Development And Validation Of Multivariable Prediction Models For In-Hospital Death, 30-Day Death, And Change In Residence After Hip Fracture Surgery And The 'Stratify-Hip' Algorithm

Aicha Goubar PhD,[1] Finbarr C. Martin MD,[1] Catherine Sackley PhD,[1] Nadine E Foster PhD,[2,3] Salma Ayis PhD,[1] Celia L Gregson PhD,[4] Ian D Cameron PhD,[5] Nicola E. Walsh PhD,[6] Katie J. Sheehan PhD[1*]

[1] School of Life Course and Population Health Sciences, Faculty of Life Science and Medicine, King's College London, UK
[2] Surgical Treatment and Rehabilitation Service (STARS) Education and Research Alliance, The University of Queensland and Metro North Health, Queensland, Australia
[3] Primary Care Centre Versus Arthritis, School of Medicine, Keele University, UK
[4] Musculoskeletal Research Unit, Translation Health Sciences, Bristol Medical School, University of Bristol, UK
[5] John Walsh Centre for Rehabilitation Research, Northern Sydney Local Health District and University of Sydney, Australia
[6] Centre for Health and Clinical Research, University of the West of England Bristol, UK

[*]**Corresponding author**
Katie J Sheehan
katie.sheehan@kcl.ac.uk

ABSTRACT

Background:

To develop and validate the stratify-hip algorithm (multivariable prediction models to predict those at low, medium, and high risk across in-hospital death, 30-day death, and residence change after hip fracture).

Methods:

Multivariable Fine-Gray and logistic regression of audit data linked to hospital records for older adults surgically-treated for hip fracture in England/Wales 2011-2014 (development n=170,411) and 2015-2016 (external validation, n=90,102). Outcomes included time to in-hospital death, death at 30-days, and time to residence change. Predictors included: age, sex, pre-fracture mobility, dementia, and pre-fracture residence (not for residence change). Model assumptions, performance, and sensitivity to missingness were assessed. Models were incorporated into the stratify-hip algorithm assigning patients to overall low (low risk across outcomes), medium (low death risk, medium/high risk of residence change), or high (high risk of in-hospital death, high/medium risk of 30-day death) risk.

Results:

For complete-case analysis, 6,780 of 141,158 patients (4.8%) died in-hospital, 8,693 of 149,258 patients (5.8%) died by 30-days, and 4,461 of 119,420 patients (3.7 %) had residence change. Models demonstrated acceptable calibration (observed:expected ratio 0.90, 0.99, and 0.94), and discrimination (area under curve 73.1, 71.1 and 71.5; Brier score 5.7, 5.3, 5.6) for in-hospital death, 30-day death, and residence change, respectively. Overall, 31%, 28%, and 41% of patients were assigned to overall low-, medium-, and high- risk. External validation and missing data analyses elicited similar findings. The algorithm is available at https://stratifyhip.co.uk.

Conclusion:

The current study developed and validated the stratify-hip algorithm as a new tool to risk stratify patients after hip fracture.

KEYWORDS

Stratification, classification, fracture neck of femur, fragility fracture, recovery

INTRODUCTION

The age standardized rate of hip fracture ranges from lows of 2/100,000 in Nigeria (women) and 35/100,000 in Ecuador (men), to highs of 574/100,000 in Denmark (women) and 290/100,000 in Denmark (men) (1). Even with surgery, up to 10% of patients die in hospital and 22% transition from living at home to care homes (2). Multidisciplinary and orthogeriatric led management is the optimal approach for acute hospital care after hip fracture, resulting in fewer deaths and transitions to care homes (risk ratio 0.88, 95% confidence interval 0.80, 0.98) (3). Early and frequent therapy input is also associated with an additional 2% of patients returning home and 4% of patients surviving to 30-days (4). Yet, a demand-capacity mismatch often limits delivery of consistent orthogeriatric care (5) and therapy services in hospital after hip fracture (6). This mismatch requires clinicians to prioritise their caseload based on perceived need. However, variation in national audit data (above what may be explained by differences in case-mix) may suggest a lack of consistency in this prioritisation (5, 6).

A stratified approach to multidisciplinary care delivery may improve efficiency and reduce inconsistencies in prioritisation by identifying groups of patients at risk of poor outcomes to be matched to different treatments, acknowledging different needs and potential benefits from healthcare professional input. To achieve this, prediction models are necessary. Most previously published models have limitations in performance and/or implementation (7). In contrast, the Nottingham Hip Fracture Score has modest discrimination and adequate calibration for death and includes predictors which clinicians can collect prior to surgery (8). The Nottingham Hip Fracture Score was developed to inform consenting procedures, timing of surgery, access to pre/post operative higher level care, and audit (8). However, it was not designed to enable stratification of patients into different risk groups to be matched to different treatments.

This study aimed to develop and validate the stratify-hip algorithm (comprised of three multivariable prediction models) using routinely collected data available on admission as a new tool to risk stratify patients after hip fracture. The algorithm sought to predict those at low, medium, and high risk across time to in-hospital death, death by 30-days, and time to change in residence, and to be able to discriminate between the groups. The stratify-hip algorithm and link to a freely available web-based app to facilitate risk prediction is provided.

METHODS

This study is reported according to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement (9). The study did not require ethical approval as it involved secondary analysis of pseudonymized data.

Source of data

The UK National Hip Fracture Database (NHFD) collates data on the characteristics of 95% of patients aged 60 years and older with hip fracture and the care they received during the acute hospital stay in England and Wales (5). Individual patient NHFD data were linked to electronic hospital records and the Office of National Statistics records for data on dementia diagnosis and death respectively. Further details on data cleaning and linkage across databases are described elsewhere (10). Data

submitted to the NHFD for 170,411 patients surgically treated for a non-pathological first hip fracture between January 1st, 2011, and December 31st, 2014, were selected for development and internal validation (follow-up to 30-days post-admission or to December 31st, 2014). Data submitted for 90,102 patients treated between January 1st, 2015, and December 31st, 2016, were selected for external (temporal) validation (follow-up to 30-days post-admission or to December 31st, 2016). Differences between patients with and without complete predictor data are presented in Supplementary File, Table S1-2. Those with missing predictor data were more dependent (greater proportion with dementia [all outcomes] and admitted from nursing/residential care [in-hospital death and 30-day death]) than those without missing data.

