Diagnostic accuracy of three types of fall risk methods for predicting falls in nursing homes

Hege Bentzen[1,2], Astrid Bergland[3] and Lisa Forsén[1]

[1]Division of Epidemiology, Norwegian Institute of Public Health, [2]Diakonhjemmet Hospital, [3]Oslo University College, Oslo, Norway

ABSTRACT. ***Aims:*** *To determine the diagnostic accuracy of three different methods for identifying individuals at high risk of falling. The St- Thomas Risk Assessment tool (STRATIFY- modified for nursing homes), staff judgment of fall risk, and previous falls remembered by the staff were evaluated. We also examined whether a combination of two of the methods would increase accuracy.* ***Materials and methods:*** *A prospective observational cohort study was carried out for 18 months. One thousand one hundred and forty-eight participants were included and assessed for fall risk. Falls among these residents were recorded from the date of inclusion to the date of death, transfer, or end of observation time. Diagnostic accuracy was evaluated in terms of sensitivity, specificity, predictive values and likelihood ratios, as well as Kaplan-Meier estimates and the Cox proportional hazard model, with time to the first fall as the dependent variable. Sensitivity, specificity, predictive value and likelihood ratios were calculated for falls within 30, 90 and 180 days of assessment for fall risk.* ***Results:*** *Five hundred and seventy (49.6%) of the 1148 residents had one or more falls during the observation period. One thousand one hundred had more than 30 days of observation, 987 more than 90 days, and 867 more than 180 days. For falls within 30 days of assessment for fall risk, sensitivity varied from 65% to 72%, specificity from 69% to 75%, positive predictive value from 31% to 35% and negative predictive value from 91% to 92%. Sensitivity and negative predictive value decreased for falls within 90 days and decreased further for falls within 180 days, whereas specificity and positive predictive value increased for all three assessment methods. Staff judgment of fall risk was the single method having the highest sensitivity but the lowest specificity. A combination of either two of them increased sensitivity to more than 80%, but decreased specificity. The positive Likelihood ratio varied from 2.24 to 2.70 and the negative Likelihood ratio from 0.41 to 0.49 for falls within 30 days. The relative risk of sustaining a fall was 2.4, 2.9 and 3.0 times higher for those assessed to be at high risk of falls compared with those assessed to be at low risk, according to STRATIFY, staff judgment of fall risk and previous falls remembered by the staff, respectively.* ***Conclusions:*** *The diagnostic accuracy of the three methods did not differ markedly. However, staff judgment had the highest sensitivity and the lowest specificity after 30, 90 and 180 days. A combination of either two of the methods showed the highest sensitivity but the lowest specificity.*

(Aging Clin Exp Res 2011; 23: 187-195)

INTRODUCTION

Falls are common among older people (1, 2) and are responsible for considerable morbidity, immobility, and mortality. More than 30% of persons 65 years or older, and more than 50% of those over the age of 85 experience one or more falls per year (1, 3-6). Both the frequency of falling and the incidence of injurious falls are far higher among nursing home residents than among community-dwelling older adults (7-9). The causes of falls among nursing home residents are complex and multifactorial, and a high number of different risk factors seems to contribute to the high incidence of falls in these settings (3). Assessments which adequately discriminate between individuals at high and low risk of falling are important, both for interventions aiming at fall prevention and for those aiming at fracture prevention in falls. So far, there are no assessment tools that stand out as "gold standards" for predicting falls in the nursing home setting (10), and few have been developed for predicting falls among older adults in such settings (11, 12). According to Oliver et al. (13), assessment tools should be validated in prospective studies; analysis of sensitivity and specificity should be applied, and they should be tested in more than one population. In addition,

Key words: Elderly, falls, nursing homes, risk assessments.

Correspondence: Hege Bentzen, PhD, Diakonhjemmet Sykehus, Postboks 23 Vinderen, 0319 Oslo, Norway.
E-mail: Hege.Bentzen@diakonsyk.no

Received November 2, 2009; accepted in revised form June 17, 2010.

they should have good face validity, inter-rater reliability, and compliance by staff, and be clear and easy to administer. It is a challenge to develop a screening tool which fulfil all these criteria. To what extent formal risk assessment tools can contribute to simply using "staff judgement" of fall risk is also uncertain (11, 14-18). Several studies have evaluated the accuracy of various types of fall risk assessments (12). However, only a few have been carried out in relation to fall recording based on a prospective design in a nursing home population (11, 19-21).

The purpose of this study was to determine the diagnostic accuracy of three different methods of identifying individuals at high risk of falling. The St- Thomas Risk Assessment tool (STRATIFY- modified for nursing homes), staff judgment of fall risk, and previous falls recalled by the staff were evaluated. We also examined whether a combination of two of them would increase accuracy, and evaluated the agreement between the three methods.

MATERIALS AND METHODS

A prospective observational cohort study was carried out within the framework of a hip protector trial in 18 Norwegian nursing homes (22). All residents living permanently in these nursing homes between May 2005 and December 2006 were asked to participate. Data from nearly all residents admitted within the study period were captured (n=1236). A few may have been missed, due to the short time between hospitalization and either death or transfer.

The Regional Committee for Ethics in Medical Research approved the study. Written consent was obtained directly from residents who were considered cognitively competent. For cognitively impaired residents, a member of the staff gave consent on their behalf, which was in accordance with recommendations from the Regional Committee for Ethics in Medical Research. Competency to provide informed consent was ascertained from the nursing staff.

Prior to intervention, participants were assessed for baseline characteristics, including fall risk. The staff member most familiar with the resident was asked to do the screening. Falls occurring among participants were recorded prospectively throughout the observation period.

