thesis [68].



ROC curves for 1-year mortality with three different model types. Xg Boost: Extreme Gradient Boosting, GLM: Generalized Linear Model, RF: Random Forest

A prediction model that is right 50% of the time, like flipping a coin, yields an AUC of 0.5 and is represented by the 45-degree line from the bottom left to the top right corner in the plot. A model that is always right yields an AUC of 1.0, and the curve would have included the whole area of the plot. In machine learning, values of 1.0 is usually indicative of a mistake or that a training set is extremely overfitted to the data. Most commonly, tools to aid in clinical decision-making will have an AUC somewhere in between 0.5 and 1.0. As mentioned previously, labeled intervals are sometimes used out of convention rather than pure statistical reasons, such as “less than 0.60 reflects poor discrimination; 0.60 to 0.75, possibly helpful discrimination; and more than 0.75, clearly useful discrimination” [25]. However, even if this is often the only measure of predictive ability in this type of literature it does not provide the whole picture [67]. The AUC can be interpreted as the probability that a model will estimate a higher value for a random positive sample than for a random negative sample, but it does not address whether the estimates are far from the actual value. As an example, a model to predict weather could have a high AUC if it estimated the

probability of rain to be 90% on rainy days and 45% on days with no rain, even if the actual frequency of rain those days was only 9% and 4,5%. So, even if the model might have a high AUC it would not always be of much help for guessing what the weather would be like. The AUC can also be misleading as a measure of the predictive value of a model in datasets with uneven distributions of positive/negatives. For instance, if a model is constructed to find the 5% in a population that has a positive outcome it could have a high AUC if all cases were labelled as negatives but would be of little value for prediction. In line with this, an important consideration when assessing ROC curves is that it is a summary the whole range of classification thresholds, while in many cases researchers are only interested in specific situations.

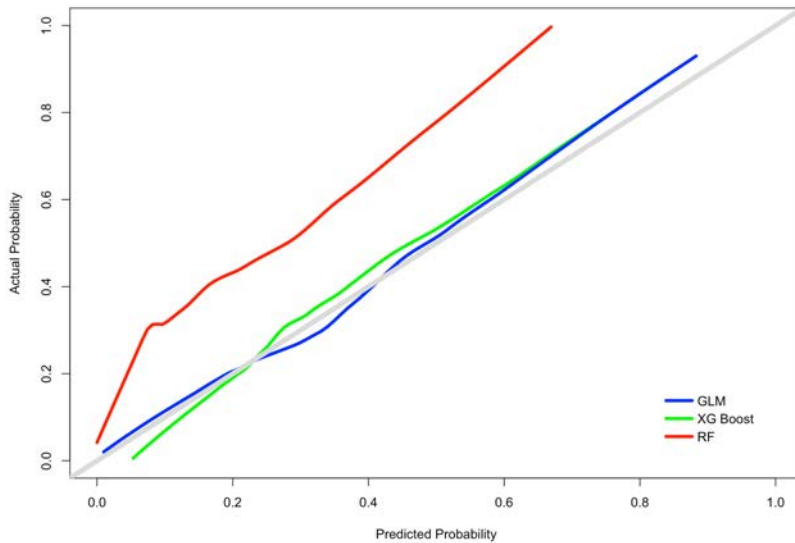
So, while ROC curves are a good measurement of discrimination for prediction models with binary outcomes, they can also be misleading as the only metric of performance.

3.2.2 Calibration

Calibration measures how estimations match the actual frequency of events. For instance, if the weather model from the example above estimated that the probability of rain was 90% even on days that only had a 9% frequency of rain, it would be poorly calibrated and not of much use for forecasting weather. Even though it might have had a high AUC from assigning a higher probability of rain to actual rainy days than to dry days.

Calibration is often discussed in several levels. The term mean calibration or “calibration in the large” measures if the overall predicted risk matches the actual event rate and can be expressed as the numerical difference between the two.

Application of flexible calibration curves provides more information. This is a plot of the actual event rate against the estimated probabilities, often divided into deciles. Below is an example with data from study 3, featuring the flexible calibration curves for the same models used in the example of ROC curves in the precious section above to predict 1 year mortality.



Calibration curves for 1-year mortality with three different model types. Xg Boost: Extreme Gradient Boosting, GLM: Generalized Linear Model, RF: Random Forest

While all three models appeared to perform similarly when considering only AUC, the Random Forest model is poorly calibrated and underestimates the actual probability.

If the estimated probabilities match the event rates for all deciles a straight diagonal 45-degree line will be created from 0 to 1.0. For example, for the decile with an event rate of 10% the probability of the outcome should be 10%. Curves that are above the diagonal line indicate underestimation and curves below indicate overestimation. This provides an opportunity to delve deeper into the predictive properties of a model. For instance, considering a decile of risk in which a model is used to decide on a certain intervention, it is important to know if over- or underestimation occurs in this area.

It is suggested that for flexible calibration curves to be stable, datasets with at least 200 events and non-events are needed, otherwise outliers and groups with very few subjects will create too much noise. For smaller datasets, numeric values of the slope and intercept of the calibration curves are recommended [69]. The slope of the curve was originally referred to as the “spread” by Cox and indicates if estimates are too conservative or too extreme in either direction. So,

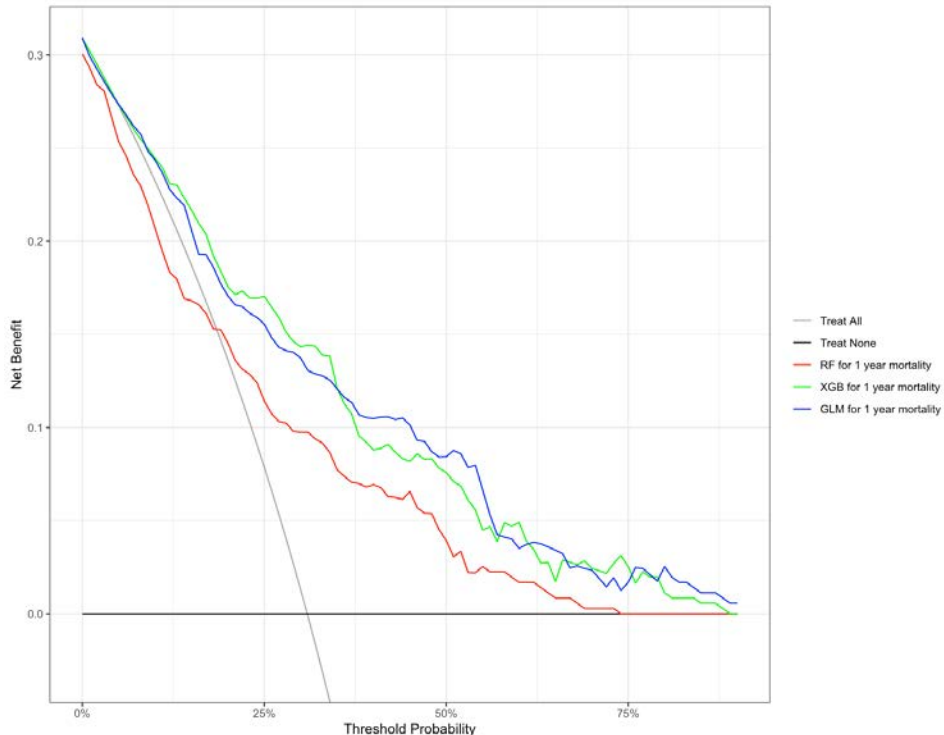
if estimations are too moderate at low risks and too exaggerated at high risk the slope will be less than 1. The target value of the slope is 1 and values above 1 indicate that the spread of the estimates is too narrow, in other words overestimating low risk and underestimating high risk. For the intercept, values larger than 0 indicate underestimation and values less than 0 indicate overestimation. However, by considering only the numeric values of slope and intercept, it is not possible to determine if a model is perfectly calibrated even if these measures are close to 1 and 0 on data sets with insufficient numbers.

