

RESEARCH PAPER

Retrospective evaluation of the world falls guidelines-algorithm in older adults

BOB VAN DE LOO^{1,2,3}, MARTIJN W. HEYMAN^{1,3}, STEPHANIE MEDLOCK^{3,4}, AMEEN ABU-HANNA^{3,4},
NATHALIE VAN DER VELDE^{2,4}, NATASJA M. VAN SCHOOR^{1,3}

¹Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, De Boelelaan 1117, Amsterdam, Netherlands

²Amsterdam UMC location University of Amsterdam, Internal Medicine, Section of Geriatric Medicine, Meibergdreef 9, Amsterdam, Netherlands

³Amsterdam Public Health research institute, Aging and Later Life, Meibergdreef 9, Amsterdam, The Netherlands

⁴Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Meibergdreef 9, Amsterdam, Netherlands

Address correspondence to: Nathalie van der Velde. Email: n.vandervelde@amsterdamumc.nl

Abstract

Background: The World Falls Guidelines (WFG) propose an algorithm that classifies patients as low-, intermediate-, and high-risk. We evaluated different operationalizations of the WFG algorithm and compared its predictive performance to other screening tools for falls, namely: the American Geriatrics Society and British Geriatrics Society (AGS/BGS) algorithm, the 3KQ on their own and fall history on its own.

Methods: We included data from 1509 adults aged ≥ 65 years from the population-based Longitudinal Aging Study Amsterdam. The outcome was ≥ 1 fall during 1-year follow-up, which was ascertained using fall calendars. The screening tools' items were retrospectively operationalized using baseline measures, using proxies where necessary.

Results: Sensitivity ranged between 30.9–48.0% and specificity ranged between 77.0–88.2%. Operationalizing the algorithm with the 3KQ instead of fall history yielded a higher sensitivity but lower specificity, whereas operationalization with the Clinical Frailty Scale (CFS) classification tree instead of Fried's frailty criteria did not affect predictive performance. Compared to the WFG algorithm, the AGS/BGS algorithm and fall history on its own yielded similar predictive performance, whereas the 3KQ on their own yielded a higher sensitivity but lower specificity.

Conclusion: The WFG algorithm can identify patients at risk of a fall, especially when the 3KQ are included in its operationalization. The CFS and Fried's frailty criteria may be used interchangeably in the algorithm's operationalization. The algorithm performed similarly compared to other screening tools, except for the 3KQ on their own, which have higher sensitivity but lower specificity and lack clinical recommendations per risk category.

Keywords: falls; prediction; stratification; screening; older people

Key Points

- The World Falls Guidelines (WFG) algorithm can help identify individuals with an elevated risk of falls.
- Operationalizing the algorithm with the Three Key Questions (3KQ) instead of the single fall history question helps identify more at-risk patients, at the cost of more overtreatment.
- Our findings suggest that the Clinical Frailty Scale and Fried's frailty criteria can be used interchangeably in the algorithm's operationalization.
- The predictive performance of the algorithm was similar to that of the other assessed screening tools, except for the Three Key Questions (3KQ) on their own, which have a higher sensitivity but lower specificity and lack clinical recommendations per risk category.

Introduction

Falls are a major growing public health issue in the older population. Approximately one out of every three older adults falls at least once each year [1]. Approximately 10% of falls result in a serious injury, such as fractures or head injuries [2]. The number of falls will likely further increase in the coming years due to the aging population and due to an increase in multimorbidity, polypharmacy and frailty [3].

Fall risk stratification is a prerequisite for efficient casefinding and resource-efficient use of fall prevention interventions. Recently, the World Falls Guidelines (WFG) Taskforce published an algorithm that clinicians can use for case finding and risk stratification [3]. Based on self-reported questions and a functional assessments, the WFG algorithm classifies patients into three risk levels: low, intermediate, or high risk [3]. The WFG provides recommendations on further risk assessment and fall prevention strategies for each risk category [3]. The development of the WFG algorithm followed the American Geriatrics Society and British Geriatrics Society (AGS/BGS) algorithm [4].

The WFG allow some flexibility in how their algorithm can be operationalized in clinical practice. For example, clinicians can use the single question about fall history ('have you fallen in the last 12 months') or, to increase sensitivity, use the more comprehensive Three Key Questions (3KQ) in case finding. The 3KQ includes the single fall history question and two additional questions about unsteadiness and worries about falling. The WFG also provide different options for determining frailty and mobility, which are included as items in the algorithm.

To date, only one cohort study evaluated the WFG algorithm and therefore little is known on the extent to which the algorithm can help identify patients with an increased risk of falling [5]. In this study, the researchers only evaluated the WFG algorithm as operationalized with the 3KQ and the Clinical Frailty Scale (CFS) classification tree. Furthermore, the researchers only compared the predictive performance of the algorithm to the single fall history question. Therefore, it is unclear how the predictive performance of the WFG algorithm compares to the AGS/BGS algorithm and the 3KQ on their own. It is of interest to compare the WFG algorithm to more practical screening strategies. The 3KQ, as well as the single fall history question, are commonly used on their own to determine whether fall preventive measures should be applied [6].

