External Validation and Further Exploration of Fall Prediction Models based on Questionnaires and Daily-life Trunk Accelerometry

Yuge Zhang (Msc)[1], Roel H.A. Weijer (PhD)[1,2], Kimberley S van Schooten (PhD)[3,4,] Sjoerd M. Bruijn (PhD)[1,5,6*], and Mirjam Pijnappels (PhD)[1]

[1] Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2] Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

[3] Neuroscience Research Australia, University of New South Wales, Sydney, Australia

[4] School of Population Health, University of New South Wales, Sydney, Australia

[5] Institute of Brain and Behavior Amsterdam, Amsterdam, The Netherlands

[6] Biomechanics Laboratory, Fujian Medical University, Quanzhou, Fujian, PR China

Author Note

Correspondence concerning this article should be addressed to Sjoerd M. Bruijn, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands.

Email: s.m.bruijn@gmail.com

Main text word count: 3867

Number of data elements: 5

Abstract

Background: Ambulatory measurements of trunk accelerations can provide valuable insight into the amount and quality of daily life activities. Such information has been used to create models that aim to identify individuals at high risk of falls. However, external validation of such prediction models is lacking, yet crucial for clinical implementation. We externally validated three previously described fall prediction models (van Schooten et al., 2015a).

Methods: Complete questionnaires and one week of trunk acceleration data were obtained in 263 community-dwelling people (mean age 71.8 years, 68.1% female). To validate models, we first used the coefficients and optimal cut-offs from original cohort, then recalibrated the original models, as well as optimized parameters based on our new cohort.

Results: Among all participants, 39.9% experienced falls during 6-month follow-up. All models showed poor precision (0.20-0.56), poor sensitivity (0.22-0.58), and good specificity (0.59-0.89). Calibration of the original models had limited effect on model performance. Using coefficients and cut-offs optimized on the external cohort also had limited benefit. Last, the odds ratios in our cohort were different from those in the original cohort and indicated that gait characteristics, except for index of harmonicity ML, were not statistically significantly associated with falls.

Conclusions: Prediction of fall risk in our cohort was not as effective as in the original cohort. Recalibration as well as optimized model parameters resulted in limited increase in accuracy. Fall prediction models are highly specific to the cohort studied. This highlights the need for large representative cohorts, preferably with an external validation cohort.

Keywords: Wearable devices, Aged, Locomotion, Accidental falls, Verification

1. Introduction

Falls among older adults are a major cause of disability and death. Each year, one in three older adults aged 65 years or over experience at least one fall (1, 2). Therefore, early identification of fall risk is necessary for timely intervention. Fall risk can be estimated by questionnaires, functional tests in a clinical setting or in a biomechanical laboratory, but all these methods have limited predictive value (3). Specifically, questionnaires have low reliability due to the subjective bias of the participants or testers (4-6). Clinical gait tests, such as Berg Balance scale, Timed Up and Go and POMA, cannot quantify fall risk well, especially in frail older populations (7-9). Gait tests in a biomechanics laboratory do not reflect exposure to environmental and task-related hazards(10, 11). Wearable sensors allow for ambulatory monitoring of the quantity and quality of daily-life activities. Several studies have shown that daily-life gait quality characteristics, assessed with a single sensor, are able to discriminate fallers from non-fallers. Examples of such characteristics are local dynamic stability, variability of gait, frequency domain parameters, and gait smoothness (12-15). However, only a few research teams focus on developing fall risk prediction models based on daily-life activities in community environment (13-16). Among these, three predictive models from van Schooten and colleagues (17) had moderate to good predictive ability for falls happening in the next 6 months, and demonstrated that daily life accelerometry contributed substantially to the identification of individuals at risk of falls (areas under curve (AUC) of 0.68, 0.71 and 0.82 respectively). These models were logistic regression models based on a relatively large (n=169) data set of the FARAO cohort, a project concerning fall-risk assessment in older adults, with independent variables obtained either from questionnaires, from daily-life accelerometry, or from a combination of both.

However, Shany and colleagues (18) pointed out that current studies of sensor-based fall risk estimation were overly optimistic due to small sample sizes, flawed modelling and validation methods, (e.g. inherent problems with overly cross-validation techniques, lack of external validation). In addition, Nouredanesh and colleagues (19) have shown that there are many inconsistencies in studies of free-living fall prediction, such as the terminology (e.g., definition of gait quality and fall/fallers), hyperparameters (locomotion windows, description index of gait quality, duration of daily-life data and sampling frequency), and data processing approaches (e.g. the cut-off thresholds to define ambulatory walking bouts). More importantly, they also indicated that none of these prediction models have been externally validated, so the repeatability and generalizability cannot be guaranteed, which hinders the establishment and implementation of a unified fall prediction model for community-dwelling older adults.

In this study, we aimed to assess the external validity of the FARAO fall prediction models (17) by using an external cohort of community-dwelling older adults. Moreover, as the ambulatory assessments specifically quantify the amount and quality of daily-life gait, we performed additional analyses that focused on walk-falls only. In addition, we calibrated the FARAO prediction models. Finally, we evaluated models based on the optimal model parameters of the new cohort and compared the odds ratio for individual variables between original and new cohorts.

2. Materials and Methods

2.1 Participants

The data in this study were available from the “Veilig in beweging blijven” (VIBE), which translates to “Safely remaining active” cohort. Participants were included if they had a Mini-Mental State Examination (MMSE) score ≥ 19 (20) and were able to walk 20 meters, if necessary with a walking aid, without dizziness or perceiving chest pain or pressure, and were able to understand the Dutch

language. The protocol was approved by the ethics committee of the faculty of Behavioural and Movement Sciences of the Vrije Universiteit Amsterdam (VCWE-2016-129). Among 441 persons who expressed interest, 283 were eligible to participate and agreed to a 12 months follow-up on falls. All participants provided signed informed consent before participation.

