Injury, Int. J. Care Injured 46 (2015) 371–377

Contents lists available at ScienceDirect

Injury

journal homepage: www.elsevier.com/locate/injury

Predicting 30-day mortality following hip fracture surgery: Evaluation of six risk prediction models

Julian Karres [*], Nicole A. Heesakkers, Jan M. Ultee, Bart C. Vrouenraets

Department of Surgery, Sint Lucas Andreas Hospital, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Accepted 7 November 2014

Keywords:
Hip fractures
Surgery
Mortality
Outcome
Scoring
Risk prediction

ABSTRACT

Introduction: While predictors for mortality after hip fracture surgery have been widely studied, research regarding risk prediction models is limited. Risk models can predict mortality for individual patients, provide insight in prognosis, and be valuable in surgical audits. Existing models have not been validated independently. The purpose of this study is to evaluate the performance of existing risk models for predicting 30-day mortality following hip fracture surgery.

Patients and methods: In this retrospective study, all consecutive hip fracture patients admitted between 2004 and 2010 were included. Predicted mortality was calculated for individual patients and compared to the observed outcome. The discriminative performance of the models was assessed using the area under the receiver operating characteristic curve (AUC). Calibration was analysed with the Hosmer–Lemeshow goodness-of-fit test.

Results: A literature search yielded six risk prediction models: the Charlson Comorbidity Index (CCI), Orthopaedic Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity (O-POSSUM), Estimation of Physiologic Ability and Surgical Stress (E-PASS), a risk model by Jiang et al., the Nottingham Hip Fracture Score (NHFS), and a model by Holt et al. The latter three models were specifically designed for the hip fracture population. All models except the O-POSSUM achieved an AUC greater than 0.70, demonstrating acceptable discriminative power. The score by Jiang et al. performed best with an AUC of 0.78, this was however not significantly different from the NHFS (0.77) or the model by Holt et al. (0.76). When applying the Hosmer–Lemeshow goodness-of-fit test, the model by Holt et al., the NHFS and the model by Jiang et al. showed a significant lack of fit ($p < 0.05$). The CCI, O-POSSUM and E-PASS did not demonstrate lack of calibration.

Discussion: None of the existing models yielded excellent discrimination (AUC > 0.80). The best discrimination was demonstrated by the models designed for the hip fracture population, however, they had a lack of fit. The NHFS shows most promising results, with reasonable discrimination and extensive validation in earlier studies. Additional research is needed to examine recalibration and to determine the best risk model for predicting early mortality following hip fracture surgery.

Introduction

Hip fractures are a common injury in the elderly and associated with high mortality and morbidity [1]. The incidence of hip fractures increases with age and consists of almost 30% of fall-related injuries in patients aged 85 years or older [2]. In the Netherlands, around 19,000 hip fractures are recorded every year, accounting for over 20% of the yearly injury-related medical costs [3,4]. Early post-operative mortality is particularly high, with reported 30-day or in-hospital mortality of 13.3% [5].

Hip fractures are generally treated by surgical means. Risk factors for adverse outcomes following hip fracture surgery have been widely studied. In a meta-analysis from 2012, Hu et al. report strong evidence for 12 predictors of mortality after hip fracture surgery, including advanced age, male gender, nursing home residence, poor preoperative mobility, higher ASA grading, poor mental state and several comorbidities [5]. Furthermore, seven moderate evidence and 12 limited evidence mortality predictors were identified, including fracture type, several serum levels and additional comorbidities. In their recent meta-analysis, Smith et al. identify nine similar strong pre-operative indicators of mortality, including age, mobility and cognitive impairment [6].

[*] Corresponding author at: Department of Surgery, Sint Lucas Andreas Hospital, PO Box 9243, 1006 AE, Amsterdam, The Netherlands. Tel.: +31 6 2142 0881; fax: +31 20 685 4014.
E-mail address: j.karres@slaz.nl (J. Karres).

http://dx.doi.org/10.1016/j.injury.2014.11.004
0020–1383/© 2014 Elsevier Ltd. All rights reserved.

Whereas mortality predictors have been studied extensively, research regarding risk prediction models is limited. Risk models can be used to calculate post-operative mortality risk for the individual patient, ideally at the time of admission. An accurate prediction of mortality can give the patient, its family and the treating physician insight in the prognosis, and could assist in clinical decision making. Furthermore, it could be used as a risk adjustment tool for baseline differences when comparing hip fracture surgery outcomes between different health care providers.

Several risk models for surgical outcome, such as the O-POSSUM and Charlson Comorbidity Index, have been applied to the hip fracture population [7,8]. In addition, a number of new, more specific prediction models have been developed, using logistic regression to determine variables associated with mortality after hip repair. These new models, however, have not been validated in independent studies. A multitude of coexisting prediction models without external validation or comparison will preclude successful application in daily practice [9,10]. With the number of hip fractures increasing worldwide, the necessity of a reliable risk prediction model is becoming more prudent.