Fall risk assessment

Fall risk assessment included STRATIFY, staff judgment of fall risk, and previous falls recalled by the staff. STRATIFY was originally developed to predict falls among elderly hospital inpatients and showed good sensitivity and specificity in the original population (23). We used a version modified for use in nursing homes (Guidelines for the prevention and management of falls in the elderly, www.laterlifetraining.co.uk/documents/DorsetHealthCare.pdf). STRATIFY consists of five questions: 1) Has the resident had a fall within the last 3 months? 2) Do you think the resident is agitated? 3) Do you think the resident is visually impaired to the extent that everyday function is affected? 4) Do you think the resident is in need of especially frequent toileting? 5) Do you think the resident has a transfer score and mobility score of 3 or 4? (based on the transfer and mobility items of the Barthel ADL index (24)). The sum score in STRATIFY ranges from zero to 5. The sum score was retrospectively calculated by the project manager, and was concealed from the staff who kept the record. A sum score of two or more was used as the cut-off value between low and high risk. Staff judgment of fall risk was a non-formal risk assessment, in which staff simply used their own clinical experience and knowledge about the resident to classify their risk of falling (15). Staff judgments of fall risk were grouped into 5 categories; How do you judge the residents risk of falling? 1) no risk, 2) very low risk, 3) small risk, 4) high risk, 5) very high risk. We used a cut-off value of four or more as high risk.

Fall definition and fall recording

A faller was defined as "a person who has fallen at least once in a given time period" (25). A fall was defined as "any event when the resident, unintentionally and regardless of cause comes to rest on the floor" (26). All falls occurring among participants were recorded from the time of inclusion to time of death, transfer, or end of study. The staff in each nursing home recorded falls prospectively on a specially developed fall documentation sheet soon after the occurrence of the fall.

Statistics

Statistical analyses were carried out with SPSS version 14.0. Differences between fallers and non-fallers were tested by the chi-squared test for categorical variables and independent sample *t*-test for means for continuous variables.

Diagnostic accuracy was evaluated in terms of sensitivity, specificity, and positive and negative predictive value as well as in terms of combining sensitivity and specificity into likelihood ratios (LR). The first fall was the reference standard. For these analyses, participants were observed for falls during a period of 30, 90 and 180 days. For the analyses for these three periods, we included only those with more than 30, 90 and 180 days of observation, respectively. Sensitivity was defined as the percentage of fallers correctly identified as at high risk. Specificity was defined as the percentage of non-fallers correctly defined as at low risk. The positive predictive value was defined as the percentage of high-risk residents who had had one or more fall, and the negative predictive value was the percentage of low risk residents who did not fall. The positive likelihood ratio (LR+) was defined as sensitivity / (100 – specificity) and the negative likelihood ratio (LR-) as (100 – sensitivity) / specificity. LR+ greater than one meant that a positive test (high fall risk) was more likely to occur in residents with falls than in those

without. LR+ less than one meant that a positive test was less likely to occur in residents with falls compared with residents without (27, 28). LR- greater than one meant that a negative test (low fall risk) was more likely to occur in residents with falls than in ones without. LR- less than one meant that a negative test was less likely to occur in residents with falls compared with those without (27, 28) (LR+: >10= high probability, 5 to 10= moderate, 2 to 5= low, 1 to 2= none, or very low. LR-: <0.1= high probability, 0.1 to 0.2= moderate, 0.2 to 0.5= low, 0.5 to 1.0= none, or very low).

To be able to include all participants independent of observation time (n=1148), the predictive value of the three assessment methods was also analysed by means of Kaplan-Meier estimates, log rank test and Cox proportional hazard model, with time to first fall as the dependent variable. The proportional hazard assumption of the Cox model was checked. The first fall was defined as a failure. Deceased, transferred, or end of study were censorings. The number of observation days for fallers was calculated from the date of inclusion to the date of the first fall. For non-fallers, the number of observation days was calculated from the date of inclusion to the end of the study, or to the date the participant died or was transferred. For adjustment of nursing home as a confounder, we used nursing home indicators as covariates. The agreement between the three fall risk assessments was analysed by means of Kappa statistics, and a value higher than 0.60 was considered as good (29).

RESULTS

Of the 1236 included residents, 71 from one nursing home were excluded due to a possible misunderstanding related to fall recording. We also excluded participants with missing information about the time of the first fall (n=17). Thus data from 1148 out of the 1236 participants was analysed in the present paper. Seven hundred and eighty-three of the 1148 residents were included when the study started, and 365 were included continuously during the study period. Mean observation time was 368 days (SD=197 days).

Demographic characteristics and fall risk assessments of the 1148 participants and among fallers and non-fallers are listed in Table 1. A larger proportion of fallers than non-fallers had a STRATIFY sum score of 2 or over, were assessed by staff to be at high risk of falling,

Table 1 - *Demographic characteristics and fall risk assessments for all participants grouped into non-fallers and fallers.*