2.2 Questionnaire related to fall risk

Descriptive characteristics such as age, weight, height, the use of a walking aid and 6-month fall history were obtained from all participants. In addition, participants independently filled out validated questionnaires and tests on fall-risk factors, i.e. cognitive function (Mini Mental State Examination; MMSE), concern about falling (16-item fall efficacy scale; FES-I (21)), LASA fall-risk profile (Longitudinal Aging Study Amsterdam) (22) and Geriatric Depression Scale (GDS-15 (23)). The LASA fall-risk profile comprised questions concerning dizziness, independence in daily life, having pets, alcohol consumption, education duration and required the measurement of hand grip strength, which was quantified using a handgrip dynamometer (TKK 5401, Takei Scientific Instruments, Tokyo, Japan). The GDS-15 is an effective scale for depressive symptoms in older populations, where 5 negative items and 10 positive items indicate symptoms of depression (24). GDS forms missing ≤20% of items were imputed according to (25). Scores of 0-4 are considered normal, depending on age, education, and complaints; 5-8 indicate mild depression; 9-11 indicate moderate depression; and 12-15 indicate severe depression. Since the GDS-30 was used in the original model, we had to convert our GDS-15 scores to GDS-30 scores. We did so by using a regression equation, in which GDS-15 scores are converted to GDS-30 scores: GDS-30 scores = 1.95*GDS-15 scores + 1.57 (26).

2.3 Assessment and analyses of daily-life gait characteristics

Conforming to the FARAO cohort study (17), participants were instructed to wear a tri-axial

accelerometer (DynaPort MoveMonitor, McRoberts, The Hague, the Netherlands) at home for 7 consecutive days, except during aquatic activities such as showering. The sensor was worn on the trunk at the level of L5 using an elastic belt, whose three axes of orientation corresponded to vertical (VT), anterior-posterior (AP) and medial-lateral (ML), ranged from −6g to +6g, with sampling frequency set to 100 Hz. We omitted the first 6 hours of the measurements from analysis, to discard any possible artefacts caused by transportation from the research team to the participants' home. To have a valid measure of the amount of daily activity, the accelerometer had to be worn for more than twelve hours each day (27). At the same time, the minimum valid days for daily physical activities are 2 days except for lying for which it is 3 days (27), so the worn time of the sensors should not be < 2 days to get valid accelerometer assessment in our study. For identification of periods of non-wearing, locomotion, sitting, lying, shuffling, and standing, we used the manufacturer's validated algorithms (28). We calculated the total duration of lying, and the number of strides, averaged over valid days, to obtain reliable, and valid indicators of the amount of daily activity.

The data analyses were performed in Matlab R2019a (Mathworks, Natwick, MA), using the Gait Analysis toolbox [https://github.com/VU-HMS/Gait-Analysis.git]. Prior to estimation of gait characteristics, data from each locomotion episode was realigned with the vertical axis, which was estimated from the direction of the mean acceleration signal (29), and anteroposterior and mediolateral directions, which were estimated by optimizing the left–right symmetry (30). Locomotion episodes were then split into epochs of 10 seconds to avoid possible sample size-related bias in outcome measures (31). Median gait characteristics were calculated over epochs of each gait episode, identical to the data length requirements of characteristics determined, i.e., logarithmic divergence rate AP (32), range VT, index of harmonicity ML (33) and sample entropy

VT (34). The mean over all episodes in a week for these measures were calculated as outcome measure (16) .

2.4 Outcome: Falls and walk-falls

Although participants were monitored prospectively for falls over a 12-month period following the baseline assessment by means of a fall diary and monthly telephone contact, we only focused on the first 6 months of follow-up for consistency with the original models. However, to explore if extended follow-up periods resulted in better model performance, we also analyzed data obtained from the 12-month fall history and 12-month prospective fall recordings; these results are reported in the supplementary materials (see supplementary tables 1-2 and figures 1-2). A fall was defined as an unintentional change in position resulting in coming to rest at a lower level or on the ground (35). Participants were classified as being a faller based on the number of prospectively monitored falls (≥1) they experienced over the 6-month period. Participants who did not record falls in their fall diary for the complete 6 months but experienced at least one fall within the recorded follow-up period were also classified as fallers.

When a fall happened, the activities and circumstances were also recorded, including walking, climbing the stair, riding a bicycle, showering, etc. Because the models were based on gait parameters, we also investigated how all models performed for people that fell specifically during walking. Participants were classified as walk-fallers if they had one record of a fall during walking.

2.5 Statistical analysis

Statistical analysis was performed in Matlab R2021a (MathWorks, Natick, MA) and R (R core team, Vienne). Unpaired t-tests were used for comparisons between fallers and non-fallers, and between walking fallers and non-fallers, separately. All continuous parameters in models were transformed to z-scores prior to analyses to facilitate interpretations.

The FARAO prediction models were all originally based on stepwise forward logistic regression analyses. Model 1 included as independent variables 6-months history of falls and GDS-30, model 2 contained accelerometry-based daily life gait quantity and quality characteristics, and model 3 combined variables from both model 1 and model 2.

To validate predictive models on our VIBE cohort data, with either all falls or walking falls as outcome, we first calculated model performance for our VIBE cohort when we used the FARAO model coefficients, original cut-offs (0.279, 0.352 and 0.396) and the mean/SDs of the FARAO cohort to calculate the z-score.

Next, we checked if recalibration (36) of the original models would improve model performance. To this aim, the agreement between the predicted probability of the models and observed binary outcomes was evaluated using a logistic regression, using predicted probability based on external cohort as independent variable and faller status as dependent variable, see equation 1

$$Actual\ values = \frac{1}{1+e^{-(\beta_0 + \beta_1 * predicted\ probablity)}} \quad [1].$$

Where the slope (β_1) evaluates the spread of estimated risks and the intercept (β_0) is an assessment of calibration-in-the-large. Slope = 1 and intercept = 0 means perfect model calibration (36). If these recalibrated models perform well, it suggests a simple scaling of the model would be enough to use the model in other populations.

Then, models with coefficients and cut-offs optimal to our VIBE dataset were evaluated to see whether performance could be improved, considering potential differences in the sample measured. Finally, we performed unpaired t-tests on the means and SD of outcome measures of the VIBE data as compared to the published FARO data, to see if potential differences in populations could explain performance. Lastly, univariate logistic regression was performed to identify the association between model variables and prospective falls, and these resulting odds ratio were compared with

189 those from the FARAO cohort (17).