This study retrospectively evaluates the WFG algorithm using data from older adults in the context of case finding. Here, we explore the use of the 3KQ, the single fall history question, and different frailty assessment instruments in the algorithm's operationalization. Additionally, we seek to compare the predictive performance of the WFG algorithm to that of other screening tools, i.e. the AGS/BGS algorithm, the 3KQ on their own and the single fall history question on its own.

Methods

Using data from a prospective cohort study, we retrospectively assessed the following screening tools: WFG algorithm with the 3KQ and Fried's frailty criteria, WFG algorithm with the 3KQ and the CFS classification tree, WFG algorithm with the single fall history question and Fried's frailty criteria, AGS/BGS algorithm, the 3KQ on its own and the single fall history question on its own. We followed the Strengthening the Reporting of Observational Studies in Epidemiology reporting guideline [7].

Study population

We used data from the Longitudinal Aging Study Amsterdam (LASA). LASA is an ongoing cohort study that focuses on the determinants, trajectories and consequences of physical, cognitive, emotional and social functioning in older adults [8]. In 1992–1993, a random sample of participants aged 55 years and over was drawn across three regions in the Netherlands. LASA was approved by the medical ethics committee of the VU University Medical Center. All participants signed informed consent. We used baseline and 1-year follow-up data from wave C (1995/1996), which had 1509 participants that were all aged 65 years and older as of January 1, 1996. Of these, 1435 (95.1%) participated in the follow-up measurements.

Outcome

Falls during a 1-year follow-up were measured using falls calendars. A fall was defined as an unintentional change in position resulting in coming to rest at a lower level or on the ground. Participants were asked to record falls every week and to return the calendar every 3 months. Participants were contacted if the calendars were not returned or filled out incorrectly. The outcome was coded as missing if loss to follow-up occurred before a fall was reported. The ascertainment and definition of falls were consistent with guidelines from the Prevention of Falls Network Europe [9]. As secondary outcomes, we report on fractures, the number of falls and the fall rate in the different risk categories in the follow-up period. Number of falls for each individual was calculated based on the number of weeks in which a fall was reported. Fall rate for each risk category was calculated as the total number of falls divided by the number of participants in that category.

Operationalization of screening tools and baseline characteristics

We operationalized the different items for the screening tools based on available data (Table 1). For the WFG and AGS/BGS, we only considered the items relevant for risk stratification in case finding, meaning that items related to management interventions were not considered. Two items in the WFG algorithm could not be operationalized as there was no data or appropriate proxy for these, i.e.: *Lying on the*

Table 1. Operationalization of items in screening tools

Screening Tool item	Item No	Operationalization
<i>Fall history</i>		
Have you fallen in the past year?	1	Fall history in the past year was assessed using the following question: <i>Did you fall in the past year?</i>
<i>Three Key Questions</i>		
Have you fallen in the past year?	2a	See item 1.
Do you feel unsteady when standing or walking?	2b	Subjective steadiness when standing or walking was not assessed. Therefore, we used data from the tandem test as a proxy for subjective steadiness. Participants who could not hold the tandem stand for at least 10 s were coded as having impaired balance [10].
Do you have worries about falling?	2c	Participants that scored ≥ 1 on a modified Falls Efficacy Scale (FES) were coded as having concerns about falling [11].
<i>World Falls Guidelines algorithm</i>		
Fall in the past 12 months?	3a	See item 1.
Injury	3b	Data with regard to injuries, only fractures were assessed. Participants were asked if they endured a fracture since their last interview.
≥ 2 falls in last year	3c	Number of falls in past 12 months was assessed using the following question: <i>How often did you fall in the past year?</i>
Frailty	3d	Frailty was operationalized using Fried's frailty phenotype and the Clinical Frailty Scale (see Supplementary File 1 for details) [12].
Lying on the floor/unable to get up	3e	Not measured
Loss of consciousness/suspected syncope	3f	Not measured
Gait & balance impaired?	3g	Participants were asked to walk 3 meters, to turn around and walk back 3 meters as quickly as possible. The gait speed was recoded into an estimated time to walk 3 meters [13]. Participants with a recoded gait speed < 0.8 m/s were coded as having abnormalities in gait or unsteadiness.
<i>American Geriatrics Society and British Geriatrics Society algorithm</i>		
Two or more falls in prior 12 months?	4a	See item 3c.
Difficulty with walking or balance?	4b	Participants were asked the following question: <i>Can you walk outside during five minutes without stopping?</i> Participants were coded as having any difficulty with walking or balance if they answered with: <i>No, I cannot; Only with help; Yes, with much difficulty; or Yes, with some difficulty.</i>
Does the person report a single fall in the past 12 months?	4c	See item 1.
Are abnormalities in gait or unsteadiness identified?	4d	See item 3g.

floor/unable to get up and *Loss of consciousness/suspected syncope* (Table 1, items 3e and 3f). Additional baseline characteristics included in the analysis were age, sex, body mass index, and educational status. Educational status was based on the number of years of education followed.