Variables	All participants n=1148	Participants with no falls n= 578	Participants with one or more falls n=570	*p*-value
Female sex (%)	72.3	73.7	70.9	0.285
Age: mean±SD	84.6 (8.1)	83.5 (9.0)	85.7 (7.1)	<0.001
Use of walking aids: Yes %	77.7	82.0	73.3	<0.001
Fracture in previous 6 months: Yes %	9.9	7.6	12.1	0.011
Memory (0-4)*: mean±SD	1.8 (1.3)	1.9 (1.4)	1.7 (1.3)	0.007
Communication (0-4)+: mean±SD	2.5 (1.4)	2.5 (1.4)	2.5 (1.3)	0.920
Barthel ADL sum score (0-20)^: mean±SD	9.6 (5.8)	8.2 (6.3)	11.1 (4.8)	<0.001
STRATIFY risk assessment tool:				
falls within previous 3 months: Yes %	32.3	18.8	45.9	<0.001
visual impairment: Yes %	21.3	20.3	22.4	0.394
frequent toileting: Yes %	28.2	15.3	21.2	0.010
agitated: Yes %	26.4	17.3	35.7	<0.001
transfer and mobility score of 3 or 4: Yes %	24.8	18.1	31.5	<0.001
STRATIFY sum score:				
sum score= 0	32.1	42.1	21.1	<0.001
sum score= 1	33.1	34.7	31.4	0.271
sum score= 2	20.2	14.3	26.3	<0.001
sum score= 3	9.4	5.9	13.2	<0.001
sum score= 4	4.4	2.0	6.8	<0.001
sum score= 5	0.7	0.2	1.3	0.076
STRATIFY sum score of 2 or over: %	35.0	22.6	47.7	<0.001
Staff judgment of fall risk:				
low	62.0	76.8	46.9	<0.001
high	38.0	23.2	53.1	
Combination of two of three methods§:				
low	52.1	65.7	36.0	
high	47.9	32.9	61.1	<0.001

*0=no memory at all, 1=large memory loss, 2=medium memory loss, 3=some memory loss, 4=no memory loss. +0=not able to, 1=considerable problems, 2=medium problems, 3=some problem, 4=no problem. ^0=very low functional ability, 20=high functional ability. §High risk resident was identified as being at high risk of falling in two of three methods.

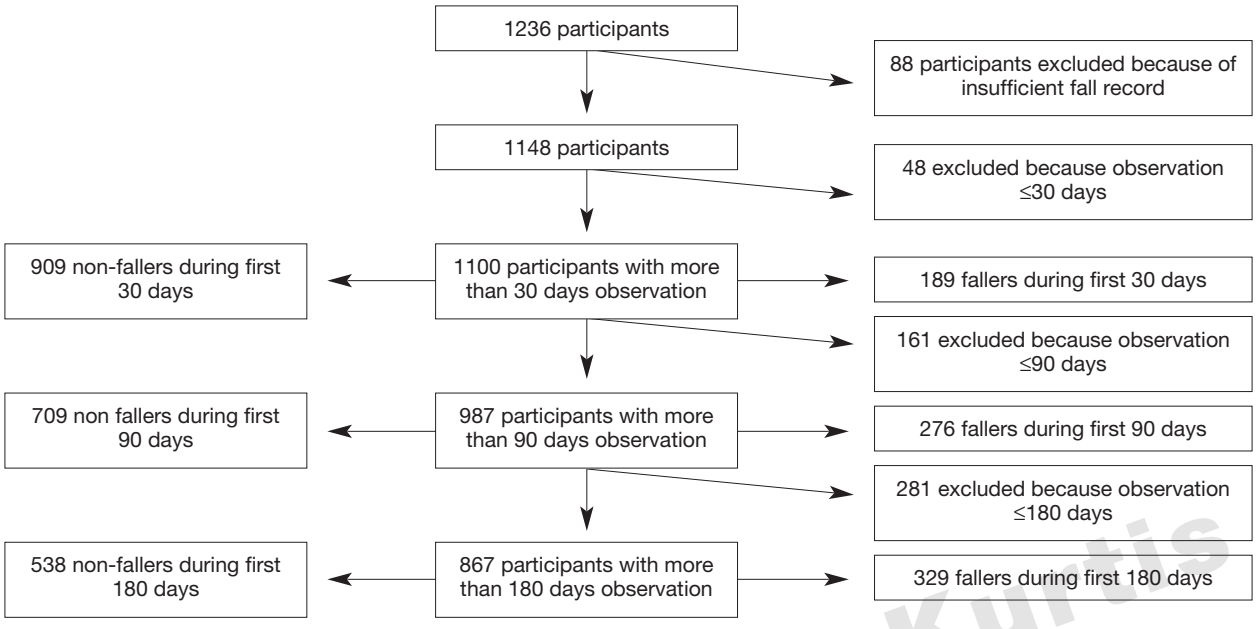

Fig. 1 - *Flow chart showing participants included in the analysis of sensitivity, specificity, predictive values and likelihood ratios.*

and had had a previous fall in the previous last 3 months, as remembered by staff.

Figure 1 shows the flow of participants included in analysis of sensitivity, specificity, predictive values and likelihood ratios at 30, 90 and 180 days after screening for fall risk. Forty-eight participants out of 1148 had less than 30 days of observation, and were excluded from the analysis at 3 months. One hundred and sixty-one participants had less than 90 days and 281 less than 180 days of observation, and were consequently excluded from the analysis at 90 and 180 days, respectively.

Five hundred and seventy (49.6%) of the 1148 par-

Table 2 - *Prognostic accuracy of "STRATIFY", "Staff judgment", "Falls in previous 3 months", and a "Combination of two of three methods" after 30, 90 and 180 days (95% CI).*