The aim of this study is to apply currently available risk prediction models on a large hip fracture cohort from our hospital, in order to identify the most accurate predictor of 30-day mortality following hip fracture surgery.

Patients and methods

Study cohort

All consecutive patients with a proximal femur fracture admitted to the *Sint Lucas Andreas* Hospital in the period from January 2004 to December 2010 were included in this study. This hospital is located in *Amsterdam* and serves an urban population. Surgical treatment was according to current guidelines; intracapsular fractures were treated with hemiarthroplasty, cannulated screws or a Dynamic Hip Screw, extracapsular fractures underwent fixation by intramedullary nailing or a Dynamic Hip Screw. Patients undergoing conservative treatment and total hip replacement were excluded, as well as patients with periprosthetic fractures or slipped capital femoral epiphysis.

Data required for risk prediction models (see section 'Results') was collected retrospectively from the medical records. Patient characteristics, including age, gender, comorbidities, pre-fracture residency, and physiological and operative data were retrieved. A contralateral fracture in the same patient on a different date was recorded as a separate case. Variables missing for more than 10% were excluded from analysis and if needed corrected for. Thirty-day mortality, defined as death within 30 days following hip fracture surgery, was verified using our hospital's administration records, clinical files, health insurance databases and via family doctors. This study and the use of clinical data have been approved by the local medical ethics committee, deeming individual informed consent to be unnecessary due to the observational character of the research.

Statistical analysis

The predictive performance of the six risk models was analysed in terms of discrimination and calibration. Discrimination is determined as the ability to distinguish between outcome groups, and can be assessed by calculating the standard receiver operating characteristic (ROC) curve [11]. The area under the ROC curve (AUC) is a measure of how well a model separates patients who experienced the designated outcome from those who did not experience the outcome. In our case, the risk prediction models should assign a higher risk of death to those patients who did not survive than to those patients who survived 30 days after hip fracture surgery. The AUC, also known as c-statistic, can be anywhere between 0.5 and 1.0, the latter indicating perfect discrimination. In mortality prediction models an AUC between 0.70 and 0.79 is considered to represent an acceptable discrimination, and an AUC between 0.80 and 0.89 is considered excellent [12]. However, a risk prediction model with good discriminative power can still produce inaccurate risk predictions if it is not well calibrated. Calibration is the assessment of how closely predictions resemble the observed outcome for a group of patients. To evaluate calibration over the entire range of prediction, a goodness-of-fit test is used, most commonly the Hosmer–Lemeshow statistic [13]. This test compares predicted and observed mortality rates across prediction deciles and determines whether the differences are greater than that expected by chance. A significant outcome of the Hosmer–Lemeshow test indicates a lack of fit. Additionally, predicted versus observed 30-day mortality was determined. Analysis was performed using SPSS version 18 (SPSS Inc., Chicago, IL, USA). ROCKIT version 1.1B was used for comparison between ROC curves [14].

Results

Risk prediction models

A literature search yielded six relevant risk prediction models for early mortality after hip fracture surgery. The first three models have been designed and validated in wide-ranging populations before being applied to the hip fracture population. The latter three have been specifically developed with the use of data from hip fracture patients. A complete overview of characteristics used by all models is shown in Table 1.

CCI – The Charlson Comorbidity Index (CCI) is a prediction model based on the classification of comorbidity, with points attributed depending on the pre-operative conditions of the patient [8]. The CCI is well-known and broadly used, and not initially designed for hip fracture patients. The accumulated points represent the total weight of the index.

O-POSSUM – The Orthopaedic version of the Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity (O-POSSUM) is well established as a tool for risk prediction in orthopaedic surgery [7,15]. The O-POSSUM uses 14 physiologic and six operative variables to predict mortality and morbidity. Between one and eight points is given for each variable. Mortality (R1) can be estimated using the equation: $\log_e \mathrm{R1}/(1-\mathrm{R1}) = -7.04 + (0.13 \times \text{physiological score}) + (0.16 \times \text{operative severity score})$.

E-PASS – The Estimation of Physiologic Ability and Surgical Stress (E-PASS) consists of a preoperative risk score (PRS) and a surgical stress score (SSS) [16,17]. Together they form the comprehensive risk score (CRS), which is calculated using the following formula: $\mathrm{CRS} = 0.328 + (0.936 \times \mathrm{PRS}) + (0.976 \times \mathrm{SSS})$. The PRS consists of age, several comorbidities, a performance index and the American Society of Anaesthesiologists (ASA) score. The SSS is based on three operative values: amount of blood loss per body weight, operation time and extent of skin incision.

