Utrecht University

Master Artificial Intelligence
Thesis (45 ECTS)

&

Value2Health

Predicting length of stay, discharge destination and mortality of patients with hip fractures

Patient flow optimization

Author
Laira Xenia Fransen
4253663

Daily supervisors
Willemijn van der Kooi
Henk Broekhuizen

First & second examiner
Krista Overvliet
Chris Janssen

July 2019

Preface

In front of you, you find the master thesis ‘Predicting length of stay, discharge destination and mortality of patients with hip fractures’. The aim of this research was to identify relevant parameters for the prediction of length of stay and discharge destination for hip fracture patients, in order to optimize the patient flow. Another goal was to develop and assess various predictive models on their predictive ability.

This thesis was written as part of my graduation internship for the Artificial Intelligence master program at Utrecht University, commissioned by Value2Health and in collaboration with SAZ (Samenwerkende Algemene Ziekenhuizen). The research and writing of this thesis took place during November 2018 – July 2019.

The research questions were devised in consultation with my daily supervisor, Willemijn van der Kooi, and initial first examiner, Sjoerd Stuit. Various methods have been applied to answer the research questions.

Hereby, I would like to thank my daily supervisors, Willemijn van der Kooi and Henk Broekhuizen, for their great guidance and support during this period. Also, I am grateful to my first examiners, Sjoerd Stuit and Krista Overvliet, for their insights and advises when decisions had to be made. In addition, I would like to thank the SAZ for their cooperation and my colleagues of Value2Health for their assistance. I also want to thank all participants for their participation. The research would not have been complete without them.

Finally, I would like to thank my boyfriend, parents and friends for their wise words and providing a sympathetic ear to bring this research and thesis to a successful ending.

I wish you much pleasure with reading my thesis.

Laira Fransen

Amsterdam, July 2019

Abstract

Background. In the Netherlands, the annual incidence of hip fractures is 88 per 100,000. A major issue in healthcare concerns the shortage of beds in hospitals, caused by a decreased flow to nursing homes. Consequently, patients have a longer hospital stay, which could lead to unnecessary complications and a longer rehabilitation period. Therefore, optimization of the patient flow is desired.

Aim. In this study, we investigated which variables are relevant for the prediction of length of stay (LOS), discharge destination and mortality. Moreover, we investigated different models on their predictive performance.

Methods. Various methods have been applied to achieve the goal, namely: literature study, interviews, model development and statistical analysis (ANOVA). We compared regression, lasso regression and random forest (RF) models with and without feature selection.

Results. This research showed that *age, fracture type* and *involvement of geriatrician* are important predictors for LOS. The most suitable model was RF without feature selection. Furthermore, it showed that *age, involvement of geriatrician* and *living situation prior to the injury* are important predictors for discharge destination. The best model was RF without feature selection. Next, it showed that *age, dementia* and *pre-surgery mobility* are important predictors for mortality. Lastly, statistical tests showed that the best models were not significantly better than all other models included in the comparison.

Conclusion. These findings suggest that RF without feature selection could be used in patient flow optimization for hip fracture patients. However, these are not statistically significant and therefore the models could be improved.

Keywords: hip fracture patients, length of stay, discharge destination, mortality, patient flow optimization, predictive models.

Table of Contents

List of Abbreviations

Abbreviation	Meaning	Additional explanation
Abs	Absolute	Describes how far a number is from zero and it is denoted as a positive number.
ADL	Activities of Daily Living	Examples of daily activities are: washing, feeding, dressing, visiting the toilet and walking.
AI	Artificial Intelligence	The simulation of human intelligence processes by machines.
AMTS	Abbreviated Mental Test Score	Test for rapidly assessing elderly patients for the possibility of dementia. The score ranges from 1 to 10. A score less than 7 is an indication for dementia.
ANOVA	Analysis of Variance	Tests whether there is an indication that multiple models are significantly different from each other.
ASA	American Society of Anesthesiologists	Indicator of the patient's health status on a scale from I to V, where ASA-score I means the patient is completely healthy and ASA-score V means the patient is dying.
AUC	Area Under the Curve	Measure which indicates how well the model can separate different classes. The value ranges from 0 to 1. The higher the score, the better the performance.
BMD	Bone mineral density	The amount of bone mineral in bone tissue. A low density could indicate osteoporosis.
BMI	Body Mass Index	A person's weight in kilograms (kg) divided by their height in meters squared (m^2).
CCI	Charlson comorbidity index	Estimates risk of readmission and mortality in patients with multiple comorbidities. The score ranges from 0 to 24. The higher the score, the higher the risk.
COPD	Chronic Obstructive Pulmonary Disease	Lung disease where the lungs of the patient are damaged. Characteristics are: long-term breathing problems and poor airflow.
Cor.	Correlation	Shows whether and how strongly variables are related. The value ranges from 0 to 1. The higher the number, the stronger the variables are correlated.
DD	Discharge destination	Describes the destination of discharge. Examples are: home, nursing home and care home.
Df	Degrees of freedom	The number of independent models minus the number of parameters used for estimation (N – 1).
DHFA	Dutch Hip Fracture Audit	Registry in the Netherlands which aims at improving quality of care for hip fracture patients.
DHS	Dynamic hip screw	Type of surgery.
DICA	Dutch Institute for Clinical Auditing	Registry which provides insight in quality of care in the Netherlands. This makes benchmarking possible.

ECI	Elixhauser comorbidity index	Estimates risk of comorbidities. The score ranges from -7 to 12. The higher the score, the higher the risk.
ELV	Eerstelijnsverblijf	Discharge destination.
F1-score	-	Weighted average of precision and recall.
FIM	Functional Independence Measure	Measures what an individual can perform under certain circumstances. The score ranges from 1 to 7. The lower the score, the more assistance is needed.
Freq.	Frequency	Measure for how often a variable occurred in 10 model runs. Score ranges from 1 to 10. The higher the number, the more important the variable.
General	General anesthesia	Anesthesia mode.
GN	Gamma nails	Type of surgery.
GRZ	Geriatrische revalidatiezorg	Discharge destination.
HA	Hemiarthroplasty	Type of surgery.
ICU	Intensive Care Unit	Hospital department.
IM	Intramedullary pen	Type of surgery.
Imp.	Weighted importance	Measure that combines the median ranking and frequency, resulting in the variable importance. The score ranges from 1 to 100. The higher the number, the more important the variable.
ISS	Injury Severity Score	Assesses trauma severity and indicates mortality, morbidity and hospitalization time. The score ranges from 1 to 75. A score greater than 15 means major trauma.
IT	Information technology	Anything related to computing technology.
KATZ	-	Index of Independence in Activities of Daily Living. The score ranges from 0 to 6. The higher the score, the more dependent the patient is.
LASSO	Least Absolute Shrinkage and Selection Operator	Prediction model.
LISS	Less invasive stabilization system	Type of surgery.
LOS	Length of stay	Refers to how many days a patient was admitted to the hospital.
MAE	Mean absolute error	Absolute average of all differences between a prediction and observation.
ME	Mean error	Average of all differences between a prediction and observation.
MRDM	Medical Research Data Management	Partner of Value2Health. MRDM processes medical data on behalf of healthcare organizations in the Netherlands.
NLR	Neutrophil-to-lymphocyte ratio	Marker of subclinical inflammation. Higher NLR can predict mortality.
Osteo	Osteoporosis	Disease that weakens the bone: the density and quality are reduced. It makes them more likely to break.
PFNA	Proximal femoral nail with anti-rotating	Type of surgery.

PPV	Positive predicted value	Proportion of positive predictions that are true positive.
Q1	Lower quartile	Splits off the lowest 25% of data from the highest 75% of the data.
Q3	Upper quartile	Splits off the highest 25% of data from the lowest 75% of the data.
Rank	Median ranking	Measure for the median value of the ranking of a variable in 10 model runs. Score ranges from 1 to 10. The lower the number, the more important the variable.
Reg.	Regression	Prediction model.
Regional	Regional anesthesia	Anesthesia mode.
RF	Random forest	Prediction model.
RMSE	Root mean squared error	Standard deviation of prediction errors.
SAZ	Samenwerkende Algemene Ziekenhuizen	A total of 28 local hospitals in the Netherlands.
SD	Standard deviation	Measures the dispersion of data relative to the mean.
SNAQ	Short Nutritional Assessment Questionnaire	Indicator for malnutrition. The score ranges from 0 to 7. A score of 2 or more leads to the additional food supplements. A score of 3 or more leads to involvement of dietician.
SNF	Skilled Nursing Facility	Facility that provides medical care by trained specialists.
Spinal	Spinal anesthesia	Anesthesia mode.
THA	Total hip arthroplasty	Type of surgery.
TTO	Time-to-operation	Refers to how many days the patient had to wait for surgery.

List of Tables

List of Figures

1. Introduction

The research presented in this thesis is executed within the scope of the running project between Value2Health and '*Samenwerkende Algemene Ziekenhuizen'* (SAZ) in the Netherlands. The SAZ-hospitals [1] aim at value-based healthcare by gaining insight in quality of care with respect to the costs. Within the SAZ-project they aim at 1) improving the objective quality of care, 2) improving the health status of the local population and 3) showing the efficiency of care by benchmarking patient outcomes [1]. Two distinct patient populations are included in the SAZ-project: patients with colon cancer and patients with a hip fracture. This thesis will only focus on the patients with a hip fracture in collaboration with six SAZ-hospitals [2].

The aim of this research is to investigate relevant predictor variables for the prediction of length of stay, discharge destination and mortality. Successful prediction could lead to optimization of the patient flow and assistance in decision support for treatment of hip fracture patients.

1.1 Background information

Worldwide, 1.6 million people are annually affected by a hip fracture [2]. In the Netherlands, the annual incidence of hip fractures is 88 per 100,000 [3,4]. Given the fact that most hip fractures occur in patients with an age above 60 years [3,4] and the fact that the population in the Netherlands is aging, it is expected that the annual incidence of hip fractures will increase to 125 per 100,000 in 2040 [3]. In 2012, 85% of hip fracture patients were initially hospitalized at the first aid of a local hospital [3].

Typically, the process for hip fracture patients is as follows. First, the patient is diagnosed at the emergency department of a hospital. Second, the patient is transferred from the emergency department to another hospital department (to wait for surgery), e.g. nursing ward, or directly to the surgery room in case of emergency. After surgery, the patient has to stay in the hospital for a few more days before hospital discharge. The last step in the process is a recovery period, in which the patient's outcome is monitored.

For the diagnoses of hip fractures, different types of hip fractures exist, which need different approaches of treatment. The fractures can be divided in two main groups: intracapsular and extracapsular fractures [5] (Figure 1). The latter exists of three types: intertrochanteric, reverse oblique and sub-trochanteric fractures [5]. Most hip fractures occur in either the femoral neck area, i.e. intracapsular, or in the intertrochanteric region [6] (Figure 1). Two types of distinct hip fractures in each region can be identified. For the fractures in the femoral neck, one can identify

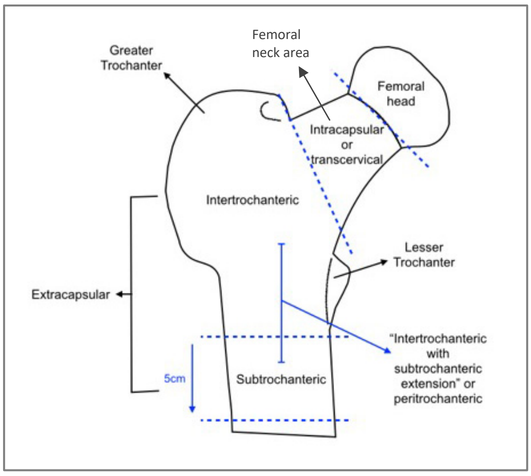

Figure 1. Classification of different parts of the hip.

---

1 Ziekenhuis Amstelland (Amstelveen), BovenIJ Ziekenhuis (Amsterdam), Wilhelmina Ziekenhuis Assen (Assen), Bravis Ziekenhuis (Bergen op Zoom), Maasziekenhuis Pantein (Beugen), Rode Kruis Ziekenhuis (Beverwijk), Ijsselland Ziekenhuis (Capelle aan den Ijssel), Het Van Weel-Bethesda Ziekenhuis (Dirksland), Slingeland Ziekenhuis – Santiz (Doetinchem), Ziekenhuis Nij Smellinghe (Drachten), Treant Zorggroep - locatie Scheper (Emmen), St. Anna Ziekenhuis (Geldrop), ADRZ (Goes), Beatrixziekenhuis (Gorinchem), Saxenburgh Groep (Hardenburg), Ziekenhuis St. Jansdal (Harderwijk), Tjongerschans Ziekenhuis Heerenveen (Heerenveen), Elkerliek Ziekenhuis (Helmond), Treant Zorggroep - locatie Bethesda (Hoogeveen), Laurentius Ziekenhuis (Roermond), Bravis Ziekenhuis (Roosendaal), Ommelander Ziekenhuis Groningen (Scheemda), Antonius Ziekenhuis Sneek (Sneek), Treant Zorggroep - locatie Refaja (Stadskanaal), ZorgSaam Zeeuws-Vlaanderen (Terneuzen), Ziekenhuis Rivierenland (Tiel), Bernhoven (Uden), SJG Weert (Weert), streekziekenhuis Koningin Beatrix – Santiz (Winterswijk), Zaans Medisch Centrum (Zaandam), LangeLand Ziekenhuis (Zoetermeer).

2 BovenIJ Ziekenhuis (Amsterdam), Bravis Ziekenhuis (Bergen op Zoom), Het Van Weel-Bethesda Ziekenhuis (Dirksland), St. Anna Ziekenhuis (Geldrop), Saxenburgh groep (Hardenburg), ZorgSaam Zeeuw-Vlaanderen (Terneuzen).

displaced and nondisplaced fractures. Whereas for the intertrochanteric region, one can identify stable and unstable fractures [7].

In most cases, a surgical treatment is preferred over no treatment. The reason for this is that without surgery, the patient will never be able to walk and therefore performing a surgery is recommended. At the highest level, there are two different types of surgical interventions to cure the patient: replacement arthroplasty and internal fixation [5]. Replacement arthroplasty can be divided in two groups: hemiarthroplasty (HA) and total hip arthroplasty (THA). In 40% of all cases a HA or THA is performed [8]. In both cases, the femoral head is replaced with a metal implant. In case of a THA, there is an additional replacement of the socket [5]. Controversially, internal fixation is not about the (total) replacement of the damaged bone with a prothesis, but rather about returning the damaged bone to its position and maintaining that position by using screws, plates or nails [5]. In the SAZ-hospitals, they mainly use HA, THA, cannulated screws, sliding screws and intramedullary (IM) hip pens as treatment for hip fracture patients.

Before discharge and after surgery, the patient stays in the hospital. The post-surgery length of stay is likely to depend on multiple factors and varies between 2 days to several weeks. Figure 2 shows the typical distribution of the total length of stay (pre- and post-surgery) of elderly hip fracture patients found in a study by Monacelli et al. [9]. When the patient is ready for the hospital discharge, there are multiple discharge destinations. In 2012 in the Netherlands, 49% of hip fracture patients was discharged to home, 24% of the patients was discharged to a nursery home, 13% to a different institution, 10% was discharged to a retirement home, rehabilitation centre or different hospital and 4% died in the hospital before discharge [3] (Figure 3).

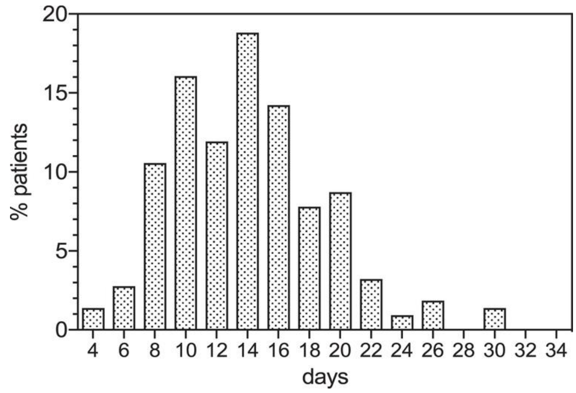

Figure 2. Distribution of length of stay.

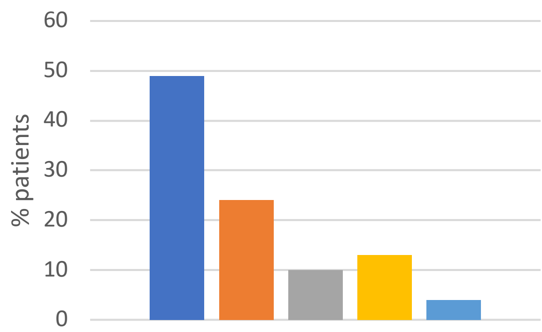

Figure 3. *Discharge destinations in the Netherlands.* *Home; nursery home; rehabilitation center/different hospital; different institution; died.*

During the period after discharge, patients have to recover from their injury and their outcome is monitored. Unfortunately, not all patients make it to (full) recovery. While some patients have to cope with complications and, therefore, have to be re-admitted in the hospital, others do not survive after surgery. The following are some facts with respect to mortality among hip fracture patients: 75% of all hip fractures occur among women, but the mortality rate among men is almost double compared to women [2]. In a study by Cornwell et al. (2004), they found an in-hospital mortality rate of 2.2% and 3.9% for femoral neck fractures and intertrochanteric fractures respectively [7]. Similarly, the mortality rate after 6 months was 21.5% for femoral neck fractures and 26.4% for intertrochanteric fractures. In general, the mortality rate in elderly patients after one year varies between 14% to 36% [10].

One major issue that can be identified in this process, concerns the shortage of beds in hospitals due to a decreased flow to, for example, nursing homes and rehabilitation centres [11]. As a result, patients have a longer stay in hospitals, which is very expensive [12]. Therefore, it would be useful to be able to predict – before surgery – the length of stay of hip fracture patients as well as their discharge destination to optimize the patient flow.

Another issue concerns the consideration to not perform surgery at all. One reason to not perform surgery is that – if the patient has a poor mental and/or physical health status and if a surgery is performed – he or she will have a long and difficult recovery process with a low success rate. In addition, the mortality rate among these patients is high and they suffer from much pain. Performing no surgery, on the other hand, usually initiates the palliative care process and then the patients dies too, nonetheless under other circumstances.

In this project, we investigate the relevant predictor variables for the prediction of length of stay, discharge destination and mortality and we aim to test them for usability in prediction with different regression and machine learning models. In case the predictions of length of stay and discharge destination are useful, this information can be used to inform the destination institution to ensure a bed is available in time and stimulate an increased patient flow. In addition, if we can accurately predict mortality, this information could help in decision making for treatment of frail hip fracture patients.

1.2 Relevance to the field of Artificial Intelligence

Artificial Intelligence (AI) in medicine started in the early 70s with an AI expert system used for treating blood infections [13]. First the program, MYCIN, diagnoses patients based on reported symptoms and medical test results [14]. After diagnosing the patient, the program recommends a treatment for the patient.

Nowadays, AI in medicine is often used to annotate clinical data, to obtain insights in patient data and to identify patient similarity [13]. Artificial Intelligence in medicine offers a lot of value since it has the potential to optimize care paths, reduce errors related to human fatigue, decrease mortality rates and to diminish medical costs [13].

AI is widely used in various fields, such as radiology and patients with cancer. However, there are more perspectives for the field of AI. In the field of hip fracture patients, AI is currently used to identify risk groups [15], for the diagnosis of fractures [16] and for the post-surgery prediction of 30-day mortality [17]. Combining the potentials of AI with the issues we formulated in Section 1.1, one can say that the application of AI in the process of hip fracture patients has the possibility to let all parties involved in the process benefit. The patient flow can improve. Patients in recovery from their injury can focus on their healing process. Hospitals have higher availability for other patients and insurances save money.

1.3 Research questions and objectives

The research question for this project is as follows: *What are relevant predictor variables to predict a) post-surgery length of stay, b) discharge destination and c) 30-day mortality for hip fracture patients and is the performance of different predictive models sufficient for implementation in practice?*

In order to answer this question, we defined three sub-questions:

1. What are relevant predictor variables to predict the post-surgery length of stay, discharge destination and mortality for hip fracture patients?
2. Which model (regression, lasso regression or random forest) is best suited to predict the post-surgery length of stay, the discharge destination and mortality[3] for hip fracture patients?
3. Is the performance of the best models sufficient to motivate clinical decisions?

[3] Note that it appeared that prediction of 30-day mortality was not possible in this study, since no reliable or complete data source was available. We explain this in Section 3.2 and Appendix 6.

2. Related literature

The aim of the literature review is to find factors associated with length of stay, discharge destination and mortality. These factors can be used in the predictive models to test their predictive ability.

2.1 Length of stay

2.1.1 Methods

Search Strategy

Table 1. Overview of categories of potential predictors for length of stay.

Category	Number of articles	Reference
Treatment	3	[18–20]
Anticoagulation	3	[21–23]
Surgery type	11	[24–34]
Patient characteristics	8	[35–42]
Frailty	2	[43,44]
Time to operation	4	[45–48]
Methodology	6	[49–54]
Orthopedic/geriatric	2	[55,56]
Other	5	[57–61]

We systematically searched literature on hip fractures and the prediction of length of stay on November 30th, 2018, using the database of PubMed. Synonyms for length of stay and corresponding Mesh terms were combined with Mesh terms for hip fracture and synonyms of prediction. The query used for the literature search can be found in Appendix 1. The initial search resulted in 1041 articles published between January 1962 to November 2018. Two filters were applied to the results. First, the articles were filtered on full-text availability, which led to the exclusion of 204 articles. Second, the articles were filtered on their publication date. Articles with a publication date older than 5 years from the moment of search were excluded to ensure relevance, quality and usability of the findings. After filtering on publication date, we excluded 466 articles. The remaining articles were screened on their title and abstract. We only included literature published in English.

In the end, 44 publications fulfilled the inclusion criteria. In order to structure the findings of the articles, we divided them in categories based on the type of data which was associated with length of stay (Table 1). Articles with parameters which cannot be categorized within one of these categories, were assigned to the class *other*, as well as articles which contain multiple parameters for different categories.

Selection of Studies

Based on title and abstract, we excluded studies performed in a non-hospital setting. Furthermore, studies in which only factors with a non-statistical association with the length of stay were found, were also excluded. Moreover, we excluded pediatric studies, since most hip fractures occur in elderly patients. We excluded articles which compared an improved workflow to the regular workflow, since these findings are difficult to generalize.

Data extraction

The selected full-text articles were reviewed, and the authors, title and year of publication were noted. Moreover, the following characteristics were listed: study population size, mean age, %-female participants, studied parameters and findings regarding the tested parameters. In addition, we noted (if available) the time-to-surgery, post-surgery length of stay (LOS) and total LOS. Lastly, we assigned a category (pre-surgery, post-surgery and both) to the parameters, depending on the moment at which they are available.

2.1.2 Results

Study Characteristics

A total of five literature reviews were found [20,25,29,31,33]. These reviews are excluded from this description of the study characteristics. A description of the study populations of the included articles can be found in Appendix 2. The number of included patients per article varied between 39 and 58,046 (median = 330). All studies included adults, with a mean age of 79.7 years. Two studies only specified ranges for age, instead of mean age [32,38]. In one study, age was not specified [54]. The mean percentage of females included in the studies was 68.2%, with a range of 49% to 84%. Two studies did not include gender in their description [49,52].

Regarding the time (relative to surgery) at which predictors were measured, we found 37, 4 and 3 articles respectively for pre-surgery, post-surgery and both.

Length of Stay Characteristics

In case the definition of LOS was not explicitly specified by the authors, it was assumed that they described the total LOS (pre- and post-surgery). In 21 of 44 articles, the post-surgery LOS was either mentioned by the authors or it could be derived from the data by using the time-to-operation (TTO) and total LOS [18,21,22,29,36,43,45,47–49,51–61]. Among these 21 publications, 15 distinct parameters can be identified which are associated with the post-surgery length of stay. Even though 23 articles did not specify the post-surgery LOS, they are still considered to contain valuable information for prediction. The reason for this is that it cannot be claimed that the association of the factors found in these studies only holds on total LOS. Since it is not investigated whether these factors are also related to post-surgery LOS only. Therefore, we can test their potential predictive power.

We will show the findings related to the post-surgery LOS per parameter category according to Table 1. LOS is used here as post-surgery length of stay, unless stated differently.

Treatment and anticoagulation

It is observed that a surgical approach results in a 2.2 day shorter LOS (mean 6.0 days) compared to a medicinal approach (mean 8.2 days) [18]. In the category of anti-coagulation, using warfarin prior to the hip fracture was associated with a longer length of stay (1.1 and 0.3 days respectively) [21,22]. Also, patients using clopidogrel prior to the injury are observed to have a longer stay compared to patients using clopidogrel in combination with aspirin [61].

Surgery type

A dynamic hip screw (DHS) is associated with a 0.9 day longer stay than gamma nails (GN), GN are associated with a 0.2 day longer stay than a proximal femoral nail with anti-rotating (PFNA) and PFNA is associated with a 2.7 day longer stay than a less invasive stabilization system (LISS) [29]. In another study, they observed that the LOS increases by 1 day, if a nail-, plate-, or screw-fixation is performed [59].