	STRATIFY			**Staff judgment of fall risk**			**Falls within the last 3 months**			**Combination of either two methods**		
	30 days	**90 days**	**180 days**	**30 days**	**90 days**	**180 days**	**30 days**	**90 days**	**180 days**	**30 days**	**90 days**	**180 days**
Sensitivity % (95% CI)*	65 (57-72)	58 (52-64)	56 (51-62)	72 (65-78)	65 (60-71)	62 (56-67)	67 (59-74)	59 (52-64)	54 (48-59)	81 (74-86)	74 (68-79)	71 (66-76)
Specificity % (95% CI)	71 (68-74)	73 (69-76)	76 (72-79)	69 (66-72)	72 (69-75)	75 (71-79)	75 (72-78)	78 (75-81)	81 (78-85)	59 (55-62)	62 (58-65)	65 (61-69)
Positive predictive value % (95% CI)	31 (27-36)	45 (40-50)	58 (53-64)	32 (28-37)	47 (42-53)	60 (54-66)	35 (30-41)	51 (46-57)	64 (58-70)	28 (24-32)	42 (38-47)	55 (51-60)
Negative predictive value % (95% CI)	91 (88-93)	82 (78-85)	74 (70-78)	92 (90-94)	84 (81-87)	76 (72-80)	92 (89-94)	83 (80-86)	74 (70-78)	94 (91-96)	86 (83-89)	79 (75-83)
LR+^ (95% CI)	2.24 (1.9-2.6)	2.12 (1.8-2.5)	2.32 (1.9-2.8)	2.30 (2.02-2.63)	2.32 (2.0-2.7)	2.49 (2.1-3.0)	2.70 (2.3-3.1)	2.70 (2.3-3.2)	2.90 (2.4-3.6)	1.94 (1.8-2.2)	1.92 (1.7-2.2)	2.05 (1.8-2.4)
LR-# (95% CI)	0.49 (0.4-0.6)	0.58 (0.5-0.7)	0.58 (0.5-0.7)	0.41 (0.3-0.5)	0.49 (0.4-0.5)	0.51 (0.4-0.6)	0.44 (0.4-0.5)	0.53 (0.5-0.1)	0.57 (0.5-0.6)	0.33 (0.3-0.5)	0.42 (0.4-0.5)	0.44 (0.4-0.5)

*Confidence interval. ^Positive likelihood ratio, interpretation of change in disease likelihood: >10=high, 5-10=moderate, 2-5=low, 1-2=none, or very low.
#Negative likelihood ratio, interpretation of change in disease likelihood: <0.1=high, 0.1-0.2=moderate, 0.2-0.5=low, 0.5-1.0=none, or very low.

ticipants had one or more falls from the time of inclusion to the end of the observation period. Seventeen percent of the 1100 with more than 30 days observation time had one or more falls within 30 days. The same numbers for 90 (n=987) and 180 (n=867) days were 28% and 38%, respectively.

Table 2 shows the sensitivity, specificity, positive and negative predictive value and likelihood ratios for the various assessment methods at 30, 90 and 180 days after assessment. Compared with 30 days, sensitivity decreased for all three methods after 90 days, and decreased still further after 180 days. Specificity increased for all three methods. "Staff judgment of fall risk" was the single assessment method showing the highest sensitivity but the lowest specificity in all 3 observation periods. The percentages of residents assessed to be at high fall risk and with falls within 30 days ranged from 31% to 35%, and the percentage of those assessed as at low risk without having falls was more than 90% for all three methods. A combination of two of the three methods had higher sensitivity but lower specificity. The LR+ for the 3 methods was between 2 and 3 and the LR- between 0.4 and 0.6, indicating that LR+ was low and LR- was low or even very low.

The relative risk of sustaining at least one fall was significantly different between those assessed at high and low risk in all three assessment methods (log rank test <0.001). The Kaplan-Meier estimates are shown in Figure 2a-c. For all methods, much of the difference between the two groups occurred within a short time after inclusion. The higher the STRATIFY sum score and the higher risk of falling as assessed by "staff judgment", the higher probability of sustaining a fall (Fig. 3a-b). Table 3 shows the Cox regression analysis of the probability of having at least one fall in unadjusted and adjusted models. The relative risk was 2.9 for those assessed at high compared with low risk of falling by "staff judgment", 2.4 times higher among those assessed at high compared with low risk according to STRATIFY, and 3.0 times higher among those who had had a fall compared with those who had not within the previous three months. The adjusted analysis did not change the result markedly. Looking at each single item in STRATIFY, previous falls were the strongest predictive factor, and only questions 1, 2 and 5 remained significant in the adjusted model. In other studies, fall history has been identified as a strong single predictor for subsequent falls (11, 20, 30, 31).

The match between staff judgment of fall risk, STRATIFY and previous falls recalled by staff is shown in Table 4. None of the Kappa values exceeded 0.62 (29), which indicates that participants classified at high and low risk were not identical in the three methods. However, the numbers of unpredicted falls with any of the tools within 30, 90 and 180 days were 7, 20 and 30, respectively. These falls did not differ in location or time with respect to predicted falls.