Jiang et al. – Jiang et al. developed a multivariate risk adjustment model based on a cohort of hip fracture patients [18]. Predicting factors are age, gender, long-term care residence and ten different comorbidities. Between 0 and 20 points is scored for each variable. Patients are divided into quartiles according to their calculated risk score, with predicted probability for in-hospital death ranging from <1% to >15%.

NHFS – The Nottingham Hip Fracture Score (NHFS) developed by Maxwell et al. predicts the probability of mortality at 30 days after hip fracture using individual clinical factors [19]. Relevant predicting factors were established by means of logistic regression and include such variables as age, gender, mini-mental test score (MMTS) and number of comorbidities. Between 0 and 4 points is scored for each variable, the total of points resulting in the NHFS. After longitudinal assessment the model was recalibrated to correct for overestimating mortality in high risk groups [20]. The predicted 30-day mortality is calculated with the formula 100/ $1 + e^{[5.012(\mathrm{NHFS} \times 0.481)]}$.

Holt et al. – Using data form the Scottish Hip Fracture Audit, Holt et al. determine several variables associated with 30 and 120-day mortality after hip fracture surgery [21]. A significant contribution to mortality is reported for age, ASA score, gender, pre-fracture residence and mobility, and fracture type. Their article proposes a formula to calculate the predicted mortality based on these pre-operative variables and their logistic regression coefficients.

Study cohort

Between January 2004 and December 2010, 1120 hip fractures were surgically treated at our hospital. A total of 24 patients (2.1%) were excluded due to incomplete medical files. In addition, 46 patients (4.1%) were lost to follow-up (Fig. 1), resulting in a total of 1050 fractures in 1017 patients. The median age was 83 (range 23–100) and most patients were female (69.2%). In our population, 30-day mortality was 8.2% and in-hospital mortality was 6.0%. Baseline characteristics are shown in Table 2.

Two patient characteristics were missing in more than 10% of cases; perioperative blood loss ($N = 731$) and pre-operative mobility ($N = 819$). These variables were excluded from analysis, and their respective risk models were corrected. Blood loss was assumed to be between 101 and 500 mL in the operative severity score of the O-POSSUM, based on the average of 359 mL in our population. In case of the E-PASS, blood loss was assumed 5.6 mL/Kg, based on the average of 450 known cases in our population. Pre-fracture mobility was excluded from the predictive model by Holt et al. by assuming mobility was without aids, unaccompanied.

Discrimination

Sensitivity and specificity of the models are shown as ROC curves in Fig. 2. All models except the O-POSSUM achieve an AUC greater than 0.70, thus demonstrating reasonable discriminative power (Table 3). The score by Jiang et al. performs best with an AUC of 0.78. This is however not significantly different from the NHFS (AUC = 0.77) or the model proposed by Holt et al. (AUC = 0.76). The E-PASS and CCI show some predictive power with an AUC of 0.72 and 0.71 respectively, but this is significantly inferior to the model developed by Jiang et al.

Table 1
Patient and treatment characteristics used by the prediction models for mortality following hip fracture surgery.

Risk model	Variables
CCI	Coronary artery disease
	Congestive heart failure
	Chronic pulmonary disease
	Peptic ulcer disease
	Peripheral vascular disease
	Mild liver disease
	Cerebrovascular disease
	Connective tissue disease
	Diabetes mellitus
	Dementia
	Hemiplegia
	Moderate to severe renal disease
	Diabetes mellitus with end organ damage
	Any tumour within 5 years
	Leukaemia
	Lymphoma
	Moderate to severe liver disease
	Metastatic solid tumour
	AIDS
O-POSSUM	Age
	Cardiac signs
	Respiratory signs
	Systolic blood pressure
	Pulse rate
	Glascow Coma Score
	Serum Urea
	Serum Na
	Serum K
	Haemoglobin
	White blood cell count
	ECG
	Operative magnitude
	Number of operations in 30 days
	Blood loss
	Contamination
	Presence of malignancy
	Timing
E-PASS	Age
	Severe heart disease
	Severe pulmonary disease
	Diabetes mellitus
	Performance status index
	ASA classification
	Intraoperative blood loss per body weight
	Operation time
	Extent of skin incision
Jiang et al.	Age
	Male sex
	Admitted from long-term care
	COPD
	Pneumonia
	Ischaemic heart disease
	Previous myocardial infarction
	Any cardiac arrhythmia
	Congestive heart failure
	Malignancy
	Malnutrition
	Any electrolyte disorder
	Renal failure
NHFS	Age
	Male sex
	Admission haemoglobin
	Admission MMTS
	Living in an institution
	Number of co-morbidities
	Malignancy
Holt et al.	Age
	ASA score
	Gender
	Pre-fracture residence
	Pre-fracture mobility
	Type of fracture

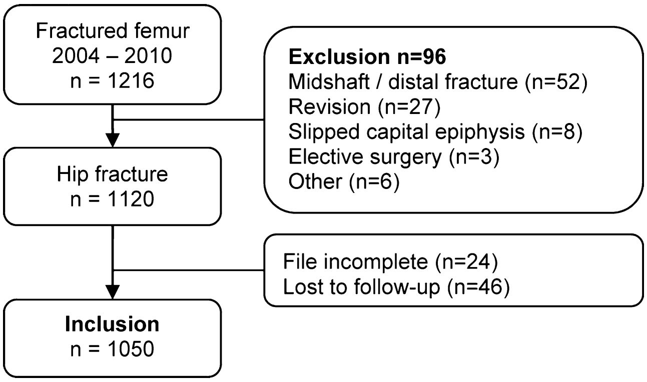

Fig. 1. Flowchart indicating patient inclusion.

