


Evaluation of Nottingham Hip Fracture Score, Age-Adjusted Charlson Comorbidity Index and the Physiological and Operative Severity Score for the enumeration of Mortality and morbidity as predictors of mortality in elderly neck of femur fracture patients

SAGE Open Medicine
Volume 8: 1–8
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2050312120918268
journals.sagepub.com/home/smo



Michael James Nelson^{1,2} , Justin Scott³ and Palvannan Sivalingam⁴

Abstract

Background: This study evaluated the use of several risk prediction models in estimating short- and long-term mortality following hip fracture in an Australian population.

Methods: Data from 195 patients were retrospectively analysed and applied to three models of interest: the Nottingham Hip Fracture Score, the Age-Adjusted Charlson Comorbidity Index and the Physiological and Operative Severity Score for enumeration of Mortality and Morbidity. The performance of these models was assessed with receiver operating characteristic curve as well as logistic regression modelling.

Results: The median age of participants was 83 years and 69% were women. Ten percent of patients were deceased by 30 days, 25% at 6 months and 31% at 12 months post-operatively. While there was no statistically significant difference between the models, the Age-Adjusted Charlson Comorbidity Index had the largest area under the receiver operating characteristic curve for within 30 day and 12 month mortality, while the Nottingham Hip Fracture Score was largest for 6-month mortality. There was no evidence to suggest that the models were selecting a specific subgroup of our population, therefore, no indication was present to suggest that using multiple models would improve mortality prediction.

Conclusions: While there was no statistically significant difference in mortality prediction, the Nottingham Hip Fracture Score is perhaps the best suited clinically, due to its ease of implementation. Larger prospective data collection across a variety of sites and its role in guiding clinical management remains an area of interest.

Keywords

Hip fracture, elderly, scoring systems, perioperative risk

Date received: 24 September 2019; accepted: 4 March 2020

Background

An increase in the incidence of hip fractures is projected in Australia, largely due to the aging population.^{1,2} Although there has been progress in reducing short-term mortality risk among elderly patients, acute care costs and long-term mortality and morbidity risk remain significant.^{2–8}

Although many scoring systems are available for use in hip fracture patients, there is relatively little data comparing the predictive value of these systems. The Age-Adjusted Charlson Comorbidity Index (AACCI),⁹ the Physiological

¹Princess Alexandra Hospital, The University of Queensland, Brisbane, QLD, Australia

²Intensive Care Unit, St Vincent's Private Hospital Northside, Chermiside, QLD, Australia

³Queensland Facility for Advanced Bioinformatics, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia

⁴Department of Anaesthesia, Princess Alexandra Hospital, The University of Queensland, Brisbane, QLD, Australia

Corresponding author:

Michael James Nelson, Intensive Care Unit, St Vincent's Private Hospital Northside, 627 Rode Rd, Chermiside, QLD 4032, Australia.

Email: Michael.nelson2@uqconnect.edu.au



and Operative Severity Score for enUmeration of Mortality and Morbidity (POSSUM)¹⁰ and the Nottingham Hip Fracture Score (NHFS) are scores of interest that have been previously evaluated in an orthopaedic setting.¹¹ To our knowledge, the Charlson Comorbidity Index is one among these systems that has been evaluated on an Australian hip fracture population.¹²

This study addresses the lack of Australian-based populations for review of these systems. It aims to identify an optimal score to assess a patient's fitness for surgery, provide risk information to patients and patient families and, most importantly, identify high-risk patients which may be of use to structure the provision of resources during these admissions.

Methods

Ethics approval

Approval was granted by the Princess Alexandra hospital Human Research and Ethics Committee and Site Governance (HREC/14/QPAH/240 and SSA/14/QPAH/201), with a waiver for patient consent granted.

This study was performed at the Princess Alexandra Hospital, a tertiary hospital in Queensland Australia. Data were stored on the Princess Alexandra Hospital information system and encrypted to preserve patient confidentiality and de-identification.

Data were entered in an IBM SPSS, version 21.0, Statistics for Windows spreadsheet (Armonk, NY) and used for most statistical analyses. Collected data were split into two sets: the calibration set composed of 130 patients and the validation set comprised of 65 patients. To validate the effectiveness of the models, the calibration set (or training set) was used for most analyses. The validation data set was reserved for analysis of the scores at an individual level. The predicted outcomes of the calibration set were then compared against the observed outcomes to assess the models' predictive power in terms of sensitivity and specificity. Patients were randomly assigned a number to generate these data sets (calibration and validation).