190 **3. Results**

191 **3.1 Descriptive Characteristics**

192 283 participants filled in questionnaires and wore an accelerometer for one week at baseline, but 5
193 participants' acceleration data could not be read. Among the remaining 278 participants, 15
194 participants were excluded due to missing more than 20% of the GDS items (n=3, 1 person missed
195 5 items and 2 persons missed 15 items), incomplete records within 6 months and no record of falls
196 (n=3), accelerometer assessment duration of less than 2 days (n=3), ≤ 50 episodes of overall
197 measurement time (n=6). Figure 1 presents a flowchart of the included fallers and non-fallers.
198 Descriptive characteristics of the participants are summarized in table 1. Of the 263 participants,
199 39.9% (n=105) experienced one or more falls during follow-up, 19% (n=50) had fallen during
200 walking. In addition, 37.6% of all participants, 43.8% of all fallers, 50% of walk-fallers and 33.5%
201 of non-fallers had a fall history in the past 6 months. The average number of prospective and past
202 falls, GDS scores, FES-I and LASA fall risk profile (score) were significantly different between
203 fallers and non-fallers, and between walk-fallers and non-fallers. Furthermore, walk-fallers had
204 significantly lower hand grip strength, and poorer independence in life, compared to non-fallers.
205 Figure 2(A) shows the main activities when a fall happened, including walking (39%), riding a
206 bicycle (13%), movement transfer (12%), going up or down the stairs (12%), doing sports (6%),
207 standing and sitting (in static positions) (4%), and other activities (15%). Among these, hiking was
208 classified as a sport in this study. Working in the garden, getting dressed, etc. were classified as
209 "Other". Furthermore, main causes of falls during walking were shown in figure 2(B), which were
210 trip/stumble (34%), balance loss (31%), slip (17%), physiological discomfort (8%), being distracted
211 (2%) and other (8%). Trip/stumble often happened on uneven roads or roads with obstacles, and

balance loss was more likely to occur when there was sudden disturbance, opposite walking height to the expected. Falling due to slips usually occurred in the bathroom or on a slippery floor. In addition, physiological discomfort included ankle sprain, sudden dizziness, drops in blood pressure, eye injury and weakened muscle strength, etc. Falling caused by distraction included chatting, looking at map, or doing other tasks when walking. Lastly, some unknown, poorly classified reasons were divided into “Other”.

3.2 Validation of predicting all falls and walk-falls based on existing models

By using optimal cut-offs (0.279, 0.352 and 0.396), coefficients and mean (SD) (for z-score) from the original cohort, we obtained model performance of the three prediction models based on only questionnaires, only accelerometry, and combined questionnaires and accelerometry. Table 2a shows the results of each prediction model, with either all falls or walk-falls as outcome (Model performance for a follow-up of 12 months based on FARAO cohort can be seen in supplementary table 2a). Model 1, with only questionnaires, performed best for predicting all falls and walk-falls, with AUC (area under the curve) of 0.62 and 0.68, respectively. Model 3 and model 2 were less effective than model 1, although model 3 was slightly better than model 2, with AUC 0.58 and 0.62 for model3, 0.52 and 0.51 for model 2. However, the precision and sensitivity in both groups were low, which meant the ability of predicting a fall was low for all models.

3.3 Model recalibration

Calibration intercepts of all models in both groups were negative, and the slope of model 2 was <1, contrary to model 1 and model 3, both of which had slopes >1 (see supplementary table 3). The performance of the calibrated models is shown in table 2b. The AUC of all calibrated models ranged from 0.51-0.68. The sensitivity of calibrated model 2 was best with 0.86, better then calibrated model 1 and model 3 in both groups.

3.4 Validation of predicting all falls and walk-falls based on the external validation dataset

The optimal cut-offs of the external validation VIBE cohort, which balanced the high sensitivity (true positive rate) and the low false positive rate (1-specificity), were 0.38, 0.40 and 0.37 in all models for predicting all falls, and 0.20, 0.23 and 0.20 for predicting walk-falls. Based on these optimal cut-offs, and the coefficients of this validation cohort, the model performances are shown in table 2c (see ROC curves in supplementary figure 3) (Model performance ROC curves for 12 months follow up based on the VIBE cohort can be seen in supplementary table 2b and figure 2, respectively). Specificity and sensitivity were still moderate for these models (ranging from 0.59-0.66). Model3 performed best in both groups (AUC=0.65 and 0.73, respectively), and then model 1 and model 2.

3.5 Comparison of cohorts and individual odds ratios

To better understand differences in model performance between the original models and models fitted on our external dataset, the comparison of descriptive statistics and odds ratio of measures for the our external VIBE cohort and original FARAO cohort (17) are presented in table 3. There were small but significant differences in age, and height, with the VIBE cohort being slightly younger and shorter. Prevalence of fall history prior to 6 months in VIBE cohort was 37.6%, which was slightly higher than 35.5% in the FARAO cohort. Cognitive function (MMSE score) of VIBE cohort was 28.2, slightly higher than 27.7 in FARAO cohort. LASA fall-risk profile score, percent of inability to use public transportation and use of walking aid in VIBE cohort were lower than those in FARAO cohort. In agreement with this, the participants of the VIBE cohort had a higher degree of living independence. Number of strides walked in our cohort was 31% higher than in the FARAO cohort, and duration of lying in our cohort was 0.9 hours less than FARAO cohort. All gait characteristics differed significantly between the VIBE and FARAO cohorts.

For both the VIBE and the FARO data, significant odds ratios for questionnaire metrics 6-months history of falls, GDS-30, FES-I, and LASA profile score were found. For the VIBE data, the only gait measure that was significantly associated with falls was the index of harmonicity ML. However, the confidence interval of the odds ratios for the other gait parameters overlapped with those in FARAO cohort, except for the logarithmic divergence rate AP.

