ORIGINAL RESEARCH

Validation of the Nottingham Hip Fracture Score (NHFS) in Patients with Hip Fracture: A Prospective Cohort Study in the Netherlands

Veronique AJIM van Rijckevorsel [ID][1]
Gert R Roukema[1]
Taco MAL Klem[2]
Tjallingius M Kuijper [ID][3]
Louis de Jong [ID][2]

On behalf of Dutch Hip Fracture Registry Collaboration

[1]Surgery Department, Maasstad Hospital, DZ Rotterdam 3079, Netherlands; [2]Surgery Department, Franciscus Hospital, PM Rotterdam 3045, Netherlands; [3]Science Board, Maasstad Hospital, DZ Rotterdam 3079, Netherlands

Introduction: The Nottingham Hip Fracture Score (NHFS) was developed to predict 30-day mortality for patients with hip fracture. This study aimed to validate the NHFS in a cohort with sufficient statistical power.

Methods: Data were extracted from a prospective hip-fracture database (FAMMI). Patients were included between January 1, 2018 and January 11, 2021. All consecutively admitted patients ≥18 years of age with a hip fracture (ie, femoral neck fracture, intertrochanteric fracture, and subtrochanteric fracture) were included. Mann–Whitney's U values were calculated to find potential miscalibration of the NHFS formula. Discrimination evaluation was performed using the concordance statistic as an equivalent to area under the receiver-operating curve.

Results: In total, 2,458 patients were included. Mean age was 80±12 years, and 66% were women (n=1,631). Median NHFS was 5 (4–6) and overall 30-day mortality 7.9% (n=195). Overall goodness of fit was tested with Pearson's $?^2$ (11.8, df 10; P=0.297). No statistically significant signs of miscalibration were found (Mann–Whitney U, P=0.08). Discrimination was tested with area under the receiver-operating curve, which was 72.1% (95% CI 68.7%–75.4%). However, observed 30-day mortality in our population of hip-fracture patients was slightly higher than the NHFS prediction.

Conclusion: The NHFS seemed to predict 30-day mortality with reasonable accuracy for patients with a hip fracture in a population within the Netherlands.

Keywords: 30-day mortality, hip fractures, NHFS, validation, prediction tool

Introduction

Hip fractures are associated with high postoperative 30-day mortality of 7%–13%.[1–4] It is of great importance to identify patients most at risk for 30-day mortality. A 30-day mortality-prediction tool should have adequate and be extensively validated before usage in daily practice. A reliable prediction tool can help to provide appropriate informed consent for patients and families, assist clinicians in determining optimum timing of surgery, and point to intervention possibilities for perioperative management. Several mortality-prediction tools have been developed to help identify patients at high risk of mortality. Among these are the Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity, the Charlson Comorbidity Index, and the Nottingham Hip Fracture Score (NHFS).[3,5–9] The NHFS has shown the most promising results in predicting 30-day mortality in hip-fracture patients.[10]