Table 2
Patient characteristics.

		N (%)
Gender	Female	727 (69.2%)
	Male	323 (30.8%)
Age in years (median, range)		83 (23–100)
Fracture type	Intracapsular	456 (43.4%)
	Extracapsular	542 (51.6%)
	Subtrochanteric	41 (3.9%)
	Pathologic	11 (1.0%)
Length of admission in days (median, range)		9 (1–181)
In hospital death		63 (6.0%)
30 day mortality		86 (8.2%)
In hospital death and/or death within 30 days		102 (9.7%)

Calibration

By applying the Hosmer–Lemeshow goodness-of-fit test, the calibration of each risk prediction model was assessed (Table 4). The O-POSSUM, CCI and E-PASS models show a good fit. The prediction models developed by Holt et al. (p = 0.002) and by Jiang et al. (p = 0.041) as well as the NHFS (p = 0.039) show a significant lack of fit. Predicted versus observed 30-day mortality is shown in Fig. 3. Most risk models underestimate 30-day mortality after hip fracture, except for the O-POSSUM. The CCI and E-PASS do not give a prediction in percentages for early mortality and are therefore not included. The expected mortality calculated by the NHFS is closest to the observed mortality in our group with a predicted-to-observed ratio of 0.85 (Table 4).

Discussion

Table 3
Discriminative performance of the risk prediction models: calculation of the area under the ROC curve.

Risk model	AUC	
Jiang et al.	0.78 (0.73–0.83)	
NHFS	0.77 (0.72–0.82)	p = 0.853
Holt et al.	0.76 (0.71–0.81)	p = 0.214
E-PASS	0.72 (0.67–0.77)	**p = 0.017**
CCI	0.71 (0.65–0.77)	**p = 0.019**
O-POSSUM	0.69 (0.63–0.74)	**p < 0.001**

Values in parentheses are 95% confidence intervals. ROC: receiver operating characteristic. AUC: Area under the ROC curve. NHFS: Nottingham Hip Fracture Score. E-PASS: Estimation of Physiologic Ability and Surgical Stress. CCI: Charlson Comorbidity Index. O-POSSUM: Orthopaedic Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity.
Bold typeface to indicate statistically significant results (p < 0.05)

This study evaluates the performance of six risk prediction models for early mortality in patients undergoing hip fracture surgery. Five of the six scoring systems resulted in acceptable discrimination when applied to our study population, while none of the models showed excellent discriminative power (AUC > 0.80). The model developed by Jiang et al. yielded the highest sensitivity and specificity with an area under the ROC curve of 0.78. This performance was not significantly different from the NHFS or the model proposed by Holt et al. Remarkably, these three models showed a significant lack of fit when applying the Hosmer–Lemeshow goodness-of-fit test. The three models with the lowest discriminative power, the O-POSSUM, the CCI and the E-PASS, did not demonstrate a significant lack of fit.

The model by Jiang et al., the NHFS and the formula by Holt et al. demonstrated the best discriminative performance with an area under the ROC curve of 0.78, 0.77 and 0.76 respectively. These three models are specifically designed for the hip fracture population. The model by Jiang et al. was published in 2005 and was developed using a cohort of 3981 hip fracture patients of the

