SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

School of Medicine and Surgery

Ph.D. program in Public Health

Cycle XXXIII

Curriculum in Health Services Research and Technology Assessment

Fall risk detection and prediction in community-dwelling older adults

Surname: CASELLI

Name: SERENA

Registration number: 835780

Tutor: Prof. Lorenzo Mantovani

Co-Tutor: Dr. Paolo Angelo Cortesi

Coordinator: Prof. Guido Grassi

Academic year 2019/2020

Abstract

The fall risk and a related injury increase with age, and with the association of neurological diseases, like Parkinson's Disease (PD) or stroke sequelae. Falls represent a significant public health issue and a fearsome event for an elderly, both for possible traumatic consequences and psychological impact, and their human, health, and material costs. Hence, fall prevention in elderly subjects at risk is a current public health priority. Epidemiological studies have identified fall risk factors, and all international guidelines recommend a multifactorial removal approach of the modifiable fall risk factors and the implementation of evidence-based effective interventions on people at risk to prevent falls. Nevertheless, at now, investigators have not used unique classifications for fall risk factors, so the use of the WHO Family of International Classifications can be the more natural and logical solution to cover the lack of a universal reference framework. Fall risk screening is the first component of effective fall prevention programs. To date, despite the use of numerous fall risk assessment tools, it is not possible to detect and predict older adult fallers with optimal diagnostic accuracy.

Thus, the aims of this thesis were: 1) to validate a fall risk serial screening algorithm with a high level of diagnostic accuracy in a sample of community-dwelling older people, also with associated neurological diseases, in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months; 2a) to assess the effect of the neurological diseases on the fall risk screening tests; 2b) to validate an International Classification of Diseases (ICD) & International Classification of Functioning (ICF) core set for the fall risk in the same population.

Using data collected in the PRE.C.I.S.A. trial about the efficacy of a tailored intervention on fall risk, we performed the following analyses: 1) to validate a fall risk serial screening algorithm we calibrated VAE, VOE1 and VOE2 measurement scales with classical techniques and Rasch analysis and we calculated the two FRAT-up on the same variables; we studied the diagnostic accuracy of the single tools and the screening algorithms, obtained with serial combinations of the calibrated scales and the two FRAT-up tools, and with logistic regression models, in the prediction of the described outcomes; we compared their properties, with a purpose of external validation; 2a) to assess the effect of neurological diseases on the screening tests we conducted a

Differential Item Functioning (DIF) analysis for the calibrated scales and a t-test comparison for the two FRAT-up; 2b) to validate the ICD&ICF core set for the fall risk we reviewed the fall risk factors and we linked them to the classification categories; we compared and integrated the core set with the existing Yen's ICF core set for falls.

The available sample from the PRE.C.I.S.A. trial was constituted by 768 older adults (female 65.3%; mean age 75.8). From 29 observed variables, we calibrated three measurement scales (VAE, VOE1, and VOE), which showed a satisfactory fit to the Rasch model (χ^2_{13}=43.4; p=0.080; χ^2_{12}=17.5, p=0.130; χ^2_6=32.9, p=0.040, respectively), including ordering of thresholds, unidimensionality, local independence, and invariance. Their reliability (Person Separation Index=0.912; 0.900; 0.800, respectively) was adequate for individual measurement (VAE and VOE1) and group measurement (VOE2). The serial combination with 'AND' rule of the calibrated scales generated fall risk serial screening algorithms, with a good level of diagnostic accuracy, in the prediction of the described outcomes in a sample of community-dwelling older people, also with PD and stroke sequelae, based on cutoffs defined using an 'ad hoc' clinical method, which considered a higher cost of false negatives compared to false positives for the specific construct (≥1 fall: sensitivity (SE)=62.4%; specificity (SP)=71.0%; diagnostic accuracy (DA)=0.672; ≥2 falls: SE=72.8%; SP=63.2%; DA=0.657; ≥3 falls: SE=79.3%; SP=60.0%; DA=0.629). We calculated the cumulative post-test probabilities obtained with the serial combination of the scales, which performed more effectively than the single tools, and we constructed additional algorithms based on logistic regression models using a parallel combination. We realized an external validation through the comparison with FRAT-up algorithms. Then, we demonstrated that the neurological disease effect on these tools' performance is minor and manageable with Rasch analysis, as for the VAE scale (splitting analysis for one testlet for the presence of DIF by neurological diseases). Finally, we validated a comprehensive ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases (103 fall risk factors, from four reviews and the already existing ICF core set for falls by Yen, and linked to 74 categories).