4. Discussion

We validated the predictive ability of three FARAO fall prediction models based on questionnaires, accelerometry of daily life gait, and both, for our external data from the VIBE cohort of community-dwelling older adults. We found that the external validity of the models developed based on the FARAO cohort was limited for predicting all falls as well as when focusing on walking falls. Subsequently, we explored why this may have been so.

When comparing the population characteristics between the two cohorts, we found that participants in the VIBE cohort were slightly younger, more often female, a little shorter, had a higher level of cognition, had less fall risk, lived more independently, use public transport more often and used less walking aids. Participant's in both studies had a normal (i.e., low) degree of depressive symptoms (24). The prevalence of falls in our study was higher than in the original study (39.9% and 34.9%, respectively), with 14.5% participants having two or more falls in the follow-up 6 months compared to 16.5% in the FARAO cohort. In addition, both cohorts had the same definition of a faller in past and follow-up 6 months when the number of falls was ≥1, respectively. Studies suggested that only one fall may be an ambiguous fall, with no apparent cause (19), and people who were 'once only fallers' were characteristically more closely related to non-fallers than 'recurrent fallers'(37) or injury-prone fallers. For example, one study indicated that the annual incidence of falls for community residents over 70 years of age may be 49% (38), and another study found the

median annual fall prevalence rates was 33.4% for older people over 50 years (2).

Since participants in the VIBE cohort reported more falls than participants in the FARAO study by Van Schooten et al., we expected that there was a difference in the method of monitoring falls that led to a wider range of falls being reported in the VIBE cohort, including relatively more falls that are not related to walking activities. Yet in both studies a combination of fall calendars and monthly phone calls were used. Possibly there was a difference in questioning, for instance related to the interpretation of the fall definition, during the phone calls that may have resulted in more falls in the VIBE cohort when compared to the FARAO cohort. We divided the causes of the falls into 7 categories and walking accounted for 39% of all falls, similar to other studies (39-41). However, it was not known if falls in the original study had a similar distribution of causes. Hence the different proportion of falls during walking may have affected the results, but the models also did not work well for the walking fallers. In post-hoc analysis, results did not change when we analyzed the 12-month follow up data available from the VIBE cohort (refer to supplementary table 2). In conclusion, there may have been differences in the classification of fallers that influenced the accuracy of the models. However, since such differences are most likely due to interpretation of the fall definition by the people who performed the monthly phone calls, we are unable to identify such potential differences to test if these affected the results.

After using the model parameters from FARAO cohort to classify fallers and non-fallers, we got moderate accuracy, and limited precision and sensitivity for all models in two groups. For model calibration, negative values of the calibration intercept suggest overestimation. Calibration slopes < 1 suggests that estimated risks are too extreme, i.e., too high for patients who are at high risk and too low for patients who are at low risk. Thus, all three models overestimated and the estimated risk of model 2 and 3, which included gait characteristics, were too extreme.

When we explored the optimal coefficients and optimal cut-offs for the VIBE cohort, we found that model 3 performed best for both all falls and walk-falls. The AUC of all models improved a little, but the precision and sensitivity remained moderate. More importantly, the model which combined the gait quality characteristics and questionnaires was better than either one alone – reinforcing the conclusion of van Schooten et al.'s study that daily-life accelerometry has value for the identification of people at risk for falls. Thus, while the same variable may yield a reasonable fall prediction model, it appears that coefficients and cut-offs may be rather specific to a given population sample (42).

To further study the predictors in the models, odds ratios of individual variables were compared with the initial FARAO cohort. Studies have shown that fall history, depressive symptoms, LASA fall-risk profile, fear of falling, and poor executive function (as assessed by TMT-A) were associated with prospective falls (22, 43, 44). Significant odds ratios of these measures in the original study confirmed this (van Schooten et al. 2015a). However, of the gait quality characteristics, only index of harmonicity ML had significant odds ratios in VIBE cohort, while other gait quality measures have been shown to be related to future falls in FARAO cohort. Specifically, lower harmonic ratio, reflecting lower gait symmetry, was shown to be related to falls by (45); higher logarithmic rate of divergence, reflecting lower local stability, was associated with future falls, according to previous findings (14, 46, 47). This was most likely due to the differences in the VIBE cohort, who were younger, and more much active than participants in the FARAO cohort.

In conclusion, prediction of fall risk in our VIBE cohort based on the FARAO models was not as effective as models in FARAO cohort (17) suggesting that these models are overfitted and have limited external validity. Despite several earlier reports indicating the importance of daily life gait

measures for predicting falls, such measures alone were of limited value in our cohort. Moreover, the fact that optimal coefficients and cut-off were different for our cohort, as well as the fact that we found almost no significant ORs for the gait parameters included in the model, suggests that this kind of fall prediction models are highly specific to the cohort studied. Nonetheless, even with these optimal coefficients and cut-offs, model performance was still poor, which highlights the need for large representative cohorts, preferably with an external validation cohort.

Acknowledgments – Author contributions: Conceptualization, MP and SMB; methodology, RW, SMB, KS and MP; formal analysis, YZ, SMB, MP, RW and KS; resources, MP and SMB; data collection, RW; data processing, YZ; writing-review and editing, YZ, RW, SMB, KS and MP; visualization, YZ; supervision, SMB and MP; project administration, MP. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by China Scholarship Council (CSC) (202009110145) to YZ; Dutch Organization for Scientific Research (NWO) (no. 91714344, 016.Vidi.178.014) to MP, RW and SMB; and Human Frontier Science Program Fellowship (LT001080/2017) to KS.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Salvà A, Bolíbar I, Pera G, Arias C. Incidence and consequences of falls among elderly people living in the community. Medicina Clínica. 2004;122(5):172-176. doi:10.1016/S0025-7753(04)74184-6.

2. Morrison A, Fan T, Sen SS, Weisenfluh L. Epidemiology of falls and osteoporotic fractures: a systematic review. Clinicoecon Outcomes Res. 2013;5:9-18. doi:10.2147/CEOR.S38721.

3. Hamacher D, Singh NB, Van Dieën JH, Heller MO, Taylor WR. Kinematic measures for assessing gait stability in elderly individuals: a systematic review. Journal of The Royal Society

Interface. 2011;8(65):1682-1698. doi:10.1098/rsif.2011.0416.