Patient characteristics and frailty

Patients with obese (Body Mass Index (BMI) > 30) are more likely to have a longer LOS (0.5 days) [36]. In addition, the frailer the patient, the longer the observed LOS [43]: robust patients compared with prefrail patients and prefrail patients compared with frail patients, have an observed shorter hospital stay of 0.9 and 1.4-days respectively. Also, gender (male) is related with a longer LOS of 2.5 days [60].

Time to operation

Despite the fact that their definition of *delay* was different, authors of different publications agreed that a delayed surgery is related with a longer hospital stay [45,47,48,61]. For example, an early surgery was specified as a TTO of <48h [45,61] and <5 days [48]. Alonso-Fernández et al., observed that per delayed day, the total LOS (time-to-surgery included) increased with 1.8 days [47]. The same result was found by Chandran et al. [57], while Basques et al. observed an increase of 1 day after a delayed TTO [59].

Methodology and orthopedic/geriatric

In the methodology category, three findings are observed. First, performing pre-surgery tests is associated with a prolonged LOS [49,52,53,60], with a range of 0.7 – 2.1 days. Second, a surgery which is led by a medical specialist results in a shorter LOS [51]. The observed decrease was 2.7 days. Third, patients who are admitted after transfer from another hospital are more likely to have a longer hospital stay compared to patients who are directly admitted to the hospital [54]. The authors observed that directly admitted patients have 4.9-day shorter LOS.

In addition, patients who had access to the services of the orthopedic ward had shorter LOS compared to patients with access to geriatric services [55,56].

Other

Another observation was that if a patient had to cope with post-surgery complications, the LOS increased with an average of 6.3 days [57]. Also, when a patient had to return to theatre, the length of stay increased with 15 days [58]. Moreover, if during surgery a general anesthesia is used, this is associated with a LOS increased by 1 day [59]. Furthermore, Ricci et al. found that admission on Thursday or Friday is related to a statistical longer LOS of 1.2 days compared to admissions in the rest of the week [60].

The findings of the 23 publications related to the total LOS have observed the following parameters associated with an increase of LOS: need of blood transfusion [19,45], general anesthesia mode [20], use of clopidogrel [23], poor mobility status [35], increased age [40,42], presence of complications [40] and higher frailty level [44]. Furthermore, we found that the patient's health status (ASA – American Society of Anesthesiologists) of score 3 or above [19,35] on a range from 1 to 5 also indicates an increase of LOS. Both, an indication for dementia (AMTS - Abbreviated Mental Test Score) [35] and not having dementia [38] were found to be associated with an increase of LOS. Furthermore, having Chronic Obstructive Pulmonary Disease (COPD) [37], Parkinson disease [39] and a high risk at comorbidities and mortality (CCI – Charlson comorbidity index) [42] were also associated with an increase of LOS. Lastly, a longer time to operation [46] and low intensity of physiotherapy [50] were associated with an increase of LOS. Lastly, multiple surgery types [24–28,30–34] were evaluated on their association with LOS.

2.1.3 Discussion

To increase the patient flow of hip fracture patients from the hospital to their discharge destination (e.g. nursing homes or rehabilitation facilities), it is useful to predict their expected post-surgery length of stay and discharge destination before the surgery is performed. In this study, we identified and discussed potential parameters on length of stay which can be used for the prediction of LOS in elderly patients.

Of 44 included articles, 21 contained (indirect) results on post-surgery LOS. There were few differences in the findings. For example, the studies on the anticoagulation drug warfarin [21,22] and pre-surgery tests [49,52,53,60] observed different lengths of stay. Also, the specification of early and delayed surgery differed among different authors [45,47,48,61].

During the selection of studies, articles with non-statistical findings were excluded. However, among these excluded articles, we found results which are in contradiction with the presented findings as well as in the results of the included articles. For example, in an article by Lott et al. [62], they found that use of anti-coagulation is associated with a significant longer LOS. However, after controlling for age, CCI and anesthesia type, their results were no longer significant. Another example of contradicting results was found in a study by Morrissey et al. [63]. By exploring different moments in time (12h, 18h, 24h, 36h), they observed that the time to operation had no statistical influence on the LOS. In addition, several other studies observed no statistical influence of age [19,35] and gender [19,35,37] with respect to length of stay. This suggests that (at least) these parameters are not guaranteed to be useful in making an accurate prediction for length of stay.

As discussed earlier, the day of admission is observed as a parameter which influences the LOS [60]. However, this might be due to policy, rather than patient characteristics. The observed effect was that admission on Thursday or Friday is associated with a longer LOS. Nevertheless, if the policy is that during weekends no surgeries are performed, and patients have a TTO of 2 days, then indeed patients admitted on Thursdays and Fridays have a longer time to surgery, and therefore a longer length of total hospital stay.

For this study, some limitations have been identified. First, the findings in the studied articles are obtained in certain settings, e.g. in patients with a particular type of hip fracture or patients with a specific operation type. Differences in these settings are not considered in this review. Therefore, it is possible that these results do not apply to all types of hip fracture patients in the prediction of LOS. A second limitation concerns the age of the study population. We excluded studies with children. Nevertheless, there were studies [20,27,42,59] that included young adults, but their mean age was still over 65 years. Also, studies that included elderly patients had a different definition of elderly age: over 50, 55, 60, 65, 70 and 80. A last limitation concerns LOS. Earlier, we mentioned that we assumed total LOS was described by the authors if they did not explicitly specify the LOS as either post-surgery LOS or total LOS [18,19,22,25,28,31,36,37,42,44,46,47,52,53,56]. As consequence, it might be the case that in this study, some results are not interpreted correctly. In other cases [18,21,22,43,45,47–49,51–58,60,61] the post-surgery LOS was inferred from the data if both time to surgery and total length of stay were available.

In this thesis we determine whether the identified parameters are indeed of (significant) importance for the prediction of post-surgery length of stay. These include: anesthesia mode, surgery type, ASA-score, pre-surgery mobility, dementia, age, frailty, time to surgery and involvement of geriatrician. Due to absence in our dataset, we excluded to following from our research: surgical/medicinal approach, need for blood transfusion, use of anticoagulation, obese, COPD, Parkinson, pre-surgery tests, transfer from other hospital, gender, day of admission and risk at comorbidities and mortality. Due to the fact that it is a post-surgery variable, we also excluded to following from our research: post-surgery hyponatremia, physiotherapy post-surgery, involvement of specialist, post-surgery complications and return to theatre. In conclusion, there are several factors contributing to LOS. Some, as age and gender, are still under debate as significant and non-significant associations have been described.

2.2 Discharge destination

2.2.1 Results

Literature search

Table 2. Overview of articles on different discharge destinations.

Discharge destination	Number of articles	Reference
Home	7	[64–70]
Skilled nursing facility	1	[64]
Nursing home	1	[71]
Rehabilitation centre	3	[70–72]
Convalescence	1	[71]
Other facilities	5	[64,73–76]

A literature study was performed on January 4th, 2019. The literature search led to the inclusion of 13 articles. The included articles are categorized according to their findings towards different discharge destinations (Table 2): home, skilled nursing facility (SNF), nursing home, rehabilitation centre, convalescence, not home (i.e. facility discharge). Three studies found factors associated with different discharge destinations [64,70,71].

Study characteristics

The number of included patients per article varied between 54 and 107,300 (median = 1,276). All studies included adults, with a mean age of 72.8 years. Two studies did not specify any details on the age of their study population, instead one study only mentioned that all patients were 18 years and older [64], while the other study only specified age as 50 years and older [76]. In addition, another study did present a median of 81 years for age [75], rather than a mean age. The mean percentage of females included in the studies was 65.9%, with a range of 43% to 85%. One study did not include details on the percentage of woman/men in their study [75]. Appendix 3 shows an overview of the study population characteristics of the included articles.

We assigned a timepoint at which the predictors with a relation to discharge destination found in the articles are available for prediction in a similar way as for LOS. This resulted in 5, 1 and 7 articles respectively for the timepoints pre-surgery, post-surgery and both.

Destination of discharge

We systematically present the findings per category according to Table 2.

Home

First, it was found that a decreased age is associated with discharge to home [66–68]. Also, a higher functional independence measure (FIM)-score – on both motoric and cognitive skills – is associated with discharge to home [68,69]. Examples of measures that contribute to the FIM-score are walking ability and level of assistance needed. Patients that are discharged to home are more likely to have a higher walking ability and lower need of assistance than patients who are not [66]. In addition, patients who were not disabled prior to their injury are also more likely to being discharged to home [70]. Furthermore, living in a rural area rather than a metropolitan [70] as well as not living alone [66] are both factors that are associated with discharge to home. Other factors that are associated with discharge to home are spinal anaesthesia mode and revision procedures [64], bowel management [65] and higher AMTS-score, lower incidence of comorbidities (amongst others COPD, diabetes, Parkinson), lower use of medication (such as anti-coagulation) and an intracapsular fracture [66].

Skilled Nursing Facility (SNF)

Factors that are associated with discharge to a SNF are increased age and being from African-American origins [64]. The same study found that a poor health status (ASA-grade greater than or equal to 3) is associated with discharge to SNF. Also, a poor health status predicted a significant reduction in the probability to a discharge to home.

Nursing home

A study by Aitken et al. [71] found that an increased age, Injury Severity Score (ISS) greater than 26/75, injury caused by a fall and having a longer length of stay are all factors that are associated with discharge to a nursing home.

Rehabilitation center

Increased age [70,72], female gender [71] and obesity [72] are three patient characteristics that are associated with discharge to a rehabilitation center. Also, admission to the Intensive Care Unit (ICU) and an increased ISS are associated with discharge to a rehabilitation center [71]. Furthermore, living alone and being unable to walk at discharge [72] as well as living in a metropolitan city, not working prior to the injury, having a more proximal injury and being privately insured [70] are also factors that are associated with discharge to a rehabilitation center.

Convalescence

Two factors that are associated with discharge to a convalescence are admission to the ICU and being transferred from a hospital to another hospital [71].

Other facilities

Lastly, we found multiple factors that are associated with discharge not to home. However, in these articles the authors did not specify the specific destination. First, increased age [64,73–75], female gender [64,74], and functionally dependence pre-injury and deficits in self-care [73–75] are associated with discharge not to home. In addition, higher BMI [64], higher ASA-score [74] and intertrochanteric fractures [75] are associated with discharge not to home. Other factors that are associated with discharge not to home are no follow-up therapy, longer LOS, marital status and impaired bladder/bowel function [73].

2.2.2 Discussion

In order to stimulate the patient flow of hip fracture patients from the hospital to their discharge destination, it is not only useful to predict the length of stay, but it is also useful to predict the patient's discharge destination.

We will shortly discuss some remarkable results. First, Kimmel et al. [70] found that patients living in a rural area are more likely to be discharged to home compared to patients who live in a metropolitan area. We think this is a surprising observation. Due to the fact that there are better facilities and the fact that there is more hygiene in a metropolitan area compared to a rural area, we would expect the exact opposite result. Furthermore, a decreased age was associated with discharge to home and an increased age was associated with facility discharge. However, an increased age was found with respect to discharge to a skilled nursing facility [64], nursing home [71], rehabilitation centre [70,72] and discharge different than home [64,73–75]. Therefore, for prediction, increased age is not expected to provide us with an accurate prediction for a specific discharge destination.

For this study, we identified multiple limitations. First, similar as for the study on length of stay, the findings in the articles found are obtained in certain settings, e.g. in patients with a particular type of hip fracture or patients with a specific surgery type. Differences in these settings are not considered in this review. Therefore, it is possible that the results do not apply to all types of hip fracture patients while using the findings for prediction of discharge destination. A second limitation of this study is that, in contrary with the study on LOS, the included articles were not filtered on their publication date. As a result, we included 6 out of 13 articles with a publication date older than 5 years from the moment of search [69–73,75]. Therefore, it is possible that the results of these articles do no longer apply.

In the research presented in this thesis, we determine whether the identified parameters are indeed of (significant) importance for the prediction of discharge destination. These include: health status (ASA-score), age, anesthesia mode, living situation prior to injury, level of assistance needed, mobility pre-surgery and type of fracture. Due to non-availability in our dataset, we excluded the following from our research: race, indication for dementia (AMTS score), use of medication, BMI, marital status, LOS, gender, severity of injury, cause of injury, type of insurance, admission to ICU and transfer from other hospital. Due to the fact that it is a post-surgery variable, we also excluded to following from our research: revision procedures, bladder/bowel function, comorbidities, mobility post-surgery and follow-up therapy.

2.3 Mortality

2.3.1 Results

Literature search

A literature study was performed on May 16th, 2019 on hip fractures and the prediction of mortality using the database of PubMed. The initial search resulted in 35 articles. After filtering on publication date (<5 years from moment of search), age of study population (adults) and language (English), we were able to exclude 17 articles. The query used for literature search can be found in Appendix 1. The remaining articles were screened on their title and abstract. We excluded studies which were not related to hip fractures and articles which described biomarkers as predictors. Also, we excluded one article since it was not available in full-text. In the end, 7 articles fulfilled the inclusion criteria.

Study characteristics

The number of included patients per article varied between 199 and 47,698 (median = 2,815). All studies included adults, with a median age of 76.4 years. One study only mentioned that their study population had an age of at least 65 years [77]. Appendix 4 shows an overview of the characteristics of the study populations of the included articles.

Mortality

First, it was found that the Elixhauser- and Charlson comorbidity indices are associated with mortality in hip fracture patients [78]. They differ in the fact that they both take different comorbidities in consideration and they have a different scoring system. However, the same study showed that their predictive power for 30-day, 1-year and 2-year mortality was low. In the contrary, a study by Toson et al. (2019) found that the Charlson comorbidity index had acceptable contribution to 30-day and 1-year mortality of hip fracture patients [77].

Furthermore, we found that an increased age, increased bone mineral density (BMD) and smoking are associated with a higher long term mortality rate (>10 years after surgery) [79]. In another study, it was found that more than 3 comorbidities, poor health status (high ASA-score), living in residential care facilities, male gender, cardiovascular- and pulmonary diseases and dependence in activities of daily living (ADL) are associated with a higher 30-day, 1-year and 3-year mortality rate [80].

Lastly, we found three studies which tested the performance of a score on the prediction of 1-year mortality [81–83]. In these scores, different factors were combined. The first study tested the Sernbo-score, which takes into account the age, living situation, walking ability and mental status of the patient [81]. The scale ranges from 8 to 20, where a lower score means a higher risk at mortality. The second study tested a frailty-score [82]. Frailty was defined as the presence of three or more of the following criteria: weight loss, poor energy, weakness, slowness and low physical activity. The last study tested a scoring system, which takes into account the age, gender and neutrophil-to-lymphocyte ratio (NLR) [83]. All these scores were found to be predictive for the 1-year mortality of hip fracture patients.

2.3.2 Discussion

In order to develop a decision support system with respect to the consideration to not perform surgery on frail hip fracture patients, it is useful to be able to predict the 30-day mortality.

For this study, we identified multiple limitations. First, similar as for the previous studies, the findings in the articles found are obtained in certain medical settings. Differences in these settings are not considered in this review. Therefore, it is possible that the results do not apply to all types of hip fracture patients while using the findings for prediction of mortality. Second, in our search, we did not have a specific definition for mortality. We are mainly interested in 30-day mortality, however, we also found variables associated with 1-year, 2-year and 3-year mortality. Consequently, it is possible that these results are not applicable for 30-day mortality. Also, regarding the three scores we found, we are not able to calculate them. The reason for this is that we do not have all required components in our dataset.

In the research presented in this thesis, we aim to determine whether the identified parameters are indeed of (significant) importance for the prediction of 30-day mortality. These include: age, no osteoporosis, living situation, mobility pre-surgery, weight loss, dependence in ADL, dementia and health status (ASA-score). Due to absence in our dataset, we excluded the following from our research: Elixhauser index, Charlson index, BMI, smoking, poor energy, weakness, slowness, cardiovascular- and pulmonary diseases, comorbidities and gender.

3. Materials and methods

3.1 Moment of prediction

The aim of developing prediction models for hip fracture patients (regarding length of stay and discharge destination), is to stimulate the patient flow from the hospitals by having a better overview of the need for beds. Hence, it is useful to have this overview of need for beds early in the process, which enables transfer nurses to immediately start arranging a place for the patient. There is not only a financial benefit from an improved flow, also, the patient receives the specific care he or she needs to rehabilitate successfully earlier, which prevents the development of unnecessary complications. Consequently, we decided that the right moment for prediction of LOS and discharge destination is before surgery, which is why we remove all post-surgery data.

The same holds for the prediction of mortality, since this prediction might be useful as decision support in the consideration whether or not to perform surgery. Therefore, the right moment for prediction for mortality is also before surgery.

The prediction of LOS and discharge destination is potentially particularly helpful in situations in which it is unsure what the LOS will be for patients who do not go home. To be more specific: patients who are in a physically weak and mentally strong condition. Usually, for this population, no arrangements are made until the patient is ready for discharge, which is too late since at that moment no bed will be available. As a consequence, they have a longer hospital stay, which can lead to additional complications and a lower success rate of rehabilitation.

3.2 Data selection

As mentioned in the introduction (Section 1), we use data from health records of six SAZ-hospitals in this project. The hospitals included are: BovenIJ Ziekenhuis *(Amsterdam)*, Bravis Ziekenhuis *(Bergen op Zoom)*, Het Van Weel-Bethesda Ziekenhuis *(Dirksland)*, St. Anna Ziekenhuis *(Geldrop)*, Saxenburgh groep *(Hardenburg)* and ZorgSaam Zeeuw-Vlaanderen *(Terneuzen)*. The original data sources can be found in the health records of the hospitals mentioned above. Data is extracted from these health records, after which it is de-identified and standardized. Note that the data is treated conform the General Data Protection Regulation (GDPR) [84–86]. Figure 4 shows an overview of the data selection process.

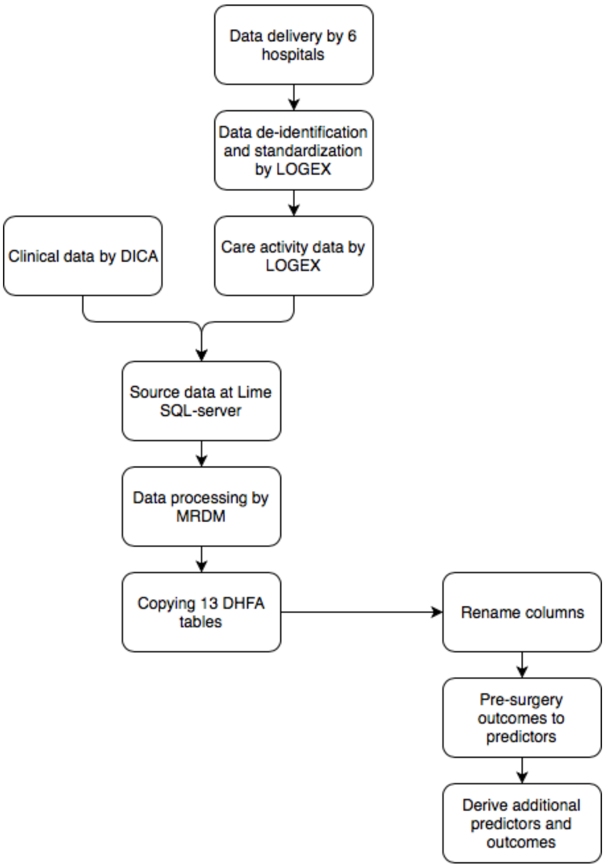

Figure 4. Overview of data selection process. *DICA, Dutch Institute for Clinical Auditing; DHFA, Dutch Hip Fracture Audit.*

The source data for this project is located at a SQL-server maintained by Lime networks, an information technology (IT) company in the Netherlands. The data in this database exists of two parts: clinical patient data and care activity data. The clinical data is collected by the Dutch Institute for Clinical Auditing (DICA) for the Dutch Hip Fracture Audit (DHFA), a quality registry which aims at improving quality of care for hip fracture patients in the Netherlands. The data is processed by Medical Research Data Management (MRDM), a Dutch company which processes medical data on behalf of healthcare organizations. The clinical patient data contains information regarding the patient history, patient characteristics and patient outcomes. The care activity data is provided by LOGEX, a health analytics company in the

Netherlands. This data contains counts per patient regarding how many times the patient had a certain treatment or procedure. The data is delivered by the hospitals every quarter of a year. All care related activities need to be registered in the health records for declaration purposes, therefore, the health records are a complete data source.

For this study, we copy all data of the DHFA registry from April 2016 till June 2018 from the database of Value2Health, which was all available data. In order to not interfere with the original DHFA datasets, we create a new registry for this project: DHFA_[studentname]. The copied datasets are stored here. The data is grouped in 13 different tables over time, since this is how the data is delivered by LOGEX. The data in each time-table contains the specific data which is available in the specific time-frame. There are 5 tables for pre-surgery data, 7 tables for post-surgery data and there is 1 table which contains all data. We only use the tables with pre-surgery data, since our moment of prediction is before surgery. The data we use contains both clinical- and care activity data. The clinical data are all outcomes. The care activity data are all predictors. After copying the data, we rename all columns. The old format of column names was DHFA_[Pred/Outc]_[variablename]. The new format of column names is DHFA_[studentname]_[Pred/Outc]_[variablename]. This is necessary to match the columns to the correct registry.

Furthermore, we change pre-surgery clinical data from outcome [Outc] to predictor [Pred_kenmerk]. The reason for this is that we want to use all pre-surgery data as predictors instead of outcomes. In addition, the clinical data corresponds to the data found in literature described in Section 2. Appendix 5 shows a list of all clinical data which is used as predictor and the meaning of the particular predictor. The *kenmerk*-extension allows us to easily separate the clinical predictors from the care activity predictors.

Next, we derive the values of two other predictors from the data, namely: age and time-to-surgery. For age this is possible, since we know the patient's year of birth and the date of surgery. For time-to-surgery this is possible, since we know the date of admission and the date of the surgery.

Due to the fact that only the outcome of discharge destination is already available in the dataset, we have to add the outcome for post-surgery length of stay. This is possible, since we know the date of surgery and the date of discharge. We also investigate which data we could use to calculate the 30-day mortality as outcome. We investigate three different approaches. However, it appeared that none of them results in a reliable or complete data source. Therefore, we are not able to make predictions for 30-day mortality neither are we able to test which parameters have a predictive power for mortality. Appendix 6 describes the findings and conclusions of this investigation.

3.3 Interviews

In Section 2 we described parameters which are in literature associated with length of stay and discharge destination. In order to validate and extend these findings, we conduct semi-structured interviews with care providers within the participating SAZ-hospitals and other professionals involved in the process of hip fracture patients, such as orthopaedic surgeons and physiotherapists. Appendix 7 describes the guiding questions for the interviews. Note that not all questions were asked during every interview. One reason is that with respect to time and focus of the interviews, we decided to skip that question.

During the interviews, we are not only interested in validating and identifying parameters that could be used for prediction, but also in the general process in different hospitals (i.e. the current state) and what predictions are considered useful. This information is systematically processed and used to define the outcomes (Section 3.5.1). Lastly, we would like to learn information about the type of patients for whom prediction is considered most valuable.

3.4 Data pre-processing

First, all available data is loaded from the server. Note that this data contains both predictors as well as all possible outcomes available in the database. Then, we remove all post-surgery data, since we are interested in prediction before surgery as explained in Section 3.1. Next, we perform multiple checks on the data: we transform NULL values to NA, exclude predictors and outcomes with more than 99% NA, remove duplicate patients and remove patients with more than 95% NA predictors.

After randomizing the order of observations, we split the data in predictors and outcomes. For the outcomes, we set unknown categories to NA and we transform categorical outcomes into factors. Whereas for the predictors we remove highly correlated predictors, transform numeric predictors into factors, normalize them, set unknown values from self-derived predictors to NA and impute missing values (using the median for continuous variables and by sampling for categorical variables). Furthermore, we select the requested outcome(s) and predictors. In case no specific outcomes or predictors are requested, we keep them all.

Lastly, before a model is trained and tested, it is possible to let it perform the requested feature selection method, e.g. correlation-based or lasso selection. In case no selection method is requested, the model uses all predictors available for training and testing. Otherwise, only the selected features will be used for training and testing. In our research we compare the performance of our models without feature selection with the performance of models with correlation-based feature selection and models with lasso selection.

3.5 Model evaluation

In order to evaluate the regression and machine learning models in this project, we split the data in a training set and a test set. The proportion is 70% and 30% respectively. This means that we use 70% of the data to train the model and 30% of the data (which the model has never seen before) for testing purposes. Also, we make sure to run each model repeatedly on independent training and test sets through randomizing the order of observations of the data. As a consequence, each run gives (slightly) different results. In the end, we calculate the mean and standard deviation of the results. Therefore, we are able to make comparisons among the different models.

3.5.1 Outcomes

For the length of stay, the initial outcome is calculated in days based on the operation date and the date of discharge. However, we not only include this continuous outcome, but also the following classifications for prediction (the boundaries are based on the interviews):

- *Short stay* (<= 2 days) versus *long stay* (>2 days).
- *Short stay* (<=2 days), *normal stay* (3-5 days) versus *long stay* (>5 days).

The reason for including both, categorical and continuous outcomes for LOS, is that it allows us to verify whether the desired prediction of the professionals, i.e. LOS in categories, is indeed more useful than continuous LOS.