As this was an exploratory analysis, we did not inflate the sample size nor was the significance value threshold adjusted (p value = 0.05) to correct for multiple hypothesis testing.

Data were retrospectively collected from patients aged 65 years and above who underwent an operation for a fractured neck of femur between January 2010 and 2013. All hip fracture surgeries in this time period were randomised and 198 individuals selected (by assigning random patient ID numbers). Three cases were excluded from the study because they did not match the age criteria (age < 65 years). Data were collected on factors related to the criteria of the three tested scores, NHFS, AACCI and POSSUM (see Appendix 1). Total possible scores for NHFS (0–10), AACCI (0–39) and POSSUM (18–132) were calculated. In addition, further physiological and operative data were collected including

type of anaesthetic used, cause of fractured neck of femur, and operation performed. The POSSUM scoring system has several criteria related to surgical pathology, which was not uniformly available. In particular, intraoperative blood loss was inconsistently recorded. An estimation was made, in line with the haemoglobin balance method described by Gao et al.¹³ where a decrease of 10 g/L from pre-operative to post-operative haemoglobin was considered equivalent to 500 mL of blood loss. Similarly, if the patient received a transfusion, each transfusion was thought equivalent to an equal volume (500 mL).

Mortality data in the form of month and year of death were collected from the Queensland Death Register and linked to within 30-day mortality, 6-month mortality and 12-month mortality. Patients not present on the register were assumed to remain alive at the time of data collection.

Performance of scores of interest

The AACCI, POSSUM and NHFS systems were analysed in several ways. Area under the receiver operating characteristic (ROC) curve was analysed for within 30-day, 6-month and 12-month mortality.¹⁴ Six- and 12-month morbidity data (complicated vs uncomplicated) were likewise assessed. Second, binary logistic regression was performed on the total score from each model for within 30-day, 6-month and 12-month mortality with performance assessed by Nagelkerke R^2 and sensitivity and specificity analysis. A cut-off value of 0.5 was taken for each score to categorise a patient into predicted deceased and alive groups (i.e. if a patient's predicted mortality was found to be >50% by any of the scores they were grouped as 'more likely to die than not').

Screening for prognostic variables

Variables within the three scoring systems of interest, in addition to anaesthetic used, cause of fractured neck of femur and operation performed were analysed. Initially, scale variables from the entire data set were analysed with correlation tables for 30-day, 6-month and 12-month mortality. Variables excluded in this process were those with p values exceeding 0.1.

Of the remaining variables a non-parametric independent sample analysis was performed, again excluding variables with p -values exceeding 0.1. Categorical variables were selected with use of crosstabs analysis, excluding variables with p -values greater than 0.1. Finally, the remaining variables from this process were individually analysed using univariate logistic regression.

Determining model variance at individual level

A Venn diagram was generated using InteractiVenn, a web-based tool developed by Heberle et al.¹⁵ on validation data, to identify if each model was predicting a similar group of

Table 1. Characteristics of participants within the calibration and validation data sets.

	Calibration data set (n = 130)	Validation data set (n = 65)
Age	84 (66–108)	82 (66–98)
Sex (% male)	37 (29%)	24 (37%)
Admission Hb (g.L ⁻¹)	116.5 (17.8)	121.3 (15.8)
Admission MMTS ≤ 6	44 (34%)	19 (29%)
Presence of CAD	47 (36%)	26 (20%)
Abnormal CXR findings	28 (22%)	13 (10%)
Presence of renal disease	16 (12%)	11 (9%)
Living in institution	47 (36%)	20 (15%)
Death within 30 days	16 (12%)	3 (5%)
Death at 6 months	30 (23%)	18 (28%)
Death at 12 months	41 (32%)	20 (31%)
Death at follow-up	68 (52%)	31 (48%)
Months to follow-up	33 (0–62)	34 (0–54)
NHFS	5 (3–8)	5 (3–8)
AACCI	5 (2–16)	5 (2–20)
POSSUM	40 (30–66)	40 (31–63)

Hb: admission haemoglobin (g.L⁻¹); MMTS: mini-mental test score (out of 10); CAD: coronary artery disease; CXR: chest x-ray; NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality. Values are mean (SD), median (IQR (range)) or number (proportion).