The described serial algorithms could constitute the first component of an effective fall prevention program in older adults, followed by the delivery of effective multifactorial and multicomponent interventions to people at risk in an outpatient 'fall clinic'. Further

projects are desirable to replicate all these findings in the context of larger, multicenter validation studies, improving the sample representativeness, and then providing an economic evaluation of the proposed screening algorithms.

List of abbreviations and acronyms

€	Euro
ABS/BGS	American Geriatrics Society/British Geriatrics Society
ACSQHC	Australian Commission on Safety and Quality in HealthCare
ADL	Activities of Daily Living
AISP	Automated Item Selection Procedure
AUC	Area Under Curve
CCI	Conference on Preventive Aspects of Chronic Disease
CCM	Chronic Care Model
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CGA	Comprehensive Geriatric Assessment
COM	Center Of Mass
DA	Diagnostic Accuracy
DIF	Differential Item Functioning
FN	False Negative
FP	False Positive
FRAT-up	Fall-Risk Assessment Tool - up
ICD	International Classification of Diseases
ICF	International Classification of Functioning, Disability, and Health
ICF-CS	ICF – Core Set
ILSA	Italian Longitudinal Study on Aging
ISS	Istituto Superiore di Sanità
ISTAT	Istituto Nazionale di Statistica
LEA	Livelli Essenziali di Assistenza

LR+	Positive Likelihood Ratio
LR-	Negative Likelihood Ratio
MA	Mokken Analysis
MHM	Monotone Homogeneity Model
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
NPV	Negative Predictive Value
NSC	National Screening Committee
PD	Parkinson's Disease
PoTP+	Positive Post-Test Probability
PoTP-	Negative Post-Test Probability
PPV	Positive Predictive Value
PRE.C.I.S.A.	PREvenzione Cadute e promozione Invecchiamento Sano e Attivo
PrTP	Pre-test Probability or prevalence of disease
PSA	Prostatic Specific Antigen
PSI	Person Separation Index
QALY	Quality Adjusted Life Years
RA	Rasch Analysis
RCT	Randomized Control Trial
RMSEA	Root Mean Square Error of Approximation
ROC	Receiving Operative Curve
SN	Sensitivity
SP	Specificity
SRMR	Standardized Root Mean Residual
TLI	Tucker Lewis Index

TN	True Negative
TP	True Positive
UK	United Kingdom
US$	United States Dollar
USA	United States of America
VAE	Valutazione Anamnestica dell'Eleggibilità
VOE	Valutazione Oggettiva dell'Eleggibilità
WHO	World Health Organization
WHO-FIC	World Health Organization – Family of International Classifications
YI	Youden Index

Contents

1 Introduction

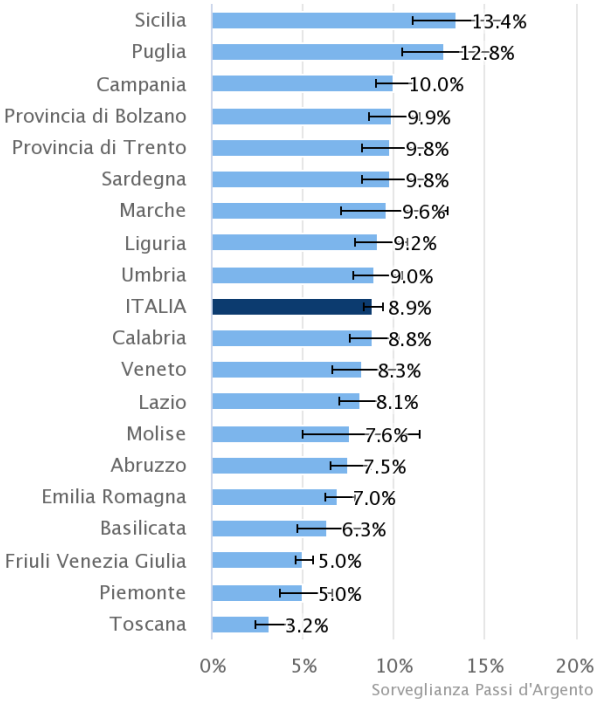
1.1 Fall risk

1.1.1 Population and epidemiology

The fall risk and of a related injury increase with age[1]. It is estimated that one elderly in three over 65 years living in the community experiments at least one fall in the following year[2], whereas, over the 80 years, this percentage grows to one in two[3, 4]. About 15% of the elderly at fall risk go through multiple falls in a year (recurrent fallers)[5, 6], and the incidence among institutionalized elderly populations is considerably even higher than that among community-dwelling elderly populations[4].