For discharge destination, the original outcome contains 6 different destinations: *home, home with care, care home, nursing home, nursing home with revalidation* and *other*. Based on the interviews, we added two extra outcomes with the following classifications:

- *Home* versus *not home.*
- *Home, home with care* versus *not home*.

In conclusion, we have three different outcomes for length of stay and three different outcomes for discharge destination, so six in total.

3.5.2 Evaluation measures

For this project, we are interested in finding relevant predictor variables for the prediction of length of stay and discharge destination and we would like to know if the performance of different models is sufficient to support clinical actions and decisions.

Since we have both categorical and continuous outcomes (Section 3.5.1), we need different outcome measures. For the categorical outcomes, we use the area under the curve (AUC), accuracy, sensitivity/recall, specificity, positive predicted value (PPV) and F1-score. We consider an AUC smaller than 0.7 as poor, AUC between 0.7 - 0.8 as acceptable, AUC between 0.8 - 0.9 as excellent and AUC greater than 0.9 as outstanding [87].

For the continuous outcome (LOS in days), we use mean error (ME), mean absolute error (MAE), median error (with Q1 = lower quartile range of 0.25 and Q3 = upper quartile range of 0.75) and root-mean-square error (RMSE).

3.5.3 Variable importance

In order to compare the importance of variables, we have invented a method by ourselves. In this way we are able to combine multiple measures which give an indication of the importance of the variable.

First, all predictive models without feature selection, return an ordered list with the top 100 most important variables for the prediction according to that model. Note that the order of this list is determined by the models themselves. All predictive models with lasso- or correlation-based feature selection, only return the selected variables for prediction.

Second, from the lists that are returned by the models themselves, we extract for all models the top 10 variables which contributed the most to the prediction. Note that there are situations in which a feature selection method is applied, where less than 10 variables are selected for prediction. In that case we only extract those variables for further interpretation.

Next, we assign a ranking to all extracted variables corresponding with the order in the list. However, since we run all models 10 times, we actually calculate the ***median ranking*** **(a)** for each of the extracted variables. The reason for this, is that it could be the case that different variables have been selected and different orderings of the variables occurred in each run. The ranking is a value between 1 and 10, where ranking = 1 means that the variable was found to be most important.

Also, we count how many times a certain variable was selected by the 10 runs of the same model, i.e. ***frequency*** **(b)** of the variable. Frequency is a value between 1 and 10, where frequency = 10 means that the variable is selected by all runs of the model. Consequently, it is assumed that a variable with a median ranking of 1 and frequency of 10 is more likely to have predictive power than a variable with a higher median ranking and lower frequency.

After transforming (a), we multiply it with (b) into a new measure called ***weighted importance*** **(c)**. The formula to calculate (c) is shown in Formula 1. The weighted importance is a value between 1 and 100. The higher the value, the more important the variable is for the prediction. By setting a threshold of importance, we are able to compare the importance of variables and to determine the true most important variables. The threshold we use is 50.

$$c = (11 - a) * b \tag{1}$$

The following is an illustration of the use of weighted importance. Assume variable X has a median ranking of 2 and a frequency of 6 and variable Y has a median ranking of 5 and a frequency of 8, then it is difficult to say which variable is more important for the prediction. One could argue that it should be variable X, since its ranking is higher. A counterargument is that the occurrence of variable X is lower and therefore, it should be variable Y. By calculating the weighted importance, we can easily solve this problem and compare both variables using both criteria. The weighted importance for variable X and Y is (11 – 2) * 6 = 54 and (11 – 5) * 8 = 48 respectively. We can conclude by saying that variable X is more important for our prediction. In addition, we exclude variable Y from future prediction for this particular outcome, since its weighted importance is below our threshold (48 < 50).

3.6 Predictive models

In Section 2 we identified multiple variables which are associated with LOS and discharge destination of hip fracture patients. In this research we are interested in comparing the performance of models using a) the identified variables in the literature study (i.e. clinical data), b) the care activity data and c) a combination of 'a' and 'b' for prediction. There are multiple types of machine learning models which can be used for prediction, such as random forests, neural networks and regression models. Below, we describe the methods used in this study. In order to find the most suitable model for each outcome (Section 3.5.1), we compare three methods (regression, lasso regression and random forests) with different settings for feature selection and input data (Table 3). This results in 54, 12 and 36 different models for regression, lasso regression and random forests (RF) respectively. For developing the models, we use RStudio version 1.2.1143 and R version 3.5.1.

Table 3. Overview of included feature selection methods and input data per methodology.

	Regression	LASSO	Random Forest
Clinical data			
No feature selection	X		X
LASSO feature selection	X	X	X
Correlation-based feature selection	X		X
Care activity data			
No feature selection	X		
LASSO feature selection	X		
Correlation-based feature selection	X		
Clinical- and care activity data			
No feature selection	X		X
LASSO feature selection	X	X	X
Correlation-based feature selection	X		X

3.6.1 Regression

For this research we start with performing a regression for the prediction of discharge destination and LOS, where the outcomes are categorical (Section 3.5.1). Therefore, we need a logistic regression for the prediction of the five categorical outcomes of discharge destination and length of stay. Next, we perform a linear regression to predict a continuous outcome of length of stay. This results in six different models. We do this for all the different data as input for the model: the clinical data, the care activity data and a combination of clinical- and care activity data. So, now we have 18 different models. As we will explain later, sometimes a feature selection method is desired. We compare the performances of our 18 models without feature selection with the performances of these models with two different feature selection methods: lasso and correlation-based. This results in a total of 54 different regression models. See Table 3 for an overview of all models. We use the R-packages `glmnet` [88] and `nnet` [89] to develop the regression models. We will elaborate on the feature selection methods in Section 3.6.4.

For the binary outcomes (length of stay: *long* versus *short*; discharge destination: *home* versus *not home*), the logistic regression model uses a Sigmoid function to map the predictions into probabilities. To be more specific: it maps any real value into a value between 0 and 1. By setting a threshold (e.g. 0.5), values above this threshold will be classified as 1 and below the threshold as 0 [90]. In our case of discharge destination, 0 represents *home* and 1 represents *not home*. Whereas for length of stay, 0 represents *short* stay and 1 represents *long stay*.

For the multinomial outcomes (length of stay: *long* versus *normal* versus *short*; discharge destination: *home* versus *home with care* versus *not home*; discharge destination: all possibilities), the logistic regression model estimates for each class the probability of the observation being in that class. Where the sum of all probabilities is equal to 1. In the end, the model assigns the class with the highest probability to the observation.

For the continuous outcome (length of stay in days), we are able to use a linear regression model. This model models the relationship between one or more variables and fits a line into the data. This makes it possible to generalize and therefore predict something that is new to the model. Since the fitted line represents future values. It is desired to minimize the error of the prediction: the distance of the true value should be as close to the predicted value as possible [91].

One major advantage of logistic regression is that is a simplistic model, since it does not require parameter tuning and it is relatively easy to implement. Also, the model is very efficient and highly interpretable, which is why it can be used as baseline model to compare the performance of more complex models. Another advantage is that the logistic regression model does not require too many computational resources and it is easy to regularize [92]. Likewise to logistic regression, the linear regression model is a very simple model and easy to interpret [93].

A disadvantage of logistic regression is that it cannot be used for the prediction of continuous outcomes. However, for this type of problems linear regression can be used. In addition, in case predictors are independent from the outcome variable or when predictors are highly correlated, the model does not perform well. Therefore, we might want to apply a feature selection method [92]. A disadvantage of linear regression is that it assumes that there is a linear relationship between the predictor and the outcome variables. It also assumes that the data is independent. Besides, the model is very sensitive to outliers in the data, which leads to a worse fit [93,94].

3.6.2 Least Absolute Shrinkage and Selection Operator (LASSO) regression

By comparing models with many input parameters, it is possible that such a model does not perform optimal. Therefore, it might be a good idea to find a subset of useful variables, i.e. apply feature selection on the variables. For *logistic* regression, it is possible to assign a penalty term to the model for having too many variables. There are three commonly used ways for penalized regression: ridge-, lasso- and elastic net regression [95]. In lasso regression, only the most important variables are included in the model, whereas for ridge- and elastic net regression (almost) all variables are still in the model. Therefore, we choose to only use the lasso regression.

As mentioned before, we are interested in comparing models with only clinical data, only care activity data and both data as input. However, based on the performance of the models from regression, we exclude only care activity data as input from the comparison. Furthermore, for the lasso regression, we exclude no feature selection and correlation-based feature selection from the runs. The reason for this is that lasso regression has its own built-in feature selection method. We do this for all outcomes mentioned in Section 3.5.1. This results in 12 different models for lasso regression. See Table 3 for an overview of all models.

The lasso regression performs two tasks: regularization and feature selection. The lasso regression model calculates for each variable its contribution to the prediction. In case a variable does not contribute enough, the variable's coefficient is set to zero and therefore excluded from the model [95]. This is called regularization. In order to control the strength of the penalty, the value of λ has to be tuned. The greater the λ-value, the more coefficients are set to zero. Consequently, the more variables are excluded from the model. During the feature selection process, the variables with a non-zero coefficient are selected for the model [96].

Using the R-package `glmnet` [88], there are two possible λ-values that can be extracted from a cross-validation: λ_{min} and λ_{1se}. The minimal lambda value corresponds to the model where the mean cross-validated error is minimalized. Likewise, the lambda value within one standard error corresponds to an error which lies within one standard error of the minimum [96]. We performed a 10-fold cross-validation to select the value of λ within one standard error and the corresponding predictor variables. Next, the model had to be re-estimated with the selected predictor variables.

One advantage of lasso regression is that it is a good method to minimize the prediction error. Another advantage of lasso regression is that it results in reduced overfitting, since the irrelevant variables for prediction are excluded from the model. Besides, this makes it easier to interpret the model. Also, lasso regression can result in a high prediction accuracy: shrinking coefficients and removing variables leads to a reduced variance without significantly increasing the bias [96]. A disadvantage is that it has a low prediction power for correlated variables.

3.6.3 Random Forest

A third method we use in our project is random forests. Similarly to the lasso regression, we exclude models with only care-activity data as input for the models. We do this for all outcomes described in Section 3.5.1. This results in 12 different models. Also, we compare the performance of these models without feature selection with the performance of these models with lasso- and correlation-based feature selection. Therefore, we have 36 models in total for RF. See Table 3 for an overview of all models. We use the R-package `ranger` [97] to develop the RF models. Table 4 shows the RF-specific settings for all RF-models. We use the default settings used by Value2Health as values for the RF-specific parameters.

Table 4. Settings for random forest models.

Number of trees	250
Number of variables available for splitting at each node	√ (number of predictors)
Minimum node size after split	1
Voting threshold	0.5
Importance	Impurity

In contrary to regression models, the prediction of a single RF is based on multiple predictions (i.e. the number of trees that have been grown): it builds multiple decision trees and combines the predictions of all trees. As a consequence, a major advantage is that the prediction is more stable and more accurate. At each node, a RF takes the best feature among a random subset of features, rather than looking for the most important variable while growing the tree. This results in a more diverse tree and therefore a better model. [98] Another advantage is that RF work well with both categorical and continuous variables. Also, RF automatically deals with missing values and in case of noise, there is little impact on the RF. [99] A disadvantage on the other hand, is that the RF model is complex and therefore not intuitively to interpret. In addition, its running time is longer than less complex techniques. [99]

To prevent the model from overfitting, it is possible apply feature selection as well. In a random forest it is easy to measure the relative importance of each feature on the prediction. The user can later choose which features to select and which features to drop based on their importance. It is possible to set some (hyper)parameters for the model to increase the models' predictive power or the speed. [98]

3.6.4 Feature selection

Lasso

As described in Section 3.6.2, the lasso regression model applies regularization in order to do feature selection. During the regularization process, the model calculates for each variable its contribution to the prediction. The coefficients of the least contributing variables are set to zero. During feature selection, the model only selects variables with a non-zero coefficient. The strength of the penalty is dependent on the chosen λ-value (λ_{min} or λ_{1se}). The greater the λ-value, the more coefficients are set to zero and therefore more variables are excluded. The value of λ_{min} and λ_{1se} can be determined by using cross-validation in the R-package `glmnet` [88].

Correlation-based

In case the correlation-based feature selection is selected, we calculate for all variables their correlation to the outcome variable. These correlations can be any real values between 0 and 1. The stronger the correlation, the higher the number. Next, we compare these correlations with a certain threshold. The variables selected for the model are those with a greater correlation than the threshold.

By trial-and-error, we ended up with a threshold of 0.2. A higher threshold would mean that more predictors would be excluded. In our case that would result in the inclusion of zero or only one predictor.

3.6.5 Online tool

We develop an online tool for the prediction of LOS (short/long stay) and discharge destination (home/not home), using Evidencio, an online platform to develop prediction models for medical decision making. We use the predictors that were found to be most helpful in prediction (of the best performing model) as input variables for a regression model.

3.7 Statistics

We perform a statistical analysis in order to determine whether it is likely that the best performing model for an outcome, based on AUC for categorical outcomes and RMSE for continuous outcomes, is indeed performing significantly better than the other models for that particular outcome.

We start by checking whether the AUC values and RMSE values are normally distributed by performing a Shapiro-Wilk test. We use $\alpha < 0.05$ as significance level. We do this to determine which test we have to use for each comparison.

For the normally distributed data, we perform a one-way Analysis of Variance (ANOVA). The independent variable are the different model types and the dependent variable are the AUC values. The one-way ANOVA test makes four assumptions [100]. First, the variances across different models should be homogeneous. We test this by making a boxplot. Second, the data should be normally distributed, which we already tested with the Shapiro-Wilk test. Third, each sample has to be drawn independently and lastly, the dependent variable should be continuous. For the non-normally distributed data, we perform a Kruskal-Wallis test. The difference between these two tests is that the one-way ANOVA test compares the mean AUC/RMSE values and the Kruskal-Wallis test compares the median values [101]. We use $\alpha < 0.05$ as significance level.

These tests tell us whether there are significant differences among the models, but not which models differ [102]. For the latter we need to perform another statistical test: a post-hoc test with Bonferroni correction on the alpha levels. For the normally distributed values, we use a pairwise t-test and for the non-normally distributed values, we use a pairwise Wilcoxon Rank Sum test. We use $\alpha < (0.05/n)$ as significance level, where *n* is the number of models that is compared in the test.

For performing the statistical tests, we use RStudio version 1.2.1143 and R version 3.5.1. The ANOVA test is called with the R-function `aov()`.

4. Results

4.1 Characteristics of the study population

A total of 643 patients were included in the research on discharge destination. In the study to length of stay, we included 634 patients. The characteristics of all patients are presented in Table 5. Median age was 83 years for both groups. There were no differences between the two groups.

Table 5. Overview of patient characteristics of the study population per outcome (discharge destination and length of stay).

	Discharge destination N=643	**Length of stay** N=634
Age (years), median (Q_1 - Q_3)	83 (74 – 88)	83 (74 – 88)
ASA-score, n (%)		
- I	29 (4.5)	28 (4.4)
- II	234 (36.4)	232 (36.6)
- III	312 (48.5)	313 (49.4)
- IV	42 (6.5)	45 (7.1)
- Unknown	26 (4.0)	16 (2.5)
TTO (days), median (Q_1 - Q_3)	1 (0 – 1)	1 (0 – 1)
Fracture type, n (%)		
- Medial femoral neck fracture	225 (35.0)	222 (35.0)
- Trochanter femoral fracture	279 (43.4)	277 (43.7)
- Sub-trochanter femoral fracture	9 (1.4)	10 (1.6)
- Unspecified	4 (0.6)	4 (0.6)
- Unknown	126 (19.6)	121 (19.1)
Procedure, n (%)		
- Conservative	10 (1.6)	-
- Hemiarthroplasty	234 (36.4)	234 (36.9)
- Cannulated hip screw	31 (4.8)	31 (3.9)
- Total arthroplasty	49 (7.6)	49 (7.7)
- Sliding hip screw	51 (7.9)	51 (8.0)
- IM pen	264 (41.1)	265 (41.8)
- Unknown	4 (0.6)	4 (0.6)
KATZ, n (%)		
- Bath/shower	225 (35.0)	220 (34.7)
- Dressing	188 (29.2)	182 (28.7)
- Toilet visit	107 (16.6)	105 (16.6)
- Use of incontinence material	167 (26.0)	162 (25.6)
- Transfer bed-chair	88 (13.7)	85 (13.4)
- Eating	49 (7.6)	47 (7.4)
- Unknown	63 (9.8)	63 (10.0)
Dementia (yes), n (%)	75 (11.7)	75 (11.8)
- Unknown	199 (30.9)	194 (30.6)

Q1, lower quartile 0.25; Q3, upper quartile 0.75; ASA: American Society for Anesthesiologists; TTO, time-to-operation; KATZ, Index of Independence in Activities of Daily Living.

Both the patient's destination of discharge and their median length of post-surgery stay are presented in Table 6. Figure 5 shows the distribution of length of stay in the study population.

Table 6. Characteristics of the outcome variables.

	Patients N = 643
Discharge destination, n (%)	
- Independent	153 (23.8)
- Independent with ADL-assistance	52 (8.1)
- Care home	26 (4.0)
- Nursing home	90 (14.0)
- Nursing home with revalidation	304 (47.3)
- Other	18 (2.8)
Length of stay (days), median (Q_1 - Q_3)	5 (3 – 7)

ADL, Activities of Daily Living; Q1, lower quartile 0.25; Q3, upper quartile 0.75.

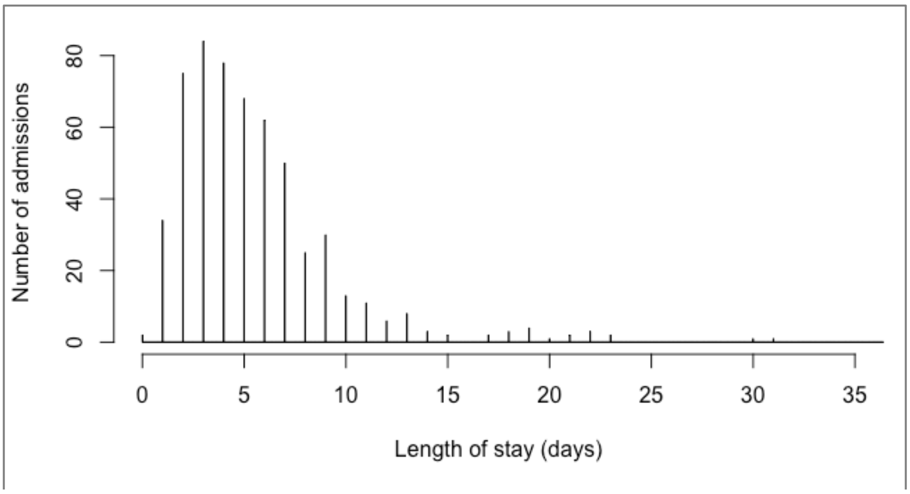

Figure 5. Distribution of post-surgery length of stay in days.

4.2 Interviews

In total, four interviews were conducted. The participants worked at different hospitals and had different functions in the process of hip fracture patients. Their professions were: surgeon, physiotherapist, orthopedic surgeon and physician assistant in training orthopedics/surgery.

Table 7 summarizes the outcomes of the interviews. The findings describe all the answers that have been given. We did not standardize the answers, due to the fact that limited number of participants were included. Some answers were given multiple times, others were only mentioned by one person. The topics described are: reasons to delay surgery in practice, desired time-to-surgery in practice, factors associated with length of stay and discharge destination, threshold of short/long stay according to participants and what the participants think that a model for post-surgery length of stay and discharge destination should predict.

Table 7. Overview of the findings of the interviews.

	Findings
Reasons to postpone surgery	Consult with specialist is required (cardiologist, pulmonologist, anesthetist); use of anticoagulation; logistic reasons; patient agreement
Time to surgery	<48h, <24h
Factors associated with length of stay	BMI; fat distribution in hip area; wound leakage; use of medication/anticoagulation; lung issues (e.g. COPD); **ASA-score**; **mobility pre-surgery**; **living situation** (alone, partner, caregiver); attitude patient; allowed to use affected side; logistics; complications post-surgery; **degree of independence** (pre- and post-surgery); **time-to-operation**; **dementia**; stabilization of patient; agreements with/availability in nursing homes/revalidation centers; involvement of general practitioner; physical and cognitive condition
Factors associated with discharge destination	Attitude patient; **living situation** (alone, partner, caregiver); **mobility**; **ASA-score**; complications post-surgery; availability of home-care; **degree of independence** (pre- and post-surgery); presence of other diseases (liver-, heart-, lung problems); live in rural/urban area; information provision pre-surgery; physiotherapy; **nutritional status**; **cognition**
Short/long stay threshold	2 days is very short; 3 days is short; >4 is long
Model for prediction length of stay should predict…	Estimate in ~2 days; 0-3 days/3-5 days/>5 days; 0-3 days/3-5 days/5-8 days/>8 days; 1/2/3/4/5/>5 days
Model for prediction discharge destination should predict…	Home/not-home; home/home with care/not-home; home/ELV/GRZ

BMI: body mass index; COPD: chronic obstructive pulmonary disease; ASA: American Society of Anesthesiologists; ELV: eerstelijnsverblijf; GRZ: geriatrische revalidatiezorg. Bold factors are the factors that were available in our data set.

An example reason to postpone surgery is the following situation: in case a patient suffers from other diseases, the patient usually visits the doctor involved in that care path before surgery is allowed. This is after admission to the hospital. Another situation would be if no surgery room is available.

We found that the hospitals, in which the participants work, aim to perform the surgery within 48 hours after admission, sometimes this is even already possible within 24 hours after admission. We can infer from this, in combination with the knowledge we have from the literature search, that a shorter time-to-surgery is most likely to be preferred over a delayed surgery.

The participants named different factors that are associated with length of stay and discharge destination. Note that, despite these are all factors listed by the participants, this list is not exhaustive. Factors related to length of stay and those that are available in our dataset are: ASA-score, pre-surgery mobility, living situation, degree of independence, time-to-operation and dementia. Factors related to discharge destination and those available in our dataset are: living situation, mobility, ASA-score, degree of independence, nutritional status and cognition (dementia).

According to the participants a post-surgery length of stay of 3 days is short and more than 4 days is long. All participants suggested different things for the model of length of stay to predict. For instance, categorization of post-surgery length of stay in classes 0-3 days/3-5 days/>5 days and categorization of post-surgery length of stay in classes 1/2/3/4/5/>5 days. For the prediction of discharge destination, the participants found that a distinction between *home* versus *not home* would suffice, possibly with additional option: home with care.

4.3 Performance predictive models

The aim of this section is to determine for each outcome the model that is best suited to predict this particular outcome. Since the models with only clinical data as input performed the best, we only show the performances of these models.

4.3.1 Length of stay

Figure 6 shows the AUC values of the regression- and RF models without feature selection, with lasso selection and correlation-based feature selection and of the lasso regression model for the prediction of length of stay (*short stay* versus *long stay*). The model, which performs best, based on AUC, is the RF model without feature selection. The mean AUC and mean accuracy of the best performing model are 0.78±0.03 and 0.82±0.02 respectively.

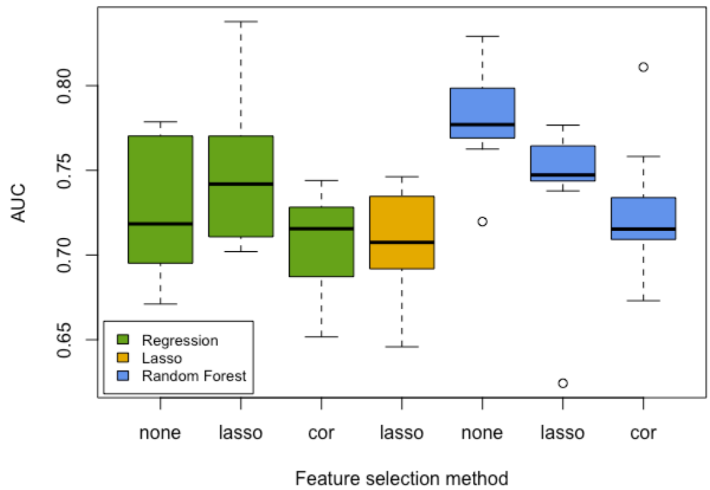

Figure 6. AUC's of different models with lasso-, correlation- and no feature selection for predicting short vs. long stay. *AUC, area under the curve; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Figure 7 shows the sensitivity, specificity, PPV, F1-score and % of patients in each class for the best performing model based on AUC. From the figure, we can read that the sensitivity of the best model is 0.88±0.03 for *long stay* and that the model is 89% sure about the classification of *long stay*. The F1-score for long stay is 0.88±0.02. However, the model is less correct (<60% of the cases) as well as less certain (<55%) about the classification of *short stay*. Furthermore, 80% of all patients was in the *long stay* class. The model performs best for this class.