Table 2. Receiver operating characteristic analysis on calibration data set (n = 130) for mortality prediction.

Risk score	<30 day		6 month		12 month	
	AUC (95% CI)	p value	AUC (95% CI)	p value	AUC (95% CI)	p value
NHFS	0.760 (0.631–0.888)	0.001	0.761 (0.665–0.856)	<0.001	0.781 (0.698–0.864)	<0.001
AACCI	0.773 (0.664–0.883)	<0.001	0.749 (0.657–0.841)	<0.001	0.791 (0.708–0.874)	<0.001
POSSUM	0.769 (0.669–0.869)	0.001	0.741 (0.638–0.845)	<0.001	0.763 (0.667–0.860)	<0.001

AUC: area under the curve; CI: confidence interval; NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality.

patients within our population that were expected to be deceased 12 months post-operatively. In addition, McNemar cross-tabulation analysis was performed on predicted outcomes between models to determine if they were significantly different from one another ($p < 0.05$).

Results

Of the 195 patients recruited in this study, 69% were women with a median age of approximately 83 years of age. On average patients had an admission haemoglobin of 118 g/L, greater than 1 comorbidity with a mean admission mini-mental test score (MMTS) of seven. Sixty-seven patients came from institutions (34%). Of all comorbidities, coronary artery disease (CAD) was by far the most common, with 73 patients (37%) having known disease. One-hundred and fifty-four patients (79%) had normal pre-operative chest radiographs. General anaesthetic was used in 147 cases (75%), with regional anaesthetic performed in 40 (21%) and a combination used in 8 (4%) of all cases. Nineteen (10%) patients were deceased by 30 days, 48 (25%) of patients were

deceased at 6 months and 61 (31%) patients were dead at 12 months post their operation date. Baseline characteristics of our population are presented in Table 1.

Ninety-six patients remained alive at the time of follow-up. The three models demonstrated areas under the curve greater than 0.7 for mortality, indicating they are capable predictors of mortality at each time point (Table 2).

Sensitivity and specificity analysis for the three pre-developed scores was completed with ROC curves for three time periods, 30-day, 6-month and 12-month mortality, graphically represented in Figures 1–3.

The AACCI had the largest area under the ROC for mortality at both 30 days and 12 months with area under the curve (AUC) of 0.773 and 0.791, respectively. The NHFS had the largest AUC for 6-month mortality of 0.761. However, the three models' areas under the ROC curve were not statistically significant from one another (Hanley & McNeil comparison $p > 0.05$; Table 3).

Sensitivity and specificity analysis via logistic regression for the scores at the various time points has been presented in Table 4 below. Within 30-day mortality sensitivity was low

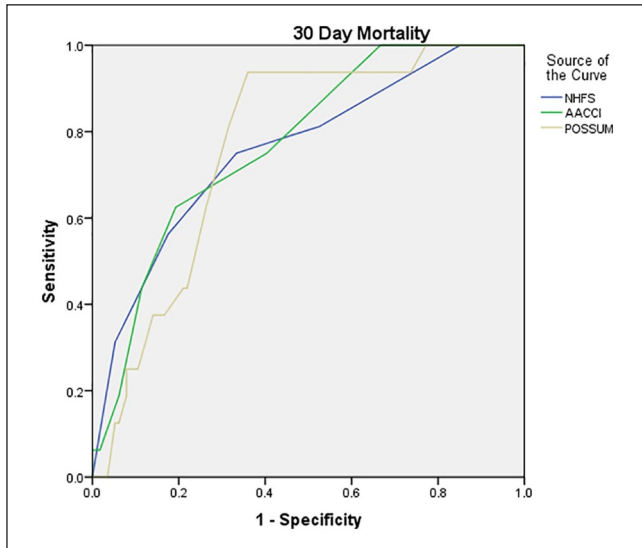


Figure 1. Receiver operating characteristic curve for within 30-day mortality prediction on calibration data set ($n = 130$). NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality.