Correspondence: Veronique AJIM van Rijckevorsel
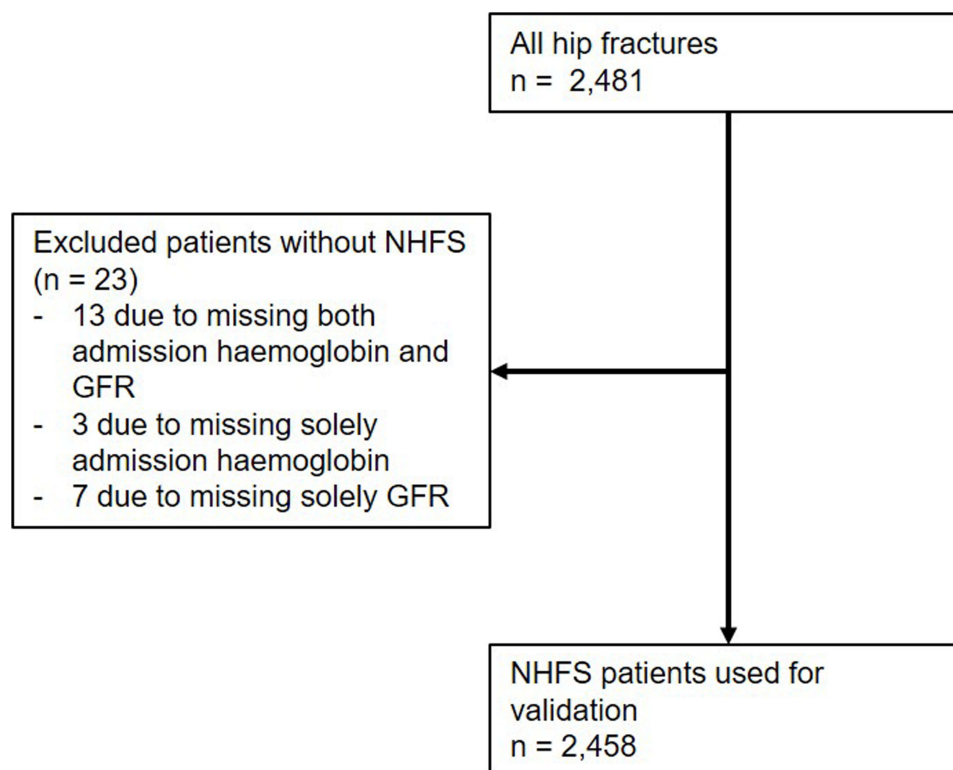Email veroniquevanrijckevorsel@gmail.com

```
                        ┌─────────────────────────┐
                        │ All hip fractures       │
                        │ n = 2,481               │
                        └─────────────────────────┘
                                     │
┌────────────────────────────────┐  │
│ Excluded patients without NHFS │◄─┤
│ (n = 23)                       │  │
│ - 13 due to missing both       │  │
│   admission haemoglobin and    │  │
│   GFR                          │  │
│ - 3 due to missing solely      │  │
│   admission haemoglobin        │  │
│ - 7 due to missing solely GFR  │  │
└────────────────────────────────┘  │
                                     ▼
                        ┌─────────────────────────┐
                        │ NHFS patients used for  │
                        │ validation              │
                        │ n = 2,458               │
                        └─────────────────────────┘
```

Figure I Flowchart of patient inclusion.

The NHFS was originally developed in 2008 by Maxwell et al as a prediction tool for 30-day mortality in patients with a hip fracture (Supplementary Table 1 and Supplementary Figure 1).[11] After its original development, the NHFS was validated by Moppett et al in 2012.[5] They found overestimation of 30-day mortality for higher and lower scores. A revised equation was thus developed, which has been calibrated in three hip-fracture cohorts (ie, Peterborough, Brighton and Nottingham).[5] Rushton et al thenvalidated this modified NHFS in a retrospective cohort of patients who had had surgical treatment for a hip fracture.[12] Simultaneously with Rushton et al, the NHFS was validated by Marufu et al in a different cohort.[13] This simultaneous recalibration and validation of the NHFS emphasizes the clinical demand for a simple but accurate prediction tool for 30-day mortality in patients with hip fractures. Development of the NHFS also demonstrates that such a prediction tool is subject to evolution of patients and care; therefore ongoing research and validations are necessary to keep the prediction tool uptodate.[13]

In the literature, the NHFS has seemed to overestimate the risk of 30-day mortality after a hip fracture and has

thus been adjusted over time.[5,13] With epidemiological changes in the hip-fracture population, ongoing research and validations are necessary to keep the prediction tool uptodate. Furthermore, the extent of applicability of the NHFS in patient populations outside the UK has to be examined more thoroughly. This study aimed to prospectively investigate the predictive value of the modified NHFS for 30-day mortality in a large (Dutch) cohort of patients with hip fractures.