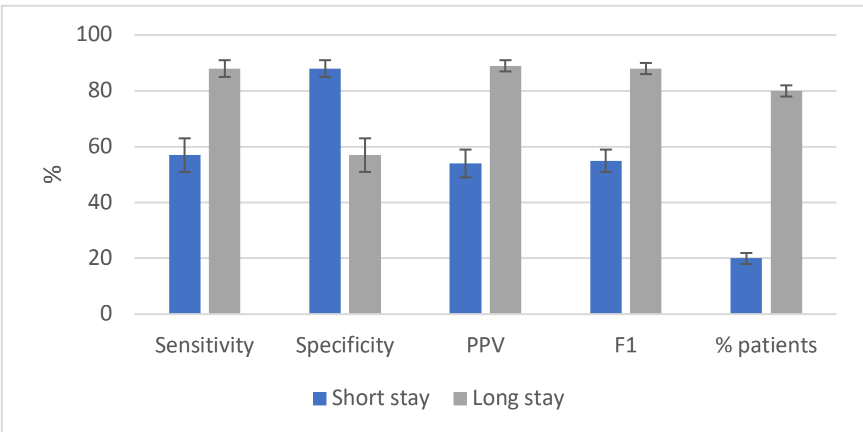

Figure 7. Overview of sensitivity, specificity, PPV, F1-score and % of patients per category for RF without feature selection for predicting of short vs. long stay. *PPV, positive predicted value; F1-score, weighted average of precision and recall.*

Figure 8 shows the AUC values of the regression- and RF models without feature selection, with lasso selection and correlation-based feature selection and of the lasso regression model for the prediction of length of stay (*short stay* versus *normal stay* versus *long stay*). The model, which performs best, based on AUC, is the regression model without feature selection. The mean AUC and mean accuracy of the best performing model are 0.68±0.02 and 0.55±0.03 respectively.

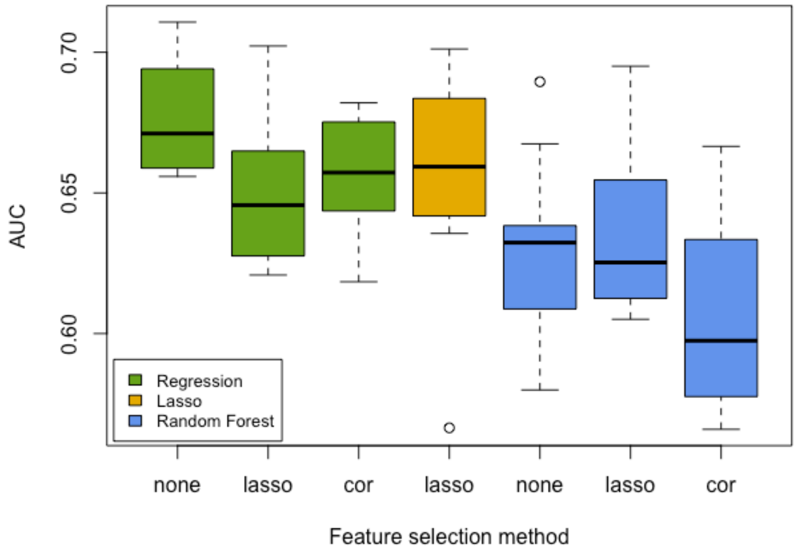

Figure 8. AUC's of different models with lasso-, correlation- and no feature selection for predicting short vs. normal vs. long stay. *AUC, area under the curve; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Figure 9 shows the sensitivity, specificity, PPV, F1-score and % of patients in each class for the best performing model based on AUC. From the figure, we can read that the sensitivity of the best model is 0.79±0.05 for *long stay* and that the model is 61% sure about the classification of *long stay*. The F1-score for *long stay* is 0.69±0.03. However, the model is less correct (<40% of cases) as well as less certain (<50%) about the classification of *short stay* and *normal stay*. The majority of patients was in *long stay* class. The model performs best for this class.

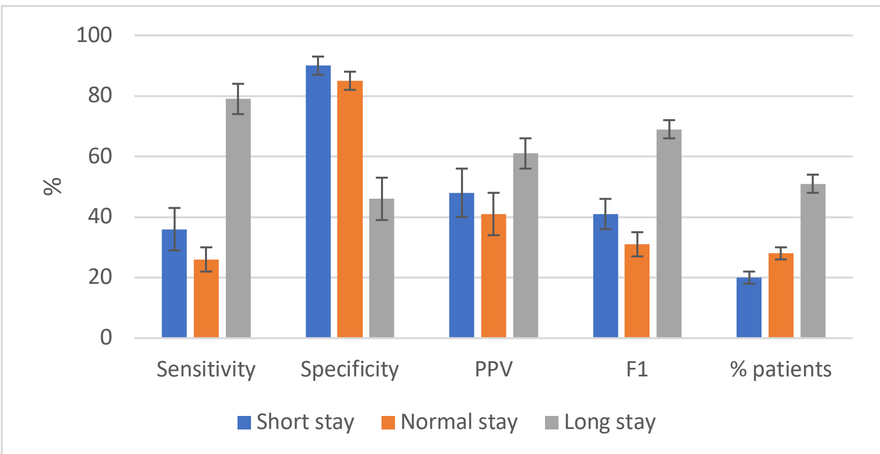

Figure 9. Overview of sensitivity, specificity, PPV, F1-score and % of patients per category for regression without feature selection for predicting of short vs. normal vs. long stay. *PPV, positive predicted value; F1-score, weighted average of precision and recall.*

Figure 10 shows the RMSE of the regression- and RF models without feature selection and correlation-based feature selection for the prediction of length of stay (continuous). The model, which performs best, based on RMSE, is the regression model without feature selection. The mean RMSE of the best performing model is 0.48±0.23. From the figure, we can see that values for all three models with lasso-selection are missing. That is due to the fact that during the lasso feature selection procedure, none or only one predictor was selected. That is too little to make a prediction.

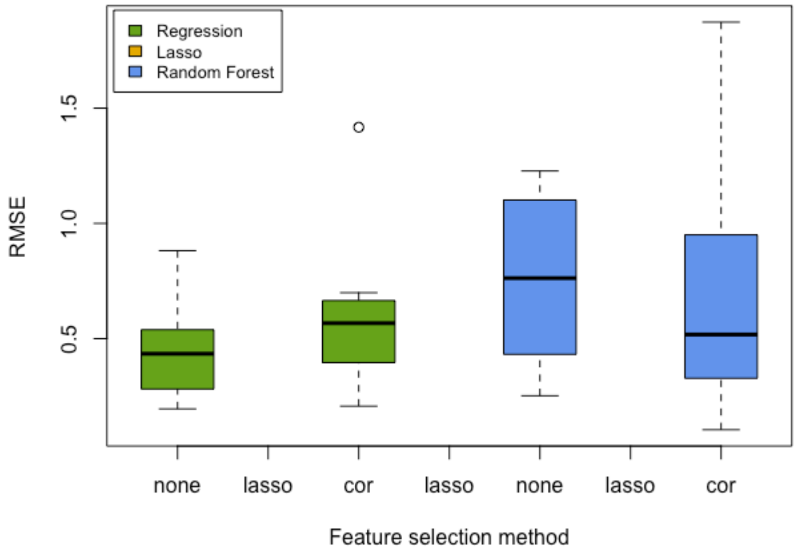

Figure 10. RMSE of different models with correlation- and no feature selection for predicting continuous length of stay. *RMSE, root mean squared error; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Table 8 shows a mean predicted length of stay of 6 days. The best model, based on RMSE, has a mean error 0.15 days and a median error of 0.93 days. The lower and upper quartiles, q0.25 and q0.75, are -10.6 and 6.6 days respectively.

Table 8. Mean prediction and error outcomes for regression without feature selection for predicting continuous length of stay.

	Value
Mean, days (S.D.)	6.05 (0.21)
Error, mean (S.D.)	0.15 (0.53)
Error, abs mean (S.D.)	3.33 (0.29)
Error, median (Q1 – Q3)	0.93 (-10.61 – 6.61)

SD, standard deviation; abs, absolute; Q1, lower quartile 0.25; Q3, upper quartile 0.75.

4.3.2 Discharge destination

Figure 11 shows the AUC values of the regression- and RF models without feature selection, with lasso selection and correlation-based feature selection and of the lasso regression model for the prediction of discharge destination (*home* versus *not home*). The model, which performs best, based on AUC, is the RF model without feature selection. The mean AUC and mean accuracy of the best performing model are 0.86±0.04 and 0.79±0.03 respectively.

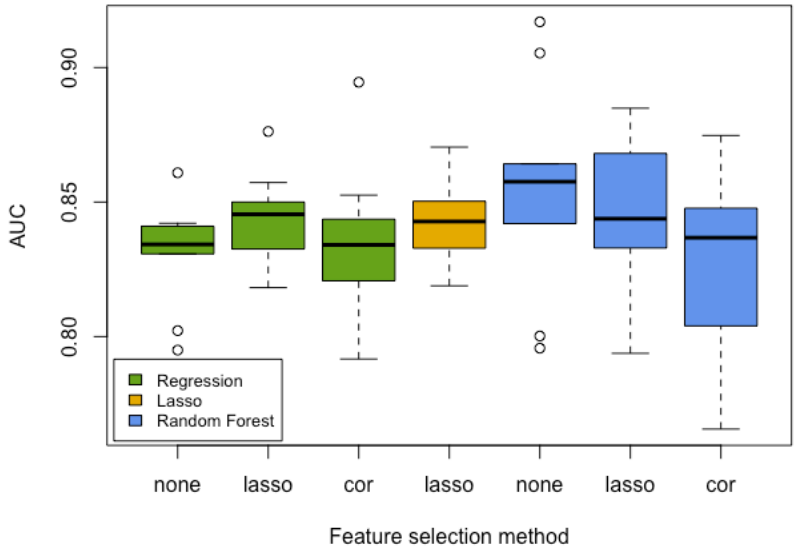

Figure 11. AUC's of different models with lasso-, correlation- and no feature selection for predicting home vs. not home. *AUC, area under the curve; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Figure 12 shows the sensitivity, specificity, PPV, F1-score and % of patients in each class for the best performing model based on AUC. From the figure, we can read that the sensitivity of the best model is 0.83±0.04 for *not home* and that the model is 85% sure about the classification of *not home*. The F1-score for *not home* is 0.84±0.03. However, the model is less certain (<70%) about the classification of *home*. Two-third of the patients was in *not home* class. The model performs best for this class.

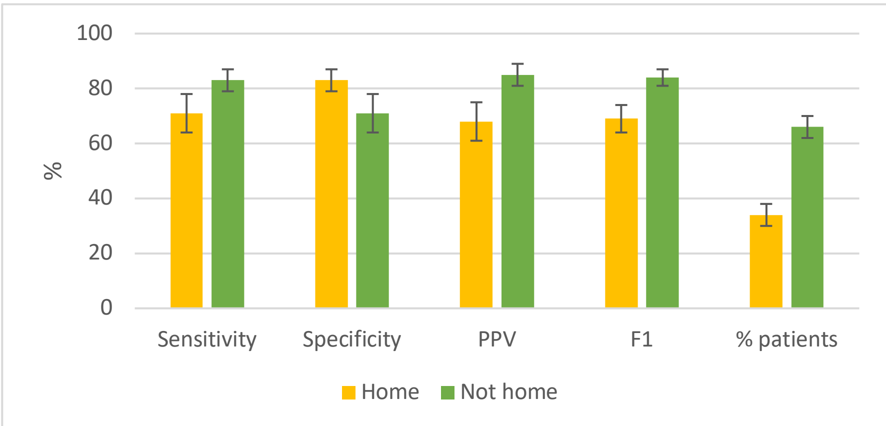

Figure 12. Overview of sensitivity, specificity, PPV, F1-score and % of patients per category for random forest without feature selection for predicting of home vs. not home. *PPV, positive predicted value; F1-score, weighted average of precision and recall.*

Figure 13 shows the AUC values of the regression- and RF models without feature selection, with lasso selection and correlation-based feature selection and of the lasso regression model for the prediction of discharge destination (*home* versus *home with care* versus *not home*). The model, which performs best, based on AUC, is the regression model with lasso selection. The mean AUC and mean accuracy of the best performing model are 0.86±0.02 and 0.78±0.02 respectively.

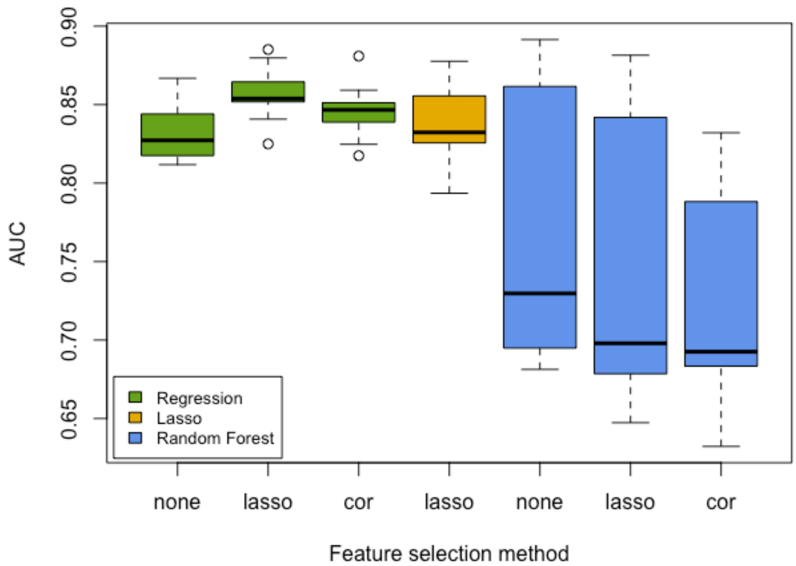

Figure 13. AUC's of different models with lasso-, correlation- and no feature selection for predicting home vs. home with care vs. not home. *AUC, area under the curve; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Figure 14 shows the sensitivity, specificity, PPV, F1-score and % of patients in each class for the best performing model based on AUC. From the figure, we can read that the sensitivity of the best model is 0.93±0.03 for *not home* and that the model is 81% sure about the classification of *not home*. The F1-score for *not home* is 0.87±0.01. However, the model is less correct (<60% and <10%) and less certain (<75% and <20%) about the classification of *home* and *home with care* respectively. The majority of the patients was in *not home* class and almost 10% of patients were in *home with care* class. The model performs best for the majority class.

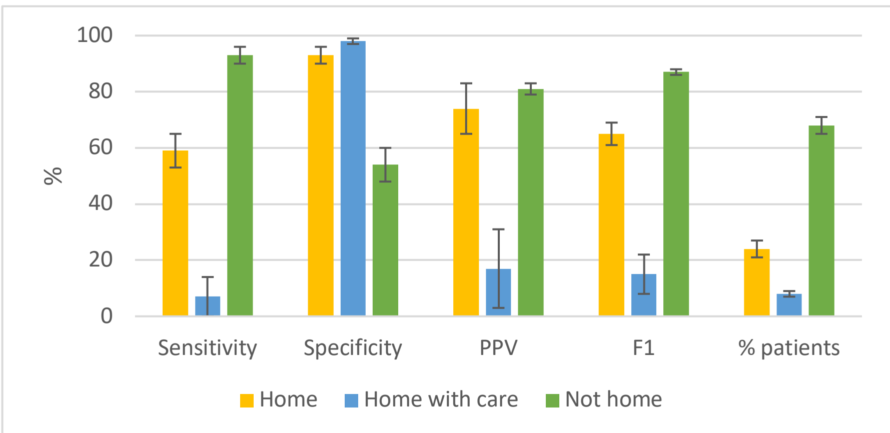

Figure 14. Overview of sensitivity, specificity, PPV, F1-score and % of patients per category for regression with lasso selection for predicting of home vs. home with care vs. not home. *PPV, positive predicted value; F1-score, weighted average of precision and recall.*

Figure 15 shows the AUC values of the regression- and RF models without feature selection, with lasso selection and correlation-based feature selection and of the lasso regression model for the prediction of discharge destination (all options). The model, which performs best, based on AUC, is the regression model with lasso selection. The mean AUC and mean accuracy of the best performing model are 0.76±0.03 and 0.58±0.03 respectively.

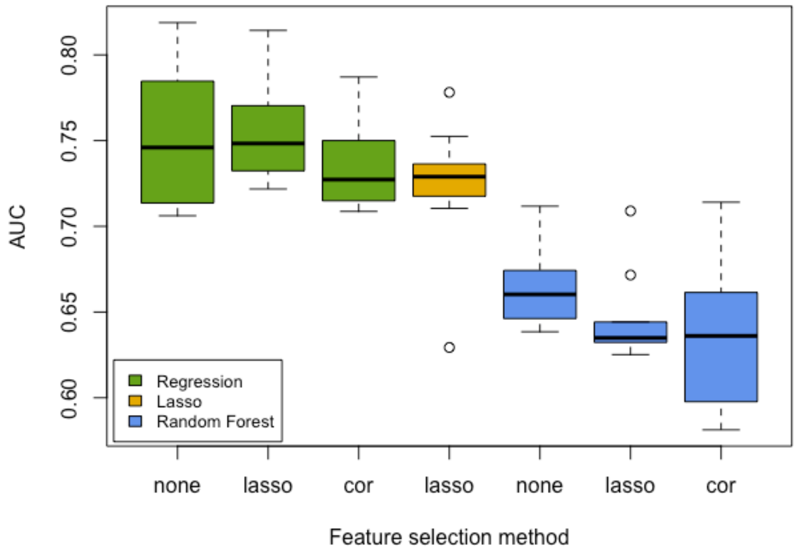

Figure 15. AUC's of different models with lasso-, correlation- and no feature selection for predicting discharge destination (all options). *AUC, area under the curve; lasso, least absolute shrinkage and selection operator; cor, correlation.*

Figure 16 shows the sensitivity, specificity, PPV, F1-score and % of patients in each class for the best performing model based on AUC. From the figure, we can read that the sensitivity of the best model is 0.79±0.04 for *nursing home with revalidation* and that the model is 60% sure about this classification. The F1-score for this is 0.68±0.04. The sensitivity and F1-score of the prediction of *home* are similar to this (0.68±0.06 and 0.67±0.05 respectively). However, the model is less correct (<30%) and less certain (<40%) about the classification of the other destinations. The majority of the patients was in *nursing home with revalidation* (47%) class and 24% of patients were in the *home* class. Note that the error-bars are relatively big in some cases (± 30-40%).

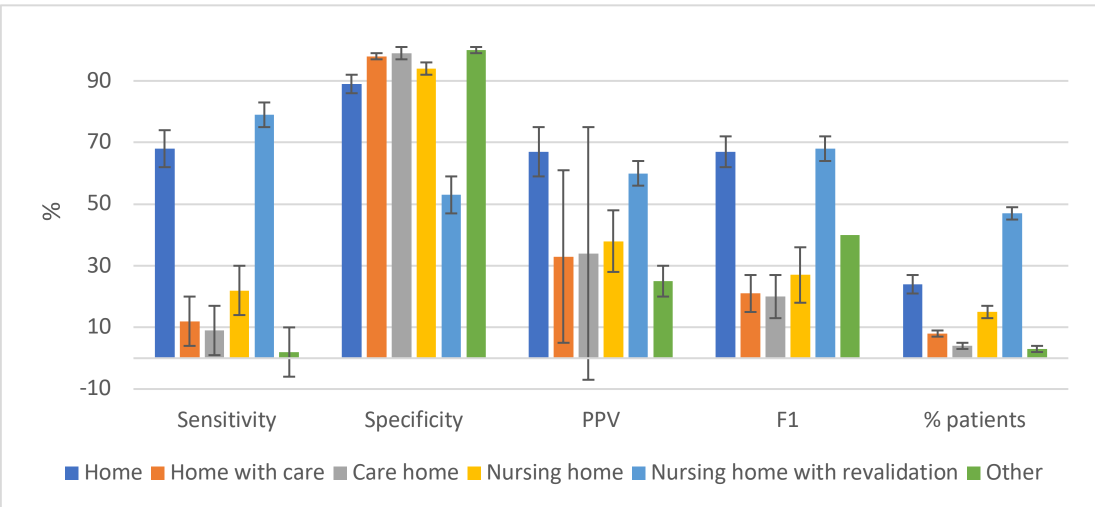

Figure 16. Overview of sensitivity, specificity, PPV, F1-score and % of patients per category for regression with lasso selection for predicting discharge destination (all options). *PPV, positive predicted value; F1-score, weighted average of precision and recall.*

4.4 Parameters

The aim of this section is to determine which parameters contribute most to the prediction of a particular outcome. We only include the results of the models which performed the best according to the findings shown in Section 4.3. An overview of the important variables of all models discussed in Section 4.3 (with clinical data as input data) is presented in Appendix 8.

4.4.1 Length of stay

Age, ASA-score, fracture type, involvement of geriatrician, patient's level of independence in activities of daily living (total KATZ-score) and type of therapy were the predictors which contributed most to the prediction whether the patient has a short/long stay (Table 9). For the prediction of short/normal/long stay, the following were the most important predictors: age, KATZ-adl-1, KATZ-adl-5 and time to surgery. Lastly, for the prediction of continuous LOS, the only (important) predictor was the patient's living situation prior to the injury.

Table 9. Overview of importance of top 10 selected variables by best performing model for predicting length of stay per outcome.

Outcome	**LOS (short/long)**			**LOS (short/normal/long)**			**LOS (continuous)**		
Model	RF - no feature selection			Reg. - no feature selection			Reg. - no feature selection		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	9	1	2						
Age	1	10	**100**	1	10	**100**			
ASA-score	6	10	**50**	8	9	27			
Dementia	10	1	1						
Fracture type	2,5	10	**85**	9	7	14			
General				8	2	6			
Geriatrician	3	10	**80**	8,5	6	15			
KATZ-adl-1				4	10	**70**			
KATZ-adl-2				9	1	2			
KATZ-adl-3				7	5	20			
KATZ-adl-5				3	10	**80**			
KATZ-adl-6	10	8	8						
KATZ-total	3,5	10	**75**						
Origin	8	10	30				1	10	**100**
Pre-mobility	8	9	27	10	5	5			
Regional				5	3	18			
SNAQ-appetite				6	7	35			
SNAQ-month				7	1	4			
SNAQ-nutrition				7,5	2	7			
SNAQ-total	10	1	1	6	9	45			
Spinal				8	1	3			
Therapy	5	10	**60**	9	2	4			
TTO	7	10	40	2	10	**90**			

*LOS, length of stay; RF, random forest; reg., regression; rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living [88]; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire. Red numbers represent the variables which have an importance greater than 50.*

4.4.2 Discharge destination

Age, fracture type, involvement of geriatrician, living situation prior to the injury and type of therapy were the predictors which contribute most to the prediction whether a patient is discharged to home or not (Table 10). For the prediction of home/home with care/not home discharge, the following were the most important predictors: age, dementia, involvement of geriatrician, KATZ-adl-6 living situation prior to the injury and total SNAQ-score. Lastly, for the prediction of all discharge destinations in this study, age, general anesthesia mode, living situation prior to the injury and time to surgery were the most important predictors.

Table 10. Overview of importance of top 10 selected variables by best performing model for predicting discharge destination per outcome.

Outcome **Model**	**Discharge destination (home/not home)** RF - no feature selection			**Discharge destination (home/care/not home)** Reg. - lasso selection			**Discharge destination (all)** Reg. - lasso selection		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	10	1	1				6	2	10
Age	1	10	**100**	4,5	10	**65**	5	9	**54**
ASA-score	6,5	10	45	6,5	10	45	6	7	35
Dementia	10	9	9	5	9	**54**	2	1	9
Fracture type	4	10	**70**	7	9	36	6	7	35
General							5	9	**54**
Geriatrician	2	10	**90**	2	10	**90**	8	5	15
KATZ-adl-6				3	9	**72**	8	3	9
KATZ-total	7	10	40	4,5	2	13	7,5	6	21
Origin	3	10	**80**	4,5	10	**65**	3	9	**72**
Pre-mobility	9	10	20				6	5	25
Pre-osteo				1	4	40	6	4	20
Regional							9	5	10
SNAQ-month				8	10	30	9	1	2
SNAQ-total				3	7	**56**	3	1	8
Spinal				3	1	8	7	3	12
Therapy	5	10	**60**	6	8	40	7	5	20
TTO	8	10	30				1	6	**60**

*RF, random forest; reg., regression; rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living [88]; osteo, osteoporosis; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire. Red numbers represent the variables which have an importance greater than 50.*

4.5 Online tool

An overview of the online tool is presented in Figure 17. By completing the form, the model calculates two predictions. First, it calculates the probability that the patient is discharged to a facility. Second, it calculates the probability that the patient has a long stay. In the example scenario in the figure, the probability for facility discharge is 96% and the probability for long hospital stay is 43%.

Figure 17. Overview of the online tool for prediction of length of stay and discharge destination.

4.6 Statistics

The aim of this section is to determine whether the best performing model (according to Section 4.3) for each outcome, was likely to perform significantly better than the other models for that particular outcome. We only include the results of models with clinical patient data as input.

The data from continuous LOS and discharge destination (*home/care/not home*) are not normally distributed (Table 11) and the data of the other outcomes are normally distributed.

Table 11. Probabilities that the AUC-values (or RMSE-values for continuous length of stay) are normally distributed.

	Shapiro-Wilk (p-value)
LOS: short/long	0.716
LOS: short/normal/long	0.137
LOS: continuous	0.001*
DD: home/not home	0.473
DD: home/care/not home	<0.001*
DD: all	0.052

*AUC, area under the curve; RMSE, root mean squared error; LOS, length of stay; DD, discharge destination; *, p-value < 0.05.*

The models for each outcome, except for discharge destination (*home/not home*), were significantly different from each other (Table 12). Furthermore, if we look at the F-value, the variation among groups is more than expected by chance for the significant outcomes tested by ANOVA.

