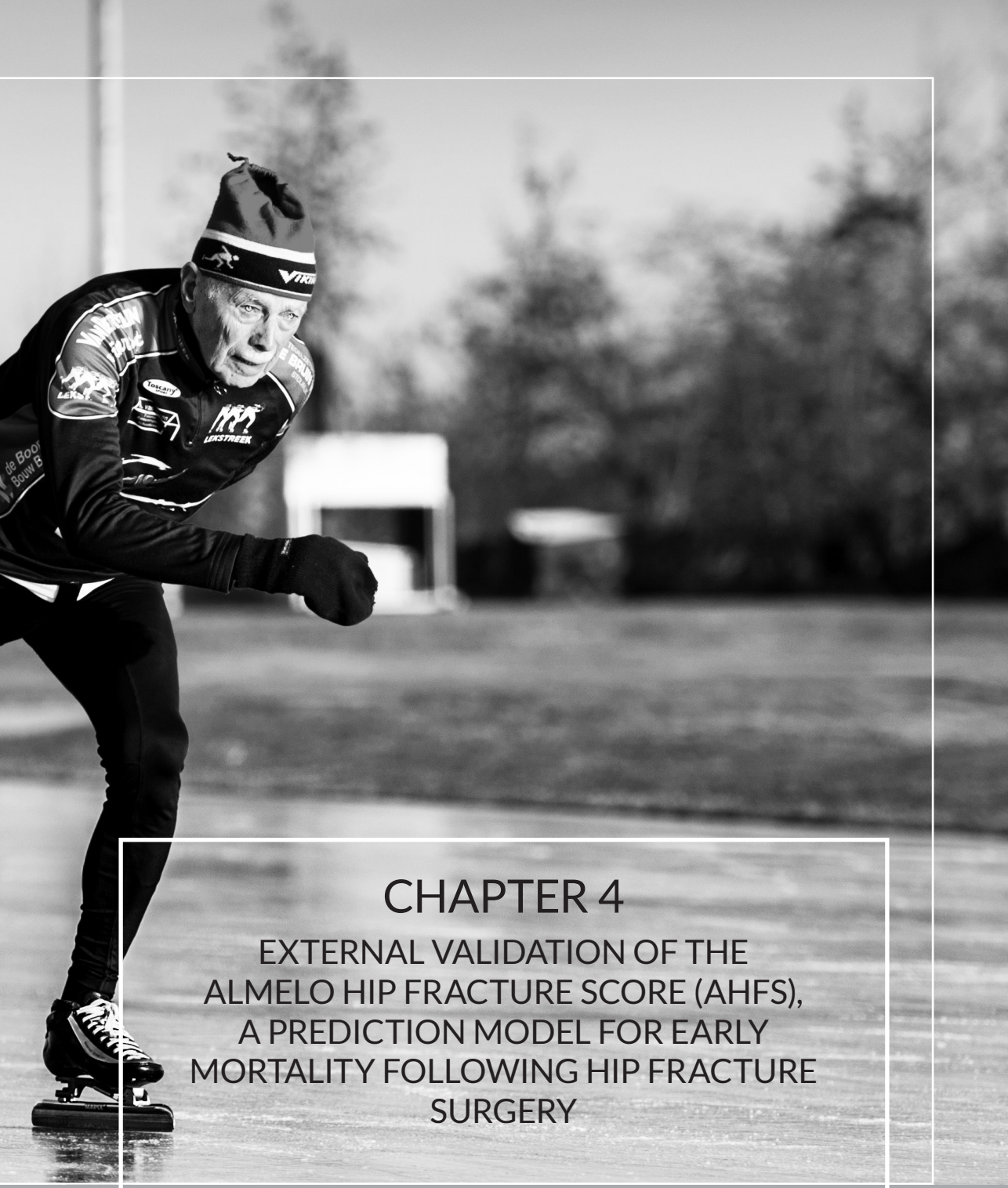


CHAPTER 4

EXTERNAL VALIDATION OF THE ALMELO HIP FRACTURE SCORE (AHFS), A PREDICTION MODEL FOR EARLY MORTALITY FOLLOWING HIP FRACTURE SURGERY

Sophie Moerman, Tim A.Wesdorp, Anne J.H. Vochteloo, Nina M.C. Mathijssen

Submitted

Abstract

Introduction

Early mortality (<30 days) in hip fracture patients is as high as 10%. Several risk assessment tools have been developed to identify patients at high risk for early mortality. Among them the Almelo Hip Fracture Score (AHFS), that was developed recently and showed promising results. Up to date, this tool has not been validated, therefore we aim to perform an external validation of the AHFS.