3.2.3 Decision Curve analysis (DCA)

Finally, there is the matter of clinical usefulness or in the example with the weather forecast, will the model guide you in a good way about when to stay home or when to bring an umbrella? In the field of predictive modeling this can be addressed with Decision Curve Analysis (DCA) [70] and while the use of this metric is emerging, it has not been utilized much in the literature concerning prediction of mortality in hip fracture patients. Returning to the example of the weather, at the extreme ends the decision is easy. If the sun is shining from blue skies the umbrella stays at home and if the clouds are grey and heavy the person using the model might stay at home. Somewhere in between the decision is more difficult, and this is one of the areas where the DCA can be useful. For a given decision an interval of probability thresholds needs to be decided and for that interval the usefulness of the model can be read either relative to always or never bringing an umbrella or relative to using other models for the same decision. In the field of medical predictive modeling this usually applies to decisions such as performing an invasive biopsy, choosing surgery or conservative treatment, administering a new treatment or simply concerning the usefulness of a general prognostic model for a wide range of decisions and information to patients. Below is the DCA for the same models as in the previous examples, to predict 1-year mortality after hip fracture with comparisons to a strategy of treating all or none instead of using the models to decide which patients to treat. The whole range of threshold probabilities is shown as one of the intended uses could be general information. Strategies of treating all or treating none is included as a default in these types of models and is sometimes labeled as intervention for all or intervention for none. The line for the strategy of "treating all" intercepts with the line for "treating none" at the event rate for the population indicating that this strategy will be effective until you have treated all

the patients with the outcome but for the rest of the population this treatment strategy will result in unnecessary treatment.

Again, the poorly calibrated Random Forest model performs worse than the other models and is worse than treating all at lower thresholds as it underestimates probabilities and labels true positives as negatives.



Decision Curve Analysis curves for 1-year mortality with three different model types. Xg Boost: Extreme Gradient Boosting, GLM: Generalized Linear Model, RF: Random Forest

3.3 Ethics

3.3.1 Overveiw of ethical approvals

Study 1 & 3: Ethical approval from Denmark where the data was collected, "Biomarkører associeret til mortalitet efter hoftebrud" with registration number: HB-2007-103.

Study 2: The study was approved by the Danish Data Protection Agency (j.nr. 2012-58-0004.), whereas ethical committee approval is not required for this type of observational study according to Danish law.

Study 4: Ethical approval from Sweden where the data was collected "Prediktionsmodeller för höftfrakturpatienter", with registration number: 2021-05298.

3.3.2 Ethical considerations

The research nurse that collected the data and the consent did not participate in the care of the patients and the patients did not meet him/ her at later stages of their stay. This way the potential subject is not in a dependent relationship with the person obtaining the consent as stated in § 27 of the Helsinki Declaration (HD) [71].

Patients who are suffering from dementia are allowed to be included in both studies. This is important, because dementia itself can be a negative prognostic factor for the function and survival after hip fracture. This is also stated as a necessary circumstance for conducting research on this group in § 30 in the HD. One of the most apparent ethical problems with this is that patients could be included without actually being able to give an informed consent as one could argue that these subjects might not fully understand the information given. Furthermore, this group of injured elderly patients is definitely in a vulnerable state when arriving to the hospital. However, this study could not be carried out in a different group and the group itself is the beneficiary so in that consideration there is no conflict with § 20 of the HD concerning medical research on vulnerable groups.

The participation in the study does not lead to discomfort and has no health risk for the patient. As such it should not conflict with § 16 in the HD concerning the importance of the objective outweighing the risks and burdens to the research subjects.

The intention of the studies is to secure the best care for the frailest patients. However, there is always risk of abuse and misuse that could apply to several aspects of the care. Patients at high risk could contrary to the intention of the studies be subject to less intensive efforts or caregivers' attitudes could be influenced by knowledge of their estimated survival so that certain costly efforts could be deemed more suitable for the groups with a higher likelihood of survival.

Furthermore, while the advent of prediction modeling and big data has influenced many fields, an important consideration will be that models created in

a medical setting for use by caregivers are not used in areas such as insurance or financial institutions.

4 Results

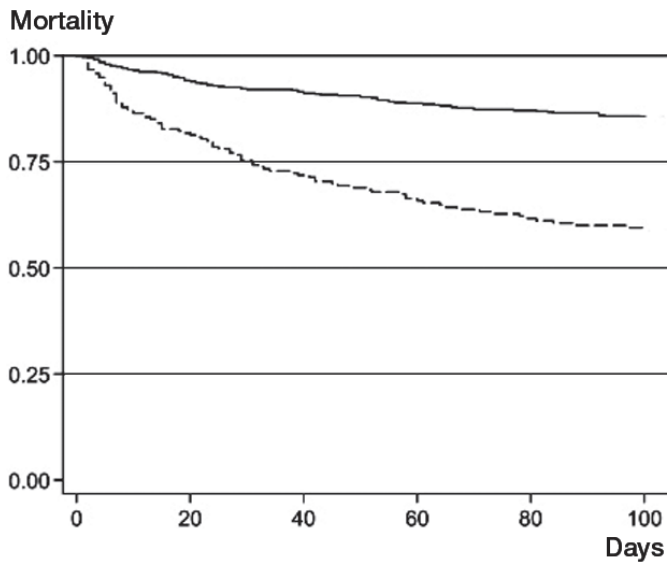
4.1 Study 1

Of the 792 patients under study, 163 (21%) died within three months. In the univariate analysis high age, high creatinine, low albumin, high potassium and low hemoglobin was associated with increased mortality. In the adjusted multivariable model age, creatinine, potassium and albumin remained as risk factors. The AUC of ROC curves were used to evaluate the predictive value of the blood samples in reference to mortality and creatinine alone had an AUC of 69% (95% CI: 0.64–0.74). The AUC of the other values under study was lower and is included in table 1 below.

	Area under ROC	95% CI
Plasma creatinine	0.69	0.64–0.74
Plasma potassium	0.61	0.55–0.66
Blood hemoglobin	0.60	0.55–0.65
Plasma albumin	0.59	0.53–0.64
Blood platelet count	0.58	0.53–0.63
Plasma CRP	0.55	0.50–0.60

Table 1, AUC values with 95% confidence intervals for 3-month mortality of 6 admission blood samples.

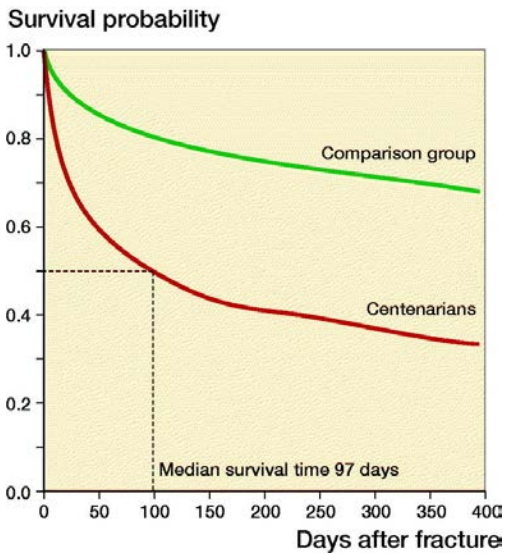
In the Kaplan Meier plot below, patients were divided based on normal or elevated admission creatinine levels and plotted against 3-month mortality and showed an almost threefold increased mortality in the elevated creatinine group.



Kaplan Meier plot of 3- month mortality in relation to creatinine levels at admission
 The dotted line is the group with elevated plasma creatinine and the solid line is the group with normal plasma creatinine.
 Excerpt from “Value of routine blood tests for prediction of mortality risk in hip fracture patients”, Mosfeldt et al., Acta Orthopaedica 2012. [72].

4.2 Study 2

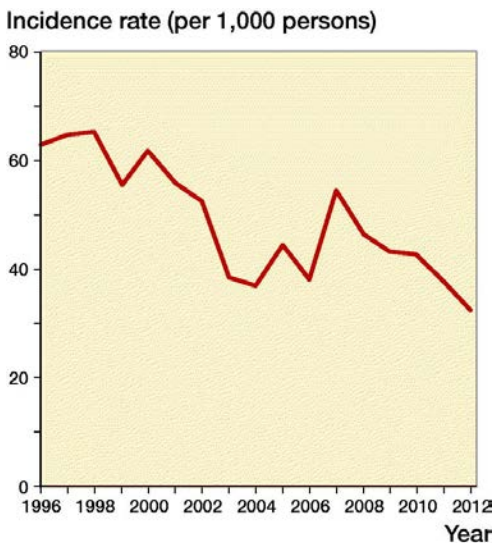
Mortality for the centenarian group under study was 34% at 30 days, 49% at 90 days, 66% at 1 year and 78% and 94% at 2 and 5 years respectively. As expected, this was higher than for the hip fracture patients in the comparison group of patients aged 70 to 99 years of age that had 11%, 19%, 31%, 42%, and 64% at the same time points.