4. Furnham A. Response bias, social desirability and dissimulation. Personality and individual differences. 1986;7(3):385-400. doi:10.1016/0191-8869(86)90014-0.

5. Malatesta D, Simar D, Dauvilliers Y, Candau R, Borrani F, Préfaut C*, et al.* Energy cost of walking and gait instability in healthy 65-and 80-yr-olds. Journal of applied physiology. 2003;95(6):2248-2256. doi: 10.1152/japplphysiol.01106.2002.

6. Winstein N. Unrealistic optimism about susceptibility to health problems: Conclusions from a community-wide simple. J Behav Med. 1987;10:481-500. doi:10.1007/BF00846146.

7. Boulgarides LK, McGinty SM, Willett JA, Barnes CW. Use of clinical and impairment-based tests to predict falls by community-dwelling older adults. Physical Therapy. 2003;83(4):328-339. doi:10.1093/ptj/83.4.328.

8. Mancini M, Horak FB. The relevance of clinical balance assessment tools to differentiate balance deficits. Eur J Phys Rehabil Med. 2010;46(2):239-248.

9. Barry E, Galvin R, Keogh C, Horgan F, Fahey T. Is the Timed Up and Go test a useful predictor of risk of falls in community dwelling older adults: a systematic review and meta- analysis. BMC Geriatrics. 2014;14(1):14. doi:10.1186/1471-2318-14-14.

10. Maki BE. Gait changes in older adults: predictors of falls or indicators of fear? Journal of the American geriatrics society. 1997;45(3):313-320.

11. Hausdorff JM, Rios DA, Edelberg HK. Gait variability and fall risk in community-living older adults: A 1-year prospective study. Archives of Physical Medicine and Rehabilitation. 2001;82(8):1050-1056. doi:10.1053/apmr.2001.24893.

12. Bourke AK, Van De Ven P, Gamble M, O'Connor R, Murphy K, Bogan E*, et al.* Assessment of waist-worn tri-axial accelerometer based fall-detection algorithms using continuous unsupervised

activities. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, 2010, pp. 2782-2785, doi:10.1109/IEMBS.2010.5626364.

13. Brodie MA, Coppens MJ, Ejupi A, Gschwind YJ, Annegarn J, Schoene D*, et al.* Comparison between clinical gait and daily-life gait assessments of fall risk in older people. Geriatrics & Gerontology International. 2017;17(11):2274-2282. doi:10.1111/ggi.12979.

14. Rispens SM, van Schooten KS, Pijnappels M, Daffertshofer A, Beek PJ, van Dieën JH. Identification of fall risk predictors in daily life measurements:Gait characteristics' reliability and association with self-reported fall history. Neurorehabilitation and Neural Repair. 2015;29(1):54-61. doi:10.1177/1545968314532031.

15. Weiss A, Brozgol M, Dorfman M, Herman T, Shema S, Giladi N*, et al.* Does the evaluation of gait quality during daily life provide insight into fall risk? A novel approach using 3-day accelerometer recordings. Neurorehabil Neural Repair. 2013;27(8):742-752. doi:10.1177/1545968313491004.

16. van Schooten KS, Pijnappels M, Rispens SM, Elders PJM, Lips P, Daffertshofer A*, et al.* Daily-ilfe gait quality as predictor of falls in older people: a 1-year prospective cohort Study. PLOS ONE. 2016;11(7):e0158623. doi:10.1371/journal.pone.0158623.

17. van Schooten KS, Pijnappels M, Rispens SM, Elders PJM, Lips P, van Dieën JH. Ambulatory fall-risk assessment: Amount and quality of daily-life gait predict falls in older adults. The Journals of Gerontology: Series A. 2015;70(5):608-615. doi:10.1093/gerona/glu225.

18. Shany T, Wang K, Liu Y, Lovell NH, Redmond SJ. Review: Are we stumbling in our quest to find the best predictor? Over-optimism in sensor-based models for predicting falls in older adults. Healthcare Technology Letters. 2015;2(4):79-88. doi:10.1049/htl.2015.0019.

19. Nouredanesh M, Godfrey A, Howcroft J, Lemaire ED, Tung J. Fall risk assessment in the wild:

397 A critical examination of wearable sensor use in free-living conditions. Gait & Posture.
398 2021;85:178-190. doi:10.1016/j.gaitpost.2020.04.010.
399 20. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the
400 cognitive state of patients for the clinician. Journal of psychiatric research. 1975;12(3):189-198.
401 doi:10.1016/0022-3956(75)90026-6.
402 21. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial
403 validation of the Falls Efficacy Scale-International (FES-I). Age and Ageing. 2005;34(6):614-619.
404 doi:10.1093/ageing/afi196.
405 22. Tromp AM, Pluijm SMF, Smit JH, Deeg DJH, Bouter LM, Lips P. Fall-risk screening test: A
406 prospective study on predictors for falls in community-dwelling elderly. Journal of Clinical
407 Epidemiology. 2001;54(8):837-844. doi:10.1016/S0895-4356(01)00349-3.
408 23. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M*, et al.* Development and validation
409 of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research.
410 1982;17(1):37-49. doi:10.1016/0022-3956(82)90033-4.
411 24. Greenberg SA. The geriatric depression scale (GDS). Best Practices in Nursing Care to Older
412 Adults. 2012;4(1):1-2.
413 25. Imai H, Furukawa TA, Kasahara Y, Ishimoto Y, Kimura Y, Fukutomi E*, et al.* Ipsative
414 imputation for a 15-item Geriatric Depression Scale in community-dwelling elderly people.
415 Psychogeriatrics. 2014;14(3):182-187. doi:10.1111/psyg.12060.
416 26. Zhang Y, Hoozemans M, Pijnappels M, Bruijn SM. A formula for calculating 30-item Geriatric
417 Depression Scale (GDS-30) scores from the 15-item version (GDS-15). bioRxiv.
418 2022:2022.2009.2022.509012. doi:10.1101/2022.09.22.509012.
419 27. van Schooten KS, Rispens SM, Elders PJM, Lips P, van Dieën JH, Pijnappels M. Assessing

physical activity in older adults: Required days of trunk accelerometer measurements for reliable estimation. Journal of Aging and Physical Activity. 2015;23(1):9-17. doi:10.1123/japa.2013-0103.