Outcomes

Outcomes included (i) time to in-hospital death, (ii) death status at 30-days post admission and (iii) time to change in residence (among those admitted from home) as key performance indicators of safe and effective care (5), and which reflect the patient priority of returning home (11). Time to in-hospital death was calculated as the number of days from admission to a coded discharge destination of death, treating discharge to another unit (loss to follow-up) or to day 30 (end of follow-up) as a censoring event, and discharge home or to nursing/residential care as a competing event. Thirty-day death was identified by a binary indicator (alive, dead) 30 days post admission. Time to change in residence (pre-fracture residence of home and discharge destination of nursing/residential care, Supplementary File, Table S3) was calculated as the number of days from admission to a residence change, treating discharge to another unit (loss to follow-up) or to day 30 (end of follow-up) as a censoring event, and return to pre-fracture residence or in-hospital death as competing events.

Predictors

Five predictors were included: age (5-year age groups from age 60 years at admission), sex (male, female), pre-fracture mobility (no functional mobility, independent indoor mobility with/without aid, independent outdoor mobility with/without aid), dementia diagnosis (International Classification of Diseases 10th Edition code F00, F01, F02, F03, or G30 during the hip fracture admission or any admission in the year prior), and pre-fracture residence (own home/sheltered housing, nursing/residential care [not for time to change in residence outcome]). Predictors were defined *a priori* following evidence review (2, 12), interviews with patients (11) and healthcare professionals (13, 14), a public and patient involvement focus group, and which were available in the dataset. To optimize simplicity of future implementation of risk stratification, the number of predictors was kept to a minimum and to those which could be feasibly collected preoperatively by any healthcare professional either directly from the patient or an informal/formal carer.

Sample size

Similarity between the external and development datasets was anticipated in terms of distribution of patients by each predictor. The extent to which the external validation dataset sample size could be justifiable to perform a validation of our models was assessed. For the observed proportion of 30-day deaths (5.8%), an estimated minimum of 13,751 patients (with 798 deaths at 30-days) was required for 1) a target standard error (SE) of the logarithm (observed/expected (O/E) ratio) of 0.245

with a target O/E ratio of 1; 2) a target area under the receiver operating characteristic curve (AUC) of 0.70 with SE(AUC) of 0.0225; and 3) and a target calibration slope of 1 with a SE of 0.051 (15).

Statistical analysis methods

Predictors were described by counts and proportions.

Time-to-event outcomes

Fine - Gray (16) regression was used to build a prediction model estimating the direction of the association between predictors and the cumulative incidence (risk) of in-hospital death and of residence change as functions of postoperative day, accounting for competing events (17). Discharges to another care setting and hospital stays exceeding 30 post-operative days were treated as right-censored observations. The model was fitted with R (18) software using *RiskRegression* packages (19) and *cmprsk* (20). The proportional hazards assumption was assessed by plotting Schoenfeld residuals against failure time with a scatterplot smoother for each covariate in the models (21). Model calibration was assessed by estimating the O/E ratio and plotting the mean predicted risk against observed risk for predicted risk deciles at 30-days from admission (22). Model discrimination was assessed by C-index statistics at 30-days (or AUC) (23). The Brier score was also calculated as the expected squared distance between predicted and observed risk at 30-days from admission (24). This score accounts for both calibration and discrimination with a lower score (scale 0-100%) indicating a higher performing model (25).

Binary outcome

A five-predictors logistic regression model was used to predict the risk of death at 30-days. The presence of influential observations was examined by visualizing the Cook's distance values (26). Multicollinearity among predictors was investigated using "vif" in the "rms" R package, which computes the variance inflation factors (VIF) (27, 28). Model calibration was assessed by quantifying the calibration slope and the O/E ratio (29, 30). The mean predicted risk against observed risk for predicted risk deciles (22) was also plotted and a generalized additive model with integrated smoothness presented (31). Discrimination was assessed by the C statistic (29) and both calibration and discrimination by the Brier score (29). The analyses were conducted with R (18) using *rms* packages (32).

Internal validation

For internal validation, 100 bootstrap samples were generated with replacement from the development dataset (33). Each sample included the same number of patients as the development dataset. Performance was assessed from each bootstrap model in each bootstrap sample (apparent performance) and the performance of each bootstrap model in the development dataset (test performance) (34). Optimism (overestimation bias often due to overfitting) was calculated as the average difference between apparent and test performance across bootstrap samples (34). Optimism adjusted measures of performance (AUC and Brier scores) were estimated by subtracting the estimate of optimism from the development model performance estimates (34).

Risk groups

Patients were clustered into three risk groups (low, medium, high) for each outcome by applying K-means clustering algorithm (an algorithm which partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean) (35). The number of groups was defined a priori to balance risk assignment with the feasibility of designing future matched treatments for each risk group. Following clustering for each outcome, patients were assigned to one of three groups: overall low (low risk across outcomes), overall medium (low risk of death but medium or high risk of change in residence), and overall high (high risk of in-hospital death, high or medium risk of 30-day death) risk across outcomes (Figure 1).

Sensitivity analyses

Multiple chained equations (MICE) was used to determine the sensitivity of findings to data missingness (36). Twenty distinct datasets were generated for efficient and stable estimates (36). Missing predictor values were replaced iteratively with values from multiple regression models within the MICE in addition to auxiliary variables Charlson comorbidity index, American Society of Anaesthesiologists Classification, deprivation (Index of Multiple Deprivation decile groups), and type of surgery (arthroplasty, hemiarthroplasty, internal fixation) to minimise bias and optimise power of the imputations (37). As in the main analysis, either Fine - Gray or logistic regression models were used, as appropriate, to predict the risk of each outcome. The optimism corrected values of AUC and Brier score for each model within the 20 imputed datasets was estimated prior to generation of pooled values and their confidence intervals (CI) across imputed datasets (38).

The influence of age groups on model performance was assessed by treating age as a continuous variable in a sensitivity analyses of the development dataset.

External validation

The three models generated in the development dataset were applied to the external validation dataset to estimate the predicted risk. Performance was estimated through the AUC and Brier score. Risk groups were subsequently generated as described above.

Model access

The final model is accessible via a freely-available web-based app. Nomograms were also generated for each outcome to be used alongside Figure 1 for settings where internet access is not available (39). Nomograms were generated with R (18) packages *rms* (32) and *cmprsk* (20).

RESULTS

Participants

Among 170,411 patients, 141,158 (83%), 119,420 (70%), and 149,258 (88%) had complete data for predictors and in-hospital death, 30-day death, and change in residence, respectively. The majority were women, admitted from home and were able to ambulate outdoors pre-fracture (Table 1). More than half were over 80 years of age, and one quarter had a diagnosis of dementia.

Development and internal validation

In-hospital death

Among 141,158 patients, 6,780 (4.8%) died in-hospital, 72,401 (51.3%) were discharged, 48,798 (34.6%) were discharged to another unit, and 13,179 (9.3%) had stays longer than 30 days. Dementia and pre-fracture residence had non-constant residuals across time indicating a potential violation to the proportional hazard assumption. However, further exploration of an interaction with time indicated no major violation for these predictors (Supplementary File, Figure S1).