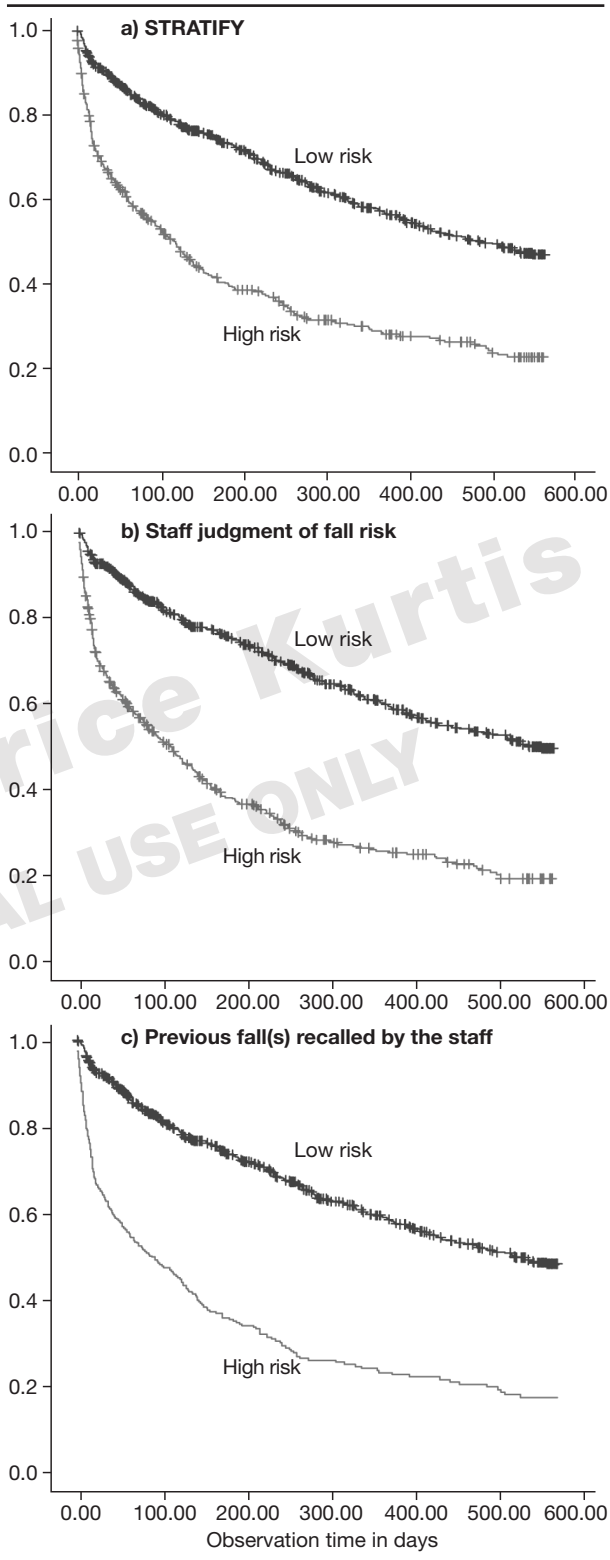

Fig. 2a-c - *Kaplan-Meier estimates of probability of sustaining a fall.*

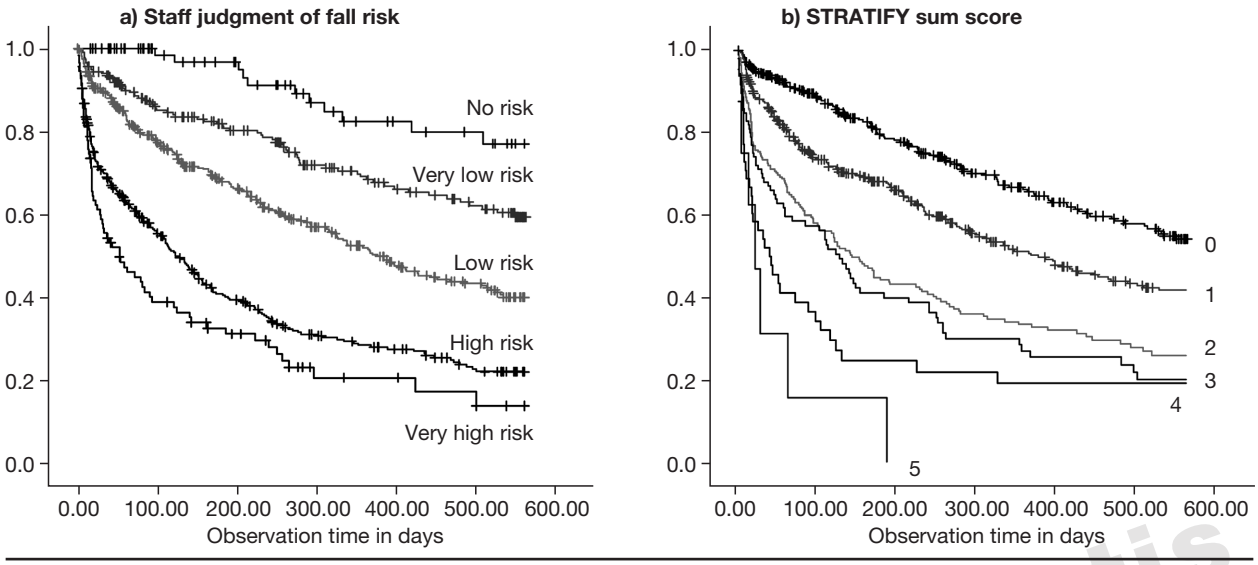

Fig. 3*a-b - Kaplan-Meier estimates of sustaining a fall according to STRATIFY sum score and staff judgment (5 categories).*

DISCUSSION

The purpose of this study was evaluation of the diagnostic accuracy of three methods for identifying nursing home residents who will sustain falls. Although other studies have reported the accuracy of these methods in different settings, this is the first large prospective study carried out in nursing homes comparing the accuracy of the three methods for assessing fall risk.

The whole point of a test is to know what probability that test has of giving the right diagnosis (32), and if it can give information beyond clinical evaluations (33). All three methods used for assessing fall risk here were easy to administer and none was time-consuming. Of the three methods, STRATIFY is probably the most time-consuming. We have shown that STRATIFY (modified for use in nursing homes), staff judgment of fall risk, and previous falls recalled by staff did not differ significantly in their ability to identify correctly residents as being at low or high risk. Within 30 days of assessment, the number of fallers correctly identified as being at high risk of falling varied from 65% to 72%, and the number of non-fallers from 69% to 75%. As may be expected, sensitivity decreased after 90 days and even more after 180 days, whereas specificity increased within the same time interval in all

Table 3 - *Cox regression of probability of sustaining a fall during observation period in unadjusted and adjusted models*.*