Statistical analysis

We assessed the predictive performance of screening tools in identifying individuals that endure one or more falls within one year after assessment. Based on available data about falls during follow-up and the operationalization of the screening tools, we calculated the number of true positives (TPs), true negatives (TNs), false positives, and false negatives. Using these, we calculated sensitivity and specificity measures for the different screening tools. Additionally, positive and negative predictive values and accuracy are presented to aid clinical interpretability of the findings. The WFG algorithm uses three risk levels and we therefore calculated its predictive performance by contrasting the risk groups in the following ways:

- Low versus high risk (TPs are those classified as high risk with a fall in follow-up and TNs are those classified as low risk with no fall in follow-up);

- Low versus intermediate to high risk (TPs are those classified as intermediate or high risk with a fall in follow-up and TNs are those classified as low risk with no fall in follow-up); and
- Low to intermediate versus high risk (TPs are those classified as high risk with a fall in follow-up and TNs are those classified as low or intermediate risk with no fall in follow-up).

We additionally assessed the predictive performance of the screening tools separately for men and women as well as for the following age strata: 65 through 69, 70 through 74, 75 through 80 and 80 years or older. Statistical analyses were performed using the R (version 4.0.2) statistical programming language.

Sensitivity analysis

As a sensitivity analysis, we tested choosing a different cut-off value for determining worries about falling. The modified Falls Efficacy Scale (FES) that was used has no validated cut-off point. In the main analysis, we opted to maximize the instrument's sensitivity by using a cut-off value of 1, similarly to Pluijm et al [14]. We tested the performance of the WFG algorithm and 3KQ when using a cut-off value of 3, as done in Tromp et al [15].

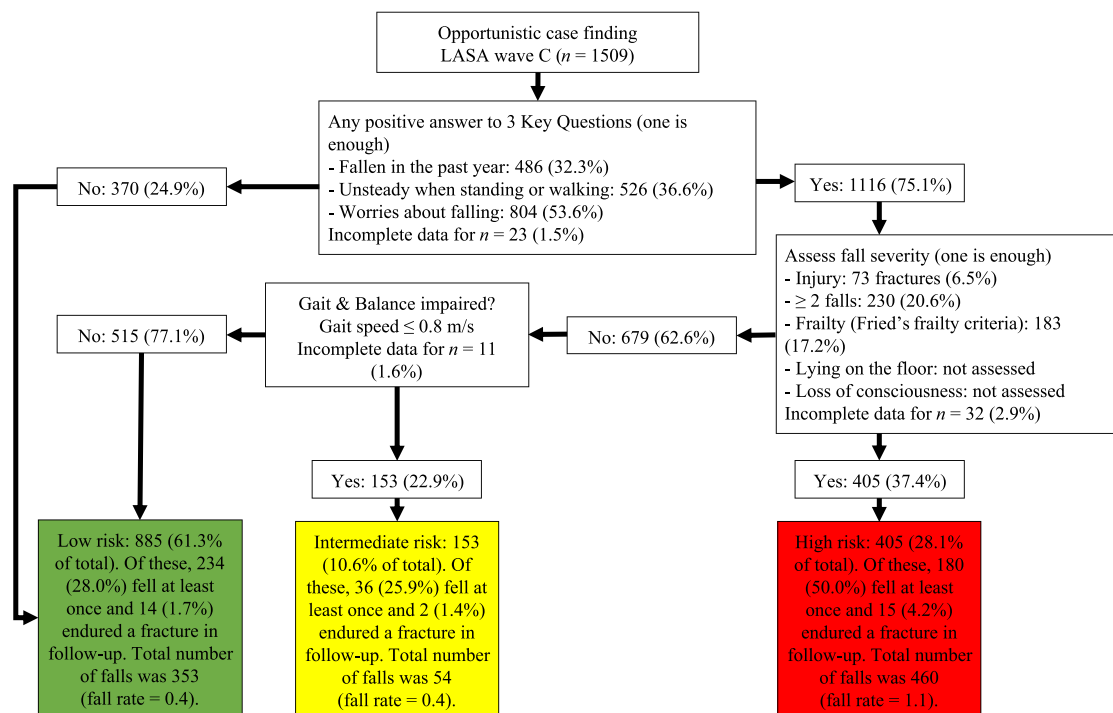


Figure 1. **Operationalization of the World Falls Guidelines algorithm using Three Key Questions and Fried's frailty criteria.** LASA = Longitudinal Aging Study Amsterdam. Unless stated otherwise, proportions are calculated based on the number of participants with complete data in the preceding step. Proportions are provided as valid percentages, meaning missing data were not included in their calculations. A total of 66 participants (4.4%) could not be classified due to incomplete data.

Results

Data of 1509 participants was included in our analyses. Of those that participated in the follow-up measurement ($n = 1435$), 1381 (91.5%) participants completed all fall calendars in the one-year follow-up with a median follow-up length of 52 months (interquartile range: 52–52 months). Of these, 468 (33.9%) reported one or more falls and 32 (2.3%) reported one or more fractures during follow-up.