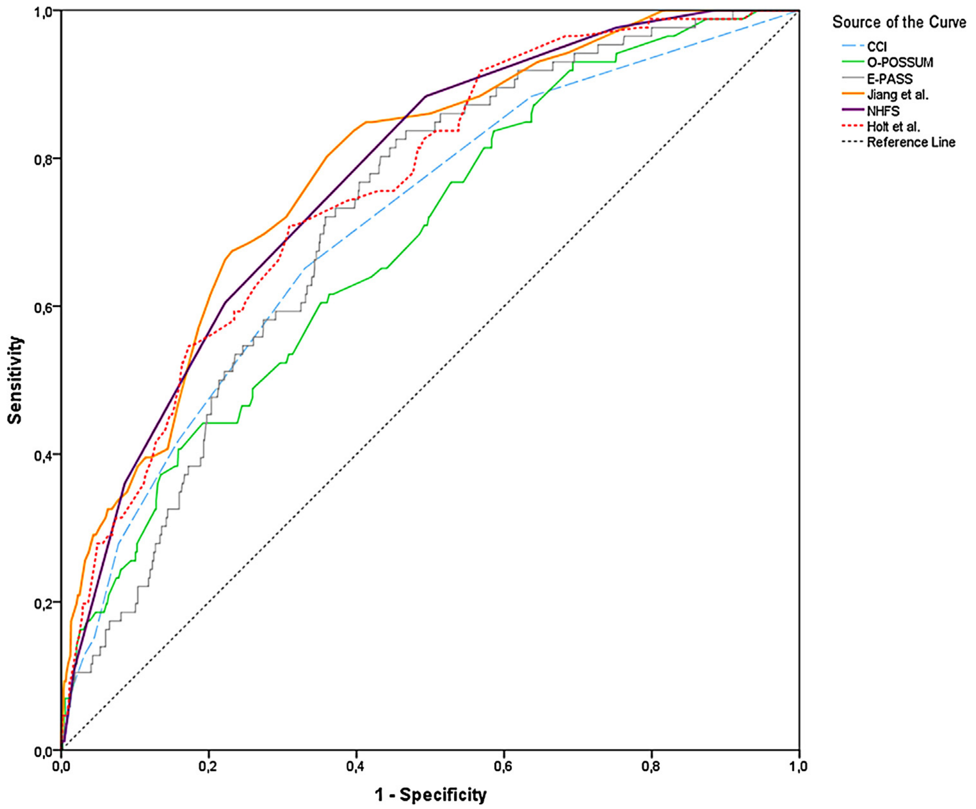

Fig. 2. Graph displaying the receiver operating characteristic curves of all six prediction models. ROC: receiver operating characteristic. CCI: Charlson Comorbidity Index. O-POSSUM: Orthopaedic Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity. E-PASS: Estimation of Physiologic Ability and Surgical Stress. NHFS: Nottingham Hip Fracture Score.

Table 4
Assessment of calibration using the Hosmer–Lemeshow goodness-of-fit test and predicted-to-observed mortality ratio.

Risk model	Hosmer–Lemeshow	Predicted-to-observed ratio
CCI	$p = 0.291$	–
O-POSSUM	$p = 0.110$	1.55
E-PASS	$p = 0.103$	–
Jiang et al.	**p = 0.041**	0.71
NHFS	**p = 0.039**	0.85
Holt et al.	**p = 0.002**	0.58

Bold typeface to indicate statistically significant results ($p < 0.05$)

Capital Health area in Canada [18]. The model uses 13 variables, including age and gender, to assess the risk of mortality after a hip fracture. It has been validated in a different hospital, albeit by the initial developers in the same paper, where it showed good discriminative power for in-hospital mortality (AUC = 0.82) as well as for 1 year mortality (AUC = 0.74) [18]. Statistical analysis regarding calibration was only reported in the derivation set for in-hospital mortality, where it demonstrated good fit (Hosmer–Lemeshow, $p > 0.50$). After this initial publication however, no further use or validation of the risk model has been reported. The NHFS was developed by Maxwell et al. in 2008 and consists of seven variables [19]. It was established by means of logistic regression using a cohort of 4967 patients, where it demonstrated a reasonable predictive performance for early mortality (AUC = 0.72) with a good fit (Hosmer–Lemeshow, $p = 0.79$). The model was later validated for 1-year mortality and as a predictor of early discharge after hip fracture surgery [22,23]. In a more recent study, multicentre validation of the NHFS using data of 7290 patients has resulted in recalibration to adjust for overestimation in high risk groups [20]. The model proposed by Holt et al. in 2008 is based on the logistic regression coefficients from the multivariate analysis of 18,817 patients from the Scottish Hip Fracture Audit [21]. In their article, Holt et al. determine variables associated with 30 and 120 day mortality after hip fracture surgery and propose a formula to estimate mortality for individual patients. Even though this formula is based on a large study sample, no analysis of its predictive performance is reported, nor has it been validated in consequent studies.

In our study cohort, the risk prediction models not specifically developed for hip fracture surgery resulted in lower discriminative power. The E-PASS, the CCI and the O-POSSUM yielded an area under the ROC curve of 0.72, 0.71 and 0.69 respectively. The E-PASS was developed in 1999 by Haga et al. as a prediction model for postoperative mortality following elective gastrointestinal surgery [24]. Several studies by Hirose et al. evaluate the E-PASS for predicting outcome after surgical repair of hip fractures, and found significant correlation between postoperative morbidity and mortality rates and the PRS and CRS ($R = 0.2$, $p < 0.01$) [16,17,25]. However, only observed-to-estimated ratios are reported and no further statistical analysis is performed, making the discriminative power and calibration of the E-PASS unclear. The Charlson Comorbidity Index was developed by Charlson et al. in 1987 and consists of several comorbidities [8]. It is one of the most widely used systems for scoring pre-operative comorbidity. A strong association of high CCI score with postoperative 30-day mortality is reported by Kirkland et al. based on 485 patients undergoing hip fracture surgery [26]. While this association is found using multivariate analysis, no further statistical assessment of the CCI is performed, leaving its predictive performance for individual cases undefined. The O-POSSUM has shown good results in predicting the outcome of general orthopaedic surgery [7]. Although it has been used successfully to predict mortality after hip fracture surgery by Wright et al. and Van Zeeland et al. (AUC = 0.83), contradictory results have been published as well

