Marufu, Takawira C. and Mannings, Alexa and Moppett, Iain K. (2015) Risk scoring models for predicting peri-operative morbidity and mortality in people with fragility hip fractures: qualitative systematic review. Injury, 46 (12). pp. 2325-2334. ISSN 0020-1383

Injury, Int. J. Care Injured 46 (2015) 2325–2334

Contents lists available at ScienceDirect

Injury

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

Review

Risk scoring models for predicting peri-operative morbidity and mortality in people with fragility hip fractures: Qualitative systematic review

Takawira C. Marufu, Alexa Mannings, Iain K. Moppett *

Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen's Medical Centre, University of Nottingham, UK

ARTICLE INFO

Article history:
Accepted 13 October 2015

Keywords:
Fragility hip fracture
Risk scoring tools
Risk stratification tools
Peri-operative
Elderly

ABSTRACT

Rationale: Accurate peri-operative risk prediction is an essential element of clinical practice. Various risk stratification tools for assessing patients' risk of mortality or morbidity have been developed and applied in clinical practice over the years. This review aims to outline essential characteristics (predictive accuracy, objectivity, clinical utility) of currently available risk scoring tools for hip fracture patients.
Methods: We searched eight databases; AMED, CINHAL, Clinical Trials.gov, Cochrane, DARE, EMBASE, MEDLINE and Web of Science for all relevant studies published until April 2015. We included published English language observational studies that considered the predictive accuracy of risk stratification tools for patients with fragility hip fracture.
Results: After removal of duplicates, 15,620 studies were screened. Twenty-nine papers met the inclusion criteria, evaluating 25 risk stratification tools. Risk stratification tools considered in more than two studies were; ASA, CCI, E-PASS, NHFS and O-POSSUM. All tools were moderately accurate and validated in multiple studies; however there are some limitations to consider. The E-PASS and O-POSSUM are comprehensive but complex, and require intraoperative data making them a challenge for use on patient bedside. The ASA, CCI and NHFS are simple, easy and inexpensive using routinely available preoperative data. Contrary to the ASA and CCI which has subjective variables in addition to other limitations, the NHFS variables are all objective.
Conclusion: In the search for a simple and inexpensive, easy to calculate, objective and accurate tool, the NHFS may be the most appropriate of the currently available scores for hip fracture patients. However more studies need to be undertaken before it becomes a national hip fracture risk stratification or audit tool of choice.

Contents

* Corresponding author at: Anaesthesia and Critical Care Section, Division of Clinical Neuroscience, University of Nottingham, Queen's Medical Centre Campus, Nottingham University Hospitals NHS Trust Derby Road, Nottingham, NG7 2UH, UK. Tel.: +44 1158230959.
E-mail address: iain.moppett@nottingham.ac.uk (I.K. Moppett).

http://dx.doi.org/10.1016/j.injury.2015.10.025

Introduction

Fragility hip fractures among the elderly constitute a significant global public health problem. Risk scoring to identify high risk patients is strongly encouraged [1]. It aims to provide prognostic information based on available patient data. This in-turn allows: (a) increased objectivity in patient outcome prediction, (b) guidance on clinical decision making during perioperative period, (c) better informed consent for patients undergoing hip fracture surgery [2], and (d) treatment optimisation to improve outcome.

Various scoring tools exist and there is uncertainty as to the most suitable tool for use in hip fracture. The ideal risk scoring tool has the following attributes: simple; ease of use; reproducible; accurate; reliable; objective and available to all patients [2]. The extent to which current hip fracture scoring systems meet these criteria is unclear. This study aims to describe the components, likely clinical utility and degree of validation of published risk scoring tools.

Materials and methods

We searched eight databases; AMED, CINHAL, Clinical Trials.gov, Cochrane, DARE, EMBASE, MEDLINE, and Web of Science. The review considered all relevant published studies on development and validation of risk stratification tools in patients with fragility hip fracture. Studies were considered using the recommended standards guidelines for reporting systematic reviews of observation studies [3]. All relevant studies worldwide in any language published from 1966 to the 30th of April 2015, inclusive were included in the review. The search strategy is outlined as supplementary data appendix 1.

Study selection and outcome definition

We defined a risk stratification tool as "a scoring system or model used to predict or adjust for either mortality or morbidity after surgery, and which contains at least two different risk factors" [1]. Eligible studies were identified by title, abstract and full-text screening independently by the authors and discrepancies resolved by consensus. Manual hand searching of first generation reference lists was performed. Data extraction was independently undertaken by TM and AM on pre-piloted database forms. We extracted data for each study against the following four facets of validity and reliability: (1) development of items: development and validation samples in same or different cohorts; random selection of samples; (2) process for validation: single centre; multicentre; international; (3) metrics of discrimination: AUROC/*c*-statistics; and (4) metrics of calibration: Hosmer–Lemeshow or Pearson chi-square statistics. Studies were assessed for methodological quality and risk of bias using Altman's [4] framework for assessing internal validity.

Data and statistical analysis

Calibration and discrimination are the two main performance measures used to evaluate individual risk scoring tools. Discrimination was reported using either the AUROC or the concordance (*c*-) statistic with AUROC of less than 0.7, 0.7–0.9 and greater than 0.9 considered to indicate poor, moderate and high tool performance respectively [1]. As AUROC was not consistently reported, the observed compared to expected outcome ratio (observed/expected (O/E)), Spearman's rank correlation and chi-squared test were also used to evaluate risk scoring tool performance.