When the WFG algorithm was operationalized with the 3KQ and Fried's frailty criteria, 885 (61.3%) participants were classified as low-risk, 153 (10.6%) as intermediate-risk, and 405 (28.1%) as high-risk (Figure 1). The risk groups differed with respect to all baseline characteristics ($P < 0.05$; Table 2). During the one-year follow-up, the respective incidences of one or more falls in the low-, intermediate- and high-risk groups were 234 (28.0%), 36 (25.9%) and 180 (50.0%). The respective incidences of fractures in the risk groups were 20 (1.9%), 2 (4.9%) and 10 (4.1%), while the total number of falls were 353 (fall rate = 0.4), 54 (fall rate = 0.4) and 460 (fall rate = 1.1).

When the WFG algorithm was operationalized using the single fall history question instead of the 3KQ, 1170 (78.4%) participants were classified as low-risk, 45 (3.0%) as intermediate-risk, and 277 (18.6%) as high-risk (Supplementary Figure 1). With this operationalization of the WFG algorithm, the respective incidences of one or more falls in low- to high-risk groups were 305 (28.3%), 15 (36.6%) and 143 (57.2%), while the respective incidences

of fractures were 20 (1.9%), 2 (4.9%) and 10 (4.1%). The respective total number of falls were 488 (fall rate = 0.4), 22 (fall rate = 0.5) and 385 (fall rate = 1.4).

We assessed the performance of the following screening tools in predicting one or more falls in one year: WFG algorithm with the 3KQ and Fried's frailty criteria; WFG algorithm with the 3KQ and the CFS classification tree; WFG algorithm with the single fall history question and Fried's frailty criteria; AGS/BGS algorithm; the 3KQ on its own and the single fall history question on its own (Table 3). When the WFG algorithm was operationalized with the 3KQ and Fried's frailty criteria, we obtained sensitivity values between 40.0% to 48.0% and specificity values between 77.0% to 79.6% by contrasting the risk groups in different ways. Use of the CFS classification tree in the WFG algorithm with 3KQ yielded similar sensitivity (range: 37.6–48.5%) and specificity values (range: 67.7–83.5%) as when Fried's frailty scale was used, with all 95% CIs overlapping. When the WFG algorithm was operationalized with the single fall history question instead of the 3KQ, we obtained lower sensitivity values (range: 30.9–34.1%) but higher specificity values (range: 85.3–88.2%). Figure 2 visualizes the operationalization of the AGS/BGS algorithm, for which we obtained a sensitivity of 44.7% and a specificity of 73.1%. Out of the assessed screening tools, use of the 3KQ alone yielded the best sensitivity (sensitivity = 86.0), but also the lowest specificity (specificity = 31.9%). Finally, we obtained respective sensitivity and specificity values of 47.9% and 75.5% for the single fall history question.

Table 2. Baseline characteristics of the different risk groups as classified by the world falls guidelines algorithm with three key questions and Fried's frailty criteria

	Overall (<i>n</i> = 1509)	Low Risk (<i>n</i> = 885)	Intermediate Risk (<i>n</i> = 153)	High Risk (<i>n</i> = 405)	<i>p</i> Value ^a	% missing
Age	75.5 [70.1, 81.5]	72.8 [68.7, 78.5]	80.0 [74.2, 83.6]	79.3 [72.9, 84.0]	<.001	0
Female sex	718 (52.6)	355 (45.2)	95 (66.4)	234 (62.4)	<.001	9.5
Education level attained (years)	9 [6, 11]	9.0 [6, 11]	6 [6, 10]	9 [6, 10]	<.001	9.5
History of 1 or more falls in the previous 12 months	486 (32.3)	152 (17.2)	45 (29.4)	277 (68.4)	<.001	0.3
Modified FES Score (0–30)	1 [0, 3]	0 [0, 2]	2 [1, 5]	3 [0, 8]	<.001	0.7
Modified FES Score \geq 1	804 (53.6)	361 (40.8)	120 (78.4)	294 (73.1)	<.001	0.7
Modified FES Score \geq 3	457 (30.5)	154 (17.4)	66 (43.1)	216 (53.7)	<.001	0.7
Body mass index	26.9 (4.4)	26.6 (3.9)	27.5 (4.9)	27.3 (4.9)	0.013	11.3
Gait speed \leq 0.8 m/s	406 (28.8)	43 (4.9)	153 (100.0)	205 (56.0)	<.001	6.4
Not able to hold tandem stand for at least 10 s	526 (36.6)	191 (21.6)	81 (53.6)	234 (61.9)	<.001	4.6
Fracture since last interview	83 (5.5)	9 (1.0)	0 (0.0)	73 (18.0)	<.001	0.3
Frailty according to Fried's frailty criteria	196 (13.6)	10 (1.1)	0 (0.0)	183 (47.8)	<.001	4.6
Frailty according to CFS classification tree	173 (11.5)	23 (2.6)	14 (9.2)	104 (25.9)	<.001	0.7

CFS = Clinical Frailty Scale; FES = Falls Efficacy Scale. Data are presented as *n* (%) or median [interquartile range]. ^a*p* Values were calculated using chi-square, Kruskal–Wallis, and ANOVA tests.