Methods
NHFS
The NHFS is based on age (≤65 years, 66–85 years, ≥86 years), sex (male), hemoglobin concentration at admission (≤10 g/dL), living in an institution, number of comorbidities (greater than or equal to two), presence of malignancy in the last 20 years, and abbreviated mental test score (AMTSof ≤6 out of 10). The AMTS is used to detect cognitive dysfunction at admission. A cutoff ≤6 is used to indicate moderate–severe cognitive dysfunction associated with dementia.[11] The AMTS was not available in the participating study centers. Geriatric assessment of cognitive dysfunction or prior diagnosis of cognitive

dysfunction was used as a proxy for the AMTS in calculating the NHFS. The NHFS can be calculated by combining the scores of the aforementioned variables (Supplementary Table 2). NHFS scores range between 0 and 10, and based on the score, 30-day mortality can be predicted. Predicted mortality for patients scoring 0 is 0.7% and increases to 45% in patients scoring 10.[5]

Patient Inclusion

Data were extracted from a prospective hip-fracture database (FAMMI trial, NL8313). Consecutively admitted patients between January 1, 2018 and January 11, 2021 in two level II trauma teaching hospitals in Rotterdam, the Netherlands were screened for eligibility. Inclusion criteria were age ≥18 years and a hip fracture, ie, femoral neck fracture, intertrochanteric fracture, or subtrochanteric fracture. Patients were included regardless of the chosen treatment (surgical or nonsurgical). Surgical treatment included placement of cannulated screw fixation, a dynamic hip screw, a femoral neck system, an intramedullary fixation (gamma nail), hip hemiarthroplasty, or total hip arthroplasty. In cases of nonsurgical treatment, patients underwent a fascia iliaca compartment nerve block or received adequate pain medication. Patients were treated nonsurgically if after careful consideration among trauma surgeon, geriatrician, patient, and family, it was estimated that a significant decrease in quality of life or death would be inevitable despite surgery. Patients were excluded for missing baseline characteristics needed to calculate the NHFS.Minimum follow-up after surgery was 30 days. The Office of National Statistics (Basisregistratie Personen) was consulted after a minimum of 30 days to extract death dates if patients were deceased. These dates were cross-checked with hospital and nursing-home records.

Patient Variables

Baseline characteristics comprised patients' age, sex, residence type, potential use of walking aids, American Society of Anesthesiologists score, medical history (malignancies and ongoing chemotherapeutic treatment, diabetes mellitus, rheumatoid arthritis, Parkinson's disease, chronic obstructive pulmonary disease, transient ischemic attack, cerebrovascular accidents, myocardial infarction and other cardiovascular diseases, and hip-fracture surgery), glomerular filtration rate (mL/min/$1.73m^2$), hemoglobin level at admission, weight and height to calculate body-mass index (kg/m^2), and medication use. A cutoff of NHFS ≥5 was

used to identify patients at high risk of 30-day mortality, as in previous literature.[4]

Statistical Analysis

Data were stored using a cloud-based clinical data–management platform, (Castor EDC, Amsterdam, Netherlands) and analyzed using Stata 14.0 (StataCorp, College Station, TX, USA). The 30-day mortality observed in the study cohort was compared to that predicted by the NHFS. Predicted mortality was calculated using the formula for the modified NHFS: 100/(1+exp [5.012NHFS0.481]).[5] All statistical tests were two-sided with a significance level of $P<0.05$. Pearson's $?^2$ was used to assess the overall goodnessoffit, using the range of possible NHFS as natural categories. Mann–Whitney U values were calculated to find potential miscalibration. Discrimination evaluation was performed using the concordance statistic as an equivalent to the area under the receiver-operating curve: 0.70–0.79 is considered to represent acceptable discrimination of a model and 0.80–0.89 excellent discrimination.[14]

Ethics Statement

The local medical research ethics committee (Maasstad Hospital, Rotterdam, Netherlands) approved the study (trail registration NL8313). All protocols were in compliance with the Declaration of Helsinki. Because of the high percentage of cognitive dysfunction and no changes made to the standard practice of care, the local ethics committee decided that patient consent to review their medical records was not required. No external funding was received for this trial, and no conflicts of interest need to be reported.