Variables	Hazard ratio (HR)	*p*-value	95% CI for HR	Hazard ratio (HR)	*p*-value	95% CI for HR
	Unadjusted			Adjusted⁺		
STRATIFY score >2	2.41	<0.001	2.03-2.84	2.35	<0.001	1.98-2.79
Staff judgment of fall risk	2.85	<0.001	2.41-3.67	2.78	<0.001	2.35-3.30
Previous fall in last 3 months#	3.00	<0.001	2.53-3.55	3.00	<0.001	2.52-3.57
Combination of two methods	2.71	<0.001	2.27-3.22	2.60	<0.001	2.18-3.10
STRATIFY-1	3.00	<0.001	2.53-3.55	2.56	<0.001	2.13-3.07
STRATIFY-2	1.91	<0.001	1.61-2.27	1.58	<0.001	1.31-1.90
STRATIFY-3	1.18	0.105	0.97-1.44	0.94	0.555	0.76-1.16
STRATIFY-4	1.27	0.022	1.03-1.55	1.18	0.128	0.95-1.46
STRATIFY-5	1.74	<0.001	1.46-2.08	1.47	<0.001	1.21-1.78

*Mean observation time: 225.6 days (SD 202.5); ⁺Adjusted for nursing home indicators, #Adjusted for nursing home indicators, and for other individual items in STRATIFY. STRATIFY-1= Has the resident had a fall in the previous 3 months? STRATIFY-2= Do you think the resident is agitated? STRATIFY-3= Do you think the resident is visually impaired to the extent that everyday function is affected? STRATIFY-4= Do you think the resident is in need of especially frequent toileting? STRATIFY-5= Do you think the resident has a transfer - and mobility score of more than 3 or 4?

Table 4 - *Match between staff judgment, STRATIFY and previous falls recalled by staff.*

	Kappa value*		
	Staff judgment	**STRATIFY**	**Falls within the last three months**
Staff judgment		0.50	
STRATIFY			0.62
Falls in previous 3 months	0.59		

*<0.20=poor; 0.21-0.40=fair; 0.41-0.60=moderate; 0.61-0.80=good; 0.81-1.00=very good.

three methods. The number of true positives for falls within 30 days varied from 31% to 35%, but increased after 90 and 180 days. The number of true negatives within the same period was more than 90%, but decreased after 90 and 180 days. For falls within 30 days, the number of false positives was high, but that of false negatives was low. The number of false positives decreased after 90 and 180 days, whereas the number of false negatives increased. The LR+ for the three methods was between 2 and 3 and the LR- between 0.4 and 0.6, which is regarded as low and very low, respectively (34). Kaplan-Meier estimates and log-rank tests showed that the relative risk of sustaining a fall was significantly different between those assessed as at high or low risk of falling in all assessment methods. The relative risk of sustaining a fall increased with the number of risk factors in STRATIFY and the higher risk of falling as assessed by staff. These results support the assumption that the risk of falling increases as the number of risk factors increases (35, 36).

One question, when analysing data from diagnostic tests or risk assessments, is to decide how accurate the test should be in order to be of clinical value, and to know to what extent a particular test result predicts the risk of disease (37). Sensitivity and specificity describe the behaviour of tests, when disease status is known. Predictive values give probabilities of disease for particular test results, but depend on the prevalence of abnormality in the sample. Predictive values can rarely be generalized beyond the study (37). Likelihood ratios allow us to assess the influence of a test result on the odds that a patient will have a disease (38), and is regarded as more useful to clinicians than sensitivity and specificity (38). Generally, a perfect test should have accuracy of 100%, although no test used in clinical practice can be expected to demonstrate this level of accuracy. The multifactorial causes of falls and predictions of a future event are complicating factors in fall risk assessments. The criteria for high accuracy of fall risk assessment tools are not well established. Oliver et al. (13) suggest 70% as a cut-off for high sensitivity and specificity. Only staff judgment of fall risk achieved sensitivity and specificity of approximately 70% which, according to Oliver et al. (13), is regarded as clinically useful. A combination of two of them achieved sensitivity of more than 80%.

Use of a formal risk assessment tool did not add any contribution to simply using staff judgment or previous falls remembered by staff. This matches the results of other studies (11, 14-18). The sensitivity and specificity of staff judgment in the present study was at the same level as that reported by Lundin-Olsson et al. (11), who found sensitivity of 60% and specificity of 71% for staff judgment of fall risk after 6 months (11). For fall history, the same numbers were 52% and 75%, respectively. Nordin et al. (20) found sensitivity for staff judgment of fall risk of 56% and sensitivity of 80% after 6 months. Myers et al. (14), using staff judgment of fall risk in an aged care and rehabilitation ward, found sensitivity, specificity, and positive and negative predictive values to be 88%, 26%, 18% and 92%.

The diagnostic accuracy of STRATIFY has been tested in several studies. However, the good diagnostic accuracy shown in the original population was not verified in other populations (39-42). The validity of STRATIFY has recently been studied in a nursing home setting (42) where 120 residents with a mean age of 74.5 years were followed for 13 weeks. It showed sensitivity and specificity of 52% and 74%, respectively at 5 weeks of observation time post-admission. The sensitivity and specificity of STRATIFY in that study remained stable after 3 months post-admission, with 50% and 76%, respectively. The sensitivity of STRATIFY declined with time after screening in our study, but in all 3 time periods was higher than that demonstrated by Wijna et al. (42), but not as good as in the original population (23). Milisen et al. (40) validated STRATIFY in a hospital setting and found that sensitivity was moderate among those admitted to geriatric wards and those aged 85 or older (67% and 69%, respectively).

Fall risk screening of participants was completed only once in our study, i.e., at the time of inclusion, and was not repeated during the intervention period. The characteristics of residents in a nursing home setting may change quite fast. How often a risk assessment should be repeated is not clear. Our data showed that the number of fallers with positive tests was highest within 30 days of assessment. This indicates that screening for fall risk should be repeated regularly. However, further research is needed to define optimal frequency intervals (43).