Picture from Mosfeldt et al. Centenarian hip fracture patients: a nationwide population-based cohort study of 507 patients. Acta Orthop. 2019 Aug;90(4):342-347.

More centenarians had CCI scores of 0 than in the comparison group indicating that the centenarians had fewer registered comorbidities. A total of 68% had a score of 0 compared to 46% in the comparison group.

During the study there was a slight annual decrease in the incidence rate of hip fractures for centenarians, estimated at 3.4%.



Picture from Mosfeldt et al. Centenarian hip fracture patients: a nationwide population-based cohort study of 507 patients. Acta Orthop. 2019 Aug;90(4):342-347.

There were slight yearly variations, but mortality remained stable with no statistically significant changes for either or 30-day or calendar year mortality.

4.3 4.3 Study 3

Mortality was 12.2%, 20.4%, 25.9%, 31.8% at 1-, 3-, 6- and 12- months respectively. The parameters chosen by the feature selection model were slightly different for the different time points but “Permanent/ registered residence”, “Admitted from”, “New Mobility Score”, ASA, potassium, creatinine, albumin, “cardiac medications y/n” and age were relevant for all models, with the addition of calcium for 3-, 6- and 12-month models and BMI and hemoglobin for 6-, and 12-month mortality. Of the models developed, the XGB models were the overall best performers. All models performed similarly with c-statistics for 1-, 3-, 6- and 12-month mortality of approximately 0.8 for all time points, see table 2 below. The only models that was fairly well calibrated for all time points was the XGB models, while the RF models was uncalibrated as measured by calibration-at-large, calibration slope, intercept and Spiegelhalters Z- test for all time points. The GLM was uncalibrated for 1-month mortality but performed better for 3-, 6- and 12-month estimations.

Prediction of mortality after hip fracture	1 month		3 months		6 months		1 year	
Overall event rate in test set	0.13		0.19		0.26		0.31	
Calibration measures	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)
XGBoost	0.11	S: 0.99 I: 0.22 0.78 (0.71-0.85)	0.21	S: 0.83 I: -0.03 0.80 (0.74-0.85)	0.25	S: 1.12 I: 0.20 0.79 (0.74-0.84)	0.30	S: 1.20 I: 0.19 0.80 (0.75-0.85)
Generalized Linear Model	0.10	S: 0.31 I: -1.02 0.79 (0.71-0.86)	0.19	S: 0.99 I: 0.22 0.79 (0.73-0.85)	0.25	S: 0.88 I: -0.01 0.78 (0.73-0.84)	0.30	S: 0.82 I: -0.08 0.79 (0.74-0.84)
Random Forest	0.09	S: 0.60 I: -0.18 0.79 (0.72-0.85)	0.04	S: 0.49 I: 0.79 0.80 (0.74-0.85)	0.07	S: 0.60 I: 1.00 0.79 (0.74-0.84)	0.14	S: 0.81 I: 1.03 0.81 (0.76-0.85)

Table 2, results for eXtreme Gradient Boosting (XGB), Random Forest (RF) and General Linear Modeling (GLM) prediction models for mortality after 1-, 3-, 6- and 12- months.

Finally, the DCA was only slightly better than the RF model for the GLM and the XGB models concerning 1-month mortality where models showed clinical benefit for the lower probability thresholds. The differences increased rather drastically as the overall results also improved for the longer time frames with clinical benefit for extended threshold probabilities. The XGB models were the overall best performers with all metrics taken into consideration. An online application using these models was developed for evaluation and educational purposes and is available at: ([https:// hipfx. shiny apps. io/ hipfx/](https://hipfx.shinyapps.io/hipfx/)).

4.4 Study 4

The mortality in the external validation set was lower than in the development set with 8.4%, 15.0%, 20.0%, 27.2% for 1-, 3-, 6- and 12-month mortality respectively. In line with this, the initial models overestimated the risk at all time points before recalibration, except for the RF model that underestimated risk for 3-, 6-, and 12-month mortality as was the case in the development study. The c-statistic was lower than in the development data but still acceptable with the XGB model as the best performer for all timepoints except for the 1-month model where the RF model had slightly higher values. As in the development data the RF model was the least calibrated and for the 1-month model that had the highest AUC of 0.74, the calibration slope was 0.55 and the intercept minus 1.02.

Overall the XGB models seemed to be the best performers so these were chosen for recalibration. The flexible calibration curves were monotonic so bootstrapped isotonic regression was chosen to recalibrate the estimations in favor of Platt scaling/ logistic regression that is more suitable for sigmoidal calibration curves.

The results of the initial application of the models as well as the results after using the recalibrated models are presented in table 3.

Prediction of mortality after hip fracture	1 month		3 months		6 months		12 months	
Overall event rate in test set	0.08		0.15		0.20		0.27	
Uncalibrated models								
Calibration measures	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)	Average predicted risk	Slope, intercept and AUC (CI)
XGBoost	0.12	S: 0.80 I: -0.80 AUC: 0.72 (0.70 - 0.74)	0.22	S: 0.95 I: -0.60 AUC: 0.74 (0.72 - 0.76)	0.30	S: 0.97 I: -0.62 AUC: 0.75 (0.73 - 0.76)	0.38	S: 1.09 I: -0.57 AUC: 0.77 (0.75 - 0.78)
Generalized Linear Model	0.12	S: 0.06 I: -2.17 AUC: 0.70 (0.67 - 0.72)	0.22	S: 0.21 I: -1.36 AUC: 0.71 (0.69 - 0.73)	0.28	S: 0.19 I: -1.10 AUC: 0.73 (0.71 - 0.74)	0.35	S: 0.14 I: -0.81 AUC: 0.74 (0.72 - 0.75)
Random Forest	0.10	S: 0.55 I: -1.02 AUC: 0.74 (0.72 - 0.76)	0.06	S: 0.40 I: -0.24 AUC: 0.73 (0.72 - 0.75)	0.09	S: 0.64 I: 0.34 AUC: 0.75 (0.73 - 0.77)	0.18	S: 0.67 I: 0.21 AUC: 0.75 (0.74 - 0.76)
After calibration with isotonic regression								
XGBoost model after recalibration	0.07	S: 0.88 I: 0.003 AUC: 0.72 (0.70 - 0.74)	0.12	S: 0.89 I: 0.07 AUC: 0.74 (0.72 - 0.76)	0.17	S: 0.90 I: 0.03 AUC: 0.75 (0.73 - 0.76)	0.24	S: 0.94 I: 0.16 AUC: 0.77 (0.75 - 0.78)

Table 3, results for external validation of eXtreme Gradient Boosting (XGB), Random Forest (RF) and General Linear Modeling (GLM) prediction models for mortality after 1-, 3-, 6- and 12-months.

Finally, the previously developed online application will be updated after peer review of the article so that the recalibrated models can be used for Swedish patients.

5 Discussion

5.1 General discussion

While previous improvements in life expectancy have stemmed from declining mortality rates among younger age groups recent advancements are mainly attributed to prolonged lifespans [73], and it is expected that populations globally will reach increasingly higher ages. As an example, the global population of persons that have reached 100 years of age or older, is expected to increase from today's levels of approximately half a million to 50 million by the year 2100 [4]. Limits to aging is a topic of ongoing debate [74], some researchers suggest a limit at approximately 115 years of age [73], others suggest that limits might continue to gradually increase. Nevertheless, regardless of potential limitations, healthcare providers can expect increasing numbers of older patients.