Results

In total, 2,481 patients were screened for eligibility, of which 23 (0.9%) were excluded due to missing data needed to calculate the NHFS (Figure 1). Data were missing due to absent laboratory tests at admission. Baseline characteristics of the included patients are shown in Table 1. Mean age was 80±12 years, 66% were women (n=1,631), and 69% had an American Society of Anesthesiologists score ≥3 (n=1,698). Median NHFS was 5 (IQR 4–6). Observed overall 30-day mortality was 7.9% (n=195).

Observed and predicted 30-day mortality for each NHFS are presented in Table 2. Observed 30-day mortality increased with higher NHFS. In total, 1,294 (53%) patients

Table 1 Baseline characteristics of study patients

	Overall (2,458) n/known n (%)
Age, years (mean ± SD)	80±12
Female sex	1,631/2,458 (66)
NHFS, median (IQR)*	5 (4–6)
ASA score*	
1	124/2,458 (5)
2	636/2,458 (26)
3	1,484/2,458 (60)
4	214/2,458 (9)
Residence	
Home	1,694/2,458 (69)
Semi-independent nursing home	195/2,458 (8)
Nursing home	558/2,458 (22)
Psychiatric unit	11/2,458 (1)
Walking aids	
None	1,240/2,292 (54)
Cane	110/2,292 (5)
Rollator	866/2,292 (38)
Wheelchair or mobility scooter	76/2,292 (3)
Comorbidities	
Dementia	591/2,458 (24)
Diabetes mellitus	444/2,458 (18)
Obesity (BMI >30)	155/1,736 (9)
Underweight (BMI <18.5)	137/1,736 (8)
Rheumatoid arthritis	114/2,458 (5)
Parkinson's	110/2,458 (4)
COPD	306/2,458 (12)
TIA	192/2,458 (8)
Cerebrovascular accident	317/2,458 (13)
Cardiovascular disease	1496/2,458 (61)
Myocardial infarction	205/2,458 (8)
Previous hip-fracture surgery	236/2,458 (10)
Active oncological status	375/2,458 (15)
Chemotherapy	12/2,458 (1)
Preoperative GFR (mean ± SD)	65±21
Severe renal disease (GFR <30)	151/2,458 (6)
Preoperative Hb (mean ± SD)	7.8±1.0
Medication	
Polypharmacy (more than four)	1,376/2,458 (56)
Inhaled steroids	215/2,458 (9)
Immunosuppressive therapy	109/2,458 (4)

Abbreviations: NHFS, Nottingham Hip Fracture Score; ASA, American Society of Anesthesiologists; BMI, body-mass index (kg/m^2); COPD, chronic obstructive pulmonary disease; TIA, transient ischemic attack; GFR, glomerular filtration rate (mL/min/1.73m^2); Hb, hemoglobin (mM/L).

Table 2 Mortality for each group compared with predicted NHFS

NHFS	Mortality (n)/total (n)	%	NHFS predicted (%)
0	0/94		0.7
1	1/141	0.7	1.1
2	0/32		1.7
3	8/337	2.4	2.7
4	26/560	4.6	4.4
5	46/596	7.7	6.9
6	66/434	15.2	10.7
7	31/209	14.8	16.2
8	13/47	27.7	23.8
9	4/8	50.0	33.6
Overall	195/2,458	7.9	NA

Abbreviation: NA, not applicable.

Overall goodness of fit was tested with Pearson's $?^2$ (11.8, *df* 10; *P*=0.297; Figure 2). No statistically significant signs of miscalibration were found (*P*=0.08). Discrimination was tested with area under the receiver-operating curve, which was 72.1% (95% CI 68.7%–75.4%). Results of validation of the modified NHFS in the original validation cohorts used by Moppett et al are presented in Supplementary Figure 2 as a visual comparison.[5]

Discussion

This prospective cohort study aimed to externally validate the NHFS with sufficient power. The NHFS was a reasonable predictor of 30-day mortality for the patients in this study cohort. No statistically significant signs of miscalibration were found (ie, no systematically deviating predictions of 30-day mortality were detected, *P*=0.08). The NHFS showed acceptable discrimination in predicting 30-day mortality, with an area under the receiver-perating curve of 72.1% (95% CI 68.7%–75.4%). For validation of the NHFS, the outcome measure of interest is 30-day