The predicted risk of in-hospital death was calculated using the "*predict.crr"* function in the *cmprsk* package which uses the formula:

$$1 - 0.9909^{\exp(LP_1)}$$

where 0.9909 is the baseline 30-day survival estimate and the linear predictor (LP_1) is equal to

$$(\beta_1 * \text{age}) + (\beta_2 * \text{sex}) + (\beta_3 * \text{prefracture mobility}) + (\beta_4 * \text{ prefracture residence}) + (\beta_5 * \text{dementia})$$

β1: 65-69 years: 0.3917426; 70-74 years: 0.6325601; 75-79 years: 0.8231233; 80-84 years: 1.2006534; 85-90 years: 1.5429587; 90-94 years: 1.8768660; 95 or more years: 2.2821073. β2: 0.6993337. β3: Indoor: 0.6518865; no function: 0.7273951. β4: -0.1337233. β5: 0.1670263.

The model was well calibrated as evidenced by a calibration plot of predicted against observed risk across deciles of predicted risk (Figure 2), with an overall calibration measured by the O/E ratio of 0.90 (from 0.89 to 1.36 across risk deciles) with a weaker fit for those in the risk groups (3%-3.3%) (4.2%-5.2%) and (5.9%-7.9%). AUC and Brier scores were similar for development and internal validation with optimism adjusted AUC and Brier scores of 73.1% (95% CI, 72.6-73.7) and 5.66% (95% CI, 5.57-5.79) respectively (Table 2).

The risk of in-hospital death was estimated at 32.0% for a man over the age of 94 years admitted from residential care with no pre-fracture mobility and a history of dementia. The risk of in-hospital death was estimated at 0.9% for a woman aged between 60 and 64 years admitted from home with outdoor mobility pre-fracture and no history of dementia.

Change in residence

Among 119,420 patients, 4,461 (3.7 %) had a change in residence, 51,178 (42.9 %) were discharged to their pre-fracture residence, 5,699 (4.8%) died in-hospital, 47,329 (39.6%) were discharged to another unit, and 10,753 (9.0%) had stays longer than 30 days. There was no evidence of violation to the proportional hazard assumption (Supplementary File, Figure S2).

The predicted risk of change in residence was calculated using the "*predict.crr"* function in the *cmprsk* package using the formula:

$$1 - 0.9800114^{\exp(LP_2)}$$

where 0. 9800114 is the baseline survival estimate for no change in residence and the linear predictor (LP_2) is equal to

$$(\beta_6 * \text{age}) + (\beta_7 * \text{sex}) + (\beta_8 * \text{prefracture mobility}) + (\beta_9 * \text{dementia})$$

β6: 65-69 years: 0.07259374 ; 70-74 years: 0.49701280 ; 75-79 years: 0.74251183; 80-84 years: 0.98564585 ; 85-90 years: 1.20026245; 90-94 years: 1.34492377; 95 or more years: 1.46828028. β7: -0.23046984. β8: Indoor: .10857114; no function: 0.48041962. β9: 0.65983936.

The model displayed a tendency towards underfitting for those in the 7th to 9th risk deciles (6.5% - 11.0%) as evidenced by a calibration plot of predicted against observed risk across deciles of predicted risk (Figure 2) with an overall calibration measured by the O/E ratio of 0.94 (range from 0.86 to 1.31 across risk deciles). AUC and Brier scores were similar for development and internal validation with optimism adjusted AUC and Brier scores of 71.5% (95% CI 70.8 – 72.5) and 5.5% (95% CI 5.4 – 5.7) respectively (Table 2).

The risk of change in residence was estimated to be 24.0% for a woman over the age of 94 years with no pre-fracture mobility and a history of dementia. The risk of a change in residence was estimated to be 1.6% for a man aged between 60 and 64 years with outdoor mobility pre-fracture and no history of dementia.

30-day death

Among 149,258 patients, 8,693 (5.8%) died by 30-days. Variance inflation factors indicated no collinearity (Supplementary File, Table S4). The standardized residual error presented in Supplementary File, Figure S3 revealed 16 data points (<0.5%) with an absolute standardized residuals above 3 which were investigated as possible outliers.

The predicted risk of in-hospital death was calculated by the formula:

$$\frac{1}{1 + exp^{-LP_3}}$$

where the linear predictor (LP_3) is equal to

$$\text{-4.6851772} + (\beta_{10} * \text{age}) + (\beta_{11} * \text{sex}) + (\beta_{12} * \text{prefracture mobility}) + (\beta_{13} * \text{prefracture residence}) + (\beta_{14} * \text{dementia})$$

β10: 65-69 years: 0.3213738; 70-74 years: 0.4358315; 75-79 years: 0.6889942; 80-84 years: 0.9946787; 85-89 years: 1.3435684; 90-94 years: 1.6353948; 95 or more years: 2.0773348. . β11: 0.7301712. β12: Indoor: 0.6327345; no function: 0.7915023. β13: 0.1929848. β14: 0.3310316.

The model displayed no tendency to under- or over-fitting as evidenced by a calibration plot of predicted against observed risk (Figure 2, Supplementary File Figure S4) with an overall calibration measured by the O/E ratio of 0.99 (ranged from 0.83 to 1.00 across risk deciles). AUC and Brier scores were similar for development and internal validation with optimism adjusted AUC and Brier scores of 71.1% (95% CI 70.6-71.6) and 5.30% (95% CI 5.20-5.40) respectively (Table 2). An optimism of 1.00 (95% CI 0.99-1.00) was identified for the calibration slope.

The risk of 30-day death was estimated to be 36.3% for a man over the age of 94 years admitted from residential care with no pre-fracture mobility and a history of dementia. The risk of in-hospital death was estimated to be 0.9% for a woman aged between 60 and 64 years admitted from home with outdoor mobility pre-fracture and no history of dementia.

Risk groups

Patients were clustered into low-, medium-, and high- predicted risk groups for each of the three outcomes using K-means clustering prior to assignment to mutually exclusive overall low (31% [n =44,364]), medium (28% [n =39,542]), and high (41% [n =57,251]), risk across all three outcomes (Figure 1). Patients in the overall low risk group were typically less than 80 years old with the majority female (66%), admitted from home (91%), with outdoor mobility pre-fracture (83%) and no dementia diagnosis (95%) (Table 3). Compared to the overall low risk group, a greater proportion of patients in the overall medium risk group were older (94% aged 80 years or more), female (99%) and had a dementia diagnosis (14%) (Table 3). Compared to the overall medium risk group, patients in the high-risk group were of a similar age (94% aged 80 years or more), however, a greater proportion were male (37%), with indoor/no mobility (69%), admitted from nursing/residential care (40%) with a dementia diagnosis (50%) (Table 3). The stratify-hip algorithm can be calculated online at: https://stratifyhip.co.uk. Nomograms for calculating the algorithm offline are available in Supplementary File, Figure S5.