The agreement between observed and predicted outcomes (calibration) was evaluated using Hosmer–Lemeshow or Pearson chi-square statistics. $P < 0.05$ reflected evidence of lack of fit [1].

Results

The search produced 15,620 articles, and 680 were eligible for abstract screening (Fig. 1). Most studies considered at the abstract stage, reported risks for sustaining hip fracture, rather than outcome following hip fracture, and 12 studies were conference abstract presentations with no full published papers and therefore were excluded leaving 43 studies for full text analysis. Of the 43 full text studies sought, 30 [5–34] met the inclusion criteria with results presented with sufficient data to evaluate the study outcomes (Table 1). Thirteen full text studies [35–48] did not have sufficient qualitative or quantitative data relevant to this review, and were excluded. All studies included in this review were cohort studies.

Quality assessment

Quality assessment for eligible studies is outlined in Table 1. Seven studies were multicentre with a maximum of nine study sites in one study [24]. Selection bias was not observed in the included studies, though ethnic origin was constrained by the demographic of the study country. Heterogeneity among included studies was observed in method of statistical analysis, variation in time frame of outcome measurements, and in the number of models assessed by individual articles.

Validation

Three forms of validation were observed across included studies; (a) internal – validation in split sample of the same study population as tool derivation cohort, (b) external – validation in

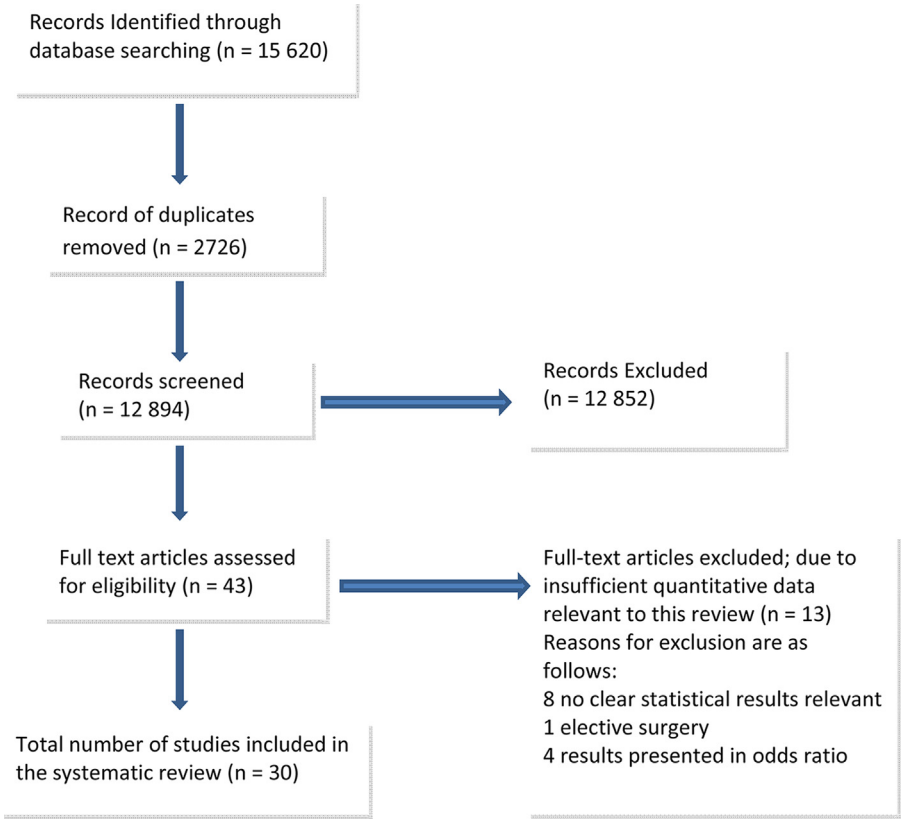

Fig. 1. Flow chart of results.

new cohort unrelated to tool derivation study at a different institute and (c) temporal – validation in new cohort from derivation study but same institution(s). Supplementary data appendix 2 summarises data for commonly used tools and shows how widely each risk stratification tool has been validated and the original tool development cohort.

Types of risk stratification tools

A total of 25 risk stratification tools (supplementary data appendix 3) were reported among all the included studies. Ten of these 24 risk scoring tools were considered in two or more studies (supplementary data appendix 4). The O-POSSUM, NHFS, E-PASS, CCI and ASA were reported in more than three studies each with a total sample of 5975, 13 977, 5832, 6 230 456 and 5411 patients respectively. The other five tools, score by Jiang et al., Risk Model for Delirium (RD), P-POSSUM, Barthel Index and Mini Mental State Examination (MMSE) were reported in two studies each.

Outcomes

The main outcomes (Table 2) were mortality, morbidity and mobility. Timing of outcome measurement ranged from in-hospital mortality or length of hospital stay to more than 1 year. Twenty-six studies reported mortality as the main outcome. In-hospital, 30 day and 1 year mortality ranged from 1.6% [25] to 9.7% [8], 6.6% [26] to 10.9% [20], and 26% [12] to 30.8% [15] respectively. Sixteen studies reported morbidity, which widely varied from 17.0% [24] to 49.6% [8].

Discrimination

The area under the receiver operating characteristic curve (AUROC) was presented in 17 studies and ranged from 0.50 to 0.87. Fifteen studies reported AUROC values for scoring systems of >0.7. None of the studies reported AUROC values >0.90.