Incidence and total amounts of hip fractures have been increasing worldwide for the last three decades [75]. It has been estimated that the amount of people at risk for an osteoporotic fracture, of which hip fractures account for 54%, will double by 2040 [76], and costs associated with osteoporosis in the EU was €37.4 billions already in 2010 and are expected to increase in a similar fashion [3]. With drastically increasing costs and emerging populations of increasingly older patients, finding new ways to utilize and interpret data that is already readily available is an interesting proposition. Finding innovative ways to work "smarter, not harder" will be paramount. The concept that future innovations and technology will solve the problems associated with population growth is sometimes labeled "technological optimism" and is widely debated in many fields [77]. One of the main points of critique is that the term is often applied loosely with no reference to what actual solutions might be and that many innovations also increase the need for resources accordingly. Furthermore, the areas that benefit from technological advances are not necessarily the same areas that provide the potential problems as technological progress is often driven by economical motivation. Without commenting on that discussion in broader terms it does shed light on some interesting aspects of this thesis. By training machine learning models using data that is already routinely collected it was possible to gain further insight about the prognosis of the patients under study without adding interventions or resources. Also, while the development of machine learning algorithms and the application of big data is in no way

connected to treating geriatric patients, in this case these patients might end up as beneficiaries anyways.

In contrast to many somatic illnesses that share physiologic traits pertinent to the disease, such as arteriosclerosis for patients with acute myocardial injury, the common denominator for the group of hip fracture patients is merely the anatomical location of the broken bone. As such, the group can theoretically include a very wide scope of patients with ages spanning over almost 50 years which makes stratifying within the group compelling.

As an example, in the study by Greve et al, an association between mortality and waiting time was only found for patients with ASA 3 and 4 [35], illustrating that some efforts might have different weight within the group.

Timing of surgery has not been explored in depth in this thesis but is an aspect of the care for these patients that in some way involves many of the orthopedic surgeons in Scandinavia and a good example of an area of the care where differentiation within the group could be beneficial. The latest rapport from RIKSHÖFT [6] describes a nationally declining trend of patients that have been operated within 24 hours, now down to 58%, and that this leads to an elevated risk for complications and increased mortality especially for the sickest patients. Resources might not always be sufficient to reach these goals for all patients but perhaps prioritizing the patients with the highest estimated risk could be advantageous. Similarly, many other issues concerning the care for hip fracture patients could be addressed by investigating stratified groups of patients rather than the whole group as anyone above a certain age could be included due to circumstances resulting in a very heterogenous group. In a recent study from the UK "*Hospital-level organizational factors associated with long-term patient outcomes and costs after hip fracture*" were examined. Amongst others, involving orthogeriatricians in the care was associated with cost savings, shorter length of stay and reduced 1-year mortality [78]. For centers where this measure is not possible for all patients from an organizational/resource standpoint, a model to estimate individualized risk might be helpful in the triage of patients for such efforts.

5.2 Study 1

"Value of routine blood tests for prediction of mortality risk in hip fracture patients"

In the first article of the thesis, we wanted to explore if routine preoperative blood samples could be used for other purposes than making sure patients are ready for surgery. The parameter with the best result for prediction of mortality was creatinine and other studies on hip fractures have found similar results as explored in a meta-analysis by the same group not included in the thesis [79]. Creatinine has also been found in many other studies on mortality as an important predictor in a wide variety of diseases or injuries, such as severe burns [80], COVID 19 [81], and *“Non-relapse mortality after allogeneic hematopoietic stem cell transplantation”*[82] to name a few. While elevated creatinine in itself is harmless, the underlying ailments that lead to this such as cardiac or renal insufficiency are both associated with high mortality rates.

As such the added physiological stress associated with a hip fracture such as dehydration from fasting before surgery, blood loss from the fracture site and surgery and hypotension associated with administered opiates might drive further deterioration of already existing renal or cardiac diseases. Interestingly in a meta-analysis of 14 combined studies, the 1-year mortality for hip fracture patients with end stage renal disease was 27% (95% CI: 18–38%)[83], similar to the 1-year mortality in Sweden and Denmark for hip fracture patients in general. It is noted in the same study however, that on subgroup analysis of patients in the age category of 75 years and above the 1-year mortality was 56%, so these results are difficult to compare to a regular population of hip fracture patients as many of the patients under study are younger and have a different set of physiological circumstances altogether.

In an article that aims to evaluate the effect of heart disease on mortality after hip fracture it was found that 6-month mortality after hip fracture was 30% and 20% in groups of patients with and without previous history of prevalent coronary heart disease respectively [84].

Given the above results, elevated creatinine might be regarded as a compound measure of several illnesses that contribute to increased mortality.

The AUC of creatinine as a predictor of 3-month mortality was 0.69 (0.64–0.74) in our study which is not far from the AUC of several other methods that have been evaluated for prediction of mortality after hip fracture. In a study by Burgos et al. [23], they evaluated the predictive ability of ASA classification, the Barthel index, the Goldman index, the *“Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity (POSSUM) scoring system*, the Charlson

index and the Visual Analogue Scale for Risk (RISK-VAS) scale. The Barthel Index was the best performer with an AUC of 0.69 (0.58 -0.79) but is much more time consuming and complicated to use than simply considering routine blood samples.

An important quality of a prediction model for clinical use is that it is applicable and easy to use in daily practice and that the constituents of the scoring system are readily available. In this respect the use of routine admission blood samples, and parameters such as age and sex are optimal and not subject to inter-observer variation in contrast to parameters like ECG and cardiac and respiratory symptoms. Also, the use of results from others tests such Mini Mental Test or Short Portable Mental Status Questionnaire (SPMSQ) as constituents in the model, requires the need to perform these tests before being able to obtain estimations.

5.3 Study 2:

“Centenarian hip fracture patients: a nationwide population-based cohort study of 507 patients”

In the second article of the thesis, we investigated temporal trends for incidence and mortality after hip fracture in centenarian patients and compared mortality and co-morbidities to a cohort of hip fracture patients aged 70- to 99. We found that mortality had remained stable over the 17-year study period but incidence had decreased, furthermore we found that the centenarians had fewer recorded comorbidities than their younger peers.

Studies of these patient groups provide insight that could be useful in treatment but also describe characteristics of a population that has survived longer than their peers. Furthermore, given that hip fracture patients are among the most resource-intensive groups in the Swedish healthcare system, accounting for 2.7% of all in-patient hospital beds in 2014 [85], it is crucial to evaluate the trajectory of centenarian hip fracture patients, as the anticipated rise in their numbers could significantly impact our healthcare systems.

Previous studies on ageing populations have generated research questions with relevance for other age groups and fields of interest, such as the studies on “the Mediterranean diet” and the proposed relationship between cholesterol and cardiovascular disease [86]. Ancel Keyes hypothesized a relationship between diet and development of coronary heart disease after observing that there was a

relatively large population of centenarians in Southern Italy and that they had a very low incidence of coronary heart disease. This was further investigated in the “Seven Countries studies” which led to the development of theories concerning the effect of saturated and unsaturated fatty acids in the diet on serum cholesterol levels.

Surprisingly and perhaps contrary to expected results we found that the centenarian hip fracture patients had fewer comorbidities than their younger peers. We also found that mortality was much higher among centenarians than among the younger age groups. This pattern has previously been termed “compression of morbidity” in articles concerning health during aging [87] and it is an important finding as it seems to indicate that in this patient group, costs of healthcare cannot be calculated by extrapolating results from the younger age groups. A larger proportion of these patients is of better health during aging and after the trauma of the hip fracture they have a much shorter remaining life expectancy.

Intuitively, it might feel logical to presume that the number of comorbidities and mortality will steadily increase with advanced age. In fact, the relevance of studying this group of patients was questioned during the review process of the article as it was considered uncontroversial and obvious what the results would be and as such there was concern that there might be a mistake in our data on comorbidities. The assumption being that both would be affected negatively by age turned out to hold concerning mortality but not with regard to co-morbidity in our study.

The term compression of morbidity was introduced by Fries and refers to the extension of disease-free years in regards to increased life span [87], and several studies on centenarians have described the same phenomenon [88, 89]. Perhaps unsurprisingly, reaching such a high age seems to be associated with living relatively disease free. In a study on aging that was unrelated to hip fractures, they grouped the centenarians according to age (100–104, 105–109, 110–119) and found that the older the age group, onset of morbidity came at increasingly higher ages, effectively decreasing the amount of time spent living with disease for these individuals [90]. As such, the impact of an ageing population on government-funded health care might be different than intuitively expected.