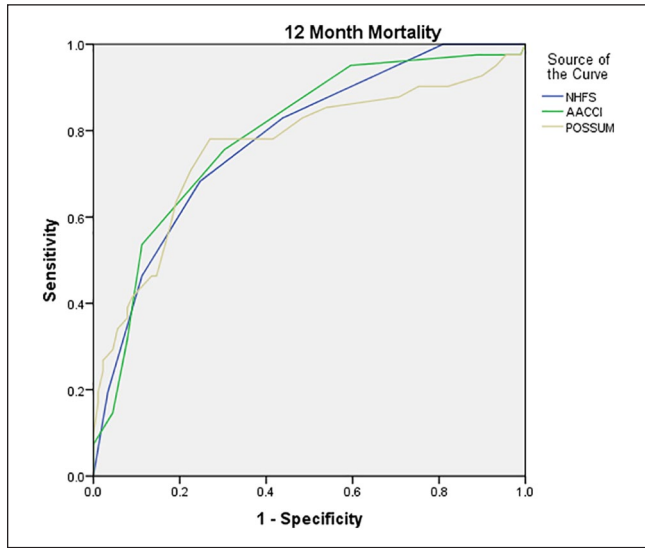


Figure 3. Receiver operating characteristic curve for 12-month mortality prediction on calibration data set ($n = 130$). NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality.

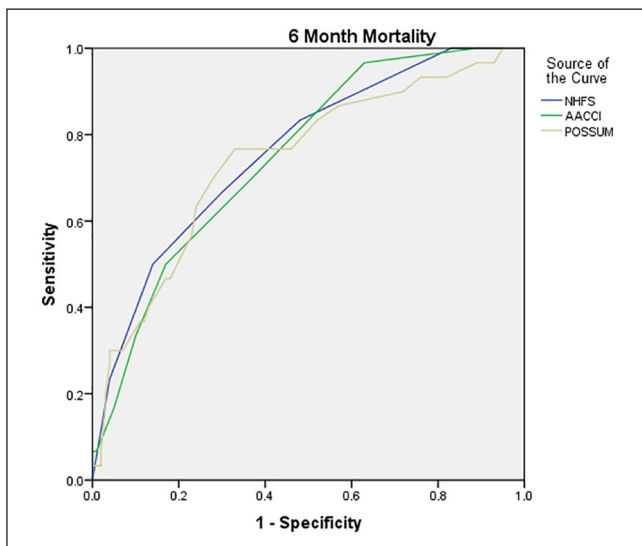


Figure 2. Receiver operating characteristic curve for 6-month mortality prediction on calibration data set ($n = 130$). NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality.

for all scores as the models were predicting most, if not all, patients to remain alive in this time period. The scores corresponding to a predicted mortality of 50% (the threshold value for determining ‘more likely to die than not’) are presented in Table 5.

Variables of significance differed slightly at each time point. For clarity, this is presented in Appendix 2.

Multivariate logistic regression was performed on the identified variables of significance, determining that the following were predictors of 12-month mortality: age, sex (male), admission haemoglobin (g/L), admission MMTS (0–10), presence of CAD, presence of renal disease and chest radiograph findings. Age, admission haemoglobin and mental test score were evaluated as both categorical and continuous variables, where the variables in their continuous form improved the predictive power of the model. Chest radiographs were classified as either normal or abnormal - where findings of cardiac changes (i.e. cardiomegaly), presence of emphysematous change, pulmonary oedema, pleural effusion, malignancy, would be classified as abnormal. The coefficients and odds ratios of these variables are shown in Table 6. Increasing age and male sex conferred a worse risk prediction, while a higher admission haemoglobin and MMTS resulted in lower risk scores.

Nagelkerke R^2 values indicate the level of variance in the data explained by each model (Table 6).

Calibration validation was performed with a Hosmer–Lemeshow goodness of fit, whereby values of less than 0.05 demonstrate a significant lack of fit. Each of the models, therefore, demonstrated a reasonable level of fit to the data (Table 7).

Venn diagram analysis reveals that the individuals within the validation population appear to be similarly identified and grouped by the four models. Of the 65 participants, 51 (79%) were placed in the same groups (alive or deceased) by the three models at 12 months post-operatively. The model predictions were analysed against one another using McNemar cross-tabulation, which determined that there was

Table 3. Area under ROC curve comparison in mortality prediction. P values presented below.

	NHFS vs AACCI	NHFS vs POSSUM	AACCI vs POSSUM
Within 30 days	0.83	0.90	0.95
6 month	0.81	0.77	0.89
12 month	0.82	0.77	0.62

ROC: receive operating curve; NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality. Using the Hanley and McNeil method on calibration data set (n = 130).