Table 3 Proportion of patients considered to be at higher risk of 30-day mortality

NHFS	Patients considered at higher risk, n (%)	Observed mortality (%)
≥4	1,854 (75)	10.0
≥5	1,294 (53)	12.4
≥6	698 (28)	16.3
≥7	264 (11)	18.2
≥8	55 (2)	30.9
≥9	8 (1)	50.0

had an NHFS ≥5 (with a 30-day mortality of 12.4%) and were identified as being at high risk of 30-day mortality based on previous literature.[4,12] Details of the NHFS and observed mortality are presented in Table 3.
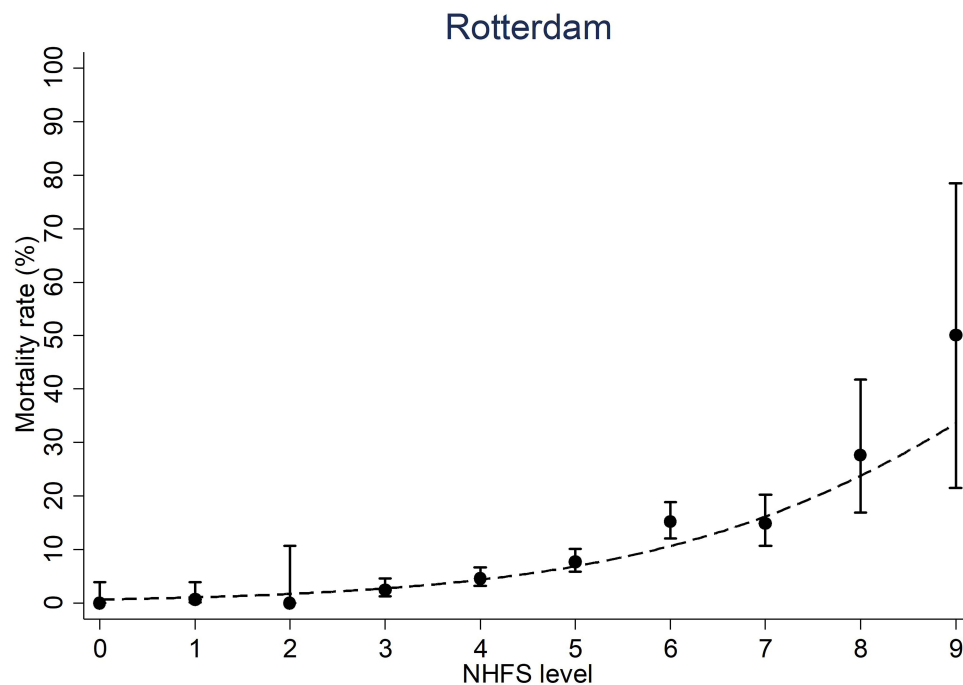
Figure 2 MortalityOobserved in Rotterdam cohort compared to modified NHFS.

mortality. As stated by Karres et al, compared to other prediction tools, such as the Charlson Comorbidity Index, Orthopaedic Physiologic and Operative Severity Score for the Enumeration of Mortality and Morbidity, and Estimation of Physiologic Ability and Surgical Stress, the NHFS shows the most promising results in predicting 30-day mortality rates.[10]

For external validation of a model, a number of statistical principles apply: validation is done in a fully different setting than that in which the development was done,[15] the minimum number of observed events in the study cohort is 100;[15,16], and the variable of interest has similar incidence or prevalence in both development and validation cohorts.[17] Three validation studies have been in conflict with these statistical principles: by Kau et al, Tilkeridis et al, and Karres et al.[10,18,19] These studies were not compliant with the minimum required observed events, reporting 15 (7.3% mortality in a sample of 212), 33 (9.5% mortality in a sample of 349), and 89 events (8.2% mortality in a sample of 1,050), respectively. Kau et al, however, found that the NHFS accurately predicted 30-day mortality for surgically treated hip-fracture patients.[18]

In Greece, most elderly adults live at home, supported by family members.[2,11,19] For this reason, Tilkeridis et al replaced the factor of institutionalization with the New

Mobility Score to calculate the NHFS; however, this version of the NHFS was not validated externally.[19] Similar adjustments to the NHFS have been made to develop new prediction tools, such as the Almelo Hip Fracture Score proposed by Nijmeijer et al.[3] However, their also had observed events included in the model (64 events in a population of 850 patients).