Method

An external validation of the AHFS was conducted in a cohort of hip fracture patients (Delft cohort). Data was prospectively collected during admission. The AHFS score was calculated for all patients over 70 years of age admitted with a hip fracture in our hospital. The characteristics of the Delft Cohort, sensitivity, specificity, positive predictive value, negative predictive and area under the curve were calculated and compared to the original Almelo cohort.

Results

422 patients of 70 years and older were included. Mortality within 30 days was 7.6%. For the high-risk cut-off point specificity was 95.4% and sensitivity was 28.1%. Specificity in the Almelo cohort was 92.5% and sensitivity 42.2%. The area under the ROC curve was 0.70 (95% CI 0.60 – 0.79).

Conclusion

This external validation showed that the AHFS was acceptable and comparable to the values in the Almelo cohort. We think that this score can be used to identify patients at high risk for early mortality.

Introduction

Hip fractures are very common among elderly, every year 1.6 million people are affected worldwide. With the current urbanization and ageing of the world's population this number is expected to grow to an even bigger number. Estimations vary between 7.3 and 21.3 million by 2050. [1, 2] Fractures of the hip have a serious effect on mortality and morbidity. The morbidity of hip fractures leads to 2.9 million disability adjusted life years in the world [3]. The mortality rate within the first 30 days is 10% and reaches up to 33% one year after surgery [4]. To be able to decrease (early) mortality it is important to correctly identify patients at high risk of early mortality.

Throughout the years, several risk assessment tools have been developed to predict early mortality [5]. The Association of Anesthetists of Great Britain and Ireland have developed the Nottingham Hip Fracture Score (NHFS), which is known as the most optimal screening tool so far [5, 6]. However, it has limited discriminative power [7]. The NHFS contains the abbreviated mental test score (AMTS), which can be time consuming and challenging to obtain in an emergency setting [8]. Therefore, the Almelo Hip Fracture Score (AHFS) was developed in the Netherlands [9]. The first results of this prediction model are promising and suggest higher specificity and sensitivity in assessing early mortality among the elderly with a hip fracture [9]. The AHFS is a more extensive and faster to obtain risk model than the NHFS. The score has two important extra variables to the original NHFS; the Parker Mobility Score and the ASA classification. The variable AMTS was replaced by cognitive frailty (yes/no), to make it easier and faster to obtain. The aim of this study is to perform an external validation of the AHFS.

Methods

Patients

We conducted a prospective observational cohort study of 525 hip fracture patients. The study did not fall under the scope of the medical research with human subject act (WMO), therefore ethical approval was not required. Information of our study for patients, or family members, was provided in a binder specially designed for hip fracture patients in our hospital. (10) All hip fracture patients were admitted to the emergency department of a 450-bed teaching hospital (Delft, the Netherlands) between January 2008 and December 2009. Patients with a fracture due to a high-energy trauma or with a pathologic fracture and patients younger than 70 years (n=103) were excluded. This cohort will be named the 'Delft cohort'. Length of follow-up was at least 3 months or until death occurred.

The AHFS was developed using data of 850 hip fracture patients aged 70 years and older admitted to the Trauma Surgery department at Hospital Group Twente (ZGT) between April 1, 2008, and October 23, 2013(10). Patients with a pathological or periprosthetic fracture were excluded, as well as patients who were referred to the orthopedic department, due to an indication for total hip replacement, and patients who died preoperatively. This cohort will be named the Almelo cohort.

Data collection

Data were collected uniformly and recorded by a standard evaluation form on the date of admission. [10]. Demographic data prospectively collected were age and sex. On admission the following clinical characteristics were prospectively obtained: serum haemoglobin measured in gram /decilitre, cognitive frailty defined as dementia (diagnosed by neurologist or geriatrician), cognitive disorders or delirium on admission. Living in an institution prior to the fracture was based on history taking. Retrospectively the hospital information system was checked to obtain data on the number of comorbidities and history of malignancy at the time of admission. The Parker Mobility Score (PMS) represents the level of mobility before fracture [11]. The total score ranges from 0 (not able to walk) to 9 (fully independent). The American Society of Anaesthesiologists Physical Status classification (ASA) was determined by the anaesthesiologist prior to surgery (range 1-4) [12]. Outcome was measured as mortality within 30 days following the hip fracture, this was defined as early mortality [9]. Mortality data was obtained from population registers of the counties as well as the hospital's information system at 30 days after hip fracture. Survival is considered as survival of 30 days or longer after hip fracture.