Calibration

Ten studies reported calibration. Eight calculated calibration using Hosmer–Lemeshow test with *P* values ranging from 0.00015 [16] to 0.79 [29]. Two studies [14,18] reported whether the model was of 'good fit' or 'poor fit' without further statistical presentation. Calibration values for one study [22] were not presented in the original paper, but were given by the authors in a subsequent letter [49].

Risk stratification tools published in more than one study

Preoperative scores

Three clinical scoring systems that use readily available *pre-operative* data were validated in multiple studies; ASA [8,18,29], CCI [7–10,12] and NHFS, [10,26–31].

American Society of Anesthesiologists physical score (ASA). The ASA is widely used as a surrogate for operative risk and grades patients according to their chronic physiological state. The three studies which considered ASA found it to be of poor to moderate discriminant accuracy (AUROC varying from 0.60 to 0.71).

Charlson Comorbidity Index (CCI). The CCI is a medical risk prediction tool, which has been adapted for surgical risk stratification. All the studies reported moderate AUROC for CCI on mortality (0.7–0.77) but poor prediction with regards to 90 days mortality (0.59).

Nottingham Hip Fracture Score (NHFS). The NHFS is a combination of seven independent predictors of mortality. AUROC values

Table 1
Characteristics of included studies.

Study ID	Risk scoring model	Timing	Country	Number of centres	Patients (*n*)	Study purpose	Validation Cohort: internal vs. external vs. temporal	Outcome	End point d – days m – months	Subject description	Selection bias
Moerman [5]	RD	Prospective	Netherlands	1	378	Validation	Internal	Morbidity	Hospital discharge	Y	N
Vochteloo [6]	RD	Prospective	Netherlands	1	378	Validation	Internal	Morbidity mortality	Hospital discharge 90 d 12 m	Y	N
Neuhaus [7]	CCI	Retrospective	USA	National	6,137,665	Validation	External	Mortality	Hospital discharge	Y	N
Burgos [8]	ASA, CCI POSSUM, Barthel Index Goldman Index	Prospective	Spain	1	232	Validation	external	Morbidity Ambulation Mortality	Hospital discharge 30 d 90 d	Y	N
Toson [9]	CCI	Retrospective	Australia	1	47,698	Validation	External	Mortality	Hospital discharge 30 d 12 m	Y	N
Karres [10]	Jiang et al., NHFS, Holt et al. E-PASS, CCI O-POSSUM	Rétrospective	Netherlands	1	1050	Validation	External	Mortality	30 d	Y	N
Dawe [11]	Sernbo Score	Prospective	UK	1	259	Validation	Internal	Mortality	30 d 12 m	Y	N
Radley [12]	CCI (Romano adaptation) CCS, Iezzoni	Retrospective	USA	1	43,811	Comparative	External	Mortality	12 m	Y	N
Bellelli [13]	New predictive risk score, MMSE, BMI Barthel Index	Prospective	Italy	1	398	Tool development	Internal	Ambulation (mobility)	Hospital discharge >12 m	Y	N
Vochteloo [14]	DHP	Prospective	Netherlands	2	435	Validation	External	Ambulation (mobility)	Hospital discharge	Y	N
Jiang [15]	New Risk score	Retrospective	Canada	2	3981	Tool development	Internal	Mortality	Hospital discharge 12 m	Y	N
Soderqvist [17]	ASA SPMSQ	Prospective	Sweden	4	1944	Comparative study	External	Mortality	Hospital discharge 120 d 24 m	N	N
Ramanathan [16]	O-POSSUM	Prospective	UK	1	1164	Validation	External	Mortality	30 d	Y	N
van Zeeland [18]	O-POSSUM ASA	Prospective	Netherlands	1	272	Validation	External	Morbidity Ambulation Mortality	Hospital discharge	N	N
Bonicoli [19]	O-POSSUM P-POSSUM	Prospective	Italy	1	134	Comparative	External	Mortality Morbidity	Hospital discharge	N	N
Wright [20]	O-POSSUM	Prospective	UK	1	230	Validation	External	Mortality Morbidity	30 d	N	N
Steinberg [21]	O-POSSUM	Retrospective	Israel	1	1770	Validation	External	Mortality	Hospital discharge	Y	N
Hirose [22]	E-PASS P-POSSUM O-POSSUM	Retrospective	Japan	8	722 Grp A 633 Grp B	Validation	External	Mortality Morbidity	Hospital discharge 30 d	N	N

Hirose [23]	E-PASS	Prospective	Japan	1	419	Validation	External	Morbidity Ambulation Mortality	Hospital discharge	N
Hirose [24]	E-PASS	Prospective	Japan	9	421 (Group A 268, B 153)	Validation	External	Ambulation Mortality Morbidity	Hospital discharge 12 m	N
Hirose [25]	E-PASS	Retrospective	Japan	7	813	Validation	External	Mortality Morbidity	Hospital discharge	N
Moppett [27]	NHFS	Prospective	UK	1	6123	Validation	Temporal	Mortality	Hospital discharge 30 d 12 m	N
Moppett [26]	NHFS	Prospective	UK	3	7290	Validation	External/ Temporal	Mortality	30 d	N
Wiles [28]	NHFS	Prospective	UK	1	6202	Validation	Temporal	Mortality	30 d 12 m	N
Maxwell [29]	NHFS, ASA Donati Score	Prospective	UK	1	4967	Tool development and validation	Internal/ External	Mortality	30 d	N
Rushton [30]	NHFS	Retrospective	UK	3	1079	Validation	External	Mortality	30 d	N
Krishnan [31]	NHFS Frailty index (FI)	Retrospective	UK	National	178	Validation	External	Mortality LOS	30 d 30 d	N
Adunsky [32]	MMSE, CDT Cognitive-FIM	Retrospective	Israel	1	143	Comparative	External	Ambulation (mobility)	Hospital discharge	N
Foss [33]	CAS NMS	Prospective	Denmark	1	426	Validation	External	Morbidity Mortality	Hospital discharge 30 d	N
Albertsson [34]	FRAMO Index	Prospective	Sweden	1	1248	Validation	Internal	Mortality	24 m	Y