Sensitivity analysis

For imputation results and where age was treated as a continuous predictor, summary performance statistics were comparable to performance estimates in the complete case analysis. Full results of sensitivity analyses are available in Supplementary file, Tables S5-6.

External (temporal) validation

Among the 90,102 patients in the validation dataset, 84,096 (93%), 87,144 (97%), and 87,414 (97%), had complete data for predictors and in-hospital death, 30-day death, and change in residence respectively. The majority were women, admitted from home and were able to ambulate outdoors pre-fracture (Table 1). More than half were over 80 years of age and one third had a diagnosis of dementia (Table 1).

In-hospital death

Among 84,096 patients, 3,752 (4.5%) died in-hospital, 44,348 (52.7%) were discharged, and 35,996 (42.8%) were censored by 30-days. Similar to the development dataset, a weaker fit was observed

for those in the risk groups (4.2%-6.8%) (Figure 2). AUC and Brier scores were comparable to the development dataset at 73.1% (95% CI, 72.7%-74.2%) and 5.26% (95% CI, 5.09%-5.43%) respectively.

Change in residence

Among 70,319 patients, 2,712 (2.1%) had a change in residence, 31,239 (24.6%) were discharged to their pre-fracture residence, 3,187 (2.5%) died in-hospital and 33,181 (26.1%) were censored by 30-days. Similar to the development dataset, the model displayed a tendency towards underfitting for those in the 7th to 9th risk deciles (6.5% - 11.0%) (Figure 2). AUC and Brier scores were comparable to the development dataset at 71.7% (95% CI 70.6 to 72.7) and 5.7% (95% CI 5.5 – 5.9) respectively.

30-day death

Among 87,414 patients, 4,790 (5.5%%) died by 30-days. Similar to the development dataset, the model displayed no tendency to under- or over-fitting as evidenced by a calibration plot of predicted against observed risk (Figure 2). AUC and Brier scores were comparable to the development dataset at 71.2% (95% CI 70.5%-0.71.9%) and 5.00% (95% CI 4.95-5.13) respectively (Table 2).

Risk groups

The distribution of patients to overall low (33% [n =27,566]), medium (31% [n =25,669]), and high (36% [n =30,861]), risk across outcomes and differences in the distribution of characteristics of patients across overall risk group were similar to the development dataset (Supplementary File, Table S7).

DISCUSSION

Main findings

This study developed and validated multivariable prediction models for in-hospital death, 30-day death and change in residence among 260,513 patients who underwent hip fracture surgery. The models enable the prediction of patients at low, medium, and high risk across the three outcomes. The models were well calibrated with acceptable discrimination between the groups consistently estimated during development, and internal and external validation. The stratify-hip algorithm (comprised of three multivariable prediction models) and link to a freely available web-based app to facilitate risk prediction is provided.

Comparison with other literature

Model performance measures for 30-day death were comparable between the current model and the Nottingham Hip Fracture Score (8). Similar to this score, older age, male sex, admission from a care home, and presence of dementia were predictive of 30-day death (8). The current model noted poorer pre-fracture mobility was also predictive of 30-day death. This predictor is not employed by the Nottingham Hip Fracture Score, which uses admission haemoglobin concentration, number of comorbidities and malignancy as its final three predictors (8). A direct comparison of model performance was not possible due to the absence of admission haemoglobin from the current dataset. For the current model, it was also noted pre-fracture mobility together with age, sex, pre-

fracture residence and dementia, was predictive of time to in-hospital death and change in residence. Although the predictors selected for inclusion in the current models have previously been associated with poor outcomes(11-13), here an algorithm is provided for considering prediction of these outcomes together at the point of admission. This outcome driven stratify-hip algorithm generated distinct risk groups whereby a patient with a given set of predictors (collected at the point of admission) is allocated to one risk group.

Interpretation

In the stratify-hip algorithm, the risk of short-term death was prioritised over changing residence. This 'outcome-driven' approach was guided by an expectation the clinical needs of patients will vary according to which outcomes they are at risk of incurring. The approach mirrors the likely priority order of in-hospital care to reduce death risk and subsequently the risk of changing residence. This is potentially controversial for those admitted from home, given a study from 2000 indicated 80% of 194 women at risk of hip fracture indicated they would rather be dead than admitted to a care home (40). K-means clustering was considered for the overall group assignment. However, this approach does not generate 'mutually exclusive' groups (a key requirement for employing the algorithm to inform a future stratified approach) and the analysis would need to be on the same population for all three outcomes (but 'changing residence' is only applicable to those admitted from home).

Given the weighting of the algorithm towards the high-risk group (Figure 1), it was surprising this group only constituted between 36% (external validation) and 41% (development) of the population. This is promising for the potential future clinical utility of the algorithm as it supports the hypothesis of distinct groups of patients within the population who may benefit from different care approaches. Indeed, the characteristics of patients exhibited an increasing level of dependency from low- to high-risk groups. For example, those in the low-risk group were the youngest and had the lowest proportion with dementia while those in the high-risk group were the oldest and had the highest proportion with dementia. This was to be expected given those with greater dependency are more likely to be at risk of short-term death (and high-risk assignment).

Future research

Predictors were selected based on existing evidence and to optimize simplicity of future implementation of the approach to risk stratification. Previously published models which incorporated more predictors yielded similar performance (7). The exception is the Orthopaedic Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity with a reported area under the curve of 83%; however, this score requires collection of physical (e.g., bloods, electrocardiogram) and operative (blood loss) severity data limiting implementation. As model performance was deemed acceptable with the five predictors identified, additional predictors were not explored here.

The type and intensity of clinical care may vary depending on the needs of an individual patient and organisational culture (3). The stratify-hip algorithm presented here is a first step in developing a stratified approach to care. The next step is to match risk groups to interventions tailored to their needs. Indeed, the risk groups likely benefit from differing in-hospital care (type, intensity, professional input) which may help to mitigate the demand-capacity mismatch for orthogeriatric and

therapy input and reduce inconsistencies in prioritisation (5, 6). For example, patients in the high risk group may require interventions which explicitly target risk factors for short-term death such as closer monitoring/managing perioperative medical complications (41), early mobilisation (42), and/or consideration of end-of-life care (43). Conversely, those not at risk of short-term death but at risk of changing residence (medium-risk group) may be more appropriate for early supported discharge (44). The goal of such an approach is to optimise outcomes across the entire population by ensuring equitable access to person-centred care, and not to ration care based on poorer prognosis. Any future stratified approaches should be assessed for feasibility, acceptability, and effectiveness in randomized controlled trials (45).