The AHFS score assesses the risk of early mortality following hip fracture in patients aged ≥70 years. The risk model consists of 9 variables: age, sex, admission serum haemoglobin, cognitive frailty, living in an institution, numbers of comorbidities, malignancy, Parker Mobility Score and ASA score (table 1). Between 0 and 7 points are scored for each variable based on the rounded-up beta coefficients associated with the original multivariate logistic regression. Data was transformed into a simple score ranging from 3-19 to predict the risk of early mortality (Table 1). The developers of the AHFS used a cut off score ≥ 13 to identify the high risk of early mortality group [9]. A cut-off point of AHFS ≤ 9 was set for the low risk group.

Table 1. Risk score form of the Almelo Hip Fracture Score

Age	≤86 years	4 points
	70-85 years	3 points
Sex	Male	1 point
	Female	0 points
Admission serum hemoglobin	≤ 10 g/dl	1 point
	> 10 g/dl	0 points
Cognitive frailty	Yes	1 point
	No	0 points
Living in an institution	Yes	1 point
	No	0 points
Number of comorbidities	≥2	1 point
	< 2	0 points
Malignancy	Yes	1 point
	No	0 points
Parker Mobility Score	≤ 5	2 points
	> 5	0 points
ASA Score	1-2	0 points
	3	3 points
	4	7 points
Sum of points		… points

Statistical analysis

Baseline characteristics in each group were collected and tested for normality using a Shapiro-Wilk test. Continuous data are shown as median with the interquartile range (IQR), in case of a non-parametric distribution. Categorical data are presented as the absolute number of subjects in each group, along with the percentages. Differences in non-parametric distributed continuous data between groups were assessed using a Mann-Whitney U-test. Categorical data were analyzed using a Chi-square or Fisher's exact test. Sensitivity analysis was performed using the original cut-off points for the AHFS of 9 and 13 (9). There was no missing data in this cohort. A receiver operating characteristics (ROC) curve was plotted by the sensitivity versus the 1-specificity. The area under the curve (AUC) was measured and the Hosmer-Lemeshow test was performed to assess the overall calibration error. All statistical analyses included two tailed tests. A P-value of 0.05 was considered to indicate statistical significance. SPSS statistics package version 24.0 for Mac (SPSS Inc., Chicago, IL, USA) was used for all analyses.

Results

A total of 422 patients were included. Baseline characteristics are described in table 2. The median age was 84.3 (IQR 79.3-89.0; range 70.3 – 101.0) years,

75.4% were female. 7.6% patients (n=32) died within 30 days after the hip fracture. Patients in the early mortality group were older (86.9 years vs. 84.1 years, *p*=0.036) and physically frail; they had a higher ASA classification, more comorbidities (>2 comorbidities 84.4% vs. 57.2%, *p*=0.002) and lower mobility scores (PMS< 5 75,0% vs. 50,0%, *p*=0.009). Less patients were institutionalized before admission in the early mortality group (21.9%) compared to the survival group (45.1%) *p*=0.015.

Table 2 Baseline characteristics

Characteristics	**All patients Delft (n=422)**	***All patients Almelo (n=850)***	**Early mortality Delft group (n=32)**	**Survival group Delft (n=390)**
Age at time of admission (years)	84.3 (79.3-89.0)	*83.0*	86.9 (81.2-93.4)	84.1(79.2-88.8)*
AHFS				
Age ≥ 86 years	181 (42.9%)	*323 (38.0%)*	16 (50%)	165 (42.3%)
Sex (male)	104 (24.6%)	*224 (26.4%)*	6 (18,8%)	98 (25.1%)
Serum haemoglobin < 10 g/dl	31 (7.3%)	*52 (6.1%)*	3 (9.4%)	28 (7.2%)
Cognitive frailty	112 (26.5%)	*293 (34.5%)*	11 (34.4%)	101 (25.9%)
Living in an institution	183 (43.4%)	*249 (29.3%)*	7 (21.9%)	176 (45.1%) *
≥ 2 comorbidities	250 (59.2%)	*450 (52.9%)*	27 (84.4%)	223 (57.2%) *
History of malignancy	64 (15.2%)	*207 (24.4%)*	5 (15.6%)	59 (15.1%)
PMS ≤ 5	219 (51.9%)	*376 (44.2%)*	24 (75%)	195 (50.0%) *
ASA 1-2	269 (63.7%)	*184 (21.7%*	11 (34.4%)	258 (66.2%)
ASA 3	128 (30.3%)	*553 (65.1%)*	13 (40.6%)	115 (29.5%)
ASA 4	25 (5.9%)	*113 (13.3%)*	8 (25.0%)	17 (4.4%)
Total AHFS points	7 (5-10)		9.50 (7-13.8)	7 (5-9)*
Mortality	32 (7.6%)	*64 (7.5%)*		
Fracture type				
Femoral neck	241 (57.1%)	*443 (52.1%)*	17 (53.2%)	224 (57.4%)
Trochanteric	163 (38.7%)	*369 (43.4%)*	13 (40.7%)	150 (38.5%)
Subtrochanteric	18 (4.3%)	*38 (4.5%)*	2 (6.3%)	16 (4.1%)
Treatment				
Osteosynthesis	229 (54.3%)		16 (50.0%)	213 (54.6%)
Arthroplasty	181 (42.9%)		11 (34.4%)	170 (43.6%)
Conservative	12 (2.8%)		5 (15.6%)	7 (1.8%)