Table 12. p-values that the models for a particular outcome are statistically different based on AUCs and RMSEs.

	ANOVA			**Kruskal-Wallis**	
	p-value	F-value	Df	p-value	Df
LOS: short/long	<0.001*	5.419	6		
LOS: short/normal/long	<0.001*	5.588	6		
LOS: continuous				0.041*	3
DD: home/not home	0.333	1.172	6		
DD: home/care/not home				<0.001*	6
DD: all	<0.001*	24.36	6		

*AUC, area under the curve; RMSE, root mean squared error; LOS, length of stay; DD, discharge destination; Df, degrees of freedom; *, p-value < 0.05.*

The RF model without feature selection performed best for the prediction of *short* versus *long stay* (Figure 6) based on AUC. After Bonferroni correction on the alpha level, this model performed significantly better than the regression model with correlation-based feature selection and the lasso model (Table 13).

Table 13. p-values that the AUCs of the RF model without feature selection compared to the AUCs of other models are statistically different for the prediction of LOS (short/long) with Bonferroni corrected alpha level.

	Pairwise t-test
	p-value
RF (none) x Reg (none)	0.016
RF (none) x Reg (lasso)	0.980
RF (none) x Reg (cor.)	<0.001*
RF (none) x LASSO (lasso)	<0.001*
RF (none) x RF (lasso)	0.357
RF (none) x RF (cor.)	0.024

*AUC, area under the curve; LOS, length of stay; none, no feature selection; lasso, lasso selection; cor., correlation-based feature selection RF, random forest; reg, regression; *, p-value < 0.0071.*

The regression model without feature selection performed best for the prediction of *short* versus *normal* versus *long stay* (Figure 8) and for the prediction of continuous LOS (Figure 10) based on AUC and RMSE respectively. After Bonferroni correction on the alpha level, the model for LOS (*short/normal/long*) performed significantly better than the RF model without feature selection and the RF model with correlation-based feature selection (Table 14).

Despite the fact that according to the Kruskal-Wallis test (Table 12) the models on continuous LOS performed significantly different from each other, this was not found after Bonferroni correction on the alpha level (Table 14).

Table 14. p-values that the AUCs and RMSEs of the regression model without feature selection compared to the AUCs and RMSEs of other models are statistically different for the prediction of LOS (short/normal/long) and continuous LOS with Bonferroni corrected alpha level.

	Pairwise t-test	**Pairwise Wilcoxon Rank Sum test**
	p-value	p-value
Reg (none) x Reg (lasso)	1.000	-
Reg (none) x Reg (cor.)	1.000	1.000
Reg (none) x LASSO (lasso)	1.000	-
Reg (none) x RF (none)	0.004*	0.069
Reg (none) x RF (lasso)	0.181	-
Reg (none) x RF (cor.)	<0.001*	1.000

*AUC, area under the curve; RMSE, root mean squared error; LOS, length of stay; none, no feature selection; lasso, lasso selection; cor., correlation-based feature selection RF, random forest; reg, regression; *, p-value < 0.0071; **, p-value < 0.0125.*

The regression model with lasso feature selection performed best for the prediction of *home/home with care/not home* (Figure 13) and for the prediction of all possible discharge destinations (Figure 15) based on AUC. After Bonferroni correction on the alpha level, the model for discharge destination (*home/home with care/not home*) performed significantly better than the RF model with correlation-based feature selection (Table 15). Moreover, we found that the regression model with correlation-based feature selection ($p = 0.003$) also performed better than the RF model with correlation-based feature selection.

Furthermore, after Bonferroni correction on the alpha level, the regression model with lasso selection on all discharge destinations performed significantly better than all RF models (Table 15). In addition, we found that the regression model without feature selection ($p < 0.001$; $p < 0.001$; $p < 0.001$), the regression model with correlation-based feature selection ($p < 0.001$; $p < 0.001$; $p < 0.001$) and the lasso model ($p < 0.001$; $p < 0.001$; $p < 0.001$) also performed significantly better than all RF models.

Table 15. p-values that the AUCs of the regression model with lasso selection compared to the AUCs of other models are statistically different for the prediction of discharge destination (home/home with care/not home) and all discharge destinations with Bonferroni corrected alpha level.

	Pairwise Wilcoxon Rank Sum test	**Pairwise t-test**
	p-value	p-value
Reg (lasso) x Reg (none)	0.144	1.000
Reg (lasso) x Reg (cor.)	1.000	1.000
Reg (lasso) x LASSO (lasso)	1.000	0.683
Reg (lasso) x RF (none)	1.000	<0.001*
Reg (lasso) x RF (lasso)	0.241	<0.001*
Reg (lasso) x RF (cor.)	0.001*	<0.001*

*AUC, area under the curve; DD, discharge destination; none, no feature selection; lasso, lasso selection; cor., correlation-based feature selection RF, random forest; reg, regression; *, p-value < 0.0071.*

5. Discussion

5.1 Main findings

In this study, we addressed the question which variables are useful for the prediction of post-surgery length of stay (LOS), discharge destination and mortality and whether the best performing model is sufficient for clinical implementation. We found that age, patient's health status (ASA score – American Society of Anesthesiologists score), fracture type, involvement of geriatrician, patient's independence in activities of daily living (total KATZ-score; ADL, activities of daily living) and type of therapy were the most important predictor variables for LOS (*short/long*). For discharge destination (*home/not home*), the most important variables were age, fracture type, involvement of geriatrician, living situation prior to the injury and type of therapy.

In both cases, a random forest (RF) model without feature selection performed the best and their area under the curve (AUC) was greater than 0.75 and 0.85 for LOS and discharge destination respectively. Therefore, based on AUC and interpretation of AUC according to literature, the model for LOS is acceptable to use as assistance tool in practice, whereas the model for discharge destination is excellent to use as assistance tool. However, based on the other measures, we see that the results are especially good for the majority groups in our data: patients with a long stay and patients who are not discharged to home.

Moreover, we found that, for predicting discharge destination, the best models of all three classifications (A: *home/not home*, B: *home/home with care/not home*, C: *home/home with care/care home/nursing home/nursing home with revalidation*) had a mean AUC greater than 0.75, of which two (A & B) had a mean AUC greater than 0.85.

Lastly, we found that it was not possible to define a reliable outcome for 30-day mortality. However, in literature we found that – amongst others – the following predictors are associated with mortality: age, living situation, pre-surgery mobility, weight loss, dependence in ADL, dementia, ASA-score, body mass index (BMI) and smoking.

5.2 Relevance

Hip fracture patients benefit from receiving the specific care they need for successful rehabilitation. This includes, for example, assistance in daily activities at home or (temporary) admission to a nursing home. However, nowadays, the patient flow from hospital to discharge destinations is not optimal due to shortage in beds in these facilities. As a consequence, patients have a longer hospital stay, which can lead to unwanted complications and a longer rehabilitation period. Therefore, optimization of the patient flow is desired. In this study we developed a prediction model which is potentially able to assist in decision making early in the process (pre-surgery) and we identified relevant variables for the prediction of length of stay and discharge destination.

5.3 Comparison with previous literature

With respect to variables that are related to length of stay and discharge destination, we found that that there are no studies which consider as many variables as we did in our study. Current studies only include one or a few variables in their investigation (Section 2). Therefore, we can compare our findings with the findings in various articles on one variable.

For predicting length of stay, we have partially similar results as described in previous literature. In literature, it was described that age [40,42,57], ASA-score [35,60], involvement of a geriatrician [55,56] and type of surgery [24–34] are associated with LOS. We found that, indeed, these factors have predictive power for the prediction of LOS. Literature also described that anesthesia mode [20,59], pre-surgery mobility [35], dementia [38], frailty [43,44] and time to surgery [45–48] are associated to LOS. However, in our study, anesthesia mode and dementia played no important role in prediction. In addition, pre-surgery mobility, frailty and time to surgery played a small role in the prediction [4]. Moreover, we found that type of fracture and total KATZ-score were important predictors. These were not found in literature.

For predicting discharge destination, we also have partially similar results as described in previous literature. In literature, it was described that age [64,66–68,70,71,73–76], type of fracture [66,70,75] and living situation prior to injury [66,70,72] are associated with discharge destination. Our study showed that these factors are important for the prediction of discharge destination. In literature, it was also described that ASA-score [64,74,76], anesthesia mode [64], level of assistance needed [66,76] and pre-surgery mobility [66,68,69,74,75] are associated with discharge destination. However, we found that ASA-score and level of assistance needed played a small role in the prediction, whereas anesthesia mode and pre-surgery mobility were not really contributing a lot to the prediction [5]. Furthermore, we found that involvement of a geriatrician, type of therapy, dementia, indication for malnutrition (total SNAQ-score) and time to surgery were important predictors for discharge destination. These were not found in literature.

There are different explanations for the deviation in our findings on length of stay and discharge destination and the findings in literature. First, we based our conclusions of important variables on a self-invented measure. This measure is not verified in literature and therefore might lead to misleading and maybe even incorrect conclusions. It is possible that we can draw different conclusions when a different approach is used for determining the important variables. This conclusion might be more similar to results in literature. Second, for the variables which we found the be important for the predictions but were not identified as important during the literature search, it is the case (as far as we know) that these were not previously tested. Consequently, those could not be found. Third, our conclusions are based on the predictions for LOS (*short* versus *long stay*) and discharge destination (*home* versus *home*). However, the findings in literature are based on other classifications. In our study, we also found that different predictors were relevant when predicting a different categorization for LOS and discharge destination (Appendix 8). Therefore, different predictors might be relevant for the classifications used in literature. Lastly, the study populations in literature are very different from the study population in this research. For instance, the literature studies were conducted with patients from different countries, such as: France, United Kingdom, Iran, United States, China and Australia. While our study was conducted with Dutch patients. It could be that for different nationalities, different predictors are important. In addition, some studies had a specific study setting which only included patients with a certain type of fracture, while we included patients with different types of hip fractures. It is possible that for one specific hip fracture type, some predictors are very relevant, while other predictors are more relevant in case more hip fracture types are included. Furthermore, our predictor set contained many variables to test on predictive power. We did not find any article which contained as many predictors as we had in our data set. Besides, it was the case that most articles tested only one or two predictors on their association with length of stay or discharge destination. Therefore, it is possible that different findings would have been found if there was a study which contained a similar predictor set as our predictor set.

Moreover, as discussed in Section 2, for several variables (time to operation, age and gender) we found contradicting statements in literature on whether those variables are associated with length of stay. In our study, we found that time-to-operation (TTO) is contributing to the prediction of both *short* versus *normal* versus *long stay* as well as *short* versus *long stay* (Table 9 & Table 10). However, for the latter, its contribution-score was below our threshold (40/100), whereas for the prior, its contribution-score was 90/100. Therefore, we have similar results as in literature, which are contradicting. We also found that the variable *age* is very important in the prediction of length of stay and there were no contradictions in this finding. For *gender*, we could not test its importance since this variable was not available in our dataset.

[4] Note that we found that some predictors were important for predicting the binomial outcome of LOS and not (very) important for the multinomial outcome of LOS and vice versa.

[5] Note that we found that some predictors were important for predicting the binomial outcome of discharge destination and not (very) important for the multinomial outcome of discharge destination and vice versa.

In addition, in literature we found that having dementia is associated with a shorter LOS compared to not having dementia. According to the experts, this was an interesting finding, since it is more difficult for people with dementia to recover from their injury. A possible explanation for this matter, is that for these patients (i.e. with dementia), it is clear that they are discharged to the nursing home where they probably already lived. Therefore, there are no difficulties with arranging a place (since they already had the place and cannot go home). As a consequence, the patient can stay in the hospital for a shorter time after surgery, despite the fact that the patient is not yet recovered. In this study, we found that dementia is, indeed, contributing to the prediction of discharge destination.

With respect to predictive models, there are currently no studies which aim to compare the performance of different predictive models with a similar patient population. Therefore, we cannot compare our findings with previous literature.

5.4 Strengths and limitations

Our study is limited in the fact that we only included articles published on PubMed in our search strategy. Therefore, it is possible that there are variables that can be used for prediction of length of stay and discharge destination, which we did not identify as they were in a different database. However, a strength of our literature search is that, in particular for the search on LOS, we already included many (44) articles. Therefore, we might conclude that we have done an extensive literature search and found most related variables. Another limitation is that, for the search on discharge destination, we also included articles which had a publication date greater than 10 years from the moment of search. As a consequence, it is possible that the findings described in these articles are no longer applicable [69,72,73]. The reason for this is that the evidence for these findings is outdated [103,104]. Whereas for the search on length of stay and mortality we only included articles with a publication date less than 5 years from the moment of search and only included articles written in English. This is also a limitation, since it is possible that we missed relevant information. Lastly, a limitation of our literature search, is that we filtered on 'full text availability' in PubMed, while we had access to the Utrecht University (UU) library. Therefore, articles which were not available in full text on PubMed, could have been accessed through the UU library. Consequently, it is possible that additional relevant variables were not found.

Another limitation of our study is that we only performed four interviews. The participants were all related to the '*Samenwerkende Algemene Ziekenhuizen'* (SAZ), since our study was performed within the scope of a running SAZ project. It might be better to include more participants to the study in order the ensure a level of information saturation is reached. A strength of conducting interviews, is that we used the expertise of the participants to verify our literature findings.

Furthermore, we applied imputation on the data. This was necessary to handle missing values. However, we should be aware that these are not true values. Therefore, the models might perform differently if values of daily practice could be used.

Moreover, in the prediction of a particular outcome, different variables were found to be important for different models (Appendix 8). In addition, for the different outcomes we compared for LOS and discharge destination (Section 3.5.1) also different predictors were important. As a consequence, it was not possible to give a definite answer to the question which predictor variables are most important for LOS and discharge destination. Furthermore, each model run contained a random train- and test set. This leads to different results for different model runs, since the models are trained and tested on different subsets of the data, each with different characteristics. By running the model once, the predictions could be based on chance, therefore we ran our models 10 times. We assumed that 10 model runs resulted in enough models to make valid conclusions. However, it is still possible that by running the model 100 or 1000 times, leads to different conclusions and that different variables would be identified as important.

Another limitation and strength of this study is that the measure we used to compare the importance of variables was self-invented. It is a limitation in the fact that we cannot be completely sure that the conclusions based on the weighted importance are valid. On the contrary, this self-invented measure allowed us to combine multiple evaluation outcomes. By using the weighted importance, we could be strict about the inclusion of important variables. If we did not use such a measure, the conclusions would have been based on guesses. So, it was important to have a measure like the weighted importance. Another strength is that we could use the variables which were the most important according to this measure. By doing this, we concluded that the performances of the models (with only the selected variables as input) were similar or better than the models with all available predictor variables as input. This indicates that the measure is useful in the selection of important variables for a particular outcome.

In addition, we found that the performance measures were particularly good for the majority classes in our study population. A possible explanation for this finding, is that it is safer for the model to classify a new patient in the majority group, since there is a bigger chance that it will be correct. This is a limitation, since there exist cases in which the doctor is interested in the minority class rather than the majority class. But then the model does not perform well. A possible solution to overcome this problem is to perform a research in which the study population is more equally divided among the different classes. However, it also is a strength that the model performs well for the majority groups. The transfer nurse has to make arrangements in case a patient is discharged to a facility. Therefore, it is useful to know whether the patient is discharged to a facility. Since *not home* is the majority group, this is something the model is good at. Also, in case a patient has a long stay, the transfer nurse does not immediately have to start making arrangements for a bed. Since, *long stay* is a majority group, this is also something the model is good at. Therefore, the transfer nurse knows she can wait a moment before arrangements have to be made.

Furthermore, we noticed that the variation in the error-bars in the prediction of all discharge destinations is relatively big for some performance measures (sensitivity, positive predicted value and F1-score) (Figure 16). A possible explanation for this is, that for *home with care* and *care home* the number of patients in that class is low. Therefore, it is possible that too few patients were in the train- or test set. Consequently, the model would not have been able to make predictions in all classes and a NA value was assigned, resulting in a large error-bar.

Lastly, a strength of our study is that we included a lot of variables in our search to find the most important variables for prediction. This enabled us to possibly find new associations, which were not previously described in literature. For example, we found that type of fracture and level of assistance needed in daily activities (total KATZ-score) were important predictors for LOS. In addition, we found that involvement of a geriatrician, type of therapy, dementia, indication for malnutrition (total SNAQ-score) and time to surgery were important predictors for discharge destination. These were not previously described in literature. On the contrary, including many variables does not necessarily lead to a better performance. Therefore, we included two feature selection methods in our comparison: lasso- and correlation-based selection.

5.5 Future research

There are multiple possibilities for further research on this topic. First, one could compare our models (regression, lasso regression and RF) with other machine learning models, e.g. neural networks (NN) and support vector machines (SVM). In case a model is developed which is more accurate and performs better than our best model, this is more useful in daily practice. In addition, it is also interesting to look into ways to improve our current models on LOS and discharge destination.

Second, it would be interesting to find a reliable data source to define *mortality* as outcome for hip fracture patients and use this outcome for training the predictive models. As predictor variables, one could use the variables associated with mortality according to Section 2.3 and test their predictive power. A prediction for (30-day) mortality is helpful in decision support in the consideration to not perform surgery on hip fracture patients with a poor mental and/or physical health status.

Third, we could not use all of the variables we found in literature that are associated with LOS and discharge destination, due to absence of these variables in our dataset. Therefore, in a future study, the predictive power of these variables (Section 2.1.3 and Section 2.2.2) could be determined.

Fourth, it would be interesting to investigate the generalizability of our models. To be more specific: to investigate whether our models could also be successfully applied to patients with other health problems, such as patients with lung- or colon cancer. For these patients, mortality is a common phenomenon [105]. Similar as for hip fracture patients, the prediction of mortality in patients with cancer, could be used as decision support method for treatment options.

Lastly, we developed an online tool for the prediction of length of stay and discharge destination. This tool could be further improved and tested on usability as a decision support system in practice [6]. For example, one could add bootstrap data to generate confidence intervals when a prediction is made. At the moment this was not possible due to the fact the there is no technical support for adding bootstrap data for categorical variables with more than two options. Another possibility to improve the tool, is to add validation data to check whether the model is also suited for other study populations. Also, due to limitations in our account, we were only able to implement the regression model instead of the RF model (which performed better). Therefore, one could ask for permission to use the technology which is required to implement the RF model.

[6] Note that it is not (yet) allowed to use the current tool in practice. For this, a license is required. In addition, one should be aware that the tool serves as support system and is not intended as an autonomous decision system.

6. Conclusion

In this study, we presented variables which are associated with length of stay, discharge destination and mortality. We found various variables which are important predictors for LOS and discharge destination: age, patient's health status, type of fracture, involvement of geriatrician, patient's independence in activities of daily living, type of therapy and living situation prior to the injury. These are mostly corresponding to what has been found in previous literature.

Furthermore, we compared several regression and machine learning models with and without feature selection in their performances, using different measures. Based on AUC, we found that a RF model without feature selection performed best for both LOS (*short* versus *long stay*) (AUC = 0.78±0.03) and discharge destination (*home* versus *not home*) (AUC = 0.86±0.04). However, these findings are not statistically significant. In addition, we found that the models perform particularly good on the majority classes: *long stay* and *not home*.

The best performing models have the potential to be used in practice. Yet, they can be further improved, and their performance might be compared to other machine learning models to find the optimal model. The predictions of these models could be used to improve the patient flow and therefore reduce healthcare costs and contribute to the rehabilitation process of hip fracture patients.

References

1. Samenwerkende Algemene Ziekenhuizen. Waardegedreven zorg. 2018.
2. Magaziner J, Chiles N, Orwig D. Recovery after Hip Fracture: Interventions and Their Timing to Address Deficits and Desired Outcomes – Evidence from the Baltimore Hip Studies. Nestle Nutr Inst Workshop Ser [Internet]. 2015;83:71–81. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5494960/
3. Landelijk netwerk acute zorg. LTR Factsheet 2012 Acute Ziekenhuisopnames Voor Heupfracturen. 2012.
4. Zorgkaart Nederland. Gebroken heup [Internet]. 2018 [cited 2018 Nov 21]. Available from: https://www.zorgkaartnederland.nl/aandoeningen/heupfractuur/artikelen
5. National Clinical Guideline Centre. The management of hip fracture in adults [Internet]. First edit. National Clinical Guideline Centre (NCGC). London: National Clinical Guideline Centre; 2011. 1-628 p. Available from: https://www.nice.org.uk/guidance/cg124/evidence/full-guideline-pdf-183081997%0Ahttps://www.nice.org.uk/guidance/cg124/evidence/full-guideline-183081997
6. Mayo Clinic staff. Hip fracture [Internet]. 2018 [cited 2018 Nov 22]. Available from: https://www.mayoclinic.org/diseases-conditions/hip-fracture/diagnosis-treatment/drc-20373472
7. Cornwall R, Gilbert MS, Koval KJ, Strauss E, Siu AL. Functional Outcomes and Mortality Vary among Different Types of Hip Fractures. Clin Orthop Relat Res [Internet]. 2004;425(425):64–71. Available from: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00003086-200408000-00009
8. Moerman S, Mathijssen NMC, Tuinebreijer WE, Vochteloo AJH, Nelissen RGHH. Hemiarthroplasty and total hip arthroplasty in 30,830 patients with hip fractures: data from the Dutch Arthroplasty Register on revision and risk factors for revision. Acta Orthop. 2018;89(5):509–14.
9. Monacelli F, Pizzonia M, Signori A, Nencioni A, Gianotti C, Minaglia C, et al. The In-Hospital Length of Stay after Hip Fracture in Octogenarians: Do Delirium and Dementia Shape a New Care Process? J Alzheimers Dis. 2018;66(1):281–8.
10. Aharonoff GB, Kenneth KJ, Skovron ML, Zuckerman JD. Hip fractures in the elderly: predictors of one year mortality. J Orthop Trauma. 1997;11(3):162–5.
11. Landeiro F, Leal J, Gray AM. The impact of social isolation on delayed hospital discharges of older hip fracture patients and associated costs. Osteoporos Int. 2016;27(2):737–45.
12. RTV Arnhem. Stijgend tekosrt ziekenhuisbedden door stokkende doorstroming ouderen. RTV Arnhem [Internet]. 2017 Dec 8 [cited 2018 Nov 21]; Available from: https://www.rtv-arnhem.nl/stijgend-tekort-ziekenhuisbedden-door-stokkende-doorstroming-ouderen/nieuws/item?1053595
13. IBM. Artifical Intelligence in medicine [Internet]. [cited 2019 Jul 2]. Available from: https://www.ibm.com/watson-health/learn/artificial-intelligence-medicine
14. Copeland B. Artificial Intelligence. Encyclopaedia Britannica, Inc. 2019.
15. Cruz AS, Lins HC, Medeiros RVA, Filho JMF, Silva SG. Artificial intelligence on the identification of risk groups for osteoporosis, a general review. Biomed Eng Online [Internet]. 2018;17(1):1–17. Available from: https://doi.org/10.1186/s12938-018-0436-1
16. D'onofrio K. New AI Accurately Predicts Hip Fractures [Internet]. 2019 [cited 2019 Jul 2]. Available from: https://www.docwirenews.com/condition-center/orthopedicscc/can-ai-predict-hip-fracture/
17. Galiatsatos D, Anastassopoulos G, Drosos G, Ververidis A, Tilkeridis K, Kasakos K. Prediction of 30-Day Mortality after a Hip Fracture Surgery Using Neural and Bayesian Networks. In: Iliadis L, Maglogiannis I, Papadopoulos H, editors. Artificial Intelligence Applications and Innovations. 10th ed. IFIP-AICT; 2014. p. 566–75.
18. Lott A, Haglin J, Belayneh R, Konda SR, Egol KA. Admitting Service Affects Cost and Length of Stay of Hip Fracture Patients. Geriatr Orthop Surg Rehabil [Internet]. 2018;9:1–5. Available from: http://journals.sagepub.com/doi/10.1177/2151459318808845%0Ahttp://www.ncbi.nlm.nih.gov/pubmed/30479850%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6249656
19. Iliopoulos E, Yousaf S, Watters H, Khaleel A. Hospital stay and blood transfusion in elderly patients with hip fractures. J Perioper Pract. 2017;27(12):288–91.
20. Van Waesberghe J, Stevanovic A, Rossaint R, Coburn M. General vs. neuraxial anaesthesia in hip fracture patients: A systematic review and meta-analysis. BMC Anesthesiol. 2017;17(1):1–25.
21. Lawrence JE, Fountain DM, Cundall-Curry DJ, Carrothers AD. Do Patients Taking Warfarin Experience Delays to Theatre, Longer Hospital Stay, and Poorer Survival After Hip Fracture? Clin Orthop Relat Res. 2017;475(1):273–9.