We assessed the performance of the screening tools according to different age and sex strata (Supplementary Table 1). In general, the sensitivity values of the screening instruments were higher in the older age strata and in women. Conversely, specificity values were generally higher in younger age strata and in men.

As a sensitivity analysis, we explored the use of a different cut-off for the modified FES scale (Supplementary Table 2). Use of a cut-off value of 3 for the FES in the WFG algorithm also resulted in similar sensitivity (range: 38.8–46.1%) and specificity values (range: 70.5–80.3%) to those obtained when a cut-off value of 1 was used. However, when the 3KQ were used on its own, choosing a cut-off value of 3 instead of 1 for the FES resulted in a higher specificity of 43.0%, but a lower sensitivity of 76.2%.

Discussion

This study evaluated the WFG algorithm. Participants classified as high-risk by the algorithm were found to have higher risk of falls as compared with those classified as low- or intermediate-risk. The risk of falls in the intermediate-risk group was found to be higher than in the low-risk group, but only when the algorithm was operationalized with the single fall history question. When the algorithm was operationalized with the 3KQ, we found a similar risk of falls in the intermediate- and low-risk groups. The algorithm's sensitivity was best when operationalized with the 3KQ whereas its specificity was best when operationalized with the single fall history question. Use of the CFS classification tree instead of Fried's frailty in the algorithm's operationalization resulted in similar predictive performance. The AGS/BGS algorithm and the single falls history question on its own yielded similar predictive performance compared to the WFG algorithm with 3KQ. The 3KQ on their own yielded a higher sensitivity but lower specificity compared with the WFG algorithm with 3KQ.

Hartley and colleagues reported mostly similar predictive performance measures for the WFG algorithm with the 3KQ and the CFS classification tree for those aged \geq 65 years [5]. Unlike the study by Hartley et al., we observed a drop in specificity when the intermediate- and high-risk groups were combined owing to the larger size of the intermediate-risk group in our study. Hartley and colleagues used the timed up and go test for assessing gait and balance impairments, which was more conservative than the gait speed test used in our study. Among participants tested for gait and balance impairments, 22.9% of our sample went on to being classified as intermediate-risk, versus 4.3% of the total sample in the study by Hartley et al [5]. Out of all screening tools, the 3KQ on their own yielded the highest sensitivity but also the lowest specificity, which is line with the findings of Burns et al [6]. Burns et al. also found the 3KQ to have higher sensitivity but lower specificity as compared with the single fall history question and the AGS/BGS algorithm. However, Burns and colleagues obtained respective sensitivity and specificity values of 68.7% and 57.9% for the 3KQ, versus 86.0% and 31.9% in our study [6]. These differences may be explained by our use of proxies and differences in outcome prevalence.

We compared the predictive performance of the WFG algorithm to the AGS/BGS algorithm, the 3KQ on their own, and the single fall history question on its own. Predictive performance of the AGS/BGS algorithm was similar to that of the WFG algorithm with the 3KQ. However, it should be noted that, unlike the AGS/BGS algorithm, the WFG algorithm prescribes targeted preventive measures for every risk level and thus minimizes the risk of undertreatment. Out of all assessed tools, use of the 3KQ on their own resulted in the highest proportion of participants being classified as being at risk, which contributed to its high sensitivity and limited specificity. Predictive performance of the single fall history question was similar to those of the WFG algorithm with 3KQ, similarly to Hartley and colleagues [5]. This demonstrates that practical screening strategies can have predictive performance similar to a more

Table 3. Performance of screening tools for predicting one or more falls in 1 year