In Italy, the Istituto Superiore di Sanità (ISS), through its national health surveillance service 'Passi d'Argento' for people over 65 years[7], affirms that in the biennium 2016-2018, 9% of the respondents (n=39.930) declared to have fallen in the 30 days preceding the interview (Figure 1). Falls are more frequent with advancing age (7% between 65-74 years to 12% in the over 85 years) and between women (10% vs. 7% for men). Sicilia and Puglia are the regions with the worst values (respectively 13.4% and 12.8%), whereas Toscana and Piemonte are those with the best values (respectively 3.2% and 5%).

Figure 1. Epidemiology of falls in Italy from 2006 to 2008 (from Passi d'Argento, Istituto Superiore di Sanità[7])



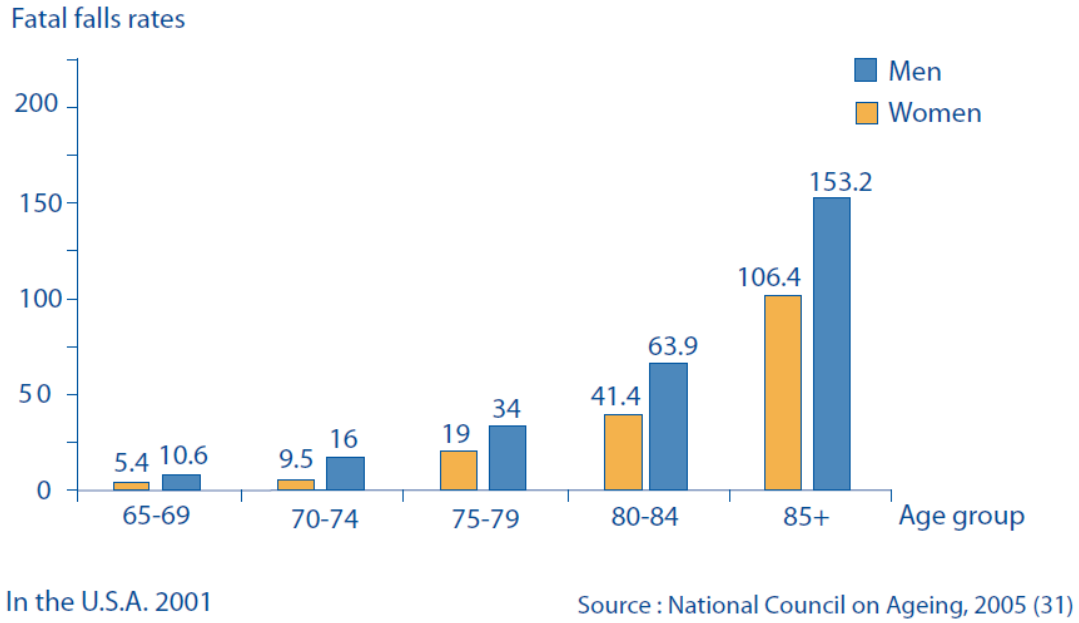
NOTES: the figure shows, in percentage, the number of people over 65 years who declared to have fallen in the 30 days preceding the interview. Both the Italian national value and the specific regional values are reported.

Falls in the elderly have various and complex implications. Falls constitute two-thirds of deaths for unintentional injuries, which are the fifth leading cause of death in older adults (after cardiovascular disease, cancer, stroke, and pulmonary disorders)[8]. Rates vary depending on the country and the studied population[9]. The fall fatality rate for people aged 65 and older in the United States of America (USA) is 36.8 per 100,000 population (46.2 for men and 31.1 for women). In contrast, in Canada, the same age group's mortality rate is 9.4 per 10,000 population[9]. In the USA, fatal fall rates increase exponentially with age for both sexes, highest at 85 years and over (Figure 2). Rates of fatal falls among men exceed that of women for all age groups despite the fewer occurrences of falls among them. This is attributed to the fact that men suffer from more co-morbid conditions than women of the same age[9].