22. Gleason LJ, Mendelson DA, Kates SL, Friedman SM. Anticoagulation management in individuals with hip fracture. J Am Geriatr Soc. 2014;62(1):159–64.
23. Zhang J, Chen X, Wang J, Liu Z, Wang X, Ren J, et al. Poor prognosis after surgery for intertrochanteric fracture in elderly patients with clopidogrel treatment: A cohort study. Med (United States). 2017;96(39):2–6.
24. Shi H, Xiao L, Wang Z. Curative effect of artificial femoral head replacement and its effect on hip joint function and complications of senile patients with femoral intertrochanteric fracture. Exp Ther Med. 2018;16(2):623–8.
25. Ma J xiong, Kuang M jie, Xing F, Zhao Y long, Chen H ting, Zhang L kai, et al. Sliding hip screw versus cannulated cancellous screws for fixation of femoral neck fracture in adults: A systematic review. Int J Surg [Internet]. 2018;52(August 2017):89–97. Available from: https://doi.org/10.1016/j.ijsu.2018.01.050
26. Kazemian GH, Manafi AR, Najafi F, Najafi MA. Treatment of intertrochanteric fractures in elderly highrisk patients: Dynamic hip screw vs. external fixation. Injury [Internet]. 2014;45(3):568–72. Available from: http://dx.doi.org/10.1016/j.injury.2013.11.020
27. Woon CYL, Moretti VM, Schwartz BE, Goldberg BA. Total Hip Arthroplasty and Hemiarthroplasty: US National Trends in the Treatment of Femoral Neck Fractures. Am J Orthop (Belle Mead NJ) [Internet]. 2018;46(6):E474–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29309466
28. Carlson VR, Ong AC, Orozco FR, Lutz RW, Duque AF, Post ZD. The Direct Anterior Approach Does Not Increase Return to Function Following Hemiarthroplasty for Femoral Neck Fracture. Orthopedics [Internet]. 2017;40(6):e1055–61. Available from: http://www.healio.com/doiresolver?doi=10.3928/01477447-20170925-08
29. Arirachakaran A, Amphansap T, Thanindratarn P, Piyapittayanun P, Srisawat P, Kongtharvonskul J. Comparative outcome of PFNA, Gamma nails, PCCP, Medoff plate, LISS and dynamic hip screws for fixation in elderly trochanteric fractures: a systematic review and network meta-analysis of randomized controlled trials. Eur J Orthop Surg Traumatol. 2017;27(7):937–52.
30. Esen E, Dur H, Ataoglu B, Ayanoglu T, Turanli S. Evaluation of proximal femoral nail-antirotation and cemented, boploar hemiarthroplasty with calcar replacement in treatment of intertrochanteric femoral fractures in terms of mortality and morbidity ratios. Eklem Hast Cerrahisi. 2017;28(1):35–40.
31. Zhang Y, Dong Q, Sun X, Hu F. External fixation versus dynamic hip screw in treatment of elderly intertrochanteric hip fractures: A systematic review and meta-analysis. J Orthop Sci [Internet]. 2016;21(6):841–6. Available from: http://dx.doi.org/10.1016/j.jos.2016.08.002
32. Jettoo P, James P. Dynamic hip screw fixation versus multiple screw fixation for intracapsular hip fracture. J Orthop Surg. 2016;24(2):146–9.
33. Shen J, Hu C, Yu S, Huang K, Xie Z. A meta-analysis of percutenous compression plate versus intramedullary nail for treatment of intertrochanteric HIP fractures. Int J Surg [Internet]. 2016;29(30):151–8. Available from: http://dx.doi.org/10.1016/j.ijsu.2016.03.065
34. Hou Z, Shi J, Ye H, Pan Z. Treatment of unstable intertrochanteric fractures with percutaneous non-contact bridging plates. Int J Surg [Internet]. 2014;12(5):538–43. Available from: http://dx.doi.org/10.1016/j.ijsu.2014.02.017
35. Richards T, Glendenning A, Benson D, Alexander S, Thati S. The independent patient factors that affect length of stay following hip fractures. Ann R Coll Surg Engl. 2018;100(7):556–62.
36. Kempegowda H, Richard R, Borade A, Tawari A, Graham J, Suk M, et al. Obesity Is Associated with High Perioperative Complications among Surgically Treated Intertrochanteric Fracture of the Femur. J Orthop Trauma. 2017;31(7):352–7.
37. Liao KM, Lu HY. A national analysis of complications following total hip replacement in patients with chronic obstructive pulmonary disease. Med (United States). 2016;95(12):1–6.
38. Harvey L, Mitchell R, Brodaty H, Draper B, Close J. The influence of dementia on injury-related hospitalisations and outcomes in older adults. Injury [Internet]. 2016;47(1):226–34. Available from: http://dx.doi.org/10.1016/j.injury.2015.09.021
39. Bliemel C, Oberkircher L, Eschbach DA, Lechler P, Balzer-Geldsetzer M, Ruchholtz S, et al. Impact of Parkinson’s disease on the acute care treatment and medium-term functional outcome in geriatric hip fracture patients. Arch Orthop Trauma Surg. 2015;135(11):1519–26.
40. Marrinan S, Pearce MS, Jiang XY, Waters S, Shanshal Y. Admission for osteoporotic pelvic fractures and predictors of length of hospital stay, mortality and loss of independence. Age Ageing. 2015;44(2):258–61.
41. Rudge JE, Kim D. New-onset hyponatraemia after surgery for traumatic hip fracture. Age Ageing.

2014;43(6):821–6.
42. Dawson-Bowling SJ, Jha S, Chettiar KK, East DJ, Gould GC, Apthorp HD. A multidisciplinary enhanced recovery programme allows discharge within two days of total hip replacement; three- to five-year results of 100 patients. HIP Int. 2014;24(2):167–74.
43. Gleason LJ, Benton EA, Alvarez-Nebreda ML, Weaver MJ, Harris MB, Javedan H. FRAIL Questionnaire Screening Tool and Short-Term Outcomes in Geriatric Fracture Patients. J Am Med Dir Assoc [Internet]. 2017;18(12):1082–6. Available from: https://doi.org/10.1016/j.jamda.2017.07.005
44. Krishnan M, Beck S, Havelock W, Eeles E, Hubbard RE, Johansen A. Predicting outcome after hip fracture: Using a frailty index to integrate comprehensive geriatric assessment results. Age Ageing. 2014;43(1):122–6.
45. Pailleret C, Ait Hamou Z, Rosencher N, Samama CM, Eyraud V, Chilot F, et al. A retrospective comparison between delayed and early hip fracture surgery in patients taking clopidogrel: same total bleeding but different timing of blood transfusion. Int Orthop. 2017;41(9):1839–44.
46. McIsaac DI, Abdulla K, Yang H, Sundaresan S, Doering P, Vaswani SG, et al. Association of delay of urgent or emergency surgery with mortality and use of health care resources: A propensity score-matched observational cohort study. Cmaj. 2017;189(27):E905–12.
47. Alonso-Fernández P, Romero E, Chung M, García-Salmones M, Cabezas P, Mora J. Delayed surgery in hip fracture patients. Can we afford it? Int J Health Plann Manage. 2017;32(4):653–9.
48. Zehir S, Zehir R, Sarak T. Early surgery is feasible in patients with hip fractures who are on clopidogrel therapy. Acta Orthop Traumatol Turc. 2015;49(3):249–54.
49. Steinberg EL, Warschawski Y, Elis J, Rotman D, Rachevsky G, Factor S, et al. Financial impact and effect on the outcome of preoperative tests for at-risk older hip fracture patients. Geriatr Gerontol Int. 2018;18(6):937–42.
50. Kimmel LA, Liew SM, Sayer JM, Holland AE. HIP4Hips (High intensity physiotherapy for hip fractures in the acute hospital setting): A randomised controlled trial. Med J Aust. 2016;205(2):73–8.
51. Rostagno C, Buzzi R, Campanacci D, Boccacini A, Cartei A, Virgili G, et al. In hospital and 3-month mortality and functional recovery rate in patients treated for hip fracture by a multidisciplinary team. PLoS One. 2016;11(7):4–11.
52. Marcantonio A, Steen B, Kain M, Bramlett K-J, Tilzey JF, Iorio R. The Clinical and Economic Impact of Preoperative Transthoracic Echocardiography in Elderly Patients with Hip Fractures. 2015;73(4):239–42. Available from: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1000086
53. Bernstein J, Roberts FO, Wiesel BB, Ahn J. Preoperative testing for hip fracture patients delays surgery, prolongs hospital stays, and rarely dictates care. J Orthop Trauma. 2016;30(2):78–80.
54. Desai SJ, Patel J, Abdo H, Lawendy AR, Sanders D. A comparison of surgical delays in directly admitted versus transferred patients with hip fractures: Opportunities for improvement? Can J Surg. 2014;57(1):40–3.
55. Frenkel Rutenberg T, Daglan E, Heller S, Velkes S. A comparison of treatment setting for elderly patients with hip fracture, is the geriatric ward superior to conventional orthopedic hospitalization? Injury [Internet]. 2017;48(7):1584–8. Available from: http://dx.doi.org/10.1016/j.injury.2017.04.049
56. Middleton M, Wan B, Da Assunçao R. Improving hip fracture outcomes with integrated orthogeriatric care: A comparison between two accepted orthogeriatric models. Age Ageing. 2017;46(3):465–70.
57. Chandran M, Tay D, Huang XF, Hao Y. The burden of inpatient care for diabetic and non-diabetic patients with osteoporotic hip fractures—does it differ? An analysis of patients recruited into a fracture liaison service in Southeast Asia. Arch Osteoporos. 2018;13(1):2–9.
58. Mamarelis G, Key S, Snook J, Aldam C. Outcomes after early return to theatre following hip hemiarthroplasty for intracapsular fracture of the femoral neck. Bone Jt J. 2017;99B(7):958–63.
59. Basques BA, Bohl DD, Golinvaux NS, Leslie MP, Baumgaertner MR, Grauer JN. Postoperative length of stay and 30-day readmission after geriatric hip fracture: An analysis of 8434 patients. J Orthop Trauma [Internet]. 2015;29(3):e115–20. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L600661565%5Cnhttp://dx.doi.org/10.1097/BOT.0000000000000222%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=15312291&id=doi:10.1097%2FBOT.0000000000000222&atitle=Postoperative+l
60. Ricci WM, Brandt A, McAndrew C, Gardner MJ. Factors affecting delay to surgery and length of stay for patients with hip fracture. J Orthop Trauma. 2015;29(3):e109–14.
61. Purushothaman B, Webb M, Weusten A, Bonczek S, Ramaskandhan J, Nanu A. Decision making on timing of surgery for hip fracture patients on clopidogrel. Ann R Coll Surg Engl. 2016;98(2):91–5.
62. Lott A, Haglin J, Belayneh R, Konda SR, Leucht P, Egol KA. Does Use of Oral Anticoagulants at the Time

of Admission Affect Outcomes Following Hip Fracture. Geriatr Orthop Surg Rehabil [Internet]. 2018;9:215145931876415. Available from: http://journals.sagepub.com/doi/10.1177/2151459318764151
63. Morrissey N, Iliopoulos E, Osmani AW, Newman K. Neck of femur fractures in the elderly: Does every hour to surgery count? Injury [Internet]. 2017;48(6):1155–8. Available from: http://dx.doi.org/10.1016/j.injury.2017.03.007
64. Rudasill SE, Dattilo JR, Liu J, Nelson CL, Kamath AF. Do illness rating systems predict discharge location, length of stay, and cost after total hip arthroplasty? Arthroplast Today [Internet]. 2018;4(2):210–5. Available from: https://doi.org/10.1016/j.artd.2018.01.004
65. Gialanella B, Ferlucci C, Monguzzi V, Prometti P. Determinants of outcome in hip fracture: role of daily living activities. Eur J Phys Rehabil Med [Internet]. 2015;51(3):253–60. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=25427525
66. Salar O, Baker PN, Forward DP, Ollivere BJ, Weerasuriya N, Moppett IK, et al. Predictors of direct home discharge following fractured neck of femur. Ann R Coll Surg Engl. 2017;99(6):444–51.
67. Ruder JA, Hart GP, Kneisl JS, Springer BD, Karunakar MA. Predictors of Functional Recovery Following Periprosthetic Distal Femur Fractures. J Arthroplasty [Internet]. 2017;32(5):1571–5. Available from: http://dx.doi.org/10.1016/j.arth.2016.12.013
68. Hayashi H, Iwai M, Matsuoka H, ... Factors affecting the discharge destination of hip fracture patients who live alone and have been admitted to an inpatient rehabilitation unit. J Phys Ther Sci [Internet]. 2016;28:1228–32. Available from: https://www.jstage.jst.go.jp/article/jpts/28/4/28_jpts-2015-1001/_article/-char/ja/
69. Kagaya H, Takahashi H, Sugawara K, Dobashi M, Kiyokawa N, Ebina H. Predicting outcomes after hip fracture repair. Am J Phys Med Rehabil. 2005;84(1):46–51.
70. Kimmel LA, Holland AE, Edwards ER, Cameron PA, De Steiger R, Page RS, et al. Discharge destination following lower limb fracture: Development of a prediction model to assist with decision making. Injury [Internet]. 2012;43(6):829–34. Available from: http://dx.doi.org/10.1016/j.injury.2011.09.027
71. Aitken LM, Burmeister E, Lang J, Chaboyer W, Richmond TS. Characteristics and outcomes of injured older adults after hospital admission. J Am Geriatr Soc. 2010;58(3):442–9.
72. De Pablo P, Losina E, Phillips CB, Fossel AH, Mahomed N, Lingard EA, et al. Determinants of discharge destination following elective total hip replacement. Arthritis Care Res. 2004;51(6):1009–17.
73. Ottenbacher KJ, Linn RT, Smith PM, Illig SB, Mancuso M, Granger C V. Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. Ann Epidemiol. 2004;14(8):551–9.
74. Gholson JJ, Pugely AJ, Bedard NA, Duchman KR, Anthony CA, Callaghan JJ. Can We Predict Discharge Status After Total Joint Arthroplasty? A Calculator to Predict Home Discharge. J Arthroplasty [Internet]. 2016;31(12):2705–9. Available from: http://dx.doi.org/10.1016/j.arth.2016.08.010
75. Kristensen MT, Foss NB, Ekdahl C, Kehlet H. Prefracture functional level evaluated by the New Mobility Score predicts in-hospital outcome after hip fracture surgery. Acta Orthop. 2010;81(3):296–302.
76. Nanjayan SK, John J, Swamy G, Mitsiou K, Tambe A, Abuzakuk T. Predictors of change in "discharge destination" following treatment for fracture neck of femur. Injury [Internet]. 2014;45(7):1080–4. Available from: http://dx.doi.org/10.1016/j.injury.2014.02.005
77. Toson B, Harvey LA, Close JCT. The ICD-10 Charlson Comorbidity Index predicted mortality but not resource utilization following hip fracture. J Clin Epidemiol [Internet]. 2015;68:44–51. Available from: http://dx.doi.org/10.1016/j.jclinepi.2014.09.017
78. Cnudde P, Rogmark C, Rolfson O, Nemes S. Low predictive power of comorbidity indices identified for mortality after acute arthroplasty surgery undertaken for femoral neck fracture. Bone Joint J. 2019;101–B(1):104–12.
79. Rikkonen T. Long-term effects of functional impairment on fracture risk and mortality in postmenopausal women. Osteoporos Int. 2018;29:2111–20.
80. Berggren M, Stenvall M, Englund U, Olofsson B, Gustafson Y. Co-morbidities , complications and causes of death among people with femoral neck fracture – a three-year follow-up study. BMC Geriatr [Internet]. 2016;16:1–10. Available from: http://dx.doi.org/10.1186/s12877-016-0291-5
81. Mellner C, Eisler T, Börsbo J, Brodén C, Morberg P, Mukka S. The Sernbo score predicts 1-year mortality after displaced femoral neck fractures treated with a hip arthroplasty. Acta Orthop. 2017;88(4):402–6.
82. Zaslavsky O, Zelber-sagi S, Gray SL, Andrea Z, Brunner RL, Wallace RB, et al. A Comparison of Frailty Phenotypes for Prediction of Mortality Incident Falls, and Hip Fractures in Older Women. J Am Geriatr Soc. 2016;64(9):1858–62.

83. Forget P, Dillien P, Engel H, Cornu O, Kock M De, Yombi JC. Use of the neutrophil - to - lymphocyte ratio as a component of a score to predict postoperative mortality after surgery for hip fracture in elderly subjects. BMC Res Notes. 2016;9:1–6.
84. LOGEX. Privacystatement Logex [Internet]. 2019 [cited 2019 May 27]. Available from: https://www.logex.com/nl/privacy
85. MRDM. Privacystatement MRDM [Internet]. 2019 [cited 2019 Jun 21]. Available from: https://mrdm.nl/privacystatement-mrdm/
86. Value2Health. Privacystatement Value2Health [Internet]. 2019 [cited 2019 Jun 21]. Available from: https://value2health.com/nl/privacystatement-value2health/
87. Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J Thorac Oncol [Internet]. 2010;5(9):1315–6. Available from: http://dx.doi.org/10.1097/JTO.0b013e3181ec173d
88. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models. J Stat Softw. 2010;33(1):1–3.
89. Venables WN, Ripley BD. Modern Applied Statistics with S-Plus [Internet]. 4th ed. Springer. New York; 2002. Available from: http://www.stats.ox.ac.uk/pub/MASS4
90. Narkhede S. Understanding Logistic Regression [Internet]. 2018 [cited 2019 May 9]. Available from: https://towardsdatascience.com/understanding-logistic-regression-9b02c2aec102
91. Al-Masri A. How Does Linear Regression Actually Work? [Internet]. 2018 [cited 2019 May 9]. Available from: https://towardsdatascience.com/how-does-linear-regression-actually-work-3297021970dd
92. Donges N. The Logistic Regression Algorithm [Internet]. 2018 [cited 2019 May 9]. Available from: https://machinelearning-blog.com/2018/04/23/logistic-regression-101/
93. Jain D. What are the advantages and disadvantages of linear regression? [Internet]. 2018 [cited 2019 May 9]. Available from: https://www.quora.com/What-are-the-advantages-and-disadvantages-of-linear-regression
94. Flom P. The Disadvantages of Linear Regression [Internet]. 2018 [cited 2019 May 9]. Available from: https://sciencing.com/disadvantages-linear-regression-8562780.html
95. Kassambara A. Penalized Logistic Regression Essentials in R: Ridge, Lasso and Elastic Net [Internet]. 2018 [cited 2019 Jan 25]. Available from: http://www.sthda.com/english/articles/36-classification-methods-essentials/149-penalized-logistic-regression-essentials-in-r-ridge-lasso-and-elastic-net/
96. Fonti V. Feature Selection using LASSO. VU Amsterdam [Internet]. 2017;1–26. Available from: http://ieeexplore.ieee.org/document/7906512/
97. Wright MN, Ziegler A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J Stat Softw [Internet]. 2017;77(1). Available from: http://arxiv.org/abs/1508.04409%0Ahttp://dx.doi.org/10.18637/jss.v077.i01
98. Donges N. The Random Forest Algorithm [Internet]. 2018 [cited 2019 May 13]. Available from: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd
99. Kumar N. Advantages and Disadvantages of Random Forest Algorithm in Machine Learning [Internet]. 2019 [cited 2019 Jul 1]. Available from: http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-random.html
100. Mackenzie RJ. One-Way vs Two-Way ANOVA: Differences, Assumptions and Hypotheses [Internet]. Technology Networks. 2018 [cited 2019 Jul 11]. Available from: https://www.technologynetworks.com/informatics/articles/one-way-vs-two-way-anova-definition-differences-assumptions-and-hypotheses-306553
101. Döring M. Comparing Measurements Across Several Groups: ANOVA [Internet]. 2018 [cited 2019 Jun 26]. Available from: https://www.datascienceblog.net/post/statistical_test/anova_one_way/
102. Analytics-for-fun. Performing ANOVA Test in R: Results and Interpretation [Internet]. 2014 [cited 2019 Jun 26]. Available from: http://www.analyticsforfun.com/2014/06/performing-anova-test-in-r-results-and.html
103. Davis P. Growing Impact of Older Articles [Internet]. 2014 [cited 2019 Jul 11]. Available from: https://scholarlykitchen.sspnet.org/2014/11/10/growing-impact-of-older-articles/
104. Shorten A. When is the evidence too old? [Internet]. Evidence based Nursing. 2013 [cited 2019 Jul 11]. Available from: https://blogs.bmj.com/ebn/2013/09/26/when-is-the-evidence-too-old/
105. CBS. Darmkanker: 5 duizend sterfgevallen per jaar [Internet]. 2016 [cited 2019 Jul 2]. Available from: https://www.cbs.nl/nl-nl/nieuws/2016/43/darmkanker-5-duizend-sterfgevallen-per-jaar
106. Katz S, Downs TD, Cash HR, Grotz RC. Progress in Development of the Index of ADL. Gerontologist. 1970;10(1–1):20–30.

Appendix 1 – query literature search

The complete query used for literature search on the prediction of length of stay for hip fracture patients on PubMed at November 30, 2018 is the following:

(prediction[All Fields] OR predict[All Fields]) AND ("length of stay"[MeSH Terms] OR ("length"[All Fields] AND "stay"[All Fields]) OR "length of stay"[All Fields]) OR ("hospitalisation"[All Fields] OR "hospitalization"[MeSH Terms] OR "hospitalization"[All Fields]) AND ("time"[MeSH Terms] OR "time"[All Fields]) AND ("hip fractures"[MeSH Terms] OR ("hip"[All Fields] AND "fractures"[All Fields]) OR "hip fractures"[All Fields] OR ("hip"[All Fields] AND "fracture"[All Fields]) OR "hip fracture"[All Fields]) AND ("loattrfull text"[sb] AND "2013/12/02"[PDat] : "2018/11/30"[PDat]).

The query for prediction of mortality on PubMed at May 16, 2019:

("hip fractures"[MeSH Terms] OR hip fracture[Title/Abstract]) AND ("mortality"[MeSH Terms] OR mortality[Title/Abstract]) AND ("death"[MeSH Terms] OR death[Title/Abstract]) AND predict[Title/Abstract] AND ("2014/05/18"[PDat] : "2019/05/16"[PDat] AND English[lang] AND "adult"[MeSH Terms])

Appendix 2 – characteristics articles on length of stay

Table A. Characteristics of study populations and findings of associated variables to length of stay.