reported in three studies 0.72 [29], 0.73 [31] and 0.77 [10] showing this score to be a moderately discriminatory tool. Calibration is adequate: the original single centre tool development reported $P = 0.79$ (Hosmer–Lemeshow) [29] and multicentre validation reported $P > 0.1$ [26] showing adequate 'goodness-of-fit' on performance of this model. One study [27] showed that increasing NHFS was negatively correlated with eventual return-to-home $r^2 = 0.949$.

Physiological scores (models incorporating pre- and intra-operative scores)

POSSUM discrimination (AUROC) ranges from 0.63 to 0.65 for 90 days mortality and ambulation, suggesting poor discrimination [8]. The orthopaedic version (O-POSSUM) has AUROC values ranging 0.62 [8] to 0.74 [21] for mortality, and 0.83 [18] for both mortality and morbidity. However, calibration appeared poor with the observed and expected ratio ranging from 0.12 to 1.19 [18]. The P-POSSUM observed and expected ratio values had a wide range 0.15 [22] and 2.17 [19] for in hospital mortality.

Five studies on E-PASS, [10,22–25] reported their results in various forms. AUROC 0.72 and calibration $P = 0.103$ [10] and O/E values ranged from 0.55 to 1.59 [22]. Four studies observed a significant positive correlation between Physiological Risk Score (PRS) and Comprehensive Risk Score (CRS) to measured outcomes but there was no significant correlation observed between the surgical stress score (SSS) and outcomes. Cost of hospitalisation was reported to be associated with SSS and CRS [23,25]. All other risk models reported in two or less studies are detailed appendix 3.

Discussion

This study provides a comprehensive review of the current evidence on a variety of risk stratification tools used in hip fracture patients. Of the 25 scoring systems identified, only five had been evaluated in more than two studies, and four outside their original centre. Of these five ASA does not perform well; despite its simplicity [1] it does not appear to be robust enough in this population. Each of the other four tools has arguments for and against its clinical utility; e.g. tool availability and the objectivity of its parameters. Some tools may be perceived as complex and less likely to be part of daily routine use.

Simplicity and availability

The Nottingham Hip Fracture Score and (NHFS) and Charlson Comorbidity Index (CCI) use readily available pre-operative data. They both have reasonable, though not excellent, discriminant characteristics for mortality and morbidity and have been validated external to their original cohort.

The CCI uses well defined comorbidities. It weights these based on severity and assigns each individual an overall risk score presenting the sum of their comorbidity weights [12]. It is a moderately discriminant tool for in-hospital morbidity and 1 year mortality. However calibration is not well described and this limits its ability as an audit tool. Functional ability and confusion (as opposed to dementia) which are known predictors of outcome following hip fracture are not included.

The NHFS is a hip fracture specific score, which has been validated for early hospital discharge [27], 30-day mortality [29] and 1 year mortality [28]. Its discriminant ability is moderate and has reasonable calibration. All the required data items are routinely collected. It uses the Abbreviated Mental Test Score (AMTS) as its assessment for cognitive impairment which may not be so widely used in countries outside the UK. There are currently no data on regarding the interchangeability of screening tests for cognitive impairment, such as MMSE or clock drawing, in this context.

Table 2
Outcomes.