Previous research highlighted challenges in communicating prognosis with patients, particularly in end-of-life care (46). These challenges were attributed to low confidence in prognostic estimates which was improved with the use of prediction models (46). The acceptability of the stratify-hip algorithm to patients, carers, and professional as a tool to support shared decision making should be explored in future research. If deemed acceptable, the algorithm may be used by professionals to help set expectations for recovery in a timely manner with patients and their informal/formal carers.

The current study focused on short-term outcomes. It may be hypothesised short-term risk of death and/or changing residence may be related to longer-term risk of mobility loss, changing residence, and/or death. Future research may assess model performance in predicting longer-term outcomes, which in turn may inform community care.

Limitations

Age was treated in 5-year increments and not as continuous with the intention of enabling paper-based implementation which may have led to a loss of power (47). However, similar model performance was noted when age was treated as continuous in a sensitivity analysis. Dementia was based on the absence or presence of a formal diagnosis code in hospital records which may be subject to under diagnosis and subsequent misclassification in our models (48). However, the data source employed here recently had the highest reporting of dementia diagnoses across three UK data sources (48) and the ascertainment rate of 25-30% is in keeping with the expected rate among older adults admitted to hospital with hip fracture (49). Date of death was not available limiting the ability to consider death at 30-days (inclusive of deaths after discharge) as a time-to-event outcome or to employ a multivariate analysis. There was a high proportion of right censored observations for time-to-event outcomes due to discharges to other care setting. These observations are unlikely independent of the predictor-outcome association given patients admitted with no mobility may be less likely discharged to inpatient rehabilitation than those admitted with mobility prefracture. Groups were defined using K-means clustering algorithm which relies on researchers assumptions to a greater extent than other methods that rely on formal tests e.g., latent class analysis (50). However, K-means clustering based on assumed three groups, led to the identification of three groups whose characteristics are likely amenable to different matched treatments which was the purpose of the algorithm. There is the potential for bias due to missing data. The sensitivity of the complete case analysis to missingness was assessed by imputation which estimated similar performance. This was not surprising given imputed datasets were predominantly comprised of complete data (88%).

The stratify-hip algorithm may not be generalisable to settings where hospital stay varies from the average UK length of stay 15 days (5). Temporal external validation was completed and yielded similar performance. Further external validation for more recent years (given implications of changes in social care funding for discharge destinations and in death rates over time), for different settings (given implications of health systems on outcomes) and using different definitions and/or measurements in similar patients is required to determine the international utility of the model.

CONCLUSIONS

The current study details the new stratify-hip algorithm (comprised of three multivariable prediction models) enabling the identification of three distinct groups (low- 31%, medium- 28% and high- 41% risk) at differing risk of poor outcome after hip fracture surgery. The multivariable prediction models were well calibrated with acceptable discrimination during development and validation. Future research should seek to develop and test the feasibility and acceptability of the algorithm to group patients and match them to interventions. External validation beyond temporal validation is also recommended.

CONFLICT OF INTEREST

KS received a grant from UK Research & Innovation Future Leaders Fellowship to support this work. This funding provides salary support for KS. KS also received funding from the National Institutes of Health Research (NIHR) and Chartered Society of Physiotherapy Charitable Trust for hip fracture health services research. KS is the Chair and CG a member of the Scientific and Publications Committee of the Falls and Fragility Fracture Audit Programme which managed the National Hip Fracture Database audit at the Royal College of Physicians. FCM was the funded (2012-2018) board chair of the Falls and Fragility Fracture programme. CS, NEF and NW receive funding from the National Institute for Health Research (NIHR). CS and NEF are NIHR Senior Investigators. Salma Ayis was funded/supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. CLG receives funding from Versus Arthritis (ref 22086). GSdP, SA, and IDC have no competing interests to declare.

ACKNOWLEDGEMENTS

This work was supported by a UKRI Future Leaders Fellowship [Grant Ref: MR/S032819/1].

KS conceived the study and led on data curation, funding acquisition, investigation, project administration, resources, software, and writing the original draft. AG led on formal analysis, methodology, validation and visualization. KS supported validation and visualisation. AG supported data curation, investigation and writing the original draft. KS, SA supported the formal analysis and supervision. KS, FCM, CS, NF, CG, IC, NW supported the methodology. AG, FCM, CS, NF, SA, CG, IC, NW supported study conception. KS, AG, FCM, CS, NF, SA, CG, IC, NW played an equal role in review and editing subsequent iterations of the draft.

Thanks to Dr Orouba Almilaji who developed the website for the stratify-hip algorithm which accompanies this manuscript. We are grateful to NHS Digital, NHS Wales Informatics Service, and the Royal College of Physician's Falls and Fragility Fracture Audit programme for providing the data used in this study. The views expressed in this publication are those of the authors and do not necessarily reflect those of the NHS or the Department of Health and Social Care. This publication is based on data collected by or on behalf of Healthcare Quality Improvement Partnership, who have no responsibility or liability for the accuracy, currency, reliability and/or correctness of this publication.

REFERENCES

1. Kanis JA, Oden A, McCloskey EV, Johansson H, Wahl DA, Cooper C, et al. A systematic review of hip fracture incidence and probability of fracture worldwide. Osteoporos Int. 2012;23(9):2239-56. https://doi.org/10.1007/s00198-012-1964-3

2. Sheehan KJ, Williamson L, Alexander J, Filliter C, Sobolev B, Guy P, Bearne LM, Sackley C. Prognostic factors of functional outcome after hip fracture surgery: a systematic review. Age and ageing. 2018 Sep 1;47(5):661-70. https://doi.org/10.1093/ageing/afy057

3. Handoll HH, Cameron ID, Mak JC, Panagoda CE, Finnegan TP. Multidisciplinary rehabilitation for older people with hip fractures. Cochrane Database Syst Rev. 2021;11:CD007125. https://doi.org/10.1002/14651858.CD007125.pub3

4. Orouba Almilaji SA, Aicha Goubar, Lauren Beaupre, Ian D Cameron, Rhian Milton-Cole, Celia L Gregson, Antony Johansen, Morten Tange Kristensen, Jay Magaziner, Finbarr C Martin, Catherine Sackley, Euan Sadler, Toby O Smith, Boris Sobolev, Katie J Sheehan. Probability of discharge home, readmission, survival and recovery following additional in-hospital physiotherapy after hip fracture surgery. Physiotherapy. 2022;Revisions under review.

5. Royal College of Physicians (2019) Falls and Fragility Fracture Audit Programme, National Hip Fracture Database Extended Report https://www.nhfd.co.uk/files/2019ReportFiles/NHFD_2019_Annual_Report_v101.pdf Accessed 16 June 2022.