Another retrospective validation study was performed by Rushton et al.[12] In their cohort, neither patients aged <60 years nor those not treated surgically were included in statistical analysis to validate the NHFS. The original cohorts used by Maxwell et al and Moppett et al included nonsurgically treated patients to develop and adjust the NHFS. Patients that do not undergo surgery for a hip fracture are naturally the frailest patients with the highest NHFS. Identifying patients for whom it is estimated that significant decrease in quality of life or death would be inevitable despite surgery is helpful in the process of well-informed shared decision-making and choosing the most appropriate treatment for hip-fracture patients. Excluding these patients from statistical analysis contributes to the finding that the NHFS has less accuracy in predicting 30-day mortality for the higher NHFS in the study by Rushton et al. This less accurate prediction can be explained by the limited number of included patients (n=42) in these higher NHFS (ie, level 8 and 9) derived by

Rushton et al.[12] Observed mortality in those with an NHFS of 8 and 9 in our study cohort was nevertheless comparable to that observed by Rushton et al. The mortality observed in these two cohorts could thus be more representative than the 30-day mortality rates predicted by the NHFS.[12] Since one of the variables of the NHFS is age, there should be no need to exclude patients according to age limits, as was done by Rushton et al. Also, a limited number of events were observed (n=79) in this cohort, which makes it uncompliant with the principles of external validation, as described previously.

The NHFS was developed for all types of hip fractures. Some of the literature, however, indicates that intracapsular hip fractures can significantly increase the risk of mortality compared to extracapsular fractures. Since fracture type is not included in the NHFS, de Jong et al performed a validation study of the modified NHFS in a cohort of patients with femoral neck fractures treated with hip hemiarthroplasty.[2] They found that the NHFS underpredicted 30-day mortality in patients aged >70 years in this cohort, as the observed 30-day mortality rate was higher. The potential role of type of surgery could explain this underprediction. Smith et al identified preoperative indicators of mortality following hip-fracture surgery. One of the factors affecting 12-month mortality was the presence of an intracapsular fracture. Compared to extracapsular fractures, the relative risk of mortality was 77% (95% CI 0.63%–0.95%),[20] possibly because a hip hemiarthroplasty is a more extensive operation than intramedullary fixation of a hip fracture with prolonged recovery.[20]

Lastly, a study by Doherty et al was published recently describing the predictive value of the NHFS for mortality, physical function, length of stay, and postoperative complications.[21] Compared to our multicenter study design with an inclusion period of 3 years, this single-center study had an inclusion period of about 10 years, limiting its generalizability. Also, its results would be subject to changes in care over the period in which enrolment of the patients was performed.

To make the NHFS more usable in daily practice, a subdivision into low- and high risk patients has been described.[4,10] Previous literature suggested an NHFS of ≥5 as a cut-off-point between low- and high risk patients.[4] In our cohort, the observed 30-day mortality is 3% for patients with an NHFS ≤4 (n=1164). However, an NHFS ≥5 identified 53% (n=1294) of the patients in our cohort as patients at high risk of 30-day mortality with an observed

mortality of 12.4%. Rushton et al report the same finding and opt for an NHFS ≥6 as cut-off-point to identify patients at high risk.[12] We on the other hand opt for an NHFS ≥8 as cut-off-point for high risk patients, since for this cut-off-point one in three patients deceased within 30 days (n=17, observed mortality 30.9%). Also, for the low risk patients with an NHFS ≤7 the observed mortality is lower than the overall observed mortality in the entire study cohort (n=178, observed mortality 7.4%).