Screening tool	TPs	TNs	% Sensitivity (95% CI)	% Specificity (95% CI)	% PPV (95% CI)	% NPV (95% CI)	% Accuracy (95% CI)
World falls guidelines algorithm using the Three Key Questions and Fried's frailty criteria							
Low versus high risk	180	601	43.5 (40.7, 46.3)	77.0 (74.6, 79.4)	50.0 (47.2, 52.8)	72.0 (69.5, 74.5)	65.4 (62.7, 68.1)
Low versus intermediate to high risk	216	601	48.0 (45.3, 50.7)	68.0 (65.5, 70.5)	43.3 (40.6, 46.0)	72.0 (69.6, 74.4)	61.2 (58.6, 63.8)
Low to intermediate versus high risk	180	704	40.0 (37.4, 42.6)	79.6 (77.4, 81.8)	50.0 (47.3, 52.7)	72.3 (69.9, 74.7)	66.3 (63.8, 68.8)
World falls guidelines algorithm using the Three Key Questions and Clinical Frailty Scale classification tree							
Low versus high risk	172	602	42.2 (39.4, 45.0)	80.4 (78.1, 82.7)	53.9 (51.0, 56.8)	71.8 (69.2, 74.4)	66.9 (64.2, 69.6)
Low versus intermediate to high risk	222	602	48.5 (45.8, 51.2)	67.7 (65.2, 70.2)	43.6 (41.0, 46.2)	71.8 (69.4, 74.2)	61.2 (58.6, 63.8)
Low to intermediate versus high risk	172	742	37.6 (35.0, 40.2)	83.5 (81.5, 85.5)	53.9 (51.2, 56.6)	72.2 (69.8, 74.6)	67.9 (65.4, 70.4)
World falls guidelines algorithm using the single fall history question							
Low versus high risk	143	773	31.9 (29.4, 34.4)	87.8 (86.0, 89.6)	57.2 (54.5, 59.9)	71.7 (69.3, 74.1)	69.0 (66.5, 71.5)
Low versus intermediate to high risk	158	773	34.1 (31.6, 36.6)	85.3 (83.4, 87.2)	54.3 (51.7, 56.9)	71.7 (69.3, 74.1)	68.0 (65.5, 70.5)
Low to intermediate versus high risk	143	799	30.9 (28.5, 33.3)	88.2 (86.5, 89.9)	57.2 (54.6, 59.8)	71.4 (69.0, 73.8)	68.8 (66.3, 71.3)
American Geriatrics Society and British Geriatrics Society algorithm							
Low versus high risk	208	662	44.7 (42.1, 47.3)	73.1 (70.8, 75.4)	46.0 (43.4, 48.6)	72.0 (69.6, 74.4)	63.5 (61.0, 66.0)
Three Key Questions	398	287	86.0 (84.2, 87.8)	31.9 (29.4, 34.4)	39.3 (36.7, 41.9)	81.5 (79.4, 83.6)	50.2 (47.5, 52.9)
Single fall history question	223	688	47.9 (45.3, 50.5)	75.5 (73.2, 77.8)	50.0 (47.4, 52.6)	73.9 (71.6, 76.2)	66.2 (63.7, 68.7)

CI = confidence interval; NPV = negative predictive value; PPV = positive predictive value; TPs = true positive cases; TNs = true negative cases.

comprehensive screening algorithm. However, this does not mean that these tools can be recommended over the WFG algorithm. This is because the 3KQ and the single falls history question provide no guidance on what interventions should be used. Moreover, similarly to the AGS/BGS algorithm, the 3KQ and the single fall history question classify patients according to only two risk levels, whereas the WFG algorithm classifies patients according to three risk levels. The intermediate-risk group may in theory help identify patients that would especially benefit from exercise interventions, such as sedentary or pre-frail older adults. Nonetheless, the use of the intermediate-risk group requires administering a physical performance test, even though the utility of the intermediate-risk group remains uncertain. Indeed, we found the low- and intermediate-risk groups to have comparable incidences of falls, numbers of fractures, and fall rates when the WFG was used with the 3KQ. When the WFG was used with the single fall history question, we observed that the intermediate-risk group had higher rates across all outcomes as compared with the low-risk group. However, under this operationalization, only 3.0% of the participants in our sample were classified as intermediate-risk, versus 10.6% when the WFG was operationalized with the 3KQ.

Our findings have important clinical implications. First, our findings illustrate how the use of different assessments in the algorithm affect how patients are categorized and subsequently treated. We recommend healthcare professionals to use the algorithm with the 3KQ instead of the falls history question, as this helps treat more at-risk patients. However, in situations where time or other resources are scarce, healthcare professionals may opt for using the single falls history question in the algorithm, as this prevents overtreatment. Furthermore, the gait speed test appears to be preferred over the timed up and go test. As noted earlier, the results of our study and those of Hartley et al. [5] together suggest that the gait speed test yields a substantially larger intermediate-risk group compared to the timed up and go test. Moreover, our findings suggest that the CFS and Fried's frailty can be used interchangeably in the algorithm's operationalization. We observed no difference in predictive performance when we used the CFS classification tree instead of Fried's frailty criteria in the algorithm's operationalization. Scores derived from the CFS classification tree show good agreement compared with those obtained using the original CFS, which requires clinical judgement [12]. Second, our findings highlight the importance of primary prevention in the low-risk group. As the WFG states 'low risk does not mean no risk'. Of participants that reported a fall in follow-up, 52.0% were classified as low-risk by the WFG algorithm with 3KQ, and 65.8% were classified as low-risk by the WFG algorithm with the single fall history question. Similarly, 55.3% of those that reported a fall in follow-up were classified as low-risk by the AGS/BGS algorithm. However, contrary to the WFG algorithm, the AGS/BGS algorithm does not prescribe any primary prevention strategies for patients classified as low-risk. Therefore, we recommend the use of the WFG algorithm.