6. Goubar A, Ayis S, Beaupre L, Cameron ID, Milton-Cole R, Gregson CL, Johansen A, Kristensen MT, Magaziner J, Martin FC, Sackley C. The impact of the frequency, duration and type of physiotherapy on discharge after hip fracture surgery: a

secondary analysis of UK national linked audit data. Osteoporosis International. 2022 Apr;33(4):839-50. https://doi.org/10.1007/s00198-021-06195-9

7. Marufu TC, Mannings A, 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):2325-34. https://doi.org/10.1016/j.injury.2015.10.025

8. 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. https://doi.org/10.1093/bja/aen236

9. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594. https://doi.org/10.7326/M14-0697

10. Sheehan KJ, Goubar A, Almilaji O, Martin FC, Potter C, Jones GD, Sackley C, Ayis S. Discharge after hip fracture surgery by mobilisation timing: secondary analysis of the UK National Hip Fracture Database. Age and ageing. 2021 Mar;50(2):415-22. https://doi.org/10.1093/ageing/afaa204

11. Southwell J, Potter C, Wyatt D, Sadler E, Sheehan KJ. Older adults' perceptions of early rehabilitation and recovery after hip fracture surgery: a UK qualitative study. Disability and Rehabilitation. 2022 Mar 13;44(6):939-46. https://doi.org/10.1080/09638288.2020.1783002

12. 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;43(4):464-71. https://doi.org/10.1093/ageing/afu065

13. Volkmer B, Sadler E, Lambe K, Martin FC, Ayis S, Beaupre L, et al. Orthopaedic physiotherapists' perceptions of mechanisms for observed variation in the

implementation of physiotherapy practices in the early postoperative phase after hip fracture: a UK qualitative study. Age Ageing. 2021;50(6):1961-70. https://doi.org/10.1093/ageing/afab131

14. Guerra S LK, Manolova G, Sadler E, Sheehan KJ. Multidisciplinary perspectives of current and optimal acute rehabilitation, a hip fracture example. Plos One. 2022; 18;17(11):e0277986. https://doi.org/10.1371/journal.pone.0277986

15. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Stat Med. 2021;40(19):4230-51. https://doi.org/10.1002/sim.9025

16. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association. 1999;94:496-509. https://doi.org/10.1080/01621459.1999.10474144

17. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat Med. 2017;36(27):4391-400. https://doi.org/10.1002/sim.7501

18. R: A language and environment for statistical computing. R Foundation for Statistical Computing [program]. 3.6.1 version. Vienna, Austria, 2019.

19. Gerds TA, Ohlendorff JS, Blanche P, Mortensen R, Wright M, Tollenaar N, Muschelli J, Mogensen UB, Ozenne B. riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks 2018 [Available from: https://CRAN.R-project.org/package=riskRegression]. Accessed 16 June 2022.

20. Gray B. cmprsk: Subdistribution analysis of competing risks. 2014. [Available from: https://cran.r-project.org/web/packages/cmprsk/cmprsk.pdf]. Accessed 16 June 2022.

21. Schoenfeld D. Partial Residuals for The Proportional Hazards Regression Model. Biometrika. 1982;69(1):239-41. https://doi.org/10.1093/biomet/69.1.239

22. Riley RD, van der Windt D, Croft P, Moons KGM. Prognosis research in healthcare. Concepts, methods, and impact. Oxford, UK: Oxford University Press; 2019.[Available from: https://oxfordmedicine.com/view/10.1093/med/9780198796619.001.0001/med-9780198796619] Accessed 16 June 2022.

23. Wolbers M, Koller MT, Witteman JC, Steyerberg EW. Prognostic models with competing risks: methods and application to coronary risk prediction. Epidemiology. 2009;20(4):555-61. https://www.jstor.org/stable/25662702

24. Schoop R, Beyersmann J, Schumacher M, Binder H. Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. Biom J. 2011;53(1):88-112. https://doi.org/10.1002/bimj.201000073

25. Gerds TA, Andersen PK, Kattan MW. Calibration plots for risk prediction models in the presence of competing risks. Stat Med. 2014;33(18):3191-203. https://doi.org/10.1002/sim.6152

26. Cook RD. Influential Observations in Linear Regression. Journal of the American Statistical Association. 1979;74:169-74. https://doi.org/10.1080/01621459.1979.10481634

27. Davis CE HJ, Bangdiwala SI, Nelson JJ. An example of dependencies among variables in a conditional logistic regression. In: Wiley, editor. Modern Statistical Methods in Chronic Disease Epidemiology. New York 1986. p. 140-7.

28. Fox J, Weisberg S. An R companion to applied regression. Sage publications; 2018 Sep 27. [Availabel from: https://toc.library.ethz.ch/objects/pdf03/z01_978-1-5443-3647-3_01.pdf] Accessed 16 June 2022.

29. Harrell Jr FE. Regression modeling strategies with applications to linear models, logistic regression, and survival analysis. Springer, editor. New York 2001.

30. Steyerberg E. Clinical prediction models: a practical approach to development, validation, and updating: Springer; 2009.

31. Wood SN. Modelling and Smoothing Parameter Estimation with Multiple Quadratic Penalties. J R Statist Soc B 2000;62:413-28. https://doi.org/10.1111/1467-9868.00240

32. Harrell Jr FE. rms: Regression Modeling Strategies. R package version 6.2-0 2021 [Available from: https://CRAN.R-project.org/package=rms]. Accessed 16 June 2022.

33. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81. https://doi.org/10.1016/S0895-4356(01)00341-9

34. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4%3C361::AID-SIM168%3E3.0.CO;2-4

35. Hartigan JA. Algorithm AS 136: A K-means clustering algorithm. Applied statistics. 1979;28:100-8. https://doi.org/10.2307/2346830

36. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med. 2011;30(4):377-99. https://doi.org/10.1002/sim.4067

37. Hardt J, Herke M, Leonhart R. Auxiliary variables in multiple imputation in regression with missing X: a warning against including too many in small sample research. BMC Med Res Methodol. 2012;12:184. https://doi.org/10.1186/1471-2288-12-184

38. Rubin DB. Multiple imputation for nonresponse in surveys. New York: John Wiley and Sons; 1987.

39. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. J Clin Oncol. 2008;26(8):1364-70. DOI 10.1200/JCO.2007.12.9791

40. Salkeld G, Cameron ID, Cumming RG, Easter S, Seymour J, Kurrle SE, et al. Quality of life related to fear of falling and hip fracture in older women: a time trade off study. BMJ. 2000;320(7231):341-6. https://doi.org/10.1136/bmj.320.7231.341