Strengths and Limitations

Our study differs from the aaforementioned studies in that it is the first validation to prospectively include all types of hip fractures in a large adult-patient cohort. The number of deceased patients in our population (n=195) was sufficient to externally validate the NHFS as per the previously mentioned principles for external validation.[16] Since similar occurrence of the variable of interest (ie, 30-day mortality) in both development and validation cohorts is desirable, overall 30-day mortality in our cohort was compared to the developmental and validation cohorts, and was 7.9%, ranging from 6.6% to 8.0% in the two cohorts.[5,11] Since these percentages were comparable to the outcome variable, our cohort was suitable to validate the NHFS. Another strength of our study is the large sample combined with the limited percentage of excluded patients (n=23, 0.9%). The real-time registration of patients during admission and follow-up prevented selection bias or observational bias.

There was, however, a difference in baseline characteristics on cognitive dysfunction compared to the cohorts described by Maxwell et al and Moppett et al.[5,11] In the participating study centers, all patients were assessed by a geriatrician at admission to distinguish cognitive dysfunction from frequently present delirium.[22–25] In these centers, it is not standard to perform such cognitive tests as the AMTS at admission, since these tests overestimate the prevalence of cognitive dysfunction when delirium is present. Assessment by a geriatrician to detect cognitive dysfunction is thus at least as reliable as the AMTS. This explains the difference in prevalence of cognitive dysfunction between our study and those of Maxwell et al and Moppett et al (24% versus 33%–45%, respectively).[5,11] These studies used an AMTS ≤6 as a cutoff for cognitive dysfunction, but when using this cutoff, mild cognitive dysfunction or delirium can be detected.[5,11] Cognitive dysfunction, as expected in dementia, is detected in cases of an AMTS ≤8.[26] This might also be the

explanation for the difference in mortality in the higher NHFS found in our cohort. When patients in our cohort had a higher NHFS and score on cognitive dysfunction, they were likely to be frailer than those who had an AMTS ≤6, since sincere cognitive dysfunction was present.

Conclusion

This study demonstrates that the NHFS is an accurate predictor of 30-day mortality for adult Dutch patients with a hip fracture. The NHFS is based on a limited number of objective criteria which makes it manageable in daily practice. However, clinically relevant variables, such as fracture type and mobility status, are not accounted for in the NHFS in its current form. This knowledge, combined with the lack of accuracy for higher NHFS values, suggests a need to develop a more precise tool to predict 30-day mortality in patients with a hip fracture.