"Staff judgment of fall risk" is connected to the staff's clinical experience and knowledge, and accuracy has been shown to vary between levels of nurses (14). Exactly what staff members include in their assessment of fall risk is unknown. Previous falls are probably important, but they are not the only factor. The moderate match between all three methods studied here indicates that staff in-

clude factors other than those captured in STRATIFY, and factors other than "previous falls" when they classify residents fall risk. This finding matches that emphasized by Nordin et al. (20).

The number of fallers and falls depends on the fall definition used (31) and the way falls are recorded (3, 44). The fall definition used in this study was restricted to events when the resident came to rest on the floor, but included falls which may have occurred as a consequence of acute medical events or external forces. Using another topographical component, such as a "lower level", as proposed by Lamb et al. (45), rather than the floor, would probably give a higher number of fallers and falls, and another result concerning the prognostic accuracy of assessment tools. Falls occurring because of an acute medical event or external forces are more difficult to predict. Consequently, including all falls, regardless of cause may have influenced our results negatively concerning the sensitivity, specificity and predictive value of the tools.

The strengths of our study were the prospective design and the high number of participants. In addition, the fact that staff kept fall records shortly after the fall provided accuracy of recorded falls and minimized the probability of recall bias. We were also able to evaluate the accuracy of the various methods within different periods, showing that sensitivity, specificity, predictive value and likelihood ratios changed over time. The fact that the sum score of STRATIFY was concealed for the staff who carried out the risk assessment decreased the possibility that the STRATIFY sum score influenced staff judgment of fall risk.

The study has some limitations. First, although no formal fall prevention interventions were implemented, we cannot exclude the possibility of a larger focus on fall prevention when falls and fracture risk were the topic of the hip protector study. This may have resulted in a lower number of falls and fallers. However, the nursing home staff were asked not to change their usual routine of fall prevention during the study period. Second, studying inter-observer agreement for the three methods was not within the scope of our study and the agreement between different observers is not known. The items in STRATIFY are prone to subjective evaluation, which compromises reproducibility between observers (41). The inter-observer agreement of STRATIFY has been reported to range from 0.742 (Cohen's Kappa) (39) to 0.78 (ICC) (41), which is regarded as good (46). "Staff judgment of fall risk" is by nature subjective, and dependent on clinical experience and knowledge. "Falls in the previous 3 months" is an objective measure, but may be prone to recall bias.

It should be emphasized that the instruments used in the present study are meant for fall risk screening. These global scores cannot lead to specific fall prevention interventions, but may be useful in identifying residents who are in need of more comprehensive assessment of fall risk or who need a hip protector.

Our study population was similar in its age and gender to that of Norwegian nursing home residents in general (47, 48). However, any generalizations to other nursing homes generally must be made with caution, because of our study setting, and because the nursing homes were not a random sample from all nursing homes in Norway.

CONCLUSIONS

The three methods evaluated here did not differ markedly in their ability to differentiate between fallers and non-fallers. Of the three methods, staff judgement of fall risk was the single risk assessment method showing the highest sensitivity. Use of a formal risk assessment tool such as STRATIFY did not improve diagnostic accuracy. The risk of sustaining a fall increased significantly with the number of risk factors in STRATIFY and with the higher degree of fall risk as assessed by staff. The match among the three methods was moderate, which indicates that they captured different risk factors for falls.

ACKNOWLEDGEMENTS

The authors wish to thank the residents and staff of participating nursing homes. Special thanks go to the study coordinators for their support and help in collecting data. Appreciation is also extended to the Norwegian Institute for Health and Rehabilitation for funding this study.

Disclaimer: The authors have no competing interests.

REFERENCES

1. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age Ageing 2006; 35: 37-41.
2. Rubenstein LZ, Josephson KR, Robbins AS. Falls in the nursing home. Ann Intern Med 1994; 121: 442-51.
3. Lord SR. Falls in older people: risk factors and strategies for prevention. Cambridge: Cambridge University Press, 2007.
4. Blake AJ, Morgan K, Bendall MJ et al. Falls by elderly people at home: prevalence and associated factors. Age Ageing 1988; 17: 365-72.
5. Campbell AJ, Reinken J, Allan BC, Martinez GS. Falls in old age: a study of frequency and related clinical factors. Age Ageing 1981; 10: 264-70.
6. Prudham D, Evans JG. Factors associated with falls in the elderly: a community study. Age Ageing 1981; 10: 141-6.
7. Luukinen H, Koski K, Honkanen R, Kivela SL. Incidence of injury-causing falls among older adults by place of residence: a population-based study. J Am Geriatr Soc 1995; 43: 871-6.
8. Luukinen H, Koski K, Hiltunen L, Kivela SL. Incidence rate of falls in an aged population in northern Finland. J Clin Epidemiol 1994; 47: 843-50.
9. Rubenstein LZ, Josephson KR, Osterweil D. Falls and fall prevention in the nursing home. Clin Geriatr Med 1996; 12: 881-902.
10. de Kinkelder A, Dierkx RI. Functional mobility tests of proclivity for falls in nursing home patients. Tijdschr Gerontol Geriatr 2001; 32: 69-73.
11. Lundin-Olsson L, Jensen J, Nyberg L, Gustafson Y. Predicting falls in residential care by a risk assessment tool, staff judgement, and history of falls. Aging Clin Exp Res 2003; 15: 51-9.
12. Scott V, Votova K, Scanlan A, Close J. Multifactorial and func-

tional mobility assessment tools for fall risk among older adults in community, home-support, long-term and acute care settings. Age Ageing 2007; 36: 130-9.