Continuous data are presented as median (interquartile range) and categorical data are described as frequency (percentage).

* Statistically significant difference between early mortality and survival group p <0.05

The AHFS was calculated for the Delft cohort, the results are shown in table 2. The median AHFS score was higher in the early mortality group 9.5 (IQR 7.0-13.8) compared to the survival group 7 (IQR 5-9) $p<0.001$. Table 3 and 4 show the results of the validity analysis of the high risk cut off point of 9 and lower and 13 and higher. Applying the low cut-off point the sensitivity was comparable (75.9% vs 78.1%) but specificity was lower (50.0% vs 72.5%) to the values found in the Almelo cohort. In the high cut-off point sensitivity (28.1 vs. 42.2%) was lower, but specificity (95.4 vs. 92.5%) was comparable to the Almelo cohort.

Table 3 mortality and survival in the different risk groups according to the AHFS

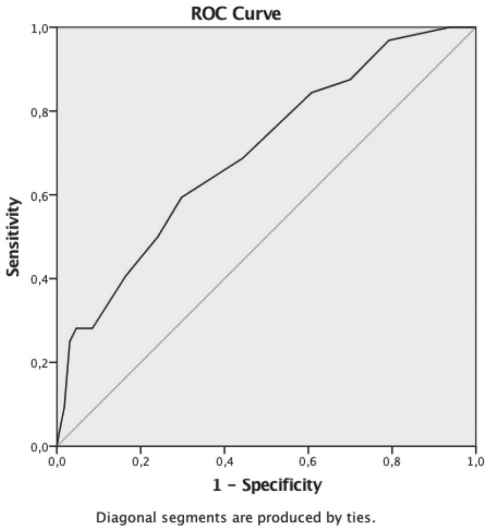
	Survived	**Died within 30 days**	**Total**
AHFS low risk (≤ 9)	296	16	312
AHFS intermediate risk (10 -12)	76	7	83
AHFS high risk (≥13)	18	9	27
Total	390	32	422

Table 4 Results of the validity analysis of the high risk AHFS for early mortality for the cut-off point of 9 and lower and 13 and higher.

	AHFS low risk ≤ 9		**AHFS high risk ≥13**	
Characteristics	Delft cohort	Almelo cohort	Delft cohort	Almelo cohort
Sensitivity	75.9%	78.1%	28.1%	42.2%
Specificity	50.0%	72.5%	95.4%	92.5%
Positive predictive value	14.5%	18.8%	33.3%	31.4%
Negative predictive value	94.9%	97.6%	94.2%	95.2%
Likelihood ratio Positive	1.51	2.35	6.11	5.62
Likelihood ratio Negative	1.84	0.33	0.71	0.62
Correlation with early mortality			0.254	n/a

The area under the ROC curves of the AHFS in the Delft cohort was 0.70 (95% CI 0.60-0.79) (**Figure 1**). The AHFS model showed a good fit between predicted and observed values (Hosmer-Lemeshow test, $p>0.76$).

Figure 1. The curved line shows the ROC curve of the Delft cohort. The diagonal indicate results no better than chance.



Discussion

The aim of this study was the external validation of the AHFS. We demonstrated that the validity of the score was generally comparable to the values found in the Almelo cohort. Therefore, we think that this score could be used to identify patients at high risk for early mortality within 30-days after hip fracture.