Reference	Title	Authors	Publication year	Population size	Age, mean (years)	Gender, f (%)	What is tested?	Conclusion
[18]	Admitting Service Affects Cost and Length of Stay of Hip Fracture Patients	Lott A, Haglin J, Belayneh R, Konda SR, Egol KA	2018	225	79.8	64	Surgical vs. medicine approach	Surgical approach associated with 2-day shorter LOS
[19]	Hospital stay and blood transfusion in elderly patients with hip fractures	Iliopoulos E, Yousaf S, Watters H, Khaleel A	2017	336	83.3	72	Blood transfusion vs. no blood transfusion	Blood transfusion (Hb < 110) associated with longer LOS
[20]	General vs. neuraxial anaesthesia in hip fracture patients: a systematic review and meta-analysis.	Van Waesberghe J, Stevanovic A, Rossaint R, Coburn M	2017	413,999			General vs. neuraxial anesthesia	Neuraxial anesthesia associated with shorter LOS
[21]	Do Patients Taking Warfarin Experience Delays to Theatre, Longer Hospital Stay, and Poorer Survival After Hip Fracture?	Lawrence JE, Fountain DM, Cundall-Curry DJ, Carrothers AD	2017	1,979	85	71	Use of warfarin vs. no use of warfarin	Use of warfarin associated with longer LOS
[22]	Anticoagulation management in individuals with hip fracture	Gleason LJ, Mendelson DA, Kates SL, Friedman SM	2014	1,080	85	76	Use of warfarin vs. no use of warfarin	Use of warfarin associated with longer LOS

[23]	Poor prognosis after surgery for intertrochanteric fracture in elderly patients with clopidogrel treatment: A cohort study	Zhang J, Chen X, Wang J, Liu Z, Wang X, Ren J, Sun T	2017	238	76	68	Use of clopidogrel vs. no use of clopidogrel	Use of clopidogrel associated with longer LOS
[24]	Curative effect of artificial femoral head replacement and its effect on hip joint function and complications of senile patients with femoral intertrochanteric fracture.	Shi H, Xiao L, Wang Z.	2018	80	73.5	49	Artificial femoral head replacement vs. internal proximal femur locking plate fixation	Artificial femoral head replacement associated with shorter LOS
[25]	Sliding hip screw versus cannulated cancellous screws for fixation of femoral neck fracture in adults: A systematic review.	Ma JX, Kuang MJ, Xing F, Zhao YL, Chen HT, Zhang LK, Fan ZR, Han C, Ma XL.	2018	594			SHS vs. CCS	SHS associated with longer LOS
[27]	Total Hip Arthroplasty and Hemiarthroplasty: US National Trends in the Treatment of Femoral Neck Fractures.	Woon CYL, Moretti VM, Schwartz BE, Goldberg BA	2017	12,757	80.8	73	THA vs. HA	THA associated with longer LOS
[28]	The Direct Anterior Approach Does Not Increase Return to Function Following Hemiarthroplasty for Femoral Neck Fracture.	Carlson VR, Ong AC, Orozco FR, Lutz RW, Duque AF, Post ZD	2017	160	82.8	61	Direct lateral vs. direct anterior approach	Direct anterior approach associated with shorter LOS

[29]	Comparative outcome of PFNA, Gamma nails, PCCP, Medoff plate, LISS and dynamic hip screws for fixation in elderly trochanteric fractures: a systematic review and network meta-analysis of randomized controlled trials.	Arirachakaran A, Amphansap T, Thanindratarn P, Piyapittayanun P, Srisawat P, Kongtharvonskul J	2017				DHS vs. PCCP vs. MSP, PFN vs. GN vs. LISS vs. Medoff	PFN associated with shortest LOS. DHS associated with longest LOS. PCCP, GN & PFN associated with shorter LOS than Medoff
[30]	Evaluation of proximal femoral nail-antirotation and cemented, bipolar hemiarthroplasty with calcar replacement in treatment of intertrochanteric femoral fractures in terms of mortality and morbidity ratios.	Esen E, Dur H, Ataoğlu MB, Ayanoğlu T, Turanlı S	2017	92	79.2	70	PFNA vs. HA	PFNA associated with shorter LOS
[31]	External fixation versus dynamic hip screw in treatment of elderly intertrochanteric hip fractures: A systematic review and meta-analysis.	Zhang Y, Dong Q, Sun X, Hu F.	2016	260			DHS vs. external fixation	DHS associated with shorter LOS
[32]	Dynamic hip screw fixation versus multiple screw fixation for intracapsular hip fracture	Jettoo P, James P.	2016	52,884	50+	74	DHS vs. MSF	MSF associated with shorter LOS
[33]	A meta-analysis of percutenous compression plate versus intramedullary nail for treatment of	Shen J, Hu C, Yu S, Huang K, Xie Z.	2016	612			PCCP vs. IMN	PCCP associated with shorter LOS

	intertrochanteric HIP fractures.							
[34]	Treatment of unstable intertrochanteric fractures with percutaneous non-contact bridging plates	Hou Z, Shi J, Ye H, Pan Z.	2014	88	77	53	NCB vs. GN	NCB associated with shorter LOS
[26]	Treatment of intertrochanteric fractures in elderly highrisk patients: dynamic hip screw vs. external fixation.	Kazemian GH, Manafi AR, Najafi F, Najafi MA.	2014	60	78	68	DHS vs. external fixation	External fixation associated with shorter LOS
[35]	The independent patient factors that affect length of stay following hip fractures.	Richards T, Glendenning A, Benson D, Alexander S, Thati S.	2018	330	82.2	78	ASA, AMTS, mobility status	ASA, AMTS and mobility status are independently associated with LOS
[36]	Obesity Is Associated With High Perioperative Complications Among Surgically Treated Intertrochanteric Fracture of the Femur.	Kempegowda H, Richard R, Borade A, Tawari A, Graham J, Suk M, Howenstein A, Kubiak EN, Sotomayor VR, Koval K, Liporace FA, Tejwani N, Horwitz DS.	2017	1,078	76	65.5	Obese vs. non-obese	BMI > 30 associated with longer LOS
[37]	A National Analysis of Complications Following Total Hip Replacement in Patients With Chronic Obstructive Pulmonary Disease.	Liao KM, Lu HY.	2016	2,426	69.9	52.5	COPD vs. no-COPD	COPD associated with longer LOS

[38]	The influence of dementia on injury-related hospitalisations and outcomes in older adults	Harvey L, Mitchell R, Brodaty H, Draper B, Close J.	2016	58,046	65+	67.5	Dementia vs. no-dementia	Dementia associated with shorter LOS
[39]	Impact of Parkinson's disease on the acute care treatment and medium-term functional outcome in geriatric hip fracture patients.	Bliemel C, Oberkircher L, Eschbach DA, Lechler P, Balzer-Geldsetzer M, Ruchholtz S, Buecking B	2015	402	81	73	Parkinson vs. no-Parkinson	Parkinson associated with longer LOS
[40]	Admission for osteoporotic pelvic fractures and predictors of length of hospital stay, mortality and loss of independence.	Marrinan S, Pearce MS, Jiang XY, Waters S, Shanshal Y.	2015	110	84	83	Age, problems on admission	Higher age and problems on admission associated with longer LOS
[41]	New-onset hyponatraemia after surgery for traumatic hip fracture.	Rudge JE, Kim D.	2014	254	82	71	Hyponatraemia vs. normonatraemic patients	Hyponatraemia associated with longer LOS
[42]	A multidisciplinary enhanced recovery programme allows discharge within two days of total hip replacement; three-to five-year results of 100 patients.	Dawson-Bowling SJ, Jha S, Chettiar KK, East DJ, Gould GC, Apthorp HD	2014	100	65	55	Age, Charlson index	Higher age and higher CI-score associated with longer LOS
[43]	Short-Term Outcomes in Geriatric Fracture Patients.	Gleason LJ, Benton EA, Alvarez-Nebreda ML, Weaver MJ, Harris MB, Javedan H	2017	175	82.3	75	Frailty level (robust, prefrail, frial)	Frailty associated with longer LOS

[44]	Predicting outcome after hip fracture: using a frailty index to integrate comprehensive geriatric assessment results.	Krishnan M, Beck S, Havelock W, Eeles E, Hubbard RE, Johansen A.	2014	178	81	74	Frailty level (high, intermediate, low)	Higher frailty level associated with longer LOS
[45]	A retrospective comparison between delayed and early hip fracture surgery in patients taking clopidogrel: same total bleeding but different timing of blood transfusion	Pailleret C, Ait Hamou Z, Rosencher N, Samama CM, Eyraud V, Chilot F, Baillard C.	2017	39	86	76.5	Early vs. delayed surgery with use of clopidogrel	Early surgery associated with shorter LOS
[46]	Association of delay of urgent or emergency surgery with mortality and use of health care resources: a propensity score-matched observational cohort study	McIsaac DI, Abdulla K, Yang H, Sundaresan S, Doering P, Vaswani SG, Thavorn K, Forster AJ	2017	15,160	57.5	51	Early vs. delayed surgery	Delayed surgery associated with longer LOS
[47]	Delayed surgery in hip fracture patients. Can we afford it?	Alonso-Fernández P, Romero E, Chung M, García-Salmones M, Cabezas P, Mora J	2017	723	84.3	78.4	Early vs. delayed surgery	Delayed surgery associated with longer LOS
[48]	Early surgery is feasible in patients with hip fractures who are on clopidogrel therapy	Zehir S, Zehir R, Sarak T.	2015	211	77.5	54.3	Early vs. delayed surgery with use of clopidogrel	Delayed surgery associated with longer LOS

[49]	Financial impact and effect on the outcome of preoperative tests for at-risk older hip fracture patients.	Steinberg EL, Warschawski Y, Elis J, Rotman D, Rachevsky G, Factor S, Salai M, Ben-Tov T	2018	2,789	82.8	-	Pre-surgery tests vs. post-surgery tests	Pre-operative test associated with longer LOS
[50]	HIP4Hips (High Intensity Physiotherapy for Hip fractures in the acute hospital setting): a randomised controlled trial.	Kimmel LA, Liew SM, Sayer JM, Holland AE.	2016	92	81.3	64	Physiotherapy vs. intensive physiotherapy	Intensive physiotherapy associated with shorter LOS
[51]	In Hospital and 3-Month Mortality and Functional Recovery Rate in Patients Treated for Hip Fracture by a Multidisciplinary Team.	Rostagno C, Buzzi R, Campanacci D, Boccacini A, Cartei A, Virgili G, Belardinelli A, Matarrese D, Ungar A, Rafanelli M, Gusinu R, Marchionni N	2016	458	83.1	66	Surgery (not) led by medicine specialist	Surgery led by specialist associated with shorter LOS
[52]	The Clinical and Economic Impact of Preoperative Transthoracic Echocardiography in Elderly Patients with Hip Fractures.	Marcantonio A, Steen B, Kain M, Bramlett KJ, Tilzey JF, Iorio R.	2015	195	80.3	-	(no) transthoracic echocardiography	Transthoracic echocardiography associated with longer LOS
[53]	Preoperative Testing for Hip Fracture Patients Delays Surgery, Prolongs Hospital Stays, and Rarely Dictates Care.	Bernstein J, Roberts FO, Wiesel BB, Ahn J.	2016	250	80.5	65.5	Pre-surgery tests vs. no pre-surgery tests	Testing associated with longer LOS

[54]	A comparison of surgical delays in directly admitted versus transferred patients with hip fractures: opportunities for improvement?	Desai SJ, Patel J, Abdo H, Lawendy AR, Sanders D.	2014	890	60+	72	Admission after transfer vs. direct admission	Admission after transfer associated with longer LOS
[55]	A comparison of treatment setting for elderly patients with hip fracture, is the geriatric ward superior to conventional orthopedic hospitalization?	Frenkel Rutenberg T, Daglan E, Heller S, Velkes S	2017	217	85.1	58.5	Orthopedic vs. geriatric ward	Orthopedic ward associated with shorter LOS
[56]	Improving hip fracture outcomes with integrated orthogeriatric care: a comparison between two accepted orthogeriatric models	Middleton M, Wan B, da Assunçao R.	2017	1,894	84	76	Geriatric vs. orthopedic services	Orthopedic services associated with shorter LOS
[57]	The burden of inpatient care for diabetic and non-diabetic patients with osteoporotic hip fractures-does it differ? An analysis of patients recruited into a fracture liaison service in Southeast Asia.	Chandran M, Tay D, Huang XF, Hao Y.	2018	389	77	84	Diabetes vs. no diabetes	Diabetes is not independently associated with LOS. Delayed surgery and post-surgery complications are associated with longer LOS.
[58]	Outcomes after early return to theatre following hip hemiarthroplasty for intracapsular fracture of the femoral neck.	Mamarelis G, Key S, Snook J, Aldam C	2017	689	83.8	74	Return to theatre vs. no return	Return to theatre associated with longer LOS

[59]	Postoperative length of stay and 30-day readmission after geriatric hip fracture: an analysis of 8434 patients.	Basques BA, Bohl DD, Golinvaux NS, Leslie MP, Baumgaertner MR, Grauer JN.	2015	8,434	84	73	Time-to-surgery, anesthesia mode, fixation	Delayed surgery, nail/plate/screw fixation and general anesthesia are associated with longer LOS
[60]	Factors affecting delay to surgery and length of stay for patients with hip fracture	Ricci WM, Brandt A, McAndrew C, Gardner MJ	2015	635	82	70	ASA, cardiac testing, day of admission, gender	Higher ASA score, performing cardiac tests, admission on Thursday/Friday and male sex are associated with longer LOS
[61]	Decision making on timing of surgery for hip fracture patients on clopidogrel	Purushothaman B, Webb M, Weusten A, Bonczek S, Ramaskandhan J, Nanu A	2016	71	83	67	Use of clopidogrel vs. clopidogrel with aspirin	Use of clopidogrel associated with longer LOS

f, female; LOS, length of stay; Hb, hemoglobin; SHS, sliding hip screw; CCS, cannulated cancellous screw; THA, total hip arthroplasty; HA, hemiarthroplasty; FNF, femoral neck fracture; DHS, dynamic hip screw; PCCP, percutaneous compression plating; MSP, Medoff sliding plate; PFN, proximal femoral nail; GN, gamma nail; LISS, Less Invasive Stabilization System; PFNA, proximal femoral nail anti-rotation; MSF, multiple screw fixation; IMN, intramedullary nail; NCB, Non-Contact Bridging; ASA, American Society of Anesthesiologists; AMTS, abbreviated mental test score; BMI, body mass index; COPD, Chronic Obstructive Pulmonary Disease; THR, total hip replacement; TTO, time to operation; CCI, Charlson comorbidity index.

Appendix 3 – characteristics articles on discharge destination

Table B. Characteristics of study populations and findings of associated variables to discharge destination.

Reference	Title	Authors	Publication year	Population size	Age (years)	Gender, f (%)	Conclusion	
							Findings	Discharge destination
[64]	Do illness rating systems predict discharge location, length of stay, and cost after total hip arthroplasty?	S.E. Rudasill, J.R. Dattilo, J. Liu, C.L. Nelson, A.F. Kamath	2018	372	18+	50.3	ASA > 3	Home or SNF
							Higher age, race (African-American)	SNF
							Revision procedures, spinal anesthesia mode	Home
							Female, increasing age/BMI	Not home
[65]	Determinants of outcome in hip fracture: role of daily living activities	Gialanella B, Ferlucci C, Monguzzi V, Prometti P	2015	204	81.5	84.8	Bowel management is ok	Home
[66]	Predictors of direct home discharge following fractured neck of femur	Salar O, Baker PN, Forward DP, Ollivere BJ, Weerasuriya N, Moppett IK, Moran CG	2017	10,044	81	73.9	Lower age, higher AMTS, lower incidence of comorbiditites (cardio-, cerero-vascular, COPD, renal, diabetes, Parkinson), lower use of medication (e.g. anti-coalgulation), not living alone, higher walking ability, lower	Home

							level of assistance needed, intracapsular fracture	
[74]	Can We Predict Discharge Status After Total Joint Arthroplasty? A Calculator to Predict Home Discharge.	Gholson JJ, Pugely AJ, Bedard NA, Duchman KR, Anthony CA, Callaghan JJ.	2016	107,300	66.5	60	Higher age, female, higher ASA, functionally dependent pre-surgery	Facility discharge
[76]	Predictors of change in 'discharge destination' following treatment for fracture neck of femur.	Nanjayan SK, John J, Swamy G, Mitsiou K, Tambe A, Abuzakuk T.	2014	1,503	83	71.1	Higher age, male, patient dependent in mobilization, higher ASA/AMTS, medical condition, delay in operation	Change in destination (home to facility)
[67]	Predictors of Functional Recovery Following Periprosthetic Distal Femur Fractures.	Ruder JA, Hart GP, Kneisl JS, Springer BD, Karunakar MA.	2017	58	80.3	79.3	Lower age	Home (living independent)
[68]	Factors affecting the discharge destination of hip fracture patients who live alone and have been admitted to an inpatient rehabilitation unit.	Hayashi H, Iwai M, Matsuoka H, Nakashima D, Nakamura S, Kubo A, Tomiyama N.	2016	54	81.3	72	Lower age, higher Functional Independence Measure motor score & cognitive	Home
[71]	Characteristics and Outcomes of Injured Older Adults After Hospital Admission	Aitken LM, Burmeister E, Lang J, Chaboyer W, Richmond TS	2010	6,069	78	58.9	Higher age, greater ISS (>26), longer LOS, injury caused by fall	Nursing home
							Admitted to ICU, female, greater ISS	Rehabilitation centre

							Admitted to ICU, transferred from hospital to trauma hospital	Convalescence
[69]	Predicting Outcomes after Hip Fracture Repair	Kagaya H, Takahashi H, Sugawara K, Dobashi M, Kiyokawa N, Ebina H	2005	141	78	70.5	Higher FIM score	Home
[72]	Determinants of Discharge Destination Following Elective Total Hip Replacement	P DE Pablo, E Losina, CB. Phillips, AH. Fossel, N Mahomed, EA. Lingard, JN. KATZ	2004	1,276	73	62	Disability to walk at discharge, higher age, obesity, living alone	Rehabilitation centre
[70]	Discharge destination following lower limb fracture: Development of a prediction model to assist with decision making	LA. Kimmel, AE. Holland, ER. Edwards, PA. Cameron, R De Steiger, RS. Page, B. Gabbe	2012	1,429	46.5	43.2	Higher age, more proximal injury, privately insured, not working prior to injury, living in metropolitan	Inpatient rehabilitation
							Living in rural area, no disability prior to injury	Home
[73]	Comparison of Logistic Regression and Neural Network Analysis Applied to Predicting Living Setting after Hip Fracture	Ottenbacher KJ, Linn RT, Smith PM, Illig SB, Mancuso M, Granger C	2004	3,708	75.5	73.7	Older patients, no follow-up therapy, impaired bowel/bladder function, deficits in self-care, marital status, LOS	Not home

[75]	Prefracture functional level evaluated by the New Mobility Score predicts in-hospital outcome after hip fracture surgery	MT Kristensen, NB Foss, C Ekdahl, H Kehlet	2010	280	81	-	Higher age, low prefracture functional level, intertrochanteric fracture	Not home

f, female; ASA, American Society of Anesthesiologists; SNF, skilled nursing facility; BMI, body mass index; AMTS, abbreviated mental test score; COPD, Chronic Obstructive Pulmonary Disease; ISS, injury severity score; LOS, length of stay; ICU, intensive care unit; FIM, functional independence measure.

Appendix 4 – characteristics articles on mortality

Table C. Characteristics of study populations and findings of associated variables to mortality.

Reference	Title	Authors	Publication year	Population size	Age (years)	Gender, f (%)	What is tested	Conclusion
[78]	Low predictive power of comorbidity indices identified for mortality after acute arthroplasty surgery undertaken for femoral neck fracture.	Bülow E, Cnudde P, Rogmark C, Rolfson O, Nemes S.	2019	42,354	79.4	74.7%	Elixhauser index, Charlson comorbidity index	Low predictive power for 30-day, 60-day, 1-year (2-year, 5-year) mortality
[79]	Long-term effects of functional impairment on fracture risk and mortality in postmenopausal women.	Rikkonen T, Poole K, Sirola J, Sund R, Honkanen R, Kröger H.	2018	2,815	59.1	100%	Age, BMD, BMI, smoking	Associated with long-term mortality
[81]	The Sernbo score predicts 1-year mortality after displaced femoral neck fractures treated with a hip arthroplasty.	Mellner C, Eisler T, Börsbo J, Brodén C, Morberg P, Mukka S.	2017	292	83	68%	Sernbo score: age, living situation, walking ability, mental status	Associated with 1-year mortality

[82]	Comparison of Frailty Phenotypes for Prediction of Mortality, Incident Falls, and Hip Fracture in Older Women.	Zaslavsky O, Zelber-Sagi S, Gray SL, LaCroix AZ, Brunner RL, Wallace RB, O'Sullivan MJ, Cochrane B, Woods NF.	2016	3,558	70.9	100%	Frailty (3 or more): weight loss, poor energy, weakness, slowness, and low physical activity	Associated with 1-year and long term mortality
[80]	Co-morbidities, complications and causes of death among people with femoral neck fracture - a three-year follow-up study.	Berggren M, Stenvall M, Englund U, Olofsson B, Gustafson Y.	2016	199	82.2	74%	Dependence in P-ADL, cardiovascular/pulmonary disease, dementia, comorbidities, dependence in walking, ASA-score, living in residential care facilities, gender	Associated with 30-day, 1-year and 3-year mortality
[83]	Use of the neutrophil-to-lymphocyte ratio as a component of a score to predict postoperative mortality after surgery for hip fracture in elderly subjects.	Forget P, Dillien P, Engel H, Cornu O, De Kock M, Yombi JC.	2016	286	84	69.4%	Scoring system: age, gender, NLR	Associated with 1-year mortality
[77]	The ICD-10 Charlson Comorbidity Index predicted mortality but not resource utilization following hip fracture.	Toson B, Harvey LA, Close JC.	2015	47,698	65+	73.1%	Charlson comorbidity index	Associated with 30-day and 1-year mortality

BMD, bone mineral density; BMI, body mass index; ADL, activities of daily living; ASA, American Society of Anesthesiologists; NLR, neutrophil-to-lymphocyte ratio.

Appendix 5 – meaning DICA predictors

In this Appendix, we elaborate on the interpretation of the selected DICA-predictors.

Affected side The side which is affected by the injury: left, right or both.

Age The patient's age at the moment of surgery.

ASA-score The American Society of Anesthesiologists (ASA) Physical Status Classification System was developed in 1941 in order to assign a physical health status to a patient before surgery. This status can be helpful in predicting operative risk. Table D describes the classes that are used to classify the patient's health.

Table D. Overview of American Society of Anesthesiologists (ASA) classification system.

Class	Description
I	Normal, healthy patient
II	Patient with mild systemic disease
III	Patient with severe systemic disease that limits activity, but is not incapacitating
IV	Patient with incapacitating systemic disease that is a constant threat to life
V	Moribund patient not expected to survive 24 hr with or without operation
E	Emergency operation of any type; E is appended to the patient's physical status

Dementia Does the patient suffer from dementia prior to the injury?

Fracture type The type of fracture the patient suffers from. Included fractures are: non-dislocated medial collum fracture, dislocated medial collum fracture, trochanter femoral fracture AO-A1, trochanter femoral fracture AO-A2, trochanter femoral fracture AO-A3 and sub-trochanter femoral fracture. Where AO stands for osteoarthritis. A1 represents a stable per-trochanter fracture, A2 represents an instable per-trochanter fracture and A3 represent an instable per-trochanter fracture with fracture line running to distal lateral cortex.

General anesthesia The anesthesia mode used during surgery was general anesthesia.

Geriatrician The involvement of a geriatrician in the patient's treatment. There are different ways to involve a geriatrician: no consultation of a geriatrician, post-operative consult, peri-operative consult at surgery/orthopedic department and intensive consultation at geriatric trauma department.

KATZ-adl-1 The patient is (not) dependent on washing him/herself.

KATZ-adl-2 The patient is (not) dependent on dressing him/herself.

KATZ-adl-3 The patient is (not) dependent on visiting the toilet.

KATZ-adl-4 The patient uses (no) incontinence material.

KATZ-adl-5 The patient is (not) dependent on moving from bed to chair.

KATZ-adl-6 The patient is (not) dependent on feeding him/herself.

KATZ-total This represents the independence of the patient in activities of daily living. The value is calculated by taking the sum of KATZ-adl-1, KATZ-adl-2, KATZ-adl-3, KATZ-adl-4, KATZ-adl-5 and KATZ-adl-6.

Origin The living situation of the patient prior to admission to the hospital. Included residences are: independent, independent with (daily) assistance, care home, nursing home, nursing home with rehabilitation and other.

Pre-mobility The patient's mobility score before the injury according to Pre-Fracture Mobility score.

Pre-osteoporosis Was the patient under treatment of osteoporosis prior to the injury?

Regional anesthesia The anesthesia mode used during surgery was regional anesthesia.

SNAQ-appetite The patient had (no) loss of appetite during the last month prior to the injury.

SNAQ-month The patient has (not) unintentionally lost 3 kg during the last month prior to the injury.

SNAQ-6months The patient has (not) unintentionally lost 6 kg during the last 6 months prior to the injury.

SNAQ-nutrition The patient had (not) used drinking- or tube feeding in the last month prior to the injury.

SNAQ-total This represents the degree of malnutrition of the patient. The value is calculated by taking the sum of SNAQ-appetite, SNAQ-month, SNAQ-6months, SNAQ-nutrition.

Spinal anesthesia The anesthesia mode used during surgery was spinal anesthesia.

Therapy The type of treatment that is used. Included therapies are: conservative approach, sliding hip screw, cannulated hip screw, IM hip pen, hemi-hip arthroplasty and total hip arthroplasty.

Time-to-operation This is the time between the date of admission and the date of surgery in days.

Appendix 6 – mortality research

We investigated different possibilities to find a reliable source for 30-day mortality as an outcome.

Mortality based on data available in DICA data

In the DICA dataset, there is a field called 'mortality_30d'. This field represents whether the patients has died within 30 days after the surgery. However, it appeared that there are a lot of missing values. Moreover, from the patients whose data was available, *only* 12 patients were labeled as deceased. Based on this, we concluded that it was not possible to train a model on the registered mortality in DICA data, let alone make reliable predictions.

Mortality based on closing reason = 2

For each treatment a patient has in the hospital, a care activity is opened. Care activities remain status 'open' until it is closed by a doctor, else it closes automatically after 120 days. There are different reasons to close a particular care activity, for example in case the patient has died. This is registered with closing reason = 2.

In addition, after consultation with hospitals, we have learnt that there currently is no link between the hospital's Electronic Health Record (EHR) and the Municipal Personal Record Database (GBA – gemeentelijke basis administratie), in which all deceased persons are registered. This means that a doctor cannot know whether a patient has died after he or she is discharged. Therefore, the care activities will only be closed with reason = 2, when a patient has died in the hospital. However, in the available data, the in-hospital mortality is low. Consequently, we are not able to use closing reason = 2 for 30-day mortality and we cannot use it for training our models. Even if the in-hospital mortality would have been high, we could not be sure that not more patients had died within 30-days.

In case there was a link between the EHR and GBA, and when a patient died after discharge, the care activity would have been automatically closed with reason = 2. Given this, and the date of death, we would have been able to extract all the patients who died within 30 days after surgery.

Mortality based on care activities

In case we would assume a patient has died, when no (new) care activities are opened after 30-days, we could possibly use this for our mortality outcome. This is due to the fact that normally hip fracture patients have multiple checkup appointments, so at least these would be registered in the care activities. However, it is not necessarily the case that a patient has died when no care activities are available after a certain moment in time. For example, a patient could have moved and visit a new doctor at another hospital. Consequently, the activities would be registered at this new hospital. Also, it is possible that there really are no care activities, and yet the patient lives. Based on this uncertainty, we concluded that this was not a reliable source for mortality either.

Appendix 7 – questions interview

Interview SAZ-specialists

Introduction

My name is Laira Fransen. For my Master's in Artificial Intelligence, I started my final project at Value2Health last November. I am also involved in the '*Waardegedreven Zorg'* program of the SAZ. During this project, I will focus on the development of prediction models for the length of stay and discharge destination of hip fracture patients and the identification of factors that are associated with these outcomes. This information can be used to inform e.g. the rehabilitation center in advance about the arrival of the patient in order to stimulate an improved patient flow.

Today, I would like to accomplish the following:

+ Obtaining (more) insight in the general process around hip fractures admitted to the hospital.
+ Identifying relevant parameters that can be used for the prediction of length of stay and discharge destination of hip fracture patients.
+ Vision of specialist concerning the cause (and possible solution) of the decreasing flow.

Do you have any questions or remarks at this moment?