41. Meehan AJ, Maher AB, Brent L, Copanitsanou P, Cross J, Kimber C, et al. The International Collaboration of Orthopaedic Nursing (ICON): Best practice nursing care standards for older adults with fragility hip fracture. Int J Orthop Trauma Nurs. 2019;32:3-26. https://doi.org/10.1016/j.ijotn.2018.11.001

42. Goubar A, Martin FC, Potter C, Jones GD, Sackley C, Ayis S, et al. The 30-day survival and recovery after hip fracture by timing of mobilization and dementia : a UK database study. Bone Joint J. 2021;103-B(7):1317-24. https://doi.org/10.1302/0301-620X.103B7.BJJ-2020-2349.R1

43. Dunn RH, Ahn J, Bernstein J. End-of-life Care Planning and Fragility Fractures of the Hip: Are We Missing a Valuable Opportunity? Clin Orthop Relat Res. 2016;474(7):1736-9. https://doi.org/10.1007/s11999-015-4675-1

44. Fox MT, Persaud M, Maimets I, Brooks D, O'Brien K, Tregunno D. Effectiveness of early discharge planning in acutely ill or injured hospitalized older adults: a systematic review and meta-analysis. BMC Geriatr. 2013;13:70. https://doi.org/10.1186/1471-2318-13-70

45. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ. 2021;374:n2061. https://doi.org/10.1136/bmj.n2061

46. Hallen SA, Hootsmans NA, Blaisdell L, Gutheil CM, Han PK. Physicians' perceptions of the value of prognostic models: the benefits and risks of prognostic confidence. Health Expect. 2015;18(6):2266-77. https://doi.org/10.1111/hex.12196

47. Greenland S. Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. Epidemiology. 1995;6(4):450-4. https://www.jstor.org/stable/3702100

48. Hayat S, Luben R, Khaw KT, Wareham N, Brayne C. Evaluation of routinely collected records for dementia outcomes in UK: a prospective cohort study. BMJ Open. 2022;12(6):e060931. doi: 10.1136/bmjopen-2022-060931

49. Seitz DP, Adunuri N, Gill SS, Rochon PA. Prevalence of dementia and cognitive impairment among older adults with hip fractures. J Am Med Dir Assoc. 2011;12(8):556-64. https://doi.org/10.1016/j.jamda.2010.12.001

50. Nagin DS, Odgers CL. Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol. 2010;6:109-38. https://doi.org/10.1146/annurev.clinpsy.121208.131413

Table 1. Characteristics of patients surgically treated for non-pathological first hip fracture for development and internal validation, and external validation, datasets.

		Inhospital death		30-day death		Change in residence[d]	
		Development and internal validation (2011-14)	External validation (2015-16)	Development and internal validation (2011-14)	External validation (2015-16)	Development and internal validation (2011-14)	External validation (2015-16)
		n=141,158	n=84,096	n=149,258	n=87,414	n=119,420	n=70,319
Age at admission (years)	60-64	4618 (3.3)	2641 (3.1)	4910 (3.3)	2772 (3.2)	4616 (3.9)	2586 (3.7)
	65-69	7867 (5.6)	5108 (6.1)	8348 (5.6)	5346 (6.1)	7711 (6.5)	4910 (7.0)
	70-74	11453 (8.1)	7239 (8.6)	12117 (8.1)	7560 (8.6)	10965 (9.2)	6847 (9.7)
	75-79	19816 (14.0)	11629 (13.8)	20927 (14.0)	12108 (13.9)	18249 (15.3)	10500 (14.9)
	80-84	31517 (22.3)	18010 (21.4)	33370 (22.4)	18698 (21.4)	27515 (23.0)	15447 (22.0)
	85-89	34903 (24.7)	20537 (24.4)	36871 (24.7)	21348 (24.4)	28337 (23.7)	16579 (23.6)
	90-94	23524 (16.7)	14001 (16.6)	24856 (16.7)	14490 (16.6)	17339 (14.5)	10225 (14.5)
	>94	7460 (5.3)	4931 (5.9)	7859 (5.3)	5092 (5.8)	4688 (3.9)	3225 (4.6)
Sex	Female	104905 (74.3)	60987 (72.5)	110927 (74.3)	63402 (72.5)	87758 (73.5)	50485 (71.8)
	Male	36253 (25.7)	23109	38331 (25.7)	24012 (27.5)	31662 (26.5)	19834

			(27.5)				(28.2)
Prefracture mobility	Outdoor mobility	86719 (61.4)	63293 (75.3)	92021 (61.7)	65861 (75.3)	83780 (70.2)	57616 (81.9)
	Indoor mobility	51573 (36.5)	19716 (23.4)	54237 (36.3)	20430 (23.4)	34338 (28.8)	12169 (17.3)
	No mobility	2866 (2.0)	1087 (1.3)	3000 (2.0)	1123 (1.3)	1302 (1.1)	534 (0.8)
Prefracture residence	Home/sheltered housing	114534 (81.1)	68522 (81.5)	121573 (81.5)	71419 (81.7)	119420(100)	70319 (100)
	Nursing/residential care	26624 (18.9)	15574 (18.5)	27685 (18.5)	15995 (18.3)	0.0	0.0
Dementia	Yes	36503 (25.9)	25224 (30.0)	38145 (25.6)	26015 (29.8)	18410 (15.4)	13844 (19.7)
Inhospital death[a]		6780 (4.8)	3752 (4.5)	6810 (4.6)	3770 (4.3)	4815 (4.0)	2744 (3.9)
30-day death[b]		8311 (5.9)	4632 (5.5)	8693 (5.8)	4790 (5.5)	5551 (4.6)	3080 (4.4)
Change in residence[c]						4461 (3.7)	2712 (3.9)

[a] Among the development dataset for model prediction of inhospital death outcome only one patient had missing 30 death outcome and missing change in residence outcome. No missing data in the validation dataset for this model.

[b] Among the development and validation datasets for the model prediction of 30 days death 7578 (5%) and 3045 (4%) have missing inhospital death outcome

[c] Among the development and validation datasets for model prediction of change in residence outcome 3276 (3%) and 1486 (2%) have missing inhospital death outcome respectively. No patient has missing 30 days death outcome.