Several of the previous studies on centenarian hip fractures have been done in slightly different ways. In a study by Mazzola [91] they also found a very high

number of centenarians with no recorded comorbidities both among those who suffered a hip fracture and a control cohort of centenarians, but there were no comparisons to younger cohorts. A smaller study by Barcelo [92] found no difference in recorded comorbidities between centenarians and nonagenarians.

Mortality after hip fractures seems to increase with age and that is in line with other articles and the generally accepted "common knowledge". Incidence decreased in line with other studies on hip fracture patients of all ages during the same period and the same region [93].

5.4 Study 3:

"Development and internal validation of a multivariable prediction model for mortality after hip fracture with machine learning techniques"

In the third article of the thesis we created prediction models for mortality 1-, 3-, 6- and 12-months after hip fracture and compared the results of three different types of machine learning techniques. As discussed in section 2.5.1 several models and systems exist using frequentist statistics but at the time of the start of this work, there was none using machine learning techniques for this purpose. In the meantime, big data, AI and machine learning have exploded and enough articles exist that there are even meta-analysis studies performed on the subject of machine learning and hip fractures as mentioned in section 2.5.3. However, machine learning models reflect the populations that the training data arise from and to the best of our knowledge there are no easy to use, generally applicable, externally validated and well calibrated models suited for the Scandinavian population. The models developed achieved good discrimination with c-statistics of approximately 0.8 for all time points and models, with the XGB model being the best calibrated on the hold-out test set.

Two of the major limitations for the study were the study size and that the development data was approximately 10 years old. As exemplified by van Calster et al. increasing sample size provides more accurate prediction models [69], so a larger data set might have yielded better results. Concerning the age of the development data, mortality rates have remained stable during the last decades so presumably the dates of data collection should not have much effect on results for our purposes. However, it is noted in the article by Meyer et al. that while incidence rates have remained stable, distribution of comorbidities have changed in the Swedish population [20], which might affect results.

5.5 Study 4:

“External validation of machine learning models for estimation of mortality 1, 3, 6 and 12 months after hip fracture”

In the 4th article of the thesis, the prediction models that were developed on a Danish population were externally validated on a Swedish population. Discrimination was good, especially for 3-, 6 and 12-month mortality, but the models were not properly calibrated, so recalibration was performed with much improved results.

As pointed out in the discussion of this article, there is a large discrepancy between the number of articles describing development of a model and the number of articles that attempt to externally validate models. It is usually considered the gold standard to externally validate a prediction model on a different population than on which it was developed to examine if results are generalizable before more widespread application. This might be more complicated than it sounds and its success dependent on several factors. If the model is trained on a population that is too small or trained insufficiently, the model might not perform well as it is trained to specifics of the training population rather than having identified characteristics that can be applied to relationships between predictors and a given outcome in a general population. Furthermore, an excellent model might perform poorly in a different population if parameters that have different characteristics in a different population are used. For instance, as mentioned in the third paper of this thesis, approximately 30% of the population in that study were residents of a nursing home and in a study of a similar population from the US, only 5% were “non-community dwellers”. Therefore, the models developed on a Danish population using living status as a predictor for mortality will most likely perform differently in a population where this parameter has a different context and other implications.

In our study the models for 1 month mortality had the worst results. One possible contributing factor to this could be that the number of cases in this group is the lowest so there was much less training data than in the other categories. While all models for all time frames display signs of overfitting from the development, this was most pronounced in the 1-month mortality models.

A more practical problem standing in the way of more widespread external validation might be limited availability of databases that contain the necessary information, and in line with this a potential limitation of this study 4 was that the

predictor, "New Mobility Score" was not present in the external validation data in its original form so that an approximation was synthesized from similar parameters. Also, a limitation could be that patients who suffered a consecutive hip fracture were kept in the study. If a second hip fracture occurred within a year after the first, the first fracture was censored (n=127). If a second fracture was suffered after more than a year had passed since the first fracture, this was counted as a new instance (n=187). In both cases, an argument could be made that the risk would be higher for the patients suffering a second fracture. However, previous studies have found that most of the increased mortality is during the first year [37], and in spite of these relatively few patients having a potentially higher risk, estimations before recalibration was higher than the event rate.

Considering the available literature in the field as discussed in article 4, there are few similar models that have been externally validated and that has included measures of moderate calibration.

6 Conclusions

Elevated creatinine was associated with almost 3-fold increased mortality after hip fracture and mortality in the quartile of patients with the highest values of creatinine was almost four times that of the patients in the lowest quartile. Creatinine had an ROC of 0.69 (0.64–0.74) which is similar to the predictive ability of many previously published results using systems such as ASA score or POSSUM. AUC for the other blood samples under study was lower and only elevated creatinine, low albumin and high potassium remained as risk factors after adjustment in a multivariate logistic regression model.

The centenarian patients under study had higher mortality at all timepoints than their younger peers but fewer registered comorbidities, supporting the compression of morbidity hypothesis for this population. Mortality remained stable during the study period but there was a yearly decrease in the incidence rate of 3.4%.

By using a combination of 10–13 predictors consisting of preoperative blood samples and anamnestic information we were able to develop machine learning models to predict 1-, 3-, 6- and 12-month mortality with good discrimination, acceptable calibration, and net benefit for a wide range of thresholds of DCA on internal validation. We evaluated RF, GLM and XGB models and achieved AUC of approximately 0.8 for all time points, but the overall best performer were the XGB models.

We externally validated the previously developed prediction models on a population of 5055 Swedish patients and again it was the XGB models that were the best calibrated and the best performers overall.

AUC was slightly lower in the external validation, the XGB models achieved 0.72, 0.75, 0.75 and 0.77 for 1-, 3-, 6-, and 12-month mortality respectively. The overall event rate was lower in the Swedish population and recalibration was performed in order to achieve well calibrated models. The updated models will be made available for healthcare professionals that have access to the necessary predictors after peer review and publication of study for 4.

7 Points of perspective

The findings in this thesis might provide new opportunities for both treatment and research to optimize results and reduce mortality after hip fractures. By using our externally validated prediction tool, resources and interventions can be focused on the groups of patients that are the most at risk so that efforts can be made where they are most needed. The input needed for the models is already available as it is part of the standard preoperative preparations for these patients and as such, incurs no inconvenience or harm to patients and no extra cost for healthcare providers. Several subjects concerning the care of hip fracture patients could be revisited with a new perspective, such as time to surgery or the effect of dedicated geriatric care as it is likely that the effect size of such interventions is higher in patients with an increased risk of mortality than what has been found with studies on hip fracture patients as a group. In turn, guidelines can be developed that aid clinical decisions and triage and help dedicate efforts to the patients that need them the most.

From a broader perspective it seems imperative that healthcare starts to catch up with the ongoing rapid development of integration of machine learning and AI based tools in computer systems developed with the use of big data. We have incredible opportunities in Scandinavia, the whole population is covered by public health care, is registered with civil service numbers and can be linked to a vast amount of meticulously collected registries spanning several decades. Most of these are however still incredulously collected on paper forms and entered in datasheets "by hand".

We collect incredible amounts of data but fail to take advantage on our patients' behalf. While companies such as Klarna, Spotify, Netflix and Google have droves of data scientists employed to maximize the use of the data entered by its users to predict outcomes and behaviors, public healthcare has yet to realize the potential. However, most people encounter this amazing potential daily in more ways than they are probably aware of. Luckily, this development is already driven by profit and the statistics are equally powerful whether it is used to; predict what product to suggest to a consumer to achieve a completed transaction based on a history of searches, or; what intervention to suggest based on a history of medical records. While the prediction models presented in this thesis just pertains to a small, isolated area of healthcare it does provide an example of

how accurate prediction algorithms can be developed based on already available data to provide an estimation that is more than the sum of its parts.