28. Dijkstra B, Kamsma Y, Zijlstra W. Detection of gait and postures using a miniaturised triaxial accelerometer-based system: accuracy in community-dwelling older adults. Age and ageing. 2010;39(2):259-262. doi:10.1016/j.apmr.2010.05.004.

29. Moe-Nilssen R. A new method for evaluating motor control in gait under real-life environmental conditions. Part 1: The instrument. Clin Biomech (Bristol, Avon). 1998;13(4-5):320-327. doi:10.1016/s0268-0033(98)00089-8.

30. Rispens SM, Pijnappels M, van Schooten KS, Beek PJ, Daffertshofer A, van Dieën JH. Consistency of gait characteristics as determined from acceleration data collected at different trunk locations. Gait & Posture. 2014;40(1):187-192. doi:10.1016/j.gaitpost.2014.03.182.

31. Bruijn SM, van Dieën JH, Meijer OG, Beek PJ. Statistical precision and sensitivity of measures of dynamic gait stability. Journal of Neuroscience Methods. 2009;178(2):327-333. doi:10.1016/j.jneumeth.2008.12.015.

32. Wolf A, Swift JB, Swinney HL, Vastano JA. Determining Lyapunov exponents from a time series. Physica D: Nonlinear Phenomena. 1985;16(3):285-317. doi:10.1016/0167-2789(85)90011-9.

33. Lamoth CJC, Beek PJ, Meijer OG. Pelvis–thorax coordination in the transverse plane during gait. Gait & Posture. 2002;16(2):101-114. doi:10.1016/S0966-6362(01)00146-1.

34. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology. 2000;278(6):H2039-H2049. doi:10.1152/ajpheart.2000.278.6.H2039.

35. Gibson MJ. The Prevention of Falls in Later Life : A Report of the Kellogg International Work

Group on the Prevention of Falls by the Elderly. Dan Med Bull. 1987;34(Suppl 4):1-24.

36. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC medicine. 2019;17(1):1-7. doi:10.1186/s12916-019-1466-7.

37. Lord SR, Ward JA, Williams P, Anstey KJ. Physiological factors associated with falls in older community-dwelling women. Journal of the American Geriatrics Society. 1994;42(10):1110-1117. doi:10.1111/j.1532-5415.1994.tb06218.x.

38. Hill K, Schwarz J, Flicker L, Carroll S. Falls among healthy, community-dwelling, older women: a prospective study of frequency, circumstances, consequences and prediction accuracy. Australian and New Zealand Journal of Public Health. 1999;23(1):41-48. doi:10.1111/j.1467-842X.1999.tb01203.x.

39. Peel NM. Epidemiology of falls in older age. Canadian Journal on Aging/La Revue canadienne du vieillissement. 2011;30(1):7-19. doi:10.1017/S071498081000070X.

40. Talbot LA, Musiol RJ, Witham EK, Metter EJ. Falls in young, middle-aged and older community dwelling adults: perceived cause, environmental factors and injury. BMC public health. 2005;5(1):1-9. doi:10.1186/1471-2458-5-86.

41. Niino N, Tsuzuku S, Ando F, Shimokata H. Frequencies and circumstances of falls in the National Institute for Longevity Sciences, Longitudinal Study of Aging (NILS-LSA). Journal of Epidemiology. 2000;10(1sup):90-94. doi:10.2188/jea.10.1sup_90.

42. Punt M, Bruijn SM, van Schooten KS, Pijnappels M, van de Port IG, Wittink H*, et al.* Characteristics of daily life gait in fall and non fall-prone stroke survivors and controls. Journal of neuroengineering and rehabilitation. 2016;13(1):1-7. doi:10.1186/s12984-016-0176-z.

43. Holtzer R, Friedman R, Lipton RB, Katz M, Xue X, Verghese J. The relationship between

specific cognitive functions and falls in aging. Neuropsychology. 2007;21(5):540. doi:10.1037/0894-4105.21.5.540.

44. Pluijm SMF, Smit JH, Tromp EAM, Stel VS, Deeg DJH, Bouter LM*, et al.* A risk profile for identifying community-dwelling elderly with a high risk of recurrent falling: results of a 3-year prospective study. Osteoporosis International. 2006;17(3):417-425. doi:10.1007/s00198-005-0002-0.

45. Doi T, Hirata S, Ono R, Tsutsumimoto K, Misu S, Ando H. The harmonic ratio of trunk acceleration predicts falling among older people: results of a 1-year prospective study. Journal of NeuroEngineering and Rehabilitation. 2013;10(1):7. doi:10.1186/1743-0003-10-7.

46. Lockhart TE, Liu J. Differentiating fall-prone and healthy adults using local dynamic stability. Ergonomics. 2008;51(12):1860-1872. doi:10.1080/00140130802567079.

47. Toebes MJP, Hoozemans MJM, Furrer R, Dekker J, van Dieën JH. Local dynamic stability and variability of gait are associated with fall history in elderly subjects. Gait & Posture. 2012;36(3):527-531. doi:10.1016/j.gaitpost.2012.05.016.