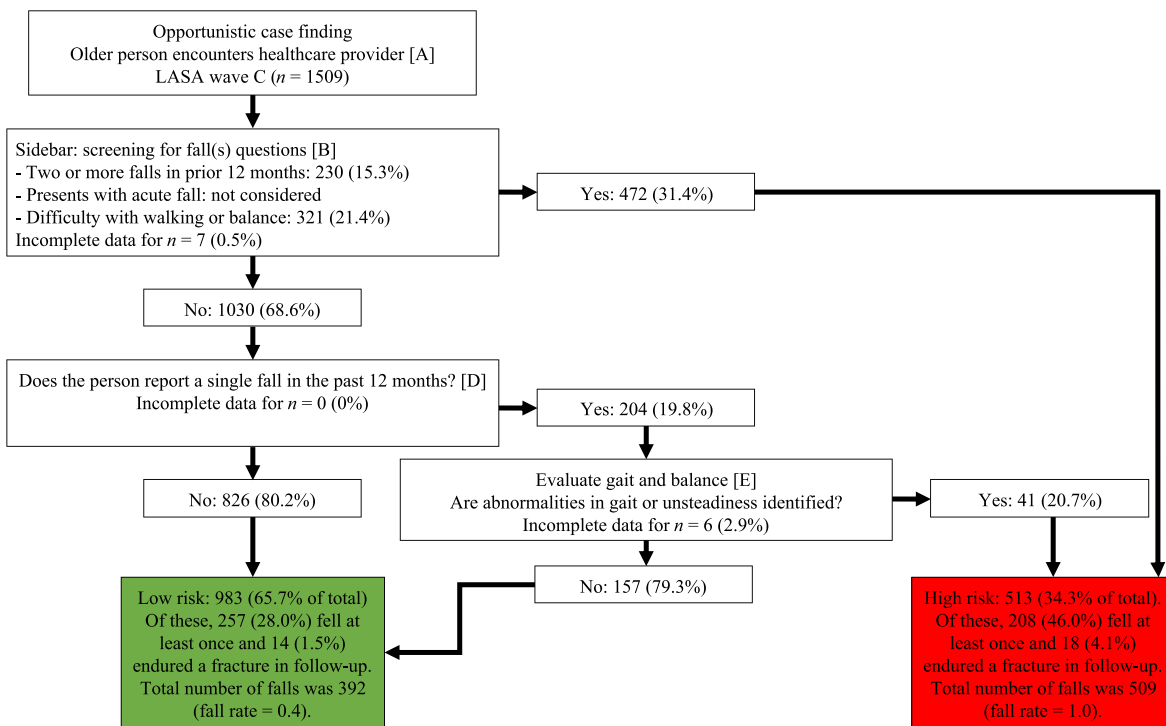


Figure 2. **Operationalization of the American Geriatrics Society and British Geriatrics Society algorithm.** LASA = Longitudinal Aging Study Amsterdam. Unless stated otherwise, proportions are calculated based on the number of participants with complete data in the preceding step. Proportions are provided as valid percentages, meaning missing data were not included in their calculations. A total of 13 participants (0.9%) could not be classified due to incomplete data.

This study has limitations. First, while retrospective studies such as ours help understand to what extent the WFG algorithm can predict falls, they provide limited insight into the algorithm's actual impact in clinical practice. As such, prospective studies across different settings are needed to assess how adopting the WFG affects health outcomes. Second, not all items in the screening tools were available in our sample and we therefore used proxies where possible. We used data from the tandem test as a proxy for the item in 3KQ on balance confidence. Consequently, we could have overestimated the number of participants classified as intermediate- or high-risk since older adults may be more likely to overestimate rather than underestimate their balance confidence [16]. Nonetheless, studies have demonstrated the tandem test to be associated with balance confidence, [17, 18] which indicates the proportion of misclassifications to be limited. Furthermore, we were unable to operationalize two of the items in the WFG algorithm (i.e. *Lying on the floor/unable to get up* and *Loss of consciousness/suspected syncope*) since we did not have data or appropriate proxies for these. In a follow-up study among a subsample of our study population (data not shown in paper), 5.5% of those that endured a fall reported lying on the floor for 15 min or longer after a fall [19]. Syncopal falls are common, accounting for 9–11% of patients that present to the emergency department due to a fall [20, 21]. Taken together, these results suggest that omitting the two items in our operationalization may have caused us to underestimate the size of high-risk group. Finally, the present study drew on data

measurements conducted between 1995 and 1997. Studies show that fall-related injuries and deaths have been increasing across Europe since 1990 [22]. Therefore, the fall rate in our sample may be lower than in the current population older adults, which could affect the generalizability of our findings.

Conclusions

Based on data of a population-based cohort study among older adults, we found that the WFG algorithm can help identify individuals with an elevated risk of falls. Operationalizing the algorithm with the 3KQ instead of the single fall history question helps identify more at-risk patients, at the cost of more overtreatment. Our findings suggest that the CFS and Fried's frailty criteria can be used interchangeably in the algorithm's operationalization. The algorithm's predictive performance was similar to that of the other assessed screening tools, except for the 3KQ on their own which have a higher sensitivity but lower specificity. However, the 3KQ can only classify patients according to two risk categories and, by itself, provides no directions with regard to further risk assessment and fall prevention strategies.

Supplementary Data: Supplementary data is available at *Age and Ageing* online.

Declaration of Conflicts of Interest: Nathalie van der Velde was a coauthor of the WFG.