Table 4. Sensitivity, specificity and positive predictive value and negative predictive value of scoring systems for mortality.

Risk score	NHFS			AACCI			POSSUM		
	<30 day (%)	6 month (%)	12 month (%)	<30 day (%)	6 month (%)	12 month (%)	<30 day (%)	6 month (%)	12 month (%)
Sensitivity	0	23.3	46.3	6.3	16.7	31.7	0	26.7	41.5
Specificity	100	96	88.8	98.2	95	92.1	99.1	96	91.5
PPV	0	63.6	65.5	33.3	50	65	0	66.7	68
NPV	100	80.7	78.2	88.2	79.2	74.5	87.6	81.4	77.1

NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality; PPV: positive predictive value; NPV: negative predictive value. Calibration data set utilised (n = 130).

Table 5. Score values corresponding to predicted mortality of 50% at varying time points via logistic regression.

Risk score	Total score		
	Within 30 day	6 month	12 month
NHFS	9	7	6
AACCI	11	9	7
POSSUM	64	54	48

NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality. Scores rounded to nearest integer. Calibration data set utilised (n = 130).

no significant difference in predictions at the individual level between models (p value > 0.05; Table 8).

Type of anaesthesia used, general versus spinal or a combination, had no impact on mortality at any time point (chi-squared p > 0.05; Table 9).

Discussion

This study evaluated the performance of three well-known predictive models for mortality in elderly patients following hip fracture. All scores assessed were predictive of mortality prediction at all time points. There was no statistically significant difference of the area under ROC curve between the models at any time point. Of the three pre-developed scores, the AACCI had the highest sensitivity and specificity at 30 days and 12 months with AUC of 0.773 and 0.791, respectively. The NHFS yielded the highest sensitivity and specificity at 6 months, AUC of 0.761. The NHFS model explained the greatest proportion of variability in the mortality data at every time point, with the largest Nagelkerke R² (Table 7).

Prior to this study, there was a lack of direct comparison of the many available mortality prediction systems in an orthopaedic setting. Burgos et al.¹⁶ evaluated six different scoring systems, including the CCI and POSSUM, in an elderly hip fracture patient population. Following this study, direct, independent comparison of multiple scoring systems on one population has been rare. Karres et al.¹⁷ investigated six prediction models, including the three we have evaluated in our study, for 30-day mortality prediction. They found that all models, except POSSUM, demonstrated an acceptable level of discriminative power for short-term mortality prediction.¹⁷

To our knowledge, our study represents one of the few independent comparative analyses of available risk models for both short- and long-term mortality prediction. Direct comparisons of our scores of interest are presented in Table 7, where the Nagelkerke R² value indicates the level of variance explained by each score. There was no statistically significant difference between each of the models.

Table 6 examines the significance of the variables of all of the scores of interest in our population. The variables that remain in Table 6 are those that correlated at a significant (p < 0.05) level in our population with 12-month mortality. It is possible that these variables could form the basis of a new scoring system specific to our population and should be examined in future research across a variety of Australian clinical sites.

In clinical practice, it is important that a score is easy to implement, quick to use and depends on as few variables as possible, while maintaining its predictive accuracy. If a score is to be used for assessment of fitness for surgery or resource allocation during surgery or in the perioperative period, it is further desirable that the score be only reliant on variables that can be obtained prior to the surgery taking place. Taking

Table 6. Multivariate logistic regression analysis of available variables.

Variable	Coefficient	Odds ratio	95% CI	p value
Age	0.087	1.091	1.011–1.178	0.026
Sex (male)	1.352	3.865	1.171–12.755	0.037
Admission Hb	−0.033	0.968	0.939–0.998	0.000
Admission MMTS	−0.262	0.769	0.669–0.885	0.010
CAD	1.353	3.870	1.374–10.898	0.026
CXR findings	1.409	4.093	1.185–14.136	0.053
Renal disease	1.452	4.160	0.985–17.577	0.026

CI: confidence interval; Hb: haemoglobin ($\text{g}\cdot\text{L}^{-1}$); MMTS: mini-mental test score (out of 10); CAD: coronary artery disease; CXR: chest x-ray.

Table 7. Logistic regression analyses of models on calibration data set ($n = 130$) for mortality prediction.