Study ID	Risk scoring model	End point	Morbidity (%) (NR = not recorded)	AUROC morbidity (95% CI)	Mortality (%)	Calibration (Hosmer–Lemeshow test) (NR = not recorded)	AUROC mortality (95% CI)
Moerman [5]	RD	Hospital discharge	NR	0.73 (0.68–0.77)	NR	NR	NR
Vochteloo [6]	RD	Hospital discharge	27	0.72(0.67–0.77)	NR	NR	NR
Neuhaus [7]	CCI	Hospital discharge	NR	NR	9	NR	CCI, ICD-9 adapted 0.767 CCI, age adjusted 0.766 CCI, updated 2011 0.768
Burgos [8]	ASA CCI POSSUM Barthel Index Goldman Index RISK-VAS	Hospital discharge 30 d 90 d	49.6	Hospital discharge: RISK-VAS 0.833 (0.757–0.910), Barthel 0.67 (0.565–0.780) Goldman 0.652 (0.522–0.781), POSSUM 0.726 (0.615–0.838) CCI 0.707 (0.602–0.811), ASA 0.675 (0.571–0.778)	11.2	NR	90 d: RISK-VAS 0.677 (0.545–0.809) Barthel 0.689 (0.584–0.794) Goldman 0.432 (0.315–0.548) POSSUM 0.635 (0.518–0.751) CCI 0.590 (0.482–0.698) ASA 0.600 (0.488–0.711)
Toson [9]	CCI	Hospital discharge 30 d 12 m	NR	NR	8.2 8.3 26.3	NR	0.72–0.76 0.72–0.75 0.69–0.75
Karres [10]	Jiang et al. NHFS Holt et al. E-PASS CCI O-POSSUM	Hospital discharge 30 d	NR	NR	6.0 8.2	$P=0.041$ $P=0.039$ $P=0.103$ $P=0.002$ $P=0.291$ $P=0.110$	0.78 (0.73–0.83) 0.77 (0.72–0.82) 0.76 (0.71–0.81) 0.72 (0.67–0.77) 0.71 (0.65–0.77) 0.69 (0.63–0.74)
Dawe [11]	Sernbo Score	30 d 12 m	NR	NR	NR	NR	30 d: 0.71 (0.65–0.76) 12 m: 0.68 (0.59–0.75)
Radley [12]	CCI (Romano adaptation) CCS Iezzoni	12 m	NR	NR	26	Overall model performance good	CCI 0.72 CCS 0.76 Iezzoni 0.73
Jiang [15]	New risk score	Hospital discharge 12 m	NR	NR	6.3 30.8	$P>0.50$ goodness of fit	0.82 0.74
Soderqvist [17]	ASA SPMSQ	Hospital discharge 120 d 24 m	NR	NR	4 16 38	NR	24 m: age, gender ASA 0.71 24 m: age, gender, SPMSQ 0.70
Ramanathan [16]	O-POSSUM	30 d	NR	NR	10	Poor fit $P<0.00015$	0.62
van Zeeland [18]	O-POSSUM ASA	Hospital discharge	NR	O-POSSUM 0.83 (0.76–0.90)	9	Good for mortality Poor for morbidity	O-POSSUM 0.83 (0.76–0.89) ASA 0.76 (0.66–0.85)
Steinberg [21]	O-POSSUM	Hospital discharge	NR	NR	NR	NR	0.63 (0.58–0.68) model without albumin levels 0.74 (0.65–0.83) model with albumin level
Moppett [26]	NHFS	30 d	NR	NR	6.6	$P>0.1$	NR
Maxwell [29]	NHFS ASA Donati Score	30 d	NR	NR	10.2	0.79	NHFS 0.719 (SE0.018) ASA 0.718 (SE 0.0163) Donati Score 0.717(SE0.0184)
Krishnan [31]	NHFS Frailty index (FI)	30 d	NR	LOS, 0.73 (0.64–0.82) LOS, 0.82 (0.75–0.89)	NHFS <5, **1.6** NHFS ≥5, **10.4** Intermediate FI, **3.4** High FI, **17.2**		NR
Albertson [34]	FRAMO Index	24 m	NR	NR	NR	NR	**0.75 (0.71–0.79)**

Studies with results presented as observed to expected ratios (*O*:*E*) and spearman's rank correlation

Bellelli [13]	O-POSSUM P-POSSUM	Hospital discharge	49.25	POSSUM 1.1	9.7	NR	POSSUM 0.81 P-POSSUM 2.17
Wright [20]	O-POSSUM	30 d	41.3	*O*:*E* 0.95	10.9	NR	*O*:*E* 1.19
Hirose [22]	E-PASS P-POSSUM O-POSSUM	Hospital discharge 30 d	Grp A 17.2 Grp B 20.2	Grp A Hospital discharge Morbidity rates increased linearly with CRS, $P=0.17$, $P\leq 0.0001$ and PRS, $P=0.17$, $P<0.0001$, but not with SSS, $P=0.01$, $P=0.8$ Grp B Hospital discharge E-PASS 1.06 30d: E-PASS 1.59 O-POSSUM 0.12	Grp A 1.7 Grp B 2.4 Grp B Hospital discharge: E-PASS 0.71, P-POSSUM 0.15 30 d: E-PASS 0.55 O-POSSUM 0.12	E-PASS in hospital (mortality 0.40, morbidity 0.65) 30 d (mortality 0.48, morbidity 0.35) P-POSSUM in hospital (mortality 0.30) O-POSSUM 30 d (mortality 0.24, morbidity 0.11)	Grp A Hospital discharge Mortality rates correlated with PRS, $P=0.16$, $P<0.0001$ and CRS $P=0.18$, $P<0.0001$ but not with SSS, $P=0.01$, $P=0.8$
Hirose [23]	E-PASS	Hospital discharge	18.4	Spearman correlation: morbidity significantly increased with both the PRS ($P=0.19$, $P=0.0001$) and CRS ($P=0.21$, $P<0.0001$), but not with the SSS ($P=0.005$, $P=0.3$) The cost of hospital stay was significantly related to the SSS ($r=0.6$, $P<0.0001$) and CRS ($r=0.4$, $P<0.0001$) (Pearson's correlation)	1.9%	NR	Spearman correlation: mortality rates correlated with PRS ($P=0.19$, $P=0.0001$) CRS ($P=0.21$, $P<0.0001$) but not with SSS ($P=0.02$, $P=0.6$)
Hirose [24]	E-PASS	Hospital discharge 12 m	Grp A 23.5, Grp B 17.0	Grp A Hospital discharge Spearman correlation In hospital morbidity correlated with the SSS $P=0.14$, $P=0.021$, CRS $P=0.13$, $P=0.030$, but not with the PRS, $P=0.09$, $P=0.08$	Grp A 2.2 Grp B 2.0 Grp A 6.8 Grp B 9.8	NR	Hospital discharge Spearman correlation In hospital mortality was correlated with SSS $P=-14$, $P=0.019$, but not with PRS, $P=0.05$, $P=0.220$ and CRS $P=0.001$, $P=0.494$
Hirose [25]	E-PASS	Hospital discharge	20	Spearman correlation Post-operative morbidity rates increased linearly and correlated significantly with (preoperative risk score (PRS)) $P=0.16$, $P<0.0001$) (comprehensive risk score (CRS) $P=0.18$, $P<0.0001$) but not surgical stress score (SSS) $P=0.06$, $P=0.07$) Cost of hospitalisation PRS ($R=0.12$, $P<0.0001$) SSS ($r=0.44$, $P<0.0001$) CRS ($r=0.23$, $P<0.0001$) (Pearson correlation)	1.6	NR	Spearman correlation mortality rates correlated with PRS ($P=0.14$, $P=0.0001$) CRS ($P=0.14$, $P=0.0001$)but not with SSS ($P=-0.03$, $P=0.4$)
Moppett [27]	NHFS	Hospital discharge 30 d 12 m	NR	Hospital discharge Increasing NHFS was negatively correlated with eventual return-to-home ($r^2=0.949$), and with the proportion of patients discharged back to their own home at 7, 14 and 21 postoperative days respectively ($r^2=0.84$, 0.94, 0.96 respectively)	 8.3 29.3	NR	NR
Wiles [28]	NHFS	30 d 12 m	NR	NR	8.3 29.3	NR	30 d survival was higher in the low risk group 96.5% vs. 86.3% ($P<0.001$) 12 m