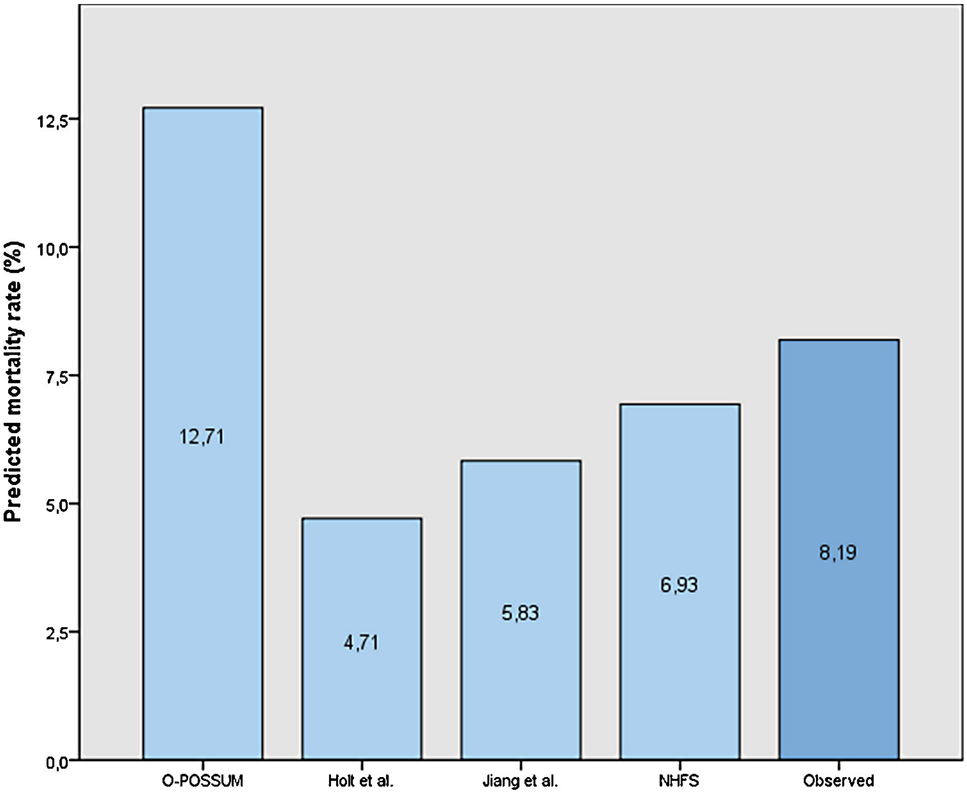

Fig. 3. Predicted vs. observed 30 day mortality. O-POSSUM: Orthopaedic Physiologic and Operative Severity Score for the enUmeration of Mortality and Morbidity. NHFS: Nottingham Hip Fracture Score.

[27–29]. In a cohort of 1164 patients, Ramanathan et al. found a poor performance of the POSSUM score with an AUC of 0.62 and a significant lack of fit [30].

There are several potential reasons why the discriminative performance in our population of the O-POSSUM, CCI and E-PASS is less adequate. Evidently, since these models are not specifically designed for risk prediction in hip fracture patients, they consist partly of variables not associated with mortality after hip fracture surgery. Moreover, operative variables used in the O-POSSUM and E-PASS will have very little differential power as a result of essentially identical surgical procedures in hip fracture treatment. The hip fracture population is homogeneous, leading to less discriminative value for such variables. While the CCI is an easy applicable model, it only consists of comorbidity, neglecting other variables of high relevance for post-operative mortality after hip fracture surgery, such as gender and pre-operative residency [5,6].

The three risk models with the highest discriminative power, the model by Jiang et al., the NHFS and the model by Holt et al., all show poor calibration when applying the Hosmer–Lemeshow test. This could be attributed to several factors. Firstly, the Hosmer–Lemeshow goodness-of-fit test has been reported to overestimate the lack of fit in big sample sizes, in which a small difference can easily become significant [31,32]. Secondly, since these three models were developed using multivariate logistic regression, their respective logistic formulas for risk calculation are complex, allowing for deviation of the curve between predicted and observed mortality. These differences will become more apparent when dividing the results into deciles as is done by the Hosmer–Lemeshow test, possibly leading to a significant lack of fit. Because it only takes one decile where the predicted and observed outcome differ enough for the Hosmer–Lemeshow test to indicate lack of fit, a statistical poor calibration does not necessarily suggest clinically significant differences between prediction by the model and observed outcome [33]. Therefore, it would be unreasonable to simply reject risk models that showed poor calibration, especially since these models resulted in significantly better discrimination. In terms of convenience, the NHFS and the model proposed by Holt et al. only require a few patient characteristics, making risk prediction easy and quick to calculate. The O-POSSUM is far less user friendly; it requires more variables and is complex, making it less suitable for daily practice.