13. Oliver D, Daly F, Martin FC, McMurdo ME. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. Age Ageing 2004; 33: 122-30.
14. Myers H, Nikoletti S. Fall risk assessment: a prospective investigation of nurses' clinical judgement and risk assessment tools in predicting patient falls. Int J Nurs Pract 2003; 9: 158-65.
15. Eagle DJ, Salama S, Whitman D, Evans LA, Ho E, Olde J. Comparison of three instruments in predicting accidental falls in selected inpatients in a general teaching hospital. J Gerontol Nurs 1999; 25: 40-5.
16. Haines TP, Hill K, Walsh W, Osborne R. Design-related bias in hospital fall risk screening tool predictive accuracy evaluations: systematic review and meta-analysis. Gerontol Series A Biol Sci Med Sci 2007; 62: 664-72.
17. Vassallo M, Poynter L, Sharma JC, Kwan J, Allen SC. Fall risk-assessment tools compared with clinical judgment: an evaluation in a rehabilitation ward. Age Ageing 2008; 37: 277-81.
18. Meyer G, Kopke S, Haastert B, Muhlhauser I. Comparison of a fall risk assessment tool with nurses' judgement alone: a cluster-randomised controlled trial. Age Ageing 2009; 38: 417-23.
19. Lundin-Olsson L, Nyberg L, Gustafson Y. The Mobility Interaction Fall chart. Physiother Res Int 2000; 5: 190-201.
20. Nordin E, Lindelof N, Rosendahl E, Jensen J, Lundin-Olsson L. Prognostic validity of the Timed Up-and-Go test, a modified Get-Up-and-Go test, staff's global judgement and fall history in evaluating fall risk in residential care facilities. Age Ageing 2008; 37: 442-8.
21. Rosendahl E, Lundin-Olsson L, Kallin K, Jensen J, Gustafson Y, Nyberg L. Prediction of falls among older people in residential care facilities by the Downton index. Aging Clin Exp Res 2003; 15: 142-7.
22. Bentzen H, Forsen L, Becker C, Bergland A. Uptake and adherence with soft- and hard-shelled hip protectors in Norwegian nursing homes: a cluster randomised trial. Osteoporos Int 2008; 19: 101-11.
23. Oliver D, Britton M, Seed P, Martin FC, Hopper AH. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. BMJ 1997; 315: 1049-53.
24. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965; 14: 61-5.
25. Rogmark C, Sernbo I, Johnell O, Nilsson JA. Incidence of hip fractures in Malmo, Sweden, 1992-1995. A trend-break. Acta Orthop Scand 1999; 70: 19-22.
26. Jensen J, Lundin-Olsson L, Nyberg L, Gustafson Y. Falls among frail older people in residential care. Scand J Public Health 2002; 30: 54-61.
27. Akobeng AK. Understanding diagnostic tests 2: likelihood ratios, pre- and post-test probabilities and their use in clinical practice. Acta Paediatr 2007; 96: 487-91.
28. Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. In Guyatt G, Rennie D, Eds. Users' guides to the medical literature. A manual for evidence-based clinical practise. Chigago: AMA Press, 2002.
29. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall, 1991.
30. Ganz DA, Bao Y, Shekelle PG, Rubenstein LZ. Will my patient fall? JAMA 2007; 297: 77-86.
31. Kiely DK, Kiel DP, Burrows AB, Lipsitz LA. Identifying nursing home residents at risk for falling. J Am Geriatr Soc 1998; 46: 551-5.
32. Altman DG, Bland JM. Diagnostic tests 2: predictive values. BMJ 1994; 309: 102.
33. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. QJM 1998; 91: 247-58.
34. Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. JAMA 1994; 271: 703-7.
35. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. N Engl J Med 1988; 319: 1701-7.
36. Tinetti ME. Clinical practice. Preventing falls in elderly persons. N Engl J Med 2003; 348: 42-9.
37. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. BMJ 2004; 329: 168-9.
38. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol 1991; 44: 763-70.
39. Coker E, Oliver D. Evaluation of the STRATIFY falls prediction tool on a geriatric unit. Outcomes Management 2003; 7: 8-14.
40. Milisen K, Staelens N, Schwendimann R et al. Fall prediction in inpatients by bedside nurses using the St. Thomas's Risk Assessment Tool in Falling Elderly Inpatients (STRATIFY) instrument: a multicenter study. J Am Geriatr Soc 2007; 55: 725-33.
41. Papaioannou A, Parkinson W, Cook R, Ferko N, Coker E, Adachi JD. Prediction of falls using a risk assessment tool in the acute care setting. BMC Med 2004; 2: 1.
42. Wijnia JW, Ooms ME, van Balen R. Validity of the STRATIFY risk score of falls in nursing homes. Prev Med 2006; 42: 154-7.
43. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. Gerontol Series A Biol Sci Med Sci 2001; 56: 761-6.
44. Hauer K, Lamb SE, Jorstad EC, Todd C, Becker C; PROFANE-Group. Systematic review of definitions and methods of measuring falls in randomised controlled fall prevention trials. Age Ageing 2006; 35: 5-10.
45. Lamb SE, Jorstad-Stein EC, Hauer K, Becker C; Prevention of Falls Network Europe and Outcomes Consensus Group. Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. J Am Geriatr Soc 2005; 53: 1618-22.
46. Domholdt E. Rehabilitation research: principles and applications. St. Louis: Elsevier Saunders, 2005.
47. Kirkevold O, Engedal K. Prevalence of patients subjected to constraint in Norwegian nursing homes. Scand J Caring Sci 2004; 18: 281-6.
48. Kirkevold O, Engedal K. The quality of care in Norwegian nursing homes. Scand J Caring Sci 2006; 20: 177-83.