Table 2 (*Continued*)

Study ID	Risk scoring model	End point	Morbidity (%) (NR = not recorded)	AUROC morbidity (95% CI)	Mortality (%)	Calibration (Hosmer–Lemeshow test) (NR = not recorded)	AUROC mortality (95% CI)
Rushton [30]	NHFS	30 d	NR	NR	7.3	NR	NR
Foss [33]	CAS NMS	Hospital discharge 30 d	21.6	Chi-squared values for CAS vs. NMS were 60.3, 20.2 for the risk of post-operative morbidity and 97.6, 68.3 for the association with patient discharge to own home	30 d: 10.3		Chi-squared values for CAS vs. NMS were 49.1, 20.0 for the association with 30 days mortality
Studies presenting ambulation as the outcome							
Burgos [8]	ASA CCI POSSUM Barthel Index Goldman Index RISK-VAS	Hospital discharge 30 d 90 d	NR	30 d: RISK-VAS 0.700, Barthel 0.737 90 d: RISK-VAS 0.700 (0.628–0.771) Barthel 0.737 (0.672–0.801) Goldman 0.567 (0.491–0.643) POSSUM 0.646 (0.573–0.718) CCI 0.634 (0.563–0.706) ASA 0.624 (0.551–0.698)	NR	NR	
Bellelli [13]	New predictive risk score MMSE BMI Barthel Index	Hospital discharge >12 m	NR	Hospital discharge: Walking independently New risk score 0.8593, MMSE 0.7685 BMI 0.4989, Barthel Index 0.8700 12 m: Walking independently New risk score 0.75, MMSE 0.7267 BMI 0.5209, Barthel Index 0.7344	NR	NR	
Vochteloo [14]	DHP	Hospital discharge	NR	Discharge location Delft Cohort (0.84, 0.79–0.88) Groningen Cohort (0.75, 0.66–0.82)	NR	NR	
Hirose [24]	E-PASS	Hospital discharge 12 m	NR	NR	NR	**Grp A** Hospital discharge Predictor variables of walking ability at discharge was significantly correlated with PRS $P = 0.34$, $P < 0.001$, CRS, $P = 0.33$, $P < 0.001$ not with SSS $P = -0.001$, $P = 0.495$ 12 m: walking ability at 1 year was significantly correlated with, PRS, $P = 0.41$, $P < 0.001$, CRS, $P = 0.40$, $P < 0.001$ not with SSS $P = -0.05$, $P = 0.236$ Grp B Hospital discharge the predicted walking ability calculated by the logarithm was significantly correlated with the actual ability at discharge $P = 0.60$, $P < 0.001$ and at 12 m: after surgery $P = 0.65$, $P < 0.001$	
Adunsky [32]	MMSE CDT Cognitive-FIM	Hospital discharge	NR	NR	$P < 0.001$	Pearson correlation between the three cognitive tests resulted in values ranging from 0.607 to 0.732	

E-PASS and O-POSSUM are comparable in their applicability and limitations. The O-POSSUM is the orthopaedic version of the original POSSUM model. Both models use weighted pre and intraoperative data; hence they cannot be used for preoperative risk prediction. They are also perhaps more complex to score with several variables; the O-POSSUM has 18 variables. E-PASS, O-POSSUM, CCI and NHFS have all been validated internationally [10].

Reported outcomes

Outcomes in the included studies were heavily biased towards mortality. It is a dichotomous variable that is clearly undesirable, objective, clinically important and easy to measure [50], and has dependence on the time frame of measurement. Mortality rates at fixed time periods were easily comparable between studies. Morbidity occurred frequently in all studies with a reported incidence of 17–49.6%. Heterogeneity in morbidity definition and classification has been observed among included studies. Most studies did not look at functional outcomes.

Prognostic variables

The range of prognostic variables used by the risk stratification tools is summarised in appendix 5. The items in the scores have face validity: they are all known independent predictors of, or surrogates for, outcome. However, some of these may be somewhat subjective. There are other predictive variables not currently included in the commonly used models (for example, red cell width distribution on admission [51], albumin [21] levels, and some inflammatory markers [52,53]) that may merit future consideration.

Study strengths and limitations

We conducted a comprehensive search strategy with strict adherence to Centre for Reviews and Dissemination (CRD) [3] systematic review guidance. Search strategy, data extraction and quality assessment was performed independently by the authors and findings were confirmed within the team.