Risk score	<30 day			6 month			12 month		
	Nagelkerke R^2	p value	Hosmer Lemeshow	Nagelkerke R^2	p value	Hosmer Lemeshow	Nagelkerke R^2	p value	Hosmer Lemeshow
NHFS	0.190	0.001	0.723	0.228	<0.001	0.758	0.289	<0.001	0.771
AACCI	0.155	0.002	0.281	0.192	<0.001	0.385	0.285	<0.001	0.090
POSSUM	0.137	0.002	0.084	0.191	<0.001	0.617	0.273	<0.001	0.078

NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for Enumeration of Mortality.

Table 8. McNemar cross-tabulation of variation between models at individual level on validation data set ($n = 65$) for 12-month mortality prediction.

	NHFS vs AACCI	NHFS vs POSSUM	AACCI vs POSSUM
McNemar p value	1.000	0.774	1.000

NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for Enumeration of Mortality.

Table 9. General anaesthetic versus regional versus combination on mortality outcome.

	Pearson chi-square	Degrees of freedom	p value
Within 30-day mortality	0.620	2	0.733
6-month mortality	4.045	2	0.132
12-month mortality	2.526	2	0.283

Results from chi-square analysis of entire data set ($n = 195$) presented above.

this into account, it would seem that the NHFS is best suited for clinical use, as it is composed of only seven variables and validated in several populations.^{11,18–21} Indeed, Marufu et al.¹⁸ determined via meta-analysis of several scores, including those assessed in this article, that the NHFS may be best suited for clinical implementation.

Comparing the application of these scores at an individual level must also be considered. Although two separate scoring systems may predict a similar mortality at a population level, the makeup of these predictions at an individual level may differ. If the models investigated in this study were predicting

mortality for different subgroups of our population, it would be reasonable to suggest the use of multiple scores to predict the mortality risk of an individual. To evaluate this, a chi-square analysis of individuals predicted to die at 12 months was performed between our available models. There was no statistically significant difference between the models, indicating that they are predicting a similar subgroup of our population at risk. This is graphically represented in a Venn diagram (Figure 4).

There remains some debate as to the utility of such scoring systems in the clinical setting. We argue that they are of use for several reasons. Scoring systems allow anaesthetists and surgeons alike to identify higher risk patients, which may alter management in the intraoperative and perioperative period, such as the implementation of multidisciplinary teams. The ‘shared-care’ model of orthogeriatric teams has been shown to improve both short- and long-term mortality.²² Given the significant cost of hip fractures in Australia and worldwide,³ identifying lower risk patients who may not require the same level of invasive monitoring in the perioperative period can have important implications from a health economic point of view. Whether varying the instituted

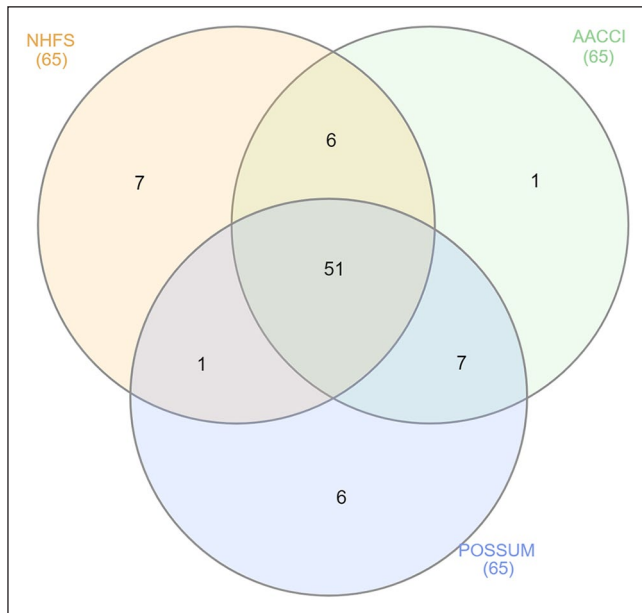


Figure 4. Venn diagram depicting the overlap of predicted categories (alive vs deceased) of the three models for an individual (n = 65).

NHFS: Nottingham Hip Fracture Score; AACCI: Age-Adjusted Charlson Comorbidity Index; POSSUM: Physiological and Operative Severity Score for enUmeration of Mortality.

measures for patients according to risk categorisation alters outcome remains to be seen and represents an area of interest for future research.