[d] For the model prediction of change in residence outcome, study population includes only patients admitted from home/sheltered housing

Table 2. Development, internal validation and external (temporal) validation of multivariable prediction models for inhospital and 30-day death, and change in residence for patients after hip fracture surgery, complete case analysis

	Development, internal validation			External validation		
	In-hospital death	Change in residence	30-day mortality	In-hospital death	Change in residence	30-day mortality
n events[a](%)	6,780 (4.8)	4,461 (3.7)	8,693 (5.8)	3,752 (4.5)	2,712 (2.1)	4,790 (5.5)
30-day CIF % (95% CI)	7.1 (6.9 – 7.2)	6.2 (6.0– 6.4)		5.9 (5.8–6.1)	6.4 (6.1 – 6.6)	
AUC[b] % (95% CI)	72.9 (72.2 – 73.7)	71.4 (70.3 – 72.7)	71.1 (70.6 – 71.6)	73.1 (72.7-74.2)	71.7 (70.6 – 72.7)	71.2 (70.6 – 72.0)
Brier Score[b] % (95% CI)	5.7 (5.5 – 5.9)	5.6 (5.3 – 5.8)	5.3 (5.2 – 5.4)	5.3 (5.1-5.4)	5.7 (5.5 – 5.9)	5.0 (5.0– 5.1)
Optimism adjusted AUC % (95% CI)	73.1 (72.6 – 73.7)	71.5 (70.8 – 72.5)	71.1 (70.5 – 71.6)			
Optimism adjusted Brier Score % (95% CI)	5.7 (5.6-5.8)	5.6 (5.4 - 5.7)	5.3 (5.2-5.4)			

CIF, cumulative incidence function; CI, confidence interval; AUC, area under the curve

[a] By 30 inpatient days for in-hospital death and change in residence.

[b] Apparent performance statistics.

Table 3. Characteristics of patients surgically treated for non-pathological first hip fracture according to the overall risk group based on outcome driven classification

		All	Low	Medium	High
		N=141,157	N=44,364	N=39,542	N=57,251
Age at admission (years)	60-64	4618 (3.3))	4590 (10.3)	9 (0.0)	19 (0.0)
	65-69	7867 (5.6)	7641 (17.2)	5 (0.0)	221 (0.4)
	70-74	11453 (8.1)	10447 (23.5)	576 (1.5)	430 (0.8)
	75-79	19816 (14.0)	15359 (34.6)	1783 (4.5)	2674 (4.7)
	80-84	31517 (22.3)	5933 (13.4)	17937 (45.4)	7647 (13.4)
	85-89	34902 (24.7)	394 (0.9)	13160 (33.3)	21348 (37.3)
	90-94	23524 (16.7)	0 (0.0)	6072 (15.4)	17452 (30.5)
	>94	7460 (5.3)	0 (0.0)	0 (0.0)	7460 (13.0)
Sex	Female	104904 (74.3)	29465 (66.4)	39212 (99.2)	36227 (63.3)
	Male	36253 (25.7)	14899 (33.6)	330 (0.8)	21024 (36.7)
Prefracture mobility	Outdoor mobility	86719 (61.4)	36895 (83.2)	34390 (87.0)	15434 (27.0)
	indoor mobility	51572 (36.5)	7152 (16.1)	4806 (12.2)	39614 (69.2)
	No mobility	2866 (2.0)	317 (0.7)	346 (0.9)	2203 (3.8)
Prefracture residence	home/sheltered housing	114533 (81.1)	40383 (91.0)	39542 (100.0)	34608 (60.4)
	Nursing/residential care	26624 (18.9)	3981 (9.0)	0 (0.0)	22643 (39.6)
Dementia	No	104655 (74.1)	42216 (95.2)	33866 (85.6)	28573 (49.9)
	Yes	36502 (25.9)	2148 (4.8)	5676 (14.4)	28678 (50.1)

FIGURE CAPTIONS

Figure 1: Stratify-hip algorithm to enable patient assignment to three overall risk groups based on predicted risk of in-hospital death, 30-day death and change in residence.

Figure 2: Calibration plots of predicted and observed risk of in-hospital death, change in residence, and 30-day death after hip fracture surgery for development and internal validation, and external validation.

Figure 1

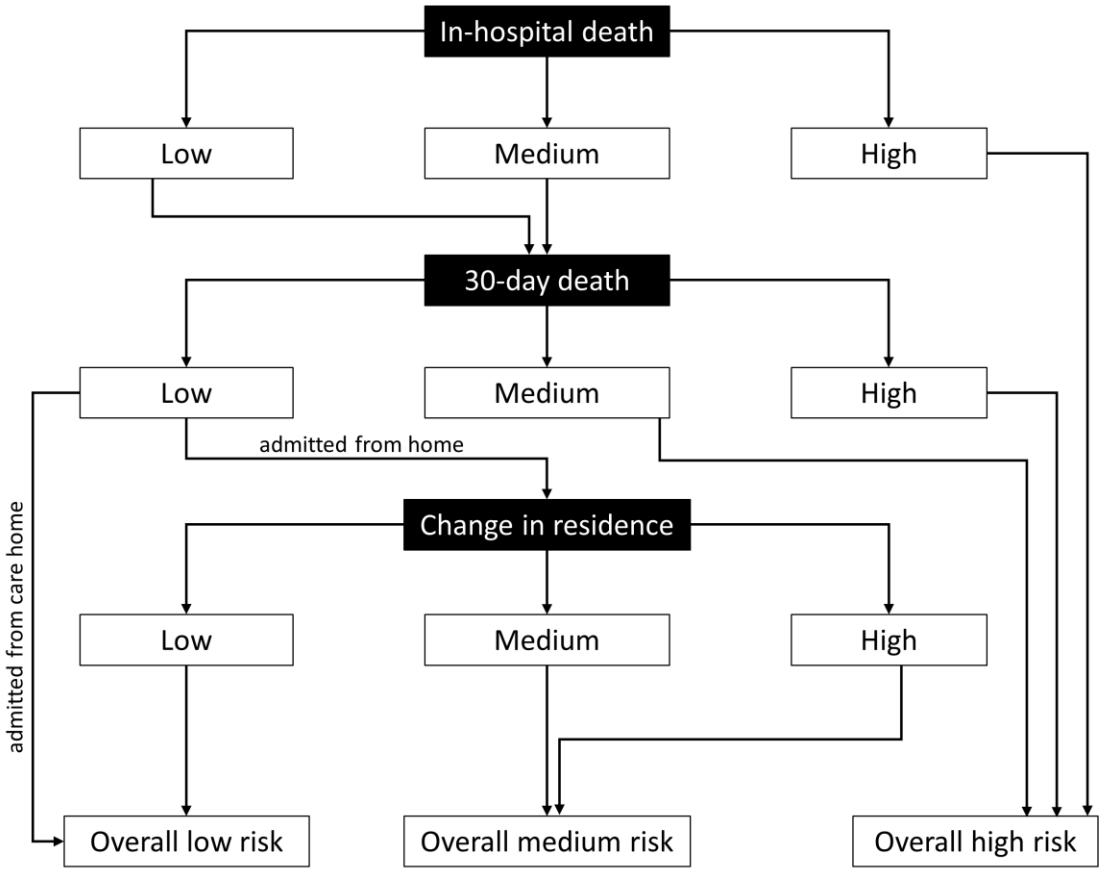

Figure 2