However the review has limitations. Heterogeneity observed in this review, within and among studies, could also have influenced our results. There was variation in outcome analysis and outcome measures. Five different statistical measures were used for predictive ability of individual risk score tools, AUROC, r^2, correlation coefficient, *O:E* and percentages. This was felt necessary to include all high quality studies, for a comprehensive over view of scoring tools available. Unfortunately this also reduced our ability to perform appropriate comparison among all the models presented. This lack of uniformity could affect clarity of which risk score is superior to the other.

Conclusions and future work

The use of risk prediction scores during the perioperative period has been accepted by clinicians as the norm, influencing important informed decision making, to help optimise individual patients' care and to support audit and service improvement. However, the predictive accuracy of risk scores could be more robust and multinational validation is currently lacking. This review has highlighted both strengths and weaknesses of the currently available risk scoring models. This study noted that all outcome measures outlined were medically oriented. Future work could consider the psychological and social dimensional factors that impede early patient discharge in this patient population.

Author contributions

The study was conceptualised by Iain Moppett (IM), Alexa Mannings (AM) and Takawira Marufu (TM). All authors participated in study screening, selection, data extraction and manuscript preparation. All three authors provided intellectual content and approved the manuscript for publication.

Funding source

The Sir Jules Thorn Charitable Trust.

Financial disclosure

This work is part of a Ph.D. sponsored project.

Acknowledgements

We acknowledge the support of library staff of the University of Nottingham, UK.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.injury.2015.10.025.

References

[1] Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MPW. Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery qualitative systematic review. Anesthesiology 2013;119:959–81.
[2] Jones HJ, de Cossart L. Risk scoring in surgical patients. Risk scoring in surgical patients. Br J Surg 1999;86:149–57.
[3] Moher D, Liberati A, Tetzlaff J, Altman DG. PRISMA group, preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009;6:e1000097. http://dx.doi.org/10.1371/journal.pmed.1000097.
[4] Altman DG. Systematic reviews of evaluations of prognostic variables. BMJ 2001;323:224–8.
[5] Moerman S, Tuinebreijer WE, de Boo M, Pilot P, Nelissen RGHH, Vochteloo AJH. Validation of the risk model for Delirium in hip fracture patients. Gen Hosp Psychiat 2012;34:153–9.
[6] Vochteloo AJ, Moerman S, van der Burg BLB, de Boo M, de Vries MR, Niesten D-D, et al. Delirium risk screening and haloperidol prophylaxis program in hip fracture patients is a helpful tool in identifying high-risk patients, but does not reduce the incidence of delirium. BMC Geriatr 2011;11:1–7.
[7] Neuhaus V, King J, Hageman MG, Ring DC. Charlson comorbidity indices and in-hospital deaths in patients with hip fractures. Clin Orthop Relat Res 2013;471:1712–9.
[8] Burgos E, Gomez-Arnau JI, Diez R, Munoz L, Fernandez-Guisasola J, del Valle SG. Predictive value of six risk scores for outcome after surgical repair of hip fracture in elderly patients. Acta Anaesth Scand 2008;52:125–31.
[9] Toson B, Harvey LA, Close JC. The ICD-10 Charlson Comorbidity Index predicted mortality but not resource utilization following hip fracture. J Clin Epidemiol 2015;68:44–51.
[10] Karres J, Heesakkers NA, Ultee JM, Vrouenraets BC. Predicting 30-day mortality following hip fracture surgery: evaluation of six risk prediction models. Injury 2014;46:371–7.
[11] Dawe EJC, Lindisfarne E, Singh T, McFadyen I, Stott P. Sernbo score predicts survival after intracapsular hip fracture in the elderly. Ann Roy Coll Surg 2013;95:29–33.
[12] Radley DC, Gottlieb DJ, Fisher ES, Tosteson ANA. Comorbidity risk-adjustment strategies are comparable among persons with hip fracture. J Clin Epidemiol 2008;61:580–7.
[13] Bellelli G, Noale M, Guerini F, Turco R, Maggi S, Crepaldi G, et al. A prognostic model predicting recovery of walking independence of elderly patients after hip-fracture surgery, an experiment in a rehabilitation unit in Northern Italy. Osteoporosis Int 2012;23:2189–200.
[14] Vochteloo AJH, Flikweert ER, Tuinebreijer WE, Maier AB, Bloem RM, Pilot P, et al. External validation of the discharge of hip fracture patients score. Int Orthop 2013;37:477–82.
[15] 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;20:494–500.