A high positive predictive value and high sensitivity are the most important features of a prediction model to correctly identify patients at risk. In the Delft cohort the positive predictive value was higher than in the Almelo cohort (33.3% vs. 31.4%). Sensitivity for the high cut-off point was higher in the Almelo cohort (42.2% vs. 28.1%). Specificity is the ability to give negative results in negative cases. Specificity in the Delft cohort was similar to the Almelo cohort (95.4% vs. 92.5%). A high negative predictive value and high specificity are important to identify patients not at risk. This is important since we expect that low risk patients might not receive the close monitoring compared to the high-risk patient. The negative predictive value for the low cut-off point was comparable to the Almelo cohort (94.9% vs 97.6%). In the Delft cohort specificity was larger than in the Almelo cohort (75.9% vs 72.5%). Together these items resulted in a lower area under the curve in the Delft cohort than in the Almelo cohort (0.70 vs 0.80). In the Delft cohort more patients were classified ASA I and II than in the Almelo cohort (63.7% vs. 22%), this resulted in a lower overall AHFS score. The observed ASA I and II classification

is considerably high in the Delft cohort compared to large international studies that showed a prevalence of 39-43% [13-15] and a large study conducted in the Netherlands showed that 40% of the hip fracture patients were classified as ASA I or II [16]. Therefore, the current low prevalence of ASA I and II patients could possibly affect the correct prediction with the original cut-off points used [9]. Despite these differences and consecutive possible underestimation, the current AUC is still considered to be acceptable [9, 17].

The mortality rate after hip fracture in elderly patients ranges between 10% at 30 days up to 33% after one year [4]. These poor outcomes highlight the importance for a valuable mortality prediction model after hip fracture surgery. First of all, to target care to the ones who need it the most. Throughout the years successful methods to decrease mortality have been published. The orthogeriatric care model is an indicative example, in this model a patient is treated using a multidisciplinary approach involving both an orthopaedic surgeon and a geriatrician [18]. In the orthogeriatric care model a 40% mortality decrease was observed in the first 30 days after hip fracture. (6) Methods to identify patients who benefit most from these expensive treatment regimens are scarce and risk scores like the AHFS can be of important use. More accurate risk scores are also helpful in shared decision making as well as patient- and family education. Recently a study was started to guide very frail patients (ASA 4-5 or not able to ambulate independently) and their family in the choice between surgery and palliative treatment. (www.frail-hip.nl) In order to guide this treatment decision, it is important to have tools to predict outcome on an individual patient level, based on larger population studies. The AHFS might be such tool to help clinicians guide this process.

Throughout the last years there is a growing number of proposed prediction models. Moreover, these prediction models are increasingly used in clinical guidelines [19]. Before prediction models can be used in clinical practice it is important that its predictive performance is empirically evaluated in a dataset that is different to the one used to develop the model (external validation) [20, 21]. As far as we are concerned, we are the first to externally validate the AHFS prediction model. A large external validation study showed that five mortality prediction models (amongst those NHFS) had an acceptable AUC in a hip fracture patient cohort [5]. However, this study reported large amounts of missing data and thereby changes made in the original models, which was due to its retrospective nature [5]. The AUC in the five studies varied between 0.69 and 0.77 and was comparable to what we observed. This implicates that that the AHFS is a valuable

risk assessment tool. A smaller external validation study of six different models found that none of the tested models had excellent discriminative power, defined as AUC > 0.80, with AUCs ranging from 0.43 to 0.68 [22]. The development cohort of the AHFS, did show an excellent discriminative power with an AUC of 0.82. Karres *et al* concluded that so far the NHFS showed the most promising results with reasonable discrimination [5]. A negative point however, is the use of the AMT which can be complicated and time consuming among cognitive impaired elderly people. By contrast, the AHFS is a more extensive but easier to obtain risk score using only cognitive frailty as a variable. Therefore, in the clinical setting the AHFS might not only be one of the best discriminative, but also easy and quicker to obtain than the current ones available.

The present study has some limitations that should be considered. First, the number of patients included in this cohort. Although 422 is a considerable number, 33 events (mortality within 39 days) is small to perform an external validation [23]. However, the reported mortality rates between the original and validation cohort was similar. Second, the variables comorbidities and history of malignancy were retrospectively collected. Nevertheless, there was no missing data on these variables after extensive investigation in the electronic patient information system. Moreover, all other patient characteristics were collected prospectively resulting in absolute completeness of data.