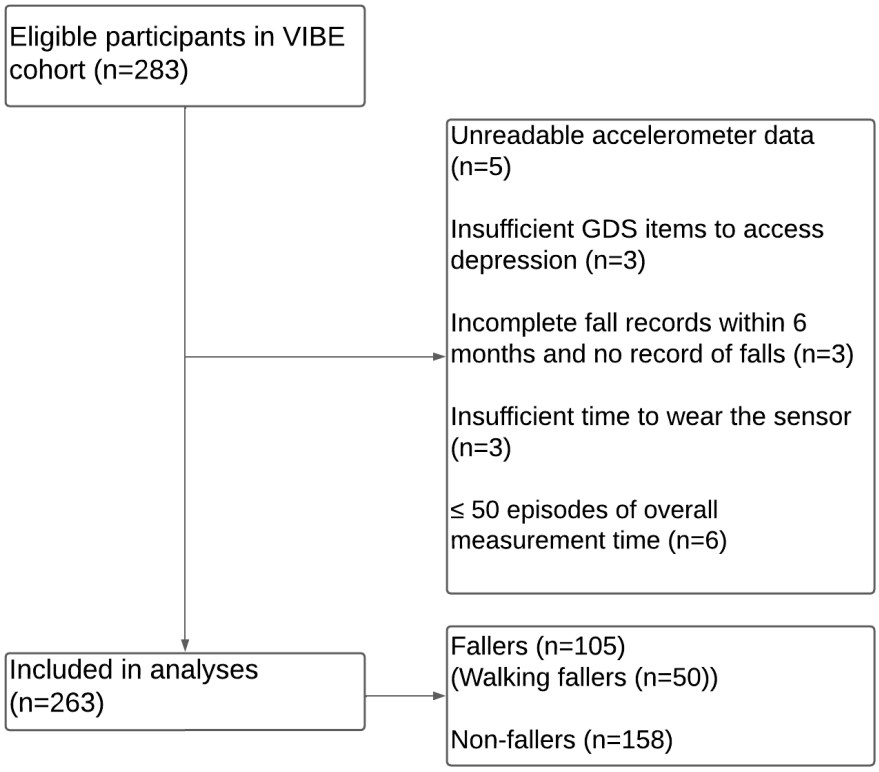

Figure 1. Flowchart of participants included for analyses. Fall records were considered incomplete when one or more months of fall records were missing.

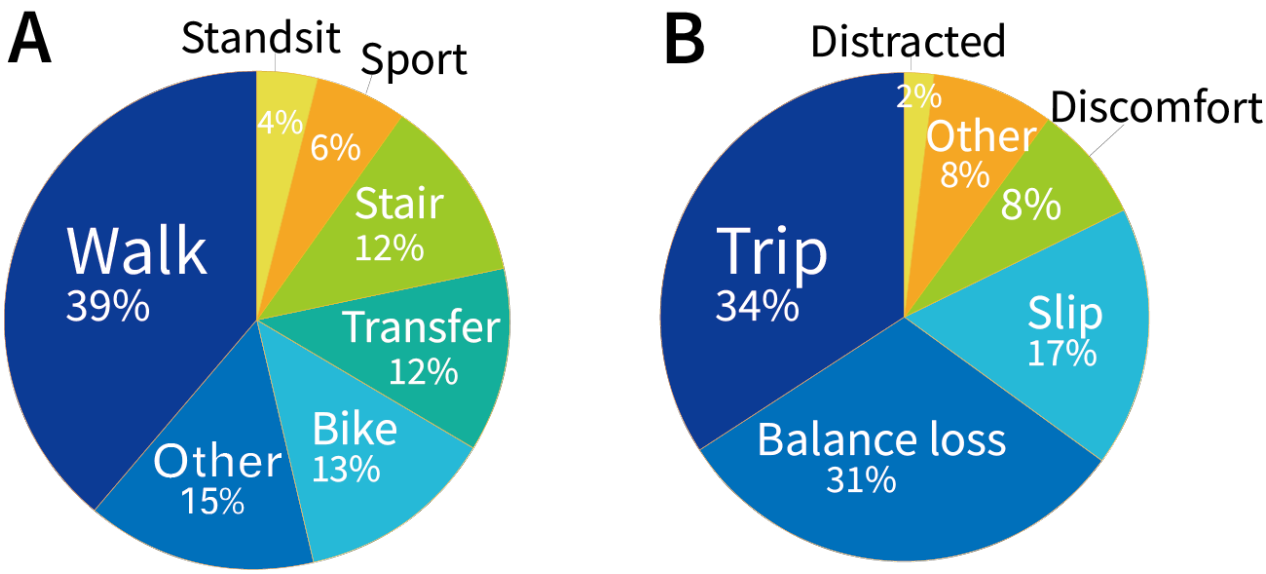

Figure 2. Main activities of all falls (A) and causes of walk-falls (B)

Table 1. Descriptive characteristics (Mean (SD)) of the VIBE participants

Characteristics	All participants	All fallers	Walk-fallers	Non-fallers
Number	263	105	50	158
Age (years)	71.8 (5.7)	71.9 (5.4)	71.9 (6.2)	71.8 (5.9)
Gender (%female)	68.1 (179)	68.6 (72)	72 (36)	67.7 (107)
Weight (kg)	74.1 (13.5)	72.6 (12.9)	70.6 (12.4)	75.1 (13.8)
Height (cm)	169.1 (8.5)	168.6 (9.1)	166.4 (8.2)	169.5 (8.1)
Number of falls in follow-up 6 months	0.6 (1.02)	1.6 (1.04) [A]	2.0 (1.3) [B]	0 [A, B]
At least one fall in follow-up 6 months (%/n)	39.9 (105)	100 (105)	100 (50)	0
Two or more falls in follow-up 6 months (%/n)	14.5 (38)	36.2 (38)	56 (28)	0
Number of falls in past 6 months	0.7 (1.4)	1 (1.9) [A]	1.4 (2.5) [B]	0.5 (0.8) [A, B]
At least one fall in past 6 months (%/n)	37.6 (99)	43.8 (46)	50 (25)	33.5 (53)
Two or more falls in past 6 months (%/n)	13.7 (36)	20 (21)	30 (15)	9.5 (15)
Hand grip strength (kg)	58.8 (16.6)	57.5 (17.5)	53.6 (17.5) [b]	59.7 (15.9) [b]
Geriatric depression score 30, GDS (score)	4.4 (4.3)	5.5 (5.3) [A]	6.5 (6.3) [B]	3.7 (3.3) [A, B]
Fall efficacy scale, FES-I(score)	20.5 (5.1)	21.4 (6.5) [a]	21.5 (6) [b]	19.9 (3.9) [a, b]
Cognitive function (MMSE score)	28.2 (1.6)	28.3 (1.6)	28.5 (1.7)	28.2 (1.6)
LASA fall risk profile (score)	3.4 (3.3)	4.4 (3.8) [A]	5.4 (4.5) [B]	2.8 (2.8) [A, B]
Inability of using transport(%/n)[1]	0.4 (1)	1 (1)	0 (0)	0 (0)
Living independently (%/n)	99.2 (261)	98.1 (103)	96 (48) [b]	100 (158) [b]
Use of a walking aid (%/n)	6.8 (18)	8.6 (9)	10 (5)	5.7 (9)

Note: [1]Either public or own transport. [A] significant difference between all fallers and non-fallers with $p<0.01$; [a] significant difference between all fallers and non-fallers with $p<0.05$; [B] significant difference between walking fallers and non-fallers with $p<0.01$; [b] significant difference between walking fallers and non-fallers with $p<0.05$.