Questions

Process

1. Could you please explain your role in the hospital and the '*Waardegedreven Zorg'* program?

2. Could you please elaborate on the process of hip fracture patients?

   Pre-operative
   a. How are patients admitted to the hospital?
   b. What are common causes of hip fractures?
   c. What are the most common types of hip fractures?
   d. What are comment treatment methods? Is there a possibility for the patient to express their preference?
   e. What is the time between admission and surgery?

   Surgery
   f. Who are involved during surgery?
   g. How much time takes an average surgery?
   h. Do you often have to deal with unexpected complications?

   Post-operative
   i. What is the recovery period of hip fracture patients? What are outliers?
   j. What causes the outliers?
   k. Where do patients go after discharge?
   l. How is the communication between the hospital and discharge destination set up? Who is responsible?
   m. Can you describe the follow-up period?

3. Can you describe (if any) the difficulties within the process?

Parameters

4. Which factors are, according to you, associated with length of stay?
5. Which factors are, according to you, associated with discharge destination?

Table E & Table F show a list with factors that are associated with length of stay and discharge destination respectively according to literature.

Table E. Factors associated with length of stay.

Factor associated with LOS	Outcome (shorter LOS)
Treatment	
Surgery vs. medicine	Surgery
(no) blood transfusion	No blood transfusion
Anesthesia mode - Neuraxial vs. general - (no) general	Neuraxial no general
(no) anti-coagulation - warfarin - chloripogrel	No anti-coalgulation
Operation type	
AFHR vs. internal proximal femur locking	AFHR
SHS vs. CCS	CCS
THA vs. HA	HA
Direct lateral vs. direct anterior	Direct anterior
DHS vs. PCCP vs. MSP vs. PFN vs. LISS vs. GN	From shortest to longest stay: PFN - LISS - PFNA - PCCP, GN, PFN - MSP -DHS
PFNA vs. HA	PFNA
DHS vs. external fixation	DHS & external fixation
DHS vs. MSF	MSF
IMN vs. PCCP	PCCP
NCB plates vs. GN	NCB plates
(no) nail/plate/screw fixation	No fixation technique
Patient characteristics	
ASA	ASA<3
AMTS	AMTS>=8
Mobility status	not poor
(no) obese	No obese
(no) COPD	No COPD
(no) dementia	Dementia
(no) Parkinson disease	No Parkinson
(no) diabetes	No diabetes
Age	Lower age
Gender	Female
Charlson comorbidity index	Lower CCI score
Frailty level - low vs. intermediate vs. high	Lower frailty level
(no) problems on admission	No problems on admission
(no) post-surgery complications	No post-ok complications
Hyponatremia vs. normonatremic	Normonatremic

Time to operation	
Early vs. delay	Early operation
Methodology	
(no) test	No tests
Physiotherapy	Intensive physiotherapy
(no) specialist involved	Specialist involved
(no) transfer	No transfer
Orthopedic vs. geriatric ward	Orthopedic wards
Other	
(no) return to theatre	No return to OK
Day of admission	Saturday-Wednesday

AFHR, artificial femoral head replacement; SHS, sliding hip screw; CCS, cannulated cancellous screw; THA, total hip arthroplasty; HA, hemiarthroplasty; DHS, dynamic hip screw; PCCP, percutaneous compression plating; MSP, Medoff sliding plating; PFN, proximal femoral nail; LISS, less invasive stabilization system; GN, gamma nails; PFNA, proximal femoral nail with anti-rotating; MSF, multiple screw fixation; IMN, intramedullary nail; NCB, non-contact bridging; ASA, American Society of Anesthesiologists; AMTS, abbreviated mental test score; COPD, Chronic Obstructive Pulmonary Disease; CCI, Charlson comorbidity index.

Table F. Factors associated with discharge destination.

Factor associated with discharge destination	Discharge destination
Revision procedures, spinal anesthesia mode, bowel management, younger age, higher AMTS, lower incidence of comorbiditites (cardiovascular, cererovascular, COPD, renal, diabetes, parkinson), lower use of medication (e.g. anti-coalgulation), not living alone, higher walking ability, lower need of assistance, intracapsular fracture, mobility score: motor & cognitive (higher), live in rural area, no disability prior to injury	Home
Female, higher BMI, older patients, no follow-up therapy, impaired bowel/bladder function, deficits in self-care, marital status, length of stay, low pre-fracture functional level, intertrochanteric fracture, admitted to ICU, transferred from hospital to trauma hospital	Not home
Increasing age, race (african american), ASA > 3	Skilled nursing facility
Higher age, female, higher ASA, functionally dependent pre-ok	Facility discharge
Older age, greater injury severity score (ISS >26), longer LOS, injury caused by fall	Nursing home
Admitted to ICU, female, greater ISS, unable to walk at discharge, older age, obesity, living alone, more proximal injury, privately insured, not working prior to injury, living in metropolitan city	Rehabilitation centre

AMTS, abbreviated mental test score; COPD, Chronic Obstructive Pulmonary Disease; BMI, body mass index; ICU, intensive care unit; ASA, American Society of Anesthesiologists; ISS, injury severity score; LOS, length of stay.

6. Do you see any remarkable/unexpected parameters? Which one(s)?
7. Can you extend this list with two-three parameters?

Length of stay

8. What do you think is the boundary of short/long post-surgery stay?
9. What is the relevance (if any) to be able to predict the exact length of stay? How many days after surgery?

Patient flow improvement

10. What possibilities do you see in the improvement of patient flow?
11. What are limitations?

Model

12. What should the model(s) be able to predict?

Thank you for your time!

Appendix 8 – variable importance all models

Table G. Overview of importance of top 10 selected variables according to 10 **regression models without feature selection** for predicting length of stay.

Outcome	LOS (short/long)			LOS (short/normal/long)			LOS (continuous)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Age	1	10	**100**	1	10	**100**			
ASA-score	8	8	24	8	9	27			
Fracture type	9	6	12	9	7	14			
General	8	3	9	8	2	6			
Geriatrician	7	8	32	8,5	6	15			
KATZ-adl-1	4	10	**70**	4	10	**70**			
KATZ-adl-2	7,5	2	7	9	1	2			
KATZ-adl-3	5	1	6	7	5	20			
KATZ-adl-5	3	10	**80**	3	10	**80**			
Origin							1	10	**100**
Pre-mobility	8	8	24	10	5	5			
Pre-osteo	10	2	2						
Regional	5	3	18	5	3	18			
SNAQ-appetite	5	7	42	6	7	35			
SNAQ-month	8	2	6	7	1	4			
SNAQ-nutrition	7	2	8	7,5	2	7			
SNAQ-total	6	7	35	6	9	45			
Spinal	10	1	1	8	1	3			
Therapy				9	2	4			
TTO	2	10	**90**	2	10	**90**			

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; osteo, osteoporosis; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation.* Appendix 5 *shows an overview of the interpretation of the predictors.* Appendix 9 *shows the KATZ questionnaire.* Appendix 10 *shows the SNAQ questionnaire.*

Table H. Overview of importance of top 10 selected variables according to 10 **regression models without feature selection** for predicting discharge destination.

Outcome	Discharge destination (home/not home)			Discharge destination (home/care/not home)			Discharge destination (all)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Age	1	10	**100**	1	10	**100**	1	10	**100**
ASA-score	8	4	12						
Fracture type	8	3	9						
General				10	1	1			
Geriatrician	9	3	6						
KATZ-adl-1	3	7	**56**	4	10	**70**	6,5	10	45

KATZ-adl-2	6	1	5	5	10	**60**	4,5	10	**65**
KATZ-adl-3	4	9	**63**	7	5	20	8	7	21
KATZ-adl-4	5	6	36	6	10	**50**	7,5	8	28
KATZ-adl-5	6	5	25	3,5	8	**60**	3	7	**56**
KATZ-adl-6	4	1	7						
KATZ-total	9	7	14				10	2	2
Origin	8	9	27						
Pre-osteo	10	4	4	9	8	16			
Regional				10	5	5			
SNAQ-6months	8	4	12	6,5	4	18	6,5	10	45
SNAQ-appetite	5	6	36	10	1	1	7	9	36
SNAQ-month	4,5	8	**52**	8	9	27	5,5	10	**55**
SNAQ-nutrition				7	5	20	6	5	25
SNAQ-total	6,5	2	9	8	4	12	9,5	2	3
Spinal				9	1	2			
Therapy	10	1	1						
TTO	2	10	**90**	2	9	**81**	2	10	**90**

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; osteo, osteoporosis; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table I. Overview of importance of top 10 selected variables according to 10 **regression models with lasso feature selection** for predicting length of stay.

Outcome	**LOS (short/long)**			**LOS (short/normal/long)**			**LOS (continuous)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side				5	1	6			
Age	1	10	**100**	2,5	10	**85**			
ASA-score				4	7	49			
Dementia				3	3	24			
Fracture type				6	7	35			
General				6	2	10			
Geriatrician	2	3	27	2	10	**90**			
KATZ-adl-6	3	1	8						
KATZ-total	2	10	**90**	6	9	45			
Origin				6	1	5			
Pre-mobility				10	1	1			
Regional				2	1	9			
SNAQ-appetite				1	4	40			
Therapy				5	4	24			

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table J. Overview of importance of top 10 selected variables according to 10 **regression models with lasso feature selection** for predicting discharge destination.

Outcome	Discharge destination (home/not home)			Discharge destination (home/care/not home)			Discharge destination (all)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	7	1	4				6	2	10
Age	1	10	**100**	4,5	10	**65**	5	9	**54**
ASA-score	5	9	**54**	6,5	10	45	6	7	35
Dementia	6,5	10	45	5	9	**54**	2	1	9
Fracture type	4	10	**70**	7	9	36	6	7	35
General							5	9	**54**
Geriatrician	2	10	**90**	2	10	**90**	8	5	15
KATZ-adl-6	7	4	16	3	9	**72**	8	3	9
KATZ-total				4,5	2	13	7,5	6	21
Origin	3,5	10	**75**	4,5	10	**65**	3	9	**72**
Pre-mobility							6	5	25
Pre-osteo	6	4	20	1	4	40	6	4	20
Regional							9	5	10
SNAQ-month				8	10	30	9	1	2
SNAQ-total	7	1	4	3	7	**56**	3	1	8
Spinal				3	1	8	7	3	12
Therapy	6	1	5	6	8	40	7	5	20
TTO							1	6	**60**

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; osteo, osteoporosis; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table K. Overview of importance of top 10 selected variables according to 10 **regression models with correlation-based feature selection** for predicting length of stay.

Outcome	LOS (short/long)			LOS (short/normal/long)			LOS (continuous)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side				5	1	6			
Dementia				4	8	**56**			
Geriatrician	3	10	**80**	3	10	**80**	1	10	**100**
KATZ-adl-1	1	1	10						
KATZ-adl-2	4,5	4	26						
KATZ-adl-3	7	9	36	7	9	36			
KATZ-adl-4	5	7	42						
KATZ-adl-5	2	9	**81**	2	9	**81**			
KATZ-adl-6	6,5	10	45	6	10	**50**			
KATZ-total	6	3	15						

Regional	3,5	10	**75**	5	10	**60**			
SNAQ-total				5	1	6			
TTO	1	10	**100**	1	10	**100**			

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table L. Overview of importance of top 10 selected variables according to 10 **regression models with correlation-based feature selection** for predicting discharge destination.

Outcome	**Discharge destination (home/not home)**			**Discharge destination (home/care/not home)**			**Discharge destination (all)**		
Predictor	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	6,5	10	45	7	9	36	7,5	4	14
Age	1	10	**100**	1	10	**100**	1	2	20
Dementia	9	8	16	9	8	16	8	10	30
Geriatrician	4	10	**70**	6	10	**50**	5	10	**60**
KATZ-adl-2							5,5	2	11
KATZ-adl-3				7,5	2	7	3,5	10	**75**
KATZ-adl-5				4	5	35	3	10	**80**
KATZ-adl-6	8	8	24	8	6	18	9	9	18
Origin	3	10	**80**	4,5	10	**65**	6	2	10
Pre-mobility				8	5	15			
Regional	6	10	**50**	3	10	**80**	2,5	10	**85**
SNAQ-nutrition	9	1	2						
SNAQ-total	5	10	**60**	7	10	40	6,5	10	45
TTO	2	10	**90**	2	10	**90**	1	10	**100**

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table M. Overview of importance of top 10 selected variables according to 10 **lasso regression models with lasso feature selection** for predicting length of stay.

Outcome	**LOS (short/long)**			**LOS (short/normal/long)**			**LOS (continuous)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side				4	3	21			
Age	1	10	**100**	5	9	**54**			
ASA-score	3	2	16	2	5	45			
Dementia				6,5	4	18			
Fracture type				4	6	42			
General				7	1	4			
Geriatrician	2	5	45	1	1	10			
KATZ-adl-5				2	1	9			
KATZ-total	2	8	**72**	2	7	**63**			

Origin				5	1	6			
Pre-mobility				10	1	1			
SNAQ-appetite				7	3	12			
Spinal	5	1	6	5	1	6			
Therapy	4	1	7	6	3	15			

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table N. Overview of importance of top 10 selected variables according to 10 **lasso regression models with lasso feature selection** for predicting discharge destination.

Outcome	**Discharge destination (home/not home)**			**Discharge destination (home/care/not home)**			**Discharge destination (all)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side							9	1	2
Age	1	10	**100**	1	10	**100**	10	1	1
ASA-score	5	10	**60**	4	10	**70**	5	8	48
Dementia	7	10	40	2	10	**90**	1	8	**80**
Fracture type	3	10	**80**	6	8	40	10	5	5
General							3	6	48
Geriatrician	2	10	**90**	4,5	10	**65**	9	4	8
KATZ-adl-6	7	5	20	6	5	25	7	3	12
KATZ-total							8,5	4	10
Origin	4	10	**70**	3,5	10	**75**	2	7	**63**
Pre-mobility							6,5	4	18
Pre-osteo	6	5	25	7,5	6	21	7	5	20
Regional							7,5	4	14
SNAQ-month							6,5	2	9
SNAQ-total	7	1	4	8	3	9	5	4	24
Spinal							8	3	9
Therapy	7	2	8	7	7	28	5	8	48
TTO							1	3	30

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; osteo, osteoporosis; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table O. Overview of importance of top 10 selected variables according to 10 **random forest models without feature selection** for predicting length of stay.

Outcome	LOS (short/long)			LOS (short/normal/long)			LOS (continuous)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	9	1	2	10	7	7	9	4	8
Age	1	10	**100**	1	10	**100**	2	9	**81**
ASA-score	6	10	**50**	5	10	**60**	5	9	**54**
Dementia	10	1	1	10	3	3	10	3	3
Fracture type	2,5	10	**85**	2	10	**90**	3	9	**72**
General							6	1	5
Geriatrician	3	10	**80**	5	10	**60**	1	9	**90**
KATZ-adl-1							10	3	3
KATZ-adl-2							9	1	2
KATZ-adl-3							5	1	6
KATZ-adl-4							3	1	8
KATZ-adl-5							8	1	3
KATZ-adl-6	10	8	8						
KATZ-total	3,5	10	**75**	6	10	**50**	6	9	45
Origin	8	10	30	9	10	20	7	9	36
Pre-mobility	8	9	27	8	10	30	4	9	**63**
SNAQ-appetite							2	1	9
SNAQ-month							7	1	4
SNAQ-total	10	1	1				9	5	10
Spinal							4	1	7
Therapy	5	10	**60**	4,5	10	**65**	9	5	10
TTO	7	10	40	3,5	10	**75**	6	9	45

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]; *SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table P. Overview of importance of top 10 selected variables according to 10 **random forest models without feature selection** for predicting discharge destination.

Outcome	Discharge destination (home/not home)			Discharge destination (home/care/not home)			Discharge destination (all)		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	10	1	1						
Age	1	10	**100**	1	10	**100**	1	10	**100**
ASA-score	6,5	10	45	7,5	10	35	10	10	10
Dementia	10	9	9	3	10	**80**	4	10	**70**
Fracture type	4	10	**70**	3	10	**80**	3	10	**80**
Geriatrician	2	10	**90**	4	10	**70**	5	10	**60**

KATZ-total	7	10	40	7	10	40	8	10	30
Origin	3	10	**80**	9	10	20	2	10	**90**
Pre-mobility	9	10	20	8	10	30	8	10	30
Therapy	5	10	**60**	5	10	**60**	6,5	10	45
TTO	8	10	30	9	10	20	7	10	40

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire.*

Table Q. Overview of importance of top 10 selected variables according to 10 **random forest models with lasso feature selection** for predicting length of stay.

Outcome	**LOS (short/long)**			**LOS (short/normal/long)**			**LOS (continuous)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side				6	1	5			
Age	1	10	**100**	1	10	**100**			
ASA-score	4	2	14	4	7	49			
Dementia				6	3	15			
Fracture type	3	1	8	2,5	6	**51**			
General				8	1	3			
Geriatrician	3	5	40	4	10	**70**			
KATZ-adl-6	3	1	8						
KATZ-total	2	9	**81**	2	10	**90**			
Pre-mobility				6	1	5			
SNAQ-appetite				9	1	2			
Therapy				6	2	10			

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; SNAQ, Short Nutritional Assessment Questionnaire. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table R. Overview of importance of top 10 selected variables according to 10 r**andom forest models with lasso feature selection** for predicting discharge destination.

Outcome	**Discharge destination (home/not home)**			**Discharge destination (home/care/not home)**			**Discharge destination (all)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Age	1	10	**100**	1	10	**100**	1	10	**100**
ASA-score	5	9	**54**				10	9	9
Dementia	6	10	**50**	2	10	**90**	5	10	**60**
Fracture type	4	10	**70**	3	9	**72**	4	10	**70**
General				6	10	**50**	8	1	3
Geriatrician	2	10	**90**	4	10	**70**	5,5	10	**55**
KATZ-adl-6	8	3	9	8	7	21	6,5	2	9
KATZ-total				4	2	14	3	8	**64**

Origin	3	10	80	6	10	50	2	10	90
pre-mobility							8	7	21
Pre-osteo	7	5	20	8,5	6	15			
SNAQ-total	7	1	4	9	5	10	10	2	2
Therapy	4	3	21	5	7	42	9	7	14
TTO							7	8	32

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; ASA, American Society for Anesthesiologists; general, general anesthesia; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; osteo, osteoporosis; SNAQ, Short Nutritional Assessment Questionnaire; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table S. Overview of importance of top 10 selected variables according to 10 r**andom forest models with correlation-based feature selection** for predicting length of stay.

Outcome	**LOS (short/long)**			**LOS (short/normal/long)**			**LOS (continuous)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side				5	5	30	5	4	24
Dementia	5	1	6	3	8	64	3	8	64
Geriatrician	1	10	100	1	10	100	1	10	100
KATZ-adl-1	9	1	2						
KATZ-adl-2	7	6	24						
KATZ-adl-3	5	9	54	6	9	45	4	8	56
KATZ-adl-4	5	7	42	5	2	12			
KATZ-adl-5	6	7	35	6	9	45	6	5	25
KATZ-adl-6	3	10	80	4	10	70	5	9	54
KATZ-total	2	3	27						
Regional	7	10	40	8	9	27	7	9	36
SNAQ-total							4,5	4	26
Spinal				6	1	5	7	3	12
TTO	3	10	80	2	10	90	2	10	90

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; spinal, spinal anesthesia; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Table T. Overview of importance of top 10 selected variables according to 10 r**andom forest models with correlation-based feature selection** for predicting discharge destination.

Outcome	**Discharge destination (home/not home)**			**Discharge destination (home/care/not home)**			**Discharge destination (all)**		
Predictors	Rank	Freq.	Imp.	Rank	Freq.	Imp.	Rank	Freq.	Imp.
Affected side	7,5	10	35	7,5	10	35	7	6	24
Age	1	10	100	1	10	100			
Dementia	5	10	60	2	10	90	1	10	100
Geriatrician	2	10	90	3	10	80	2	10	90

KATZ-adl-2							5,5	2	11
KATZ-adl-3				8	1	3	5	7	42
KATZ-adl-5				8	7	21	6	9	45
KATZ-adl-6	7,5	10	35	9	9	18	7	10	40
Origin	3	10	**80**	4,5	10	**65**			
Pre-mobility				4	4	28			
Regional	9	10	20	10	7	7	8,5	10	25
SNAQ-nutrition							8	2	6
SNAQ-total	6	10	**50**	6,5	10	45	4	10	**70**
TTO	4	10	**70**	5	10	**60**	3,5	10	**75**

*Rank, median ranking of variable; freq., frequency of variable in top 10; imp., weighted importance = (11 - rank) * freq.; KATZ-adl, Index of Independence in Activities of Daily Living* [106]*; regional, regional anesthesia; SNAQ, Short Nutritional Assessment Questionnaire; TTO, time-to-operation. Appendix 5 shows an overview of the interpretation of the predictors. Appendix 9 shows the KATZ questionnaire. Appendix 10 shows the SNAQ questionnaire.*

Appendix 9 – KATZ

Katz-schaal voor BASALE activiteiten van het dagelijkse leven (ADL)

Definities

De ADL-index is gebaseerd op de beoordeling van de functionele (on)afhankelijkheid van personen voor zich wassen, zich kleden, WC-bezoek, zich verplaatsen binnenhuis, continentie en zich voeden. Specifieke definities van (on)afhankelijkheid zijn als volgt te interpreteren: onafhankelijkheid betekent het uitvoeren van de functie zonder supervisie, of verbale of fysieke ondersteuning van derden, tenzij het uitdrukkelijk wordt gespecificeerd. De beoordeling wordt gebaseerd op de actuele status en niet op de mogelijkheden van de persoon. Dit betekent dat een persoon die weigert om een opdracht uit te voeren, beschouwd wordt als afhankelijk voor deze opdracht, ook al vermoedt men dat hij er wel toe in staat is.

WASSEN (gewoon wassen of het nemen van een douche of bad)	☐ Onafhankelijk: wast zichzelf volledig onafhankelijk of wordt slechts geholpen voor één onderdeel (bv. het wassen van de rug of een gehandicapt lichaamsdeel). ☐ Afhankelijk: heeft hulp nodig bij het wassen van meer dan één lichaamsdeel; heeft hulp nodig om in en uit het bad te komen of wast zichzelf helemaal niet.
KLEDEN	☐ Onafhankelijk: neemt zelf de kledingsstukken uit de kast of lade, kleedt zichzelf aan en kan losse kledingsstukken zonder problemen aandoen. Het vastbinden van de schoenveters wordt niet beoordeeld. ☐ Afhankelijk: kleedt zichzelf niet aan of slaagt er slechts in om zich gedeeltelijk aan te kleden.
WC-BEZOEK	☐ Onafhankelijk: kan zich zonder hulp verplaatsen naar of van het toilet, zich neerzetten en rechtkomen van het toilet en zichzelf reinigen. Hij kan eventueel zelfstandig 's nachts een bedpan of urinaal gebruiken. Gebruik van mechanische hulpmiddelen is toegelaten. ☐ Afhankelijk: De patiënt kan zich niet zonder gedeeltelijke hulp (stimulatie, controle) van derden verplaatsen naar of van het toilet en/of zich neerzetten en rechtkomen van het toilet. Hij heeft hulp nodig bij het gebruik van een bedpan.
VERPLAATSEN BINNENSHUIS	☐ Onafhankelijk: kan zich volledig zelfstandig in en uit een bed of een fauteuil verplaatsen (mag hiervoor mechanische hulpmiddelen gebruiken). ☐ Afhankelijk: heeft hulp nodig om in en uit een bed of fauteuil te komen; doet geen zelfstandige verplaatsingen.
CONTINENTIE	☐ Onafhankelijk: De patiënt heeft geen enkel probleem, noch voor urine noch voor feces. ☐ Afhankelijk: er bestaat een gedeeltelijke of volledige incontinentie voor urine of feces. Voor partiële of volledige controle van de continentie worden lavementen, katheters of een bedpan/urinaal gebruikt.
VOEDEN	☐ Onafhankelijk: neemt het voedsel zelf van het bord en eet zelfstandig. Voorbereidende handelingen zoals het snijden van vlees of het boteren van brood worden niet geëvalueerd. ☐ Afhankelijk: heeft hulp nodig bij de voeding of moet gevoed worden (ook artificieel).

Score		
	A	*Onafhankelijk voor de 6 items*
	B	*Afhankelijk voor 1 van de 6 items*
	C	*Afhankelijk voor wassen en 1 bijkomend item*
	D	*Afhankelijk voor wassen, kleden en 1 bijkomend item*
	E	*Afhankelijk voor wassen, kleden, WC-bezoek en 1 bijkomend item*
	F	*Afhankelijk voor wassen, kleden, WC-bezoek, verplaatsen en 1 bijkomend item*
	G	*Afhankelijk voor alle items*
	Ander	*Afhankelijk voor ten minste 2 functies maar niet te klasseren als C, D, E of F*

Appendix 10 – SNAQ

SNAQ

Short Nutritional Assessment Questionnaire

• **Did you lose weight unintentionally?**	
More than 6 kg in the last 6 months	●●●
More than 3 kg in the last month	●●
• **Did you experience a decreased appetite over the last month?**	●
• **Did you use supplemental drinks or tube feeding over the last month?**	●

● no intervention

●● moderately malnourished; nutritional intervention

●●● severely malnourished; nutritional intervention and treatment dietician