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/

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