Abbreviations

AMTS, abbreviated mental test score; CCI, Charlson Comorbidity Index; NHFS, Nottingham Hip Fracture Score.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Dubljanin Raspopovic E, Markovic Denic L, Marinkovic J, et al. Early mortality after hip fracture: what matters? *Psychogeriatrics*. 2015;15 (2):95–101. doi:10.1111/psyg.12076
2. de Jong L, Mal Klem T, Kuijper TM, Roukema GR. Validation of the nottingham hip fracture score (NHFS) to predict 30-day mortality in patients with an intracapsular hip fracture. *Orthop Traumatol Surg Res*. 2019;105(3):485–489. doi:10.1016/j.otsr.2019.02.004
3. Nijmeijer WS, Folbert EC, Vermeer M, Slaets JP, Hegeman JH. Prediction of early mortality following hip fracture surgery in frail elderly: the almelo hip fracture score (AHFS). *Injury*. 2016;47 (10):2138–2143. doi:10.1016/j.injury.2016.07.022
4. Wiles MD, Moran CG, Sahota O, Moppett IK. Nottingham hip fracture score as a predictor of one year mortality in patients undergoing surgical repair of fractured neck of femur. *Br J Anaesth*. 2011;106 (4):501–504. doi:10.1093/bja/aeq405
5. Moppett IK, Parker M, Griffiths R, Bowers T, White SM, Moran CG. Nottingham hip fracture score: longitudinal and multi-assessment. *Br J Anaesth*. 2012;109(4):546–550. doi:10.1093/bja/aes187
6. Ramanathan TS, Moppett IK, Wenn R, Moran CG. POSSUM scoring for patients with fractured neck of femur. *Br J Anaesth*. 2005;94 (4):430–433. doi:10.1093/bja/aei064
7. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383. doi:10.1016/0021-9681(87)90171-8
8. Mohamed K, Copeland GP, Boot DA, et al. An assessment of the POSSUM system in orthopaedic surgery. *J Bone Joint Surg Br*. 2002;84(5):735–739. doi:10.1302/0301-620X.84B5.0840735
9. Jonsson MH, Bentzer P, Turkiewicz A, Hommel A. Accuracy of the physiological and operative severity score for the enUmeration of mortality and morbidity score and the nottingham risk score in hip fracture patients in sweden - A prospective observational study. *Acta Anaesthesiol Scand*. 2018;62(8):1057–1063. doi:10.1111/aas.13131
10. Karres J, Heesakkers NA, Ultee JM, Vrouenraets BC. Predicting 30-day mortality following hip fracture surgery: evaluation of six risk prediction models. *Injury*. 2015;46(2):371–377. doi:10.1016/j.injury.2014.11.004
11. 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):511–517. doi:10.1093/bja/aen236
12. Rushton PR, Reed MR, Pratt RK. Independent validation of the nottingham hip fracture score and identification of regional variation in patient risk within england. *Bone Joint J*. 2015;97-B(1):100–103. doi:10.1302/0301-620X.97B1.34670
13. Marufu TC, White SM, Griffiths R, Moonesinghe SR, Moppett IK. Prediction of 30-day mortality after hip fracture surgery by the nottingham hip fracture score and the surgical outcome risk tool. *Anaesthesia*. 2016;71(5):515–521. doi:10.1111/anae.13418
14. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. 2nd ed. Wiley J, Sons I. 2000.
15. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245–247. doi:10.1016/j.jclinepi.2015.04.005
16. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol*. 2005;58(5):475–483. doi:10.1016/j.jclinepi.2004.06.017
17. Riley RD, Ensor J, Snell KI, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. *BMJ*. 2016;353:i3140. doi:10.1136/bmj.i3140
18. Kau CY, 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 Singap*. 2014;43(9):448–453.
19. Tilkeridis K, Ververidis A, Kiziridis G, et al. Validity of nottingham hip fracture score in different health systems and a new modified version validated to the greek population. *Med Sci Monit*. 2018;24:7665–7672. doi:10.12659/MSM.909943
20. 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):464–471. doi:10.1093/ageing/afu065
21. Doherty WJ, Stubbs TA, Chaplin A, et al. Prediction of postoperative outcomes following hip fracture surgery: independent validation and recalibration of the nottingham hip fracture score. *J Am Med Dir Assoc*. 2020;22:663–669.

22. de Jong L, van Rijckevorsel VAJIM, Raats JW, Klem TMAL, Kuijper TM, Roukema GR. Delirium after hip hemiarthroplasty for proximal femoral fractures in elderly patients: risk factors and clinical outcomes. *Clin Interv Aging*. 2019;14:427–435. doi:10.2147/CIA.S189760

23. Witlox J, Eurelings LS, de Jonghe JF, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA*. 2010;304(4):443–451. doi:10.1001/jama.2010.1013

24. Krogseth M, Watne LO, Juliebø V, et al. Delirium is a risk factor for further cognitive decline in cognitively impaired hip fracture patients. *Arch Gerontol Geriatr*. 2016;64:38–44. doi:10.1016/j.archger.2015.12.004

25. Dolan MM, Hawkes WG, Zimmerman SI, et al. Delirium on hospital admission in aged hip fracture patients: prediction of mortality and 2-year functional outcomes. *J Gerontol a Biol Sci Med Sci*. 2000;55 (9):M527–34. doi:10.1093/gerona/55.9.M527

26. Lam SC, Wong YY, Woo J. Reliability and validity of the abbreviated mental test (hong kong version) in residential care homes. *J Am Geriatr Soc*. 2010;58(11):2255–2257. doi:10.1111/j.1532-5415.2010.03129.x