Table 2. Model performance of fallers in follow-up 6 months for a) models based on the FARAO cohort, b) the calibrated FARAO model performance, c) models based on VIBE cohort

a) model based on FARAO cohort	All falls			Walk-falls		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Accuracy	0.58	0.50	0.49	0.63	0.53	0.75
Precision	0.48	0.37	0.40	0.34	0.20	0.49
Sensitivity	0.50	0.36	0.55	0.58	0.32	0.34
Specificity	0.64	0.59	0.45	0.64	0.59	0.89
AUC	0.62	0.50	0.53	0.68	0.49	0.59
b) calibrated FARAO model performance	**All falls**			**Walk-falls**		
	Calibrated Model1	**Calibrated Model2**	**Calibrated Model3**	**Calibrated Model1**	**Calibrated Model2**	**Calibrated Model3**
Optimal cut-offs	0.36	0.39	0.45	0.23	0.24	0.30
Accuracy	0.59	0.48	0.63	0.71	0.57	0.75
Precision	0.49	0.42	0.59	0.42	0.27	0.47
Sensitivity	0.53	0.86	0.28	0.50	0.48	0.36
Specificity	0.63	0.22	0.87	0.78	0.59	0.87
AUC	0.62	0.52	0.58	0.68	0.51	0.62
c) model based on VIBE cohort	**All falls**			**Walk-falls**		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Optimal cut-offs	0.38	0.40	0.37	0.20	0.23	0.20
Accuracy	0.62	0.53	0.57	0.66	0.54	0.6
Precision	0.53	0.43	0.47	0.38	0.27	0.33
Sensitivity	0.56	0.53	0.58	0.66	0.54	0.62
Specificity	0.66	0.53	0.56	0.66	0.54	0.59
AUC	0.63	0.53	0.65	0.69	0.56	0.73

Table 3. Mean (SD) and odds ratios (OR) for VIBE cohort (n=263) and FARAO cohort (n=169) (17)

Measures	Prevalence or mean (SD) for VIBE cohort	Prospective Falls OR (95% CI)	Prevalence or mean (SD) for FARAO cohort	Prospective Falls OR (95% CI)
Demographic information				
Age (years)	71.8 (5.7) ***	1.02 (0.8-1.3)	75.4 (6.8)	1.09 (0.8-1.5)
Gender (% female)	68.1 **	1.04 (0.6-1.8)	52.1	1.08 (0.6-2.06)
Weight (kg)	74.1 (13.5)	0.8 (0.6-1.1)	74.1 (14.2)	1.2 (0.9-1.7)
Height (cm)	169.1 (8.5) *	0.9 (0.7-1.1)	170.8 (8.9)	1.2 (0.9-1.7)
Questionnaires				
6-months history of falls	37.6%	1.6 (1.2-2.2) **	35.5%	3.1 (1.6–6.1) **
Hand grip strength (kg)	58.8 (16.6)	0.9 (0.7-1.1)	58.4 (20.5)	0.97 (0.7–1.4)
Geriatric depression score, GDS-30	4.4 (4.3)	1.5 (1.2-2.0) **	4.6 (4.2)	1.6 (1.1–2.2) **
Fall efficacy scale, FES-I (score)	20.5 (5.1)	1.3 (1.1-1.7) *	20.8 (4.4)	1.1 (1.0–1.1) *
Cognitive function (MMSE score)	28.2 (1.6) **	1.1 (0.9-1.4)	27.7 (2.2)	0.9 (0.7-1.3)
LASA fall-risk profile (score)	3.4 (3.3) ***	1.7 (1.3-2.2) **	5.5 (4.7)	1.6 (1.1–2.2) **
Inability to use public transportation	0.4% ***	N/A	7.1%	2.8 (0.8–9.5)
Living independently	99.2% ***	N/A	88.1%	1.0 (0.5-2.1)
Use of walking aid	6.8% ***	1.6 (0.6-4.1)	17.2%	1.4 (0.6–3.2)
Daily activities (per day)				
Number of strides (N)	4023 (2053) ***	1.1 (0.8-1.4)	3060 (1527)	0.9 (0.7–1.3)
Duration of lying (h)	9.2 (4.1) *	1.0 (0.8-1.3)	10.1 (2.01)	0.8 (0.6-1.1)
Gait characteristics				
Logarithmic divergence rate AP (/stride)	1.5 (0.3) ***	0.9 (0.7-1.2)	1.9 (0.3)	1.7 (1.2–2.4) **
Range VT (m/s^2)	9.1 (1.9) ***	0.9 (0.7-1.1)	11.8 (5.6)	0.4 (0.2-0.8) **
Index of harmonicity ML	0.4 (0.2) ***	1.3 (1.0-1.7) *	2.0 (0.3)	1.4 (1.0-2.0) *
Sample entropy VT	0.2 (0.03) ***	0.9 (0.7-1.1)	0.3 (0.1)	1.4 (1.0-2.0)

Note: AP = anteroposterior direction; ML = mediolateral direction; VT = vertical direction; N/A = not applicable with no solution (Infinity); Continuous parameters are z-transformed. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Significance indicators in the column for mean VIBE data indicate significant differences between the mean (SD) of the VIBE and FARAO data, as tested using an unpaired t-test. Significance indicators in the OR columns indicate significance of the OR.