As such some of the most obvious first steps towards implementation of basic machine learning would be to have the medical charts system setup so that it could be queried as a database. Furthermore, collection of the necessary data for registries should not have to rely on manually cross referencing and collecting but rather set up so that it is collected automatically. Prediction models for a multitude of outcomes could be developed and integrated in the medical charts systems so that estimations are done automatically, and models are continuously being updated with the data that is being added to further improve accuracy. The speech recognition software already available and used for much of this thesis could be used to present secretaries with a draft of the doctor's notes along with the sound file to drastically decrease the time and efforts needed to write it, possibly freeing up resources to employ data scientists to benefit health care by utilizing the data that is already present in our medical charts. Ultimately the goal for that type of integration must be to let computers do calculations and data processing instead of functioning as big typewriters and filing cabinets and letting doctors and health care professionals make informed decisions and care for patients.

While talking robots dressed as nurses might be in our future someday much of the technology labeled AI is already in use everywhere around us waiting to be used for the benefit of patients.

8 Acknowledgements

Karl-Åke Jansson, my supervisor for the thesis who also got me started at Karolinska and supported me in so many ways over the years. I am forever grateful for your understanding and help towards me and my family when you were my boss as well as my supervisor.

Henrik L. Jørgensen, one of the smartest persons I have ever meet. Always in a cheerful mood and with a keen interest in anything challenging. I am truly grateful for everything you have thought me, much of it might not even have been on purpose but by being such an inspiration.

Jes Bruun Lauritzen, who got me started with research in the first place, sent me to my first international congress and a highly appreciated support over the years. Like Henrik Jørgensen, always cheerful and inspiring.

Henrik Lundblad, my mentor in deformity surgery and a role model in most other things! Forever grateful for all the good times, all the patience and for teaching me pretty much everything I know about surgery and much of what I know about a list of things that became too long to include here. Anyone else would have given up on this impossible task long ago!

All the colleagues at Karolinska, especially the Orthopedic Department in Solna where I spend my days, I am so glad to work in an environment where I can be inspired daily by your efforts and I am so honored to take part in all the amazing things we do and the good times we have. Richard Wallensten for always helping colleagues out and spreading enthusiasm about orthopedics, and for the very much appreciated help in proofreading this any many other colleague's thesis. Gunnar Sandersjö for all the friendly support and his ambitions for our department.

The colleagues at the Orthopedic Department at Groote Schuur Hospital in Cape Town for their hospitality. It was a life changing experience and I hope to go back someday! Thanks to Vieslav Boscovic for the insider tips, to Tim Sluis for teaching me about shaping surfboards and last but not least to Maritz Laubscher for teaching me about frames and above all for the friendships.

Jonathan A. Forsberg for opening my eyes to machine learning and getting me started on that path. I also hope we will meet on a mountaintop or remote island someday!

The Limb Lengthening and Reconstructive Surgery (LLRS) Nordic group for our continuous efforts in sharing knowledge the Scandinavian way.

All my amazing friends in Oslo, Copenhagen and Stockholm (and around the world). I started writing individual texts, but it got a little bit out of hand. You guys are the best! I will give my acknowledgements to you in person, especially if I forgot your name below.

Erlend Tenden Øverbø (30+ years of friendship and still going strong, hoping for 30+ more!), Morten Stavseth, Helge Sund, Anette Søholt, Anneli Kjellberg, Peter Stemann Andersen, Jonas Haukaas, Daniel Adams-Ray, David Razzaz, Sebastian McCallum and Johan Gulled, you are all very special to me and I am so thankful for our friendships.

Mikkel Mosfeldt for being a beloved brother and my oldest friend! And of course, his band of Norwegian Mosfeldt's.

My dear mother Live Hov, longevity and vitality personified.

Fanny Bergström for the love and support and for putting up with my crazy talk!

Finally, Iris and Edith Mosfeldt, literally the meaning of life and my company for the best times I have ever had. I love you girls!

9 References

1. Roser , Max (2019) The global population pyramid: How global demography has changed and what we can expect for the 21st century. Our World in Data
2. Dong Y, Zhang Y, Song K, Kang H, Ye D, Li F (2023) What was the Epidemiology and Global Burden of Disease of Hip Fractures From 1990 to 2019? Results From and Additional Analysis of the Global Burden of Disease Study 2019. Clin Orthop Relat Res 481:1209-1220
3. Hernlund E, Svedbom A, Ivergard M, Compston J, Cooper C, Stenmark J, McCloskey EV, Jonsson B, Kanis JA (2013) Osteoporosis in the European Union: medical management, epidemiology and economic burden. A report prepared in collaboration with the International Osteoporosis Foundation (IOF) and the European Federation of Pharmaceutical Industry Associations (EFPIA). Archives of osteoporosis 8:136
4. Robine JM, Cubaynes S (2017) Worldwide demography of centenarians. Mech Ageing Dev 165:59-67
5. Milman S, Barzilai N (2015) Dissecting the Mechanisms Underlying Unusually Successful Human Health Span and Life Span. Cold Spring Harb Perspect Med 6:a025098
6. RIKSHÖFT.SE Swedish National Hip Fracture Registry (RIKSHÖFT) 2023.
7. Kyriacou H, Khan WS (2021) Important perioperative factors, guidelines and outcomes in the management of hip fracture. J Perioper Pract 31:140-146
8. Garden RS (1961) LOW-ANGLE FIXATION IN FRACTURES OF THE FEMORAL NECK. The Journal of Bone & Joint Surgery British Volume 43-B:647-663
9. Parker MJ, Kendrew J, Gurusamy K (2011) Radiological predictive factors in the healing of displaced intracapsular hip fractures. A clinical study of 404 cases. Hip Int 21:393-398
10. Evans EM (1949) THE TREATMENT OF TROCHANTERIC FRACTURES OF THE FEMUR. The Journal of Bone & Joint Surgery British Volume 31-B:190-203

11. Meinberg EG, Agel J, Roberts CS, Karam MD, Kellam JF (2018) Fracture and Dislocation Classification Compendium–2018. *J Orthop Trauma* 32 Suppl 1:S1–S170
12. sjukvård NsfkH-o (2024) Nationellt vårdprogram för höftfraktur. <https://www.nationelltklinisktkunskapsstod.se/globalassets/nkk/nationell/media/dokument/kunskapsstod/varprogram/nationellt-varprogram-for-hoftfraktur.pdf>
13. Borgström F, Karlsson L, Ortsäter G, et al. (2020) Fragility fractures in Europe: burden, management and opportunities. *Arch Osteoporos* 15:59
14. Odén A, McCloskey EV, Johansson H, Kanis JA (2013) Assessing the impact of osteoporosis on the burden of hip fractures. *Calcif Tissue Int* 92:42–49
15. Kanis JA, Odén A, McCloskey EV, Johansson H, Wahl DA, Cooper C, Life IOFWGoEaQo (2012) A systematic review of hip fracture incidence and probability of fracture worldwide. *Osteoporos Int* 23:2239–2256
16. Bui M, Nijmeijer WS, Hegeman JH, Witteveen A, Groothuis-Oudshoorn CGM (2023) Systematic review and meta-analysis of preoperative predictors for early mortality following hip fracture surgery. *Osteoporos Int*
17. Åhman R, Siverhall PF, Snygg J, Fredrikson M, Enlund G, Björnström K, Chew MS (2018) Determinants of mortality after hip fracture surgery in Sweden: a registry-based retrospective cohort study. *Sci Rep* 8:15695
18. Haentjens P, Magaziner J, Colón-Emeric CS, Vanderschueren D, Milisen K, Velkeniers B, Boonen S (2010) Meta-analysis: Excess Mortality After Hip Fracture Among Older Women and Men. *Annals of Internal Medicine* 152:380–390
19. Jantzen C, Madsen CM, Lauritzen JB, Jørgensen HL (2018) Temporal trends in hip fracture incidence, mortality, and morbidity in Denmark from 1999 to 2012. *Acta Orthop* 89:170–176
20. Meyer AC, Ek S, Drefahl S, Ahlbom A, Hedström M, Modig K (2021) Trends in Hip Fracture Incidence, Recurrence, and Survival by Education and Comorbidity: A Swedish Register-based Study. *Epidemiology* 32:425–433
21. Smith T, Pelpola K, Ball M, Ong A, Myint PK (2014) Pre-operative indicators for mortality following hip fracture surgery: a systematic review and meta-analysis. *Age Ageing* 43:464–471
22. Norrington-Agerskov D, Laulund AS, Lauritzen JB, Duus BR, van der Mark S, Mosfeldt M, Jørgensen HL (2013) Metaanalysis of risk factors for mortality in patients with hip fracture. *Dan Med J* 60:A4675
23. Burgos E, Gomez-Arnau JI, Diez R, Munoz L, Fernandez-Guisasola J, Garcia del Valle S (2008) Predictive value of six risk scores for outcome after surgical repair of hip fracture in elderly patients. *Acta Anaesthesiol Scand* 52:125–131