Declaration of Sources of Funding: This work was supported by funding from The Netherlands Organization for Health Research and Development (ZonMw, Grant 848017004), The Hague. The sponsor played no part in the design, execution, analysis and interpretation of data, or writing of the study.

The Longitudinal Aging Study Amsterdam is supported by a grant from the Netherlands Ministry of Health, Welfare and Sport, Directorate of Long-Term Care.

Data Availability: The data underlying this study are available from the Longitudinal Aging Study Amsterdam (LASA). Data of LASA may be requested for research purposes. More information on data requests can be found on the LASA website: www.lasa-vu.nl.

References

1. Ganz DA, Latham NK. Prevention of falls in community-dwelling older adults. *N Engl J Med* 2020;**382**:734–43.
2. Tinetti ME. Preventing falls in elderly persons. *N Engl J Med* 2003;**348**:42–9.
3. Montero-Odasso M, van der Velde N, Martin FC *et al*. World guidelines for falls prevention and management for older adults: A global initiative. *Age Ageing* 2022;**51**:afac205.
4. Panel on prevention of falls in older persons, American Geriatrics Society and British geriatrics society. Summary of the updated American Geriatrics Society/British geriatrics society clinical practice guideline for prevention of falls in older persons. *J Am Geriatr Soc* 2011;**59**:148–57.
5. Hartley P, Forsyth F, Rowbotham S *et al*. The use of the world guidelines for falls prevention and Management's risk stratification algorithm in predicting falls in the Irish longitudinal study on ageing (TILDA). *Age Ageing* 2023;**52**:afad129.
6. Burns ER, Lee R, Hodge SE. *et al*. Validation and comparison of fall screening tools for predicting future falls among older adults. *Arch Gerontol Geriatr.* 2022;**101**:104713.
7. von EE, Altman DG, Egger M. *et al*. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *PLoS Med.* 2007;**4**:e296.
8. Hoogendijk EO, Deeg DJH, de Breij S *et al*. The longitudinal aging study Amsterdam: Cohort update 2019 and additional data collections. *Eur J Epidemiol* 2020;**35**:61–74.
9. Lamb SE, Jørstad-Stein EC, Hauer K *et al*. 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.
10. Guralnik JM, Simonsick EM, Ferrucci L *et al*. A short physical performance battery assessing lower extremity function: Association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol* 1994;**49**:M85–94.
11. Tromp AM, Pluijm SM, Smit JH. *et al*. Fall-risk screening test: A prospective study on predictors for falls in community-dwelling elderly. *J Clin Epidemiol.* 2001;**54**:837–44.
12. Theou O, Pérez-Zepeda MU, van der Valk AM *et al*. A classification tree to assist with routine scoring of the clinical frailty scale. *Age Ageing* 2021;**50**:1406–11.
13. Schaap LA, van Schoor NM, Lips P *et al*. Associations of sarcopenia definitions, and their components, with the incidence of recurrent falling and fractures: The longitudinal aging study Amsterdam. *J Gerontol Ser A* 2018;**73**:1199–204.
14. Pluijm SMF, Smit JH, Tromp EAM *et al*. A risk profile for identifying community-dwelling elderly with a high risk of recurrent falling: Results of a 3-year prospective study. *Osteoporos Int J Establ Result Coop Eur Found Osteoporos Natl Osteoporos Found USA* 2006;**17**:417–25.
15. Tromp AM, Smit JH, Deeg DJ *et al*. Predictors for falls and fractures in the longitudinal aging study Amsterdam. *J Bone Miner Res Off J Am Soc Bone Miner Res* 1998;**13**:1932–9.
16. Ickert EC, Hughes T, Berg-Carramusa CA *et al*. Overestimation of balance ability among older adults at risk for falls. *J Aging Health* 2023; 8982643231186630. <https://doi.org/10.1177/08982643231186630>.
17. Hile ES, Brach JS, Perera S *et al*. Interpreting the need for initial support to perform tandem stance tests of balance. *Phys Ther* 2012;**92**:1316–28.
18. Cho B, Scarpace D, Alexander NB. Tests of stepping as indicators of mobility, balance, and fall risk in balance-impaired older adults. *J Am Geriatr Soc* 2004;**52**:1168–73.
19. Stel VS, Smit JH, Pluijm SMF *et al*. Consequences of falling in older men and women and risk factors for health service use and functional decline. *Age Ageing* 2004;**33**:58–65.
20. Bhangu J, Hall P, Devaney N *et al*. The prevalence of unexplained falls and syncope in older adults presenting to an Irish urban emergency department. *Eur J Emerg Med* 2019;**26**:100.
21. Blomaard LC, Mooijaart SP, van Meer LJ *et al*. Geriatric screening, fall characteristics and 3- and 12 months adverse outcomes in older patients visiting the emergency department with a fall. *Scand J Trauma Resusc Emerg Med* 2021;**29**:43.
22. Hartholt KA, van der Velde N, Looman CWN *et al*. Trends in fall-related hospital admissions in older persons in the Netherlands. *Arch Intern Med* 2010;**170**:905–11.

Received 22 February 2024; editorial decision 20 August 2024