There are several limitations to this research. Because of the retrospective nature of this study, not all data were available for our dataset. Two characteristics were missing for more than 10%: peroperative blood loss and pre-operative mobility. The corresponding models were corrected for these variables, in order to reduce the potential loss of discriminative power on the total performance of the model. The mini-mental test score used by the NHFS was not available for our population. To correct this, points were attributed in case of a medical history of dementia. Since dementia is well documented in our records and associated with mortality in hip fracture patients, any effect on the predictive performance of the NFHS would be minimal [5,6].

To our knowledge, this is the first comprehensive and independent evaluation of available risk models for early mortality following hip fracture surgery. Comparison of multiple risk models should preferably be carried out in an independent sample by investigators other than those who originally proposed the models [9,10,34]. In 2008, Burgos et al. compared six predictors for incidence of post-operative complications, ambulation after a 3-month period and 90-day mortality after hip fracture surgery [35]. Although accurate predictors for the ability to walk and for serious complications were reported, none of the models yielded any predictive power for mortality (AUC < 0.70). For this analysis however, the models specifically designed for hip fracture patients were not available, making the evaluation incomplete.

With growing numbers of hip fractures each year, the importance of a reliable model for predicting mortality is evident. Implementation of such a model becomes possible with digital patient records, creating easy access to risk estimation for all hip fracture patients. Moreover, these models can assist in comparison of surgical outcomes. The newly proposed models are easy to use and result in good discriminative performance, so it would be unreasonable to discard them solely based on an apparent lack of fit. Because these models only use pre-operative variables, they can estimate mortality risk for the hip fracture patient at the time of admission, and may even be of assistance in clinical decision making. Of these new risk scoring systems, the NHFS is most promising, being the only model with extensive validation, albeit by the initial developers, where it showed good calibration [20]. Although several models we evaluated showed acceptable discrimination, none of them achieved an AUC over 0.80, leaving room for improvement in predicting mortality after hip fracture surgery. Additional prospective research should be undertaken to assess a possible need for recalibration to fit geographical differences and to evaluate the best overall risk prediction model.

Conclusions

In conclusion, the perfect risk model for predicting mortality following hip fracture surgery does not exist. In our evaluation of six existing prediction models, the best discriminative performance was demonstrated by the models that were specifically designed for the hip fracture population: the model by Jiang et al., the NHFS and the model by Holt et al. None of these three models however showed excellent discrimination. Furthermore, they resulted in a significant lack of fit when applying the Hosmer–Lemeshow goodness-of-fit test. Up to now, the NHFS shows most promise of all proposed models, with a good discriminative performance in our population and comprehensive validation in recent follow-up studies. Further research should be carried out to examine potential recalibration needs and to determine the best risk model for predicting mortality following hip fracture surgery.

Conflict of interest

The authors declare that they have no conflicts of interest.

Acknowledgements

The authors would like to acknowledge Piet J. Kostense, PhD, Department of Epidemiology and Biostatistics, VU University Medical Center Amsterdam, for his assistance to the statistical work, and Hendrik A. Marsman, MD, PhD, Department of Surgery, Academic Medical Centre Amsterdam, for his contribution to the proof-reading of the manuscript.

References

[1] Zuckerman JD. Hip fracture. N Engl J Med 1996;334(23):1519–25.
[2] Hartholt KA, van Beeck EF, Polinder S, van der Velde N, van Lieshout EM, Panneman MJ, et al. Societal consequences of falls in the older population: injuries, healthcare costs, and long-term reduced quality of life. J Trauma 2011;71(3):748–53.
[3] Meerding WJ, Mulder S, van Beeck EF. Incidence and costs of injuries in the Netherlands. Eur J Public Health 2006;16(3):272–8.
[4] Statistics Netherlands. Hip fractures 2010-2011 Statline CBS [2014 [cited 2014 May 27]; Available from: URL: http://statline.cbs.nl/StatWeb/publication/default.aspx?DM=SLNL&PA=71859NED&D1=4&D2=0&D3=0&D4=157&D5=26-30&STB=G2%2cG3%2cG4%2cG1%2cT&CHARTTYPE=1&VW=T
[5] Hu F, Jiang C, Shen J, Tang P, Wang Y. Preoperative predictors for mortality following hip fracture surgery: a systematic review and meta-analysis. Injury 2011.
[6] Smith T, Pelpola K, Ball M, Ong A, Myint PK. Pre-operative indicators for mortality following hip fracture surgery: a systematic review and meta-analysis. Age Ageing 2014.