24. Quach LH, Jayamaha S, Whitehouse SL, Crawford R, Pulle CR, Bell JJ (2020) Comparison of the Charlson Comorbidity Index with the ASA score for predicting 12-month mortality in acute hip fracture. *Injury* 51:1004–1010
25. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G (2017) Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 318:1377–1384
26. Steyerberg EW, Vergouwe Y (2014) Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 35:1925–1931
27. Hou M, Zhang Y, Chen AC, Liu T, Yang H, Zhu X, He F (2021) The effects of dementia on the prognosis and mortality of hip fracture surgery: a systematic review and meta-analysis. *Aging Clin Exp Res* 33:3161–3172
28. Bai J, Zhang P, Liang X, Wu Z, Wang J, Liang Y (2018) Association between dementia and mortality in the elderly patients undergoing hip fracture surgery: a meta-analysis. *J Orthop Surg Res* 13:298
29. Ioannidis I, Mohammad Ismail A, Forssten MP, Ahl R, Cao Y, Borg T, Mohseni S (2022) The mortality burden in patients with hip fractures and dementia. *Eur J Trauma Emerg Surg* 48:2919–2925
30. Iosifidis M, Iliopoulos E, Panagiotou A, Apostolidis K, Traios S, Giantsis G (2016) Walking ability before and after a hip fracture in elderly predict greater long-term survivorship. *J Orthop Sci* 21:48–52
31. Hu F, Jiang C, Shen J, Tang P, Wang Y (2012) Preoperative predictors for mortality following hip fracture surgery: a systematic review and meta-analysis. *Injury* 43:676–685
32. Nijmeijer WS, Folbert EC, Vermeer M, Slaets JP, Hegeman JH (2016) Prediction of early mortality following hip fracture surgery in frail elderly: The Almelo Hip Fracture Score (AHFS). *Injury* 47:2138–2143
33. Holt G, Smith R, Duncan K, Finlayson DF, Gregori A (2008) Early mortality after surgical fixation of hip fractures in the elderly: an analysis of data from the scottish hip fracture audit. *J Bone Joint Surg Br* 90:1357–1363
34. Daugaard CL, Jørgensen HL, Riis T, Lauritzen JB, Duus BR, van der Mark S (2012) Is mortality after hip fracture associated with surgical delay or admission during weekends and public holidays? A retrospective study of 38,020 patients. *Acta Orthop* 83:609–613
35. Greve K, Modig K, Talbäck M, Bartha E, Hedström M (2020) No association between waiting time to surgery and mortality for healthier patients with hip fracture: a nationwide Swedish cohort of 59,675 patients. *Acta Orthop* 91:396–400
36. Greve K, Ek S, Bartha E, Modig K, Hedström M (2023) Waiting more than 24 hours for hip fracture surgery is associated with increased risk of

adverse outcomes for sicker patients: a nationwide cohort study of 63,998 patients using the Swedish Hip Fracture Register. *Acta Orthop* 94:87–96

37. Haentjens P, Magaziner J, Colón-Emeric CS, Vanderschueren D, Milisen K, Velkeniers B, Boonen S (2010) Meta-analysis: excess mortality after hip fracture among older women and men. *Ann Intern Med* 152:380–390
38. Abelleira Lastoria DA, Benny CK, Smith T, Hing CB (2023) Outcomes of hip fracture in centenarians: a systematic review and meta-analysis. *Eur Geriatr Med* 14:1223–1239
39. Maxwell MJ, Moran CG, Moppett IK (2008) Development and validation of a preoperative scoring system to predict 30 day mortality in patients undergoing hip fracture surgery. *Br J Anaesth* 101:511–517
40. Wiles MD, Moran CG, Sahota O, Moppett IK (2011) Nottingham Hip Fracture Score as a predictor of one year mortality in patients undergoing surgical repair of fractured neck of femur. *Br J Anaesth* 106:501–504
41. Pugely AJ, Martin CT, Gao Y, Klocke NF, Callaghan JJ, Marsh JL (2014) A risk calculator for short-term morbidity and mortality after hip fracture surgery. *J Orthop Trauma* 28:63–69
42. Jiang HX, Majumdar SR, Dick DA, Moreau M, Raso J, Otto DD, Johnston DWC (2005) Development and Initial Validation of a Risk Score for Predicting In-Hospital and 1-Year Mortality in Patients With Hip Fractures. *Journal of Bone and Mineral Research* 20:494–500
43. Copeland GP, Jones D, Walters M (1991) POSSUM: a scoring system for surgical audit. *Br J Surg* 78:355–360
44. Bürtin F, Ludwig T, Leuchter M, Hendricks A, Schafmayer C, Philipp M (2023) More than 30 Years of POSSUM: Are Scoring Systems Still Relevant Today for Colorectal Surgery? *J Clin Med* 13:
45. Mohamed K, Copeland GP, Boot DA, Casserley HC, Shackelford IM, Sherry PG, Stewart GJ (2002) An assessment of the POSSUM system in orthopaedic surgery. *J Bone Joint Surg Br* 84:735–739
46. Yang G, Cui G, Liu Y, Guo J, Yue C (2023) O-POSSUM and P-POSSUM as predictors of morbidity and mortality in older patients after hip fracture surgery: a meta-analysis. *Arch Orthop Trauma Surg* 143:6837–6847
47. Söderqvist A, Ekström W, Ponzer S, Pettersson H, Cederholm T, Dalén N, Hedström M, Tidermark J, Stockholm Hip Fracture G (2009) Prediction of mortality in elderly patients with hip fractures: a two-year prospective study of 1,944 patients. *Gerontology* 55:496–504
48. Kau CY, Kwek EB (2014) Can preoperative scoring systems be applied to Asian hip fracture populations? Validation of the Nottingham Hip Fracture Score (NHFS) and identification of preoperative risk factors in hip fractures. *Ann Acad Med Singap* 43:448–453

49. Liu F, Liu C, Tang X, Gong D, Zhu J, Zhang X (2023) Predictive Value of Machine Learning Models in Postoperative Mortality of Older Adults Patients with Hip Fracture: A Systematic Review and Meta-analysis. *Arch Gerontol Geriatr* 115:105120
50. Jonsson MH, Bentzer P, Turkiewicz A, Hommel A (2018) Accuracy of the Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity score and the Nottingham risk score in hip fracture patients in Sweden – A prospective observational study. *Acta Anaesthesiol Scand* 62:1057-1063
51. Karres J, Heesackers NA, Ultee JM, Vrouwenraets BC (2015) Predicting 30-day mortality following hip fracture surgery: evaluation of six risk prediction models. *Injury* 46:371-377
52. DeBaun MR, Chavez G, Fithian A, Oladeji K, Van Rysselberghe N, Goodnough LH, Bishop JA, Gardner MJ (2021) Artificial Neural Networks Predict 30-Day Mortality After Hip Fracture: Insights From Machine Learning. *J Am Acad Orthop Surg* 29:977-983
53. Forssten MP, Bass GA, Ismail AM, Mohseni S, Cao Y (2021) Predicting 1-Year Mortality after Hip Fracture Surgery: An Evaluation of Multiple Machine Learning Approaches. *J Pers Med* 11:
54. Cary MP, Jr., Zhuang F, Draelos RL, Pan W, Amarasekara S, Douthit BJ, Kang Y, Colón-Emeric CS (2021) Machine Learning Algorithms to Predict Mortality and Allocate Palliative Care for Older Patients With Hip Fracture. *J Am Med Dir Assoc* 22:291-296
55. Li Y, Chen M, Lv H, Yin P, Zhang L, Tang P (2021) A novel machine-learning algorithm for predicting mortality risk after hip fracture surgery. *Injury* 52:1487-1493
56. Lei M, Han Z, Wang S, Han T, Fang S, Lin F, Huang T (2023) A machine learning-based prediction model for in-hospital mortality among critically ill patients with hip fracture: An internal and external validated study. *Injury* 54:636-644
57. Oosterhoff JHF, Savelberg A, Karhade AV, Gravesteijn BY, Doornberg JN, Schwab JH, Heng M (2022) Development and internal validation of a clinical prediction model using machine learning algorithms for 90 day and 2 year mortality in femoral neck fracture patients aged 65 years or above. *Eur J Trauma Emerg Surg* 48:4669-4682
58. Harris AHS, Trickey AW, Eddington HS, Seib CD, Kamal RN, Kuo AC, Ding Q, Giori NJ (2022) A Tool to Estimate Risk of 30-day Mortality and Complications After Hip Fracture Surgery: Accurate Enough for Some but Not All Purposes? A Study From the ACS-NSQIP Database. *Clin Orthop Relat Res* 480:2335-2346