Our findings suggest that there is no difference in post-operative mortality (short or long term) when comparing general versus regional anaesthesia. This is in keeping with previous literature concluding there is no difference in mortality or post-operative complications including myocardial infarction, cerebrovascular incidents or delirium dependent on anaesthesia used.^{23–26}

The limitations of our study largely revolve around its relatively small sample size and retrospective design. Intraoperative blood loss was poorly recorded, resulting in the use of an estimate of this variable. Unfortunately, the data relating to the post-operative recovery period was lacking, mostly due to the loss of contact details of the patients. This study was limited to a single centre, with a relatively small data set. To address these concerns, a prospective study over a variety of clinical sites would be beneficial to further compare the available scores in an Australian population. The formation of a nationwide registry, such as present in the United Kingdom, would greatly help in this endeavour.

In conclusion, though all the pre-developed scores were capable of predicting both short- and long-term mortality, the NHFS demonstrates qualities that facilitate its ease of implementation in clinical practice, therefore it is our preferred score of choice. The formation of an Australian registry would greatly assist in future research endeavours which should aim

to prospectively identify an optimal scoring system for use in this population and determine its impact in helping clinicians guide management and resource distribution.

Authors contributions

P.S. formed the aims and methodology for the article and provided literature analysis of pre-existing data. M.J.N. gathered, analysed and interpreted data utilised in this study. J.S. played a vital role in the statistical analysis of the data. M.J.N. was a major contributor in writing of the manuscript. All authors contributed to editing and final drafting. All authors read and approved the final manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval and consent to participate

Approval was granted by the Princess Alexandra hospital Human Research and Ethics Committee and Site Governance (HREC/14/QPAH/240 and SSA/14/QPAH/201), with a waiver for patient consent granted. Ethical approval for this study was obtained from Metro South Health Human Research Ethics Committee – HREC/14/QPAH/240.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Informed consent

Waiver of consent approved.

ORCID iD

Michael James Nelson  <https://orcid.org/0000-0003-4607-7639>

Availability of data and material

The data sets generated and/or analysed during the current study are not publicly available as this was not a condition put forward in initial ethics approval, but are available from the corresponding author on reasonable request.

References

1. Williams N, Hardy BM, Tarrant S, et al. Changes in hip fracture incidence, mortality and length of stay over the last decade in an Australian major trauma centre. *Arch Osteoporos* 2013; 8: 150.
2. Stephens AS, Toson B and Close JC. Current and future burden of incident hip fractures in New South Wales, Australia. *Arch Osteoporos* 2014; 9: 200.
3. Ireland AW and Kelly PJ. Total length of stay, costs and outcomes at final discharge for admitted patients with hip fracture: linked episode data for Australian veterans and war widows. *Intern Med J* 2013; 43(12): 1280–1286.
4. Fisher AA, O'Brien ED and Davis MW. Trends in hip fracture epidemiology in Australia: possible impact of bisphosphonates and hormone replacement therapy. *Bone* 2009; 45(2): 246–253.