[16] Ramanathan TS, Moppett IK, Wenn R, Moran CG. POSSUM scoring for patients with fractured neck of femur. Br J Anaesth 2005;94:430–3.
[17] Soderqvist A, Ekstrom W, Ponzer S, Pettersson H, Cederholm T, Dalen N, et al. Prediction of mortality in elderly patients with hip fractures: a two-year prospective study of 1,944 patients. Gerontology 2009;55:496–504.
[18] van Zeeland MLP, Genovesi IPO, Mulder JWR, Strating PR, Glas AS, Engel AF. POSSUM predicts hospital mortality and long-term survival in patients with hip fractures. J Trauma 2011;70:E67–72.
[19] Bonicoli E, Parchi P, Piolanti N, Andreani L, Niccolai F, Lisanti M. Comparison of the POSSUM score and P-POSSUM score in patients with femoral neck fracture. Musculoskelet Surg 2014;98:201–4.
[20] Wright DM, Blanckley S, Stewart GJ, Copeland GP. The use of orthopaedic POSSUM as an audit tool for fractured neck of femur. Injury 2008;39:430–5.
[21] Steinberg EL, Amar E, Sagy Y, Rath E, Kadar A, Sternheim A. The impact of serum albumin and serum protein levels on POSSUM score of patients with proximal femur fractures. Injury 2014;45:1928–31.
[22] 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;467:3327–33.
[23] 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 Traum Surg 2008;128:1447–52.
[24] Hirose J, Ide J, Yakushiji T, Abe Y, Nishida K, Maeda S, et al. Prediction of postoperative ambulatory status 1 year after hip fracture surgery. Arch Phys Med Rehab 2010;91:67–72.
[25] 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;466:2833–41.
[26] Moppett IK, Parker M, Griffiths R, Bowers T, White SM, Moran CG. Nottingham hip fracture score: longitudinal and multicentre assessment. Br J Anaesth 2012;109:546–50.
[27] 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;41:322–6.
[28] 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;106:501–4.
[29] 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;101:511–7.
[30] Rushton P, Reed M, Pratt R. Independent validation of the Nottingham hip fracture score and identification of regional variation in patient risk within England. Bone Joint J 2015;97:100–3.
[31] 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;43:122–6.
[32] Adunsky A, Fleissig Y, Levenkrohn S, Arad M, Noy S. A comparative study of Mini-Mental Test, Clock Drawing Task and Cognitive-FIM in evaluating functional outcome of elderly hip fracture patients. Clin Rehabil 2002;16:414–9.
[33] Foss NB, Kristensen MT, Kehlet H. Prediction of postoperative rehabilitation in hip fracture morbidity, mortality and patients: the cumulated ambulation score. Clin Rehabil 2006;20(33):701–8.
[34] Albertsson DM, Mellstrom D, Petersson C, Eggertsen R. Validation of a 4-item score predicting hip fracture and mortality risk among elderly women. Ann Fam Med 2007;5:48–56.
[35] 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;26:461–7.
[36] 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;81:296–302.
[37] Ringdal GI, Ringdal K, Juliebo V, Wyller TB, Hjermstad MJ, Loge JH. Using the Mini-Mental State Examination to screen for delirium in elderly patients with hip fracture. Dement Geriatr Cogn 2011;32:394–400.
[38] de Souza RC, Pinheiro RS, Coeli CM, de Camargo KR. The Charlson Comorbidity Index (CCI) for adjustment of hip fracture mortality in the elderly: analysis of the importance of recording secondary diagnoses. Cad Saude Publica 2008;24:315–22.
[39] Michel JP, Klopfenstein C, Hoffmeyer P, Stern R, Grab B. Hip fracture surgery: is the pre-operative American Society of Anesthesiologists (ASA) score a predictor of functional outcome? Aging Clin Exp Res 2002;14:389–94.
[40] Gold A, Sever R, Lerman Y, Salai M, Justo D. Admission Norton scale scores (ANSS) and postoperative complications following hip fracture surgery in the elderly. Arch Gerontol Geriat 2012;55:173–6.
[41] Thomas M, Eastwood H. Re-evaluation of two simple prognostic scores of outcome after proximal femoral fractures. Injury 1996;27:111–5.
[42] Zuckerman JD, Koval KJ, Aharonoff GB, Hiebert R, Skovron ML. A functional recovery score for elderly hip fracture patients: I. Development. J Orthop Trauma 2000;14:20–5.
[43] Zuckerman JD, Koval KJ, Aharonoff GB, Skovron ML. A functional recovery score for elderly hip fracture patients: II. Validity and reliability. J Orthop Trauma 2000;14:26–30.
[44] Hershkovitz A, Jacubovski OS, Bot MA, Oshry V, Brill S. Clock drawing and rehabilitation outcome in hip fracture patients. Disabil Rehabil 2010;32:2113–7.
[45] 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;60:61–70.
[46] Wang TJ, Zhang BH, Gu GS. Evaluation of POSSUM scoring system in the treatment of osteoporotic fracture of the hip in elder patients. Chin J Traumatol 2008;11:89–93.
[47] Cullen DJ, Apolone G, Greenfield S, Guadagnoli E, Cleary P. ASA physical status and age predict morbidity after three surgical procedures. Ann Surg 1994;220:3–9.
[48] Kristensen MT, Jakobsen TL, Nielsen JW, Jorgensen LM, Nienhuis RJ, Jonsson LR. Cumulated Ambulation Score to evaluate mobility is feasible in geriatric patients and in patients with hip fracture. Dan Med J 2012;59:1–5.
[49] Hirose J. Reply to letter to the Editor: new equations for predicting postoperative risk in patients with hip fracture. Clin Orthop Relat Res 2010;468:1706–7.
[50] Moonesinghe SR, Mythen MG, Grocott MP. High-risk surgery: epidemiology and outcomes. Anesth Analg 2011;112:891–901.
[51] Garbharran U, Chinthapalli S, Hopper I, George M, Back DL, Dockery F. Red cell distribution width is an independent predictor of mortality in hip fracture. Age Ageing 2012;6:1–4.
[52] Kieffer WKM, Rennie CS, Gandhe AJ. Preoperative albumin as a predictor of one-year mortality in patients with fractured neck of femur. Ann Roy Coll Surg 2013;95:26–8.
[53] Laulund AS, Lauritzen JB, Duus BR, Mosfeldt M, Jorgensen HL. Routine blood tests as predictors of mortality in hip fracture patients. Injury 2012;43:1014–20.