59. Leopold SS, Porcher R, Gebhardt MC, Gioe TJ, Manner PA, Rimnac CM, Wongworawat MD (2020) Editorial: Opposites Attract at CORR®–Machine Learning and Qualitative Research. *Clin Orthop Relat Res*. United States, pp 2193–2196
60. Lex JR, Di Michele J, Koucheiki R, Pincus D, Whyne C, Ravi B (2023) Artificial Intelligence for Hip Fracture Detection and Outcome Prediction: A Systematic Review and Meta-analysis. *JAMA Netw Open* 6:e233391
61. Collins GS, Reitsma JB, Altman DG, Moons KG (2015) Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Bmj* 350:g7594
62. Mainz J, Hess MH, Johnsen SP (2019) The Danish unique personal identifier and the Danish Civil Registration System as a tool for research and quality improvement. *Int J Qual Health Care* 31:717–720
63. Ekegren CL, Edwards ER, Page R, Hau R, de Steiger R, Bucknill A, Liew S, Oppy A, Gabbe BJ (2016) Twelve-month mortality and functional outcomes in hip fracture patients under 65 years of age. *Injury* 47:2182–2188
64. Thoors O, Mellner C, Hedström M (2021) Good clinical outcome for the majority of younger patients with hip fractures: a Swedish nationwide study on 905 patients younger than 50 years of age. *Acta Orthop* 92:292–296
65. Stekhoven DJ, Bühlmann P (2012) MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28:112–118
66. Breiman L (2001) Random forests. *Machine learning* 45:5–32
67. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Topic Group 'Evaluating diagnostic tests and prediction models' of the Si (2019) Calibration: the Achilles heel of predictive analytics. *BMC Med* 17:230
68. Mosfeldt M, Jørgensen HL, Lauritzen JB, Jansson K (2024) Development and Internal Validation of a Multivariable Prediction Model for Mortality After Hip Fracture with Machine Learning Techniques. *Calcif Tissue Int*
69. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW (2016) A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 74:167–176
70. Vickers AJ, van Calster B, Steyerberg EW (2019) A simple, step-by-step guide to interpreting decision curve analysis. *Diagn Progn Res* 3:18
71. World Medical A (2013) World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 310:2191–2194
72. Mosfeldt M, Pedersen OB, Riis T, Worm HO, Mark S, Jørgensen HL, Duus BR, Lauritzen JB (2012) Value of routine blood tests for prediction of mortality risk in hip fracture patients. *Acta Orthop* 83:31–35

73. Dong X, Milholland B, Vijg J (2016) Evidence for a limit to human lifespan. *Nature* 538:257–259
74. Lenart A, Vaupel JW (2017) Questionable evidence for a limit to human lifespan. *Nature* 546:E13–E14
75. Feng JN, Zhang CG, Li BH, Zhan SY, Wang SF, Song CL (2024) Global burden of hip fracture: The Global Burden of Disease Study. *Osteoporos Int* 35:41–52
76. Oden A, McCloskey EV, Kanis JA, Harvey NC, Johansson H (2015) Burden of high fracture probability worldwide: secular increases 2010–2040. *Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 26:2243–2248
77. Gonella F, Almeida CMVB, Fiorentino G, Handayani K, Spanò F, Testoni R, Zucaro A (2019) Is technology optimism justified? A discussion towards a comprehensive narrative. *Journal of Cleaner Production* 223:456–465
78. Baji P, Patel R, Judge A, et al. (2023) Organisational factors associated with hospital costs and patient mortality in the 365 days following hip fracture in England and Wales (REDUCE): a record-linkage cohort study. *Lancet Healthy Longev* 4:e386–e398
79. Laulund AS, Lauritzen JB, Duus BR, Mosfeldt M, Jorgensen HL (2012) Routine blood tests as predictors of mortality in hip fracture patients. *Injury* 43:1014–1020
80. Ma QM, Tang WB, Li XJ, et al. (2024) [Multicenter retrospect analysis of early clinical features and analysis of risk factors on prognosis of elderly patients with severe burns]. *Zhonghua Shao Shang Yu Chuang Mian Xiu Fu Za Zhi* 40:249–257
81. Alizad G, Ayatollahi AA, Shariati Samani A, Samadzadeh S, Aghcheli B, Rajabi A, Nakstad B, Tahamtan A (2023) Hematological and Biochemical Laboratory Parameters in COVID-19 Patients: A Retrospective Modeling Study of Severity and Mortality Predictors. *Biomed Res Int* 2023:7753631
82. Weise G, Massoud R, Krause R, Heidenreich S, Janson D, Klyuchnikov E, Wolschke C, Zeck G, Kröger N, Ayuk F (2024) Development and Validation of a Concise Objectifiable Risk Evaluation Score for Non-Relapse Mortality after Allogeneic Hematopoietic Stem Cell Transplantation. *Cancers (Basel)* 16:
83. Shimamura Y, Kuniyoshi Y, Ueta H, Miyauchi T, Yamamoto M, Tsujimoto Y (2023) Mortality After Hip and Spine Fractures in Patients With End-Stage Kidney Disease: A Systematic Review and Meta-Analysis. *Cureus* 15:e49636
84. Robbins JA, Buzkova P, Barzilay JI, et al. (2023) Mortality Following Hip Fracture in Older Adults With and Without Coronary Heart Disease. *Am J Med* 136:789–795 e782

85. regioner SSKo (2020) Att följa omställningen till Nära vård – En översikt av mått och indikationer.
86. Minelli P, Montinari MR (2019) The Mediterranean Diet And Cardioprotection: Historical Overview And Current Research. *J Multidiscip Healthc* 12:805–815
87. Fries JF (1980) Aging, natural death, and the compression of morbidity. *N Engl J Med* 303:130–135
88. Ailshire JA, Beltrán-Sánchez H, Crimmins EM (2015) Becoming centenarians: disease and functioning trajectories of older US Adults as they survive to 100. *J Gerontol A Biol Sci Med Sci* 70:193–201
89. Sebastiani P, Sun FX, Andersen SL, Lee JH, Wojczynski MK, Sanders JL, Yashin A, Newman AB, Perls TT (2013) Families Enriched for Exceptional Longevity also have Increased Health-Span: Findings from the Long Life Family Study. *Front Public Health* 1:38
90. Andersen SL, Sebastiani P, Dworkis DA, Feldman L, Perls TT (2012) Health span approximates life span among many supercentenarians: compression of morbidity at the approximate limit of life span. *J Gerontol A Biol Sci Med Sci* 67:395–405
91. Mazzola P, Rea F, Merlino L, Bellelli G, Dubner L, Corrao G, Pasinetti GM, Annoni G (2016) Hip Fracture Surgery and Survival in Centenarians. *J Gerontol A Biol Sci Med Sci* 71:1514–1518
92. Barcelo M, Francia E, Romero C, Ruiz D, Casademont J, Torres OH (2018) Hip fractures in the oldest old. Comparative study of centenarians and nonagenarians and mortality risk factors. *Injury* 49:2198–2202
93. Abrahamsen B, Vestergaard P (2010) Declining incidence of hip fractures and the extent of use of anti-osteoporotic therapy in Denmark 1997–2006. *Osteoporos Int* 21:373–380