[7] Mohamed K, Copeland GP, Boot DA, Casserley HC, Shackleford IM, Sherry PG, et al. An assessment of the POSSUM system in orthopaedic surgery. J Bone Joint Surg Br 2002;84(5):735–9.

[8] Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987;40(5):373–83.

[9] Collins GS, Moons KG. Comparing risk prediction models. BMJ 2012;344:3186.

[10] Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. J Clin Epidemiol 2008;61(11):1085–94.

[11] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143(1):29–36.

[12] Hosmer DW, Lemeshow S. Applied logistic regression. New York, USA: Wiley; 2004.

[13] Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115(1): 92–106.

[14] Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. Med Decis Making 1998;18(1):110–21.

[15] Copeland GP, Jones D, Walters M. POSSUM: a scoring system for surgical audit. Br J Surg 1991;78(3):355–60.

[16] Hirose J, Mizuta H, Ide J, Nakamura E, Takada K. E-PASS for predicting postoperative risk with hip fracture: a multicenter study. Clin Orthop Relat Res 2008;466(11):2833–41.

[17] Hirose J, Mizuta H, Ide J, Nomura K. Evaluation of estimation of physiologic ability and surgical stress (E-PASS) to predict the postoperative risk for hip fracture in elder patients. Arch Orthop Trauma Surg 2008;128(12):1447–52.

[18] Jiang HX, Majumdar SR, Dick DA, Moreau M, Raso J, Otto DD, et al. Development and initial validation of a risk score for predicting in-hospital and 1-year mortality in patients with hip fractures. J Bone Miner Res 2005;20(3): 494–500.

[19] 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. Br J Anaesth 2008;101(4):511–7.

[20] Moppett IK, Parker M, Griffiths R, Bowers T, White SM, Moran CG. Nottingham Hip Fracture Score: longitudinal and multi-assessment. Br J Anaesth 2012;109(4):546–50.

[21] Holt G, Smith R, Duncan K, Finlayson DF, Gregori A. Early mortality after surgical fixation of hip fractures in the elderly: an analysis of data from the scottish hip fracture audit. J Bone Joint Surg Br 2008;90(10):1357–63.

[22] Wiles MD, Moran CG, Sahota O, Moppett IK. Nottingham Hip Fracture Score as a predictor of one year mortality in patients undergoing surgical repair of fractured neck of femur. Br J Anaesth 2011;106(4):501–4.

[23] Moppett IK, Wiles MD, Moran CG, Sahota O. The Nottingham Hip Fracture Score as a predictor of early discharge following fractured neck of femur. Age Ageing 2012;41(3):322–6.

[24] Haga Y, Ikei S, Ogawa M. Estimation of Physiologic Ability and Surgical Stress (E-PASS) as a new prediction scoring system for postoperative morbidity and mortality following elective gastrointestinal surgery. Surg Today 1999;29(3):219–25.

[25] Hirose J, Ide J, Irie H, Kikukawa K, Mizuta H. New equations for predicting postoperative risk in patients with hip fracture. Clin Orthop Relat Res 2009;467(12):3327–33.

[26] Kirkland LL, Kashiwagi DT, Burton MC, Cha S, Varkey P. The Charlson Comorbidity Index Score as a predictor of 30 day mortality after hip fracture surgery. Am J Med Qual 2011;26(6):461–7.

[27] van Zeeland ML, Genovesi IP, Mulder JW, Strating PR, Glas AS, Engel AF. POSSUM predicts hospital mortality and long-term survival in patients with hip fractures. J Trauma 2011;70(4):E67–72.

[28] Wright DM, Blanckley S, Stewart GJ, Copeland GP. The use of orthopaedic POSSUM as an audit tool for fractured neck of femur. Injury 2008;39(4):430–5.

[29] Young W, Seigne R, Bright S, Gardner M. Audit of morbidity and mortality following neck of femur fracture using the POSSUM scoring system. N Z Med J 2006;119(1234):1986.

[30] Ramanathan TS, Moppett IK, Wenn R, Moran CG. POSSUM scoring for patients with fractured neck of femur. Br J Anaesth 2005;94(4):430–3.

[31] Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer–Lemeshow test revisited. Crit Care Med 2007;35(9):2052–6.

[32] Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer–Lemeshow goodness-of-fit test for the logistic regression model. J Epidemiol Biostat 2000;5(4):251–3.

[33] Marcin JP, Romano PS. Size matters to a model's fit. Crit Care Med 2007;35(9):2212–3.

[34] Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. BMJ 2012;344:3318.

[35] 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 Anaesthesiol Scand 2008;52(1):125–31.