In conclusion, we showed that the AHFS can predict mortality as accurate as was suggested by Nijmeijer et al.[9]. The AHFS is a reliable, feasible and easy-to-use instrument to predict 30-day mortality after hip fracture surgery. A better prediction of patients at risk may be beneficial in reducing the high early mortality rate after hip fracture surgery. Moreover, the model may also be useful in managing expectations and education of hip fracture patients- and their families.

References

1. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 1997;7(5):407-13.
2. Johnell O, Kanis JA. An estimate of the worldwide prevalence and disability associated with osteoporotic fractures. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2006;17(12):1726-33.
3. Johnell O, Kanis JA. An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. Osteoporosis international : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA. 2004;15(11):897-902.

4. Roche JJ, Wenn RT, Sahota O, Moran CG. Effect of comorbidities and postoperative complications on mortality after hip fracture in elderly people: prospective observational cohort study. BMJ (Clinical research ed). 2005;331(7529):1374.
5. Karres J, Heesakkers NA, Ultee JM, Vrouenraets BC. Predicting 30-day mortality following hip fracture surgery: evaluation of six risk prediction models. Injury. 2015;46(2):371-7.
6. Maxwell MJ, Moran CG, Moppett IK. Development and validation of a preoperative scoring system to predict 30 day mortality in patients undergoing hip fracture surgery. British journal of anaesthesia. 2008;101(4):511-7.
7. Marufu TC, White SM, Griffiths R, Moonesinghe SR, Moppett IK. Prediction of 30-day mortality after hip fracture surgery by the Nottingham Hip Fracture Score and the Surgical Outcome Risk Tool. Anaesthesia. 2016;71(5):515-21.
8. Hodkinson HM. Evaluation of a mental test score for assessment of mental impairment in the elderly. Age and ageing. 1972;1(4):233-8.
9. Nijmeijer WS, Folbert EC, Vermeer M, Slaets JP, Hegeman JH. Prediction of early mortality following hip fracture surgery in frail elderly: The Almelo Hip Fracture Score (AHFS). Injury. 2016;47(10):2138-43.
10. Vochteloo AJH ND, Cornelisse H, De Vries MR, Bloem RM, Pilot P. Voor elke heup een rode map. . Medisch contact. 2009(4):158-62.
11. Parker MJ, Palmer CR. A new mobility score for predicting mortality after hip fracture. The Journal of bone and joint surgery British volume. 1993;75(5):797-8.
12. Owens WD, Felts JA, Spitznagel EL, Jr. ASA physical status classifications: a study of consistency of ratings. Anesthesiology. 1978;49(4):239-43.
13. Gjertsen JE, Fenstad AM, Leonardsson O, Engesaeter LB, Karrholm J, Furnes O, et al. Hemiarthroplasties after hip fractures in Norway and Sweden: a collaboration between the Norwegian and Swedish national registries. Hip international : the journal of clinical and experimental research on hip pathology and therapy. 2014;24(3):223-30.
14. Gremillet C, Jakobsson JG. Acute hip fracture surgery anaesthetic technique and 30-day mortality in Sweden 2016 and 2017: A retrospective register study. F1000Research. 2018;7:1009.
15. Ruths S, Baste V, Bakken MS, Engesaeter LB, Lie SA, Haugland S. Municipal resources and patient outcomes through the first year after a hip fracture. BMC health services research. 2017;17(1):144.
16. Moerman S, Mathijssen NMC, Tuinebreijer WE, Vochteloo AJH, Nelissen R. 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 orthopaedica. 2018;89(5):509-14.
17. Hosmer DW LS, Sturdivant RX. Applied logistic regression. Hoboken, NJ, USA: John Wiley & Sons, Inc. ; 2000.
18. Grigoryan KV, Javedan H, Rudolph JL. Orthogeriatric care models and outcomes in hip fracture patients: a systematic review and meta-analysis. Journal of orthopaedic trauma. 2014;28(3):e49-55.
19. Rabar S, Lau R, O'Flynn N, Li L, Barry P. Risk assessment of fragility fractures: summary of NICE guidance. BMJ (Clinical research ed). 2012;345:e3698.
20. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ (Clinical research ed). 2009;338:b605.
21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ (Clinical research ed). 2015;350:g7594.
22. Burgos E, Gomez-Arnau JI, Diez R, Munoz L, Fernandez-Guisasola J, Garcia del Valle S. Predictive value of six risk scores for outcome after surgical repair of hip fracture in elderly patients. Acta anaesthesiologica Scandinavica. 2008;52(1):125-31.
23. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. Journal of clinical epidemiology. 2005;58(5):475-83.