5. Hollingworth W, Todd CJ and Parker MJ. The cost of treating hip fractures in the twenty-first century: short report. *Osteoporos Int* 1996; 6(Suppl. 2): 13–15.
6. Sanders KM, Nicholson GC, Ugoni AM, et al. Health burden of hip and other fractures in Australia beyond 2000. Projections based on the Geelong Osteoporosis Study. *Med J Aust* 1999; 170(10): 467–470.
7. Kammerlander C, Gosch M, Kammerlander-Knauer U, et al. Long-term functional outcome in geriatric hip fracture patients. *Arch Orthop Trauma Surg* 2011; 131(10): 1435–1444.
8. Cooper C. The crippling consequences of fractures and their impact on quality of life. *Am J Med* 1997; 103(2A): 12S–17S; discussion 17S–19S.
9. Neuhaus V, King J, Hageman MG, et al. Charlson comorbidity indices and in-hospital deaths in patients with hip fractures. *Clin Orthop Relat Res* 2013; 471(5): 1712–1719.
10. Copeland GP, Jones D and Walters M. POSSUM: a scoring system for surgical audit. *Br J Surg* 1991; 78(3): 355–360.
11. Maxwell MJ, Moran CG and 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): 511–517.
12. Scandol JP, Toson B and Close JC. Fall-related hip fracture hospitalisations and the prevalence of dementia within older people in New South Wales, Australia: an analysis of linked data. *Injury* 2013; 44(6): 776–783.
13. Gao FQ, Li ZJ, Zhang K, et al. Four methods for calculating blood-loss after total knee arthroplasty. *Chin Med J* 2015; 128(21): 2856–2860.
14. Hanley JA and McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148(3): 839–843.
15. Heberle H, Meirelles GV, da Silva FR, et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 2015; 16: 169.
16. Burgos E, Gomez-Arnau JI, Diez R, et al. Predictive value of six risk scores for outcome after surgical repair of hip fracture in elderly patients. *Acta Anaesthesiol Scand* 2008; 52(1): 125–131.
17. Karres J, Heesakkers NA, Ultee JM, et al. Predicting 30-day mortality following hip fracture surgery: evaluation of six risk prediction models. *Injury* 2015; 46(2): 371–377.
18. Marufu TC, Mannings A and Moppett IK. Risk scoring models for predicting peri-operative morbidity and mortality in people with fragility hip fractures: qualitative systematic review. *Injury* 2015; 46(12): 2325–2334.
19. Wiles MD, Moran CG, Sahota O, et al. 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): 501–504.
20. Moppett IK, Wiles MD, Moran CG, et al. The Nottingham Hip Fracture Score as a predictor of early discharge following fractured neck of femur. *Age Ageing* 2012; 41(3): 322–326.
21. Kau CY and Kwek EB. Can preoperative scoring systems be applied to Asian hip fracture populations? Validation of the Nottingham Hip Fracture Score (NHFS) and identification of preoperative risk factors in hip fractures. *Ann Acad Med Singapore* 2014; 43(9): 448–453.
22. Grigoryan KV, Javedan H and Rudolph JL. Orthogeriatric care models and outcomes in hip fracture patients: a systematic review and meta-analysis. *J Orthop Trauma* 2014; 28(3): e49–e55.
23. Ilango S, Pulle RC, Bell J, et al. General versus spinal anaesthesia and postoperative delirium in an orthogeriatric population. *Australas J Ageing* 2016; 35(1): 42–47.
24. Guay J, Parker MJ, Gajendragadkar PR, et al. Anaesthesia for hip fracture surgery in adults. *Cochrane Database Syst Rev* 2016; 2: CD000521.
25. Basques BA, Bohl DD, Golinvaux NS, et al. General versus spinal anaesthesia for patients aged 70 years and older with a fracture of the hip. *Bone Joint J* 2015; 97-B(5): 689–695.
26. Zuo D, Jin C, Shan M, et al. A comparison of general versus regional anaesthesia for hip fracture surgery: a meta-analysis. *Int J Clin Exp Med* 2015; 8(11): 20295–20301.

Appendix 1. Grouped characteristics present in tested models. Note variables below do not reflect all variables present in each model.

Variables	Scoring system		
	POSSUM	AACCI	NHFS
Age	✓	✓	✓
Gender	×	×	✓
Examination findings (i.e. vital signs)	✓	×	×
Institutional status (i.e. lives in nursing home)	×	×	✓
Number of comorbidities	×	×	✓
Individual comorbidities	×	✓	×
Presence of malignancy	✓	✓	✓
Mental test score	×	×	✓
Glasgow coma score	✓	×	×
Blood test results (i.e. serum urea)	✓	×	✓
ECG	✓	×	×
Chest radiograph findings	✓	×	×
Operation-related data	✓	×	×

Appendix 2. Identified variables of significance for mortality prediction at various time points.

Variables	Within 30 days	6 month	12 month
Age	✓	✓	✓
Sex	✓	✓	✓
Admission Hb	✓	✓	✓
Admission MMTS	✓	✓	✓
Radiograph findings	✓	✓	✓
Cause of fracture	✓	✓	✓
Dementia	✓	×	×
CHF	✓	✓	×
Renal impairment	✓	×	✓
Number of comorbidities	×	✓	×
ECG findings	×	×	✓
CAD	×	×	✓

Hb: admission haemoglobin (g.L⁻¹); MMTS: mini-mental test score (out of 10); CAD: coronary artery disease; CXR: chest x-ray; COAD: chronic obstructive airways disease; CHF: congestive heart failure.