The Italian ISTAT, Istituto Nazionale di Statistica, declares that in 2014, among victims of domestic accidents of people over 65, the incidence of falls was 76.9% (81% if women over 65 years)[10]. In 2017, the same institute reported that 3,690 people over 65

years have died for an accidental fall on the total deaths for accidental falls (4,097), which constitute, in turn, the 1% of the total deaths[11].

Figure 2. Fatal falls rate by age and sex group in 2001 in the USA (from the WHO Global Report on Falls Prevention in Older Age[9])



NOTES: the figure shows fatal falls by 5-year age group and sex in the USA 2001.

1.1.2 Burden of falls

1.1.2.1 Consequences of falls

As reported by Rubenstein in her works[4, 8], a key issue of concern is not merely the high incidence of falls in elderly persons - young children and athletes certainly have an even higher incidence of falls - but rather the 'combination of this high incidence and high susceptibility to injury'. This propensity for fall-related injuries is caused by a high prevalence of clinical diseases (e.g., osteoporosis) and age-related physiologic changes (e.g., slowed protective reflexes) that make even a relatively mild fall particularly dangerous.

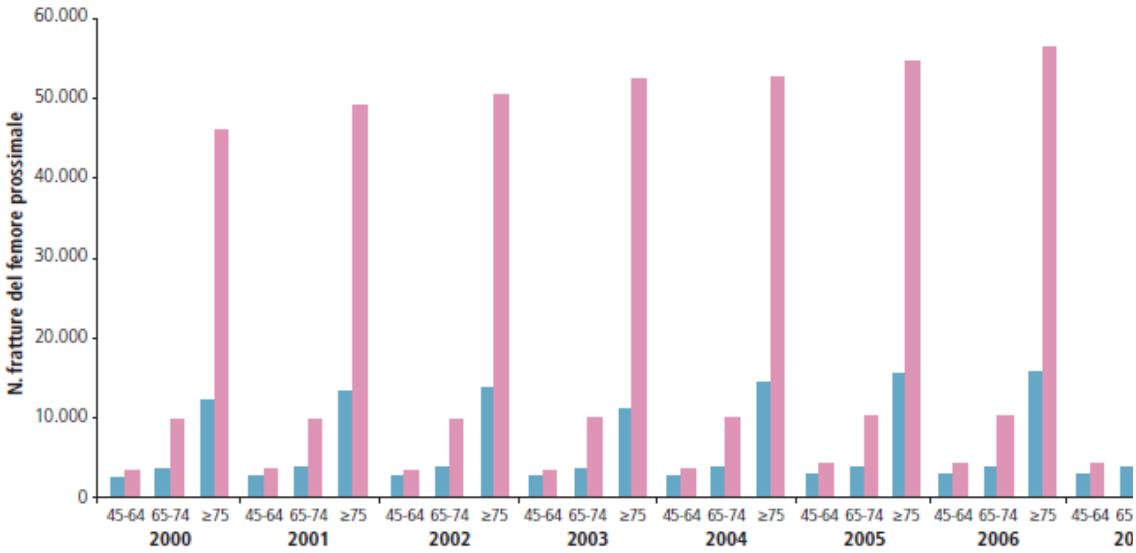
Although most falls produce no severe injury, community surveys reported over half of the falls result in minor injuries (bruising, abrasions, slight lacerations) that usually do not require medical treatment[2, 4]. Generally, 20% of falls require medical intervention, 5-10% result in a serious injury, such as a fracture, head injury, severe laceration, and 7% of elderly who have reported a fracture die. Falls are also the second cause of

traumatic brain injury and spinal cord injury[2, 4, 9, 12]. An important outcome of falls is the femur fracture, whose incidence is 10.8% in people over 65 years and grows to 12.9% for those over 74, until 14.2% over 79. When the elderly falls, connecting to this event, he/she may experience dehydration, pressure ulcers, rhabdomyolysis, hypothermia, and pneumonia if he/she remained abandoned on the ground for a time[12].

In Italy, in the period between 2016 and 2018, 19.3% of the elderly needed to be hospitalized at least for one day due to a fall[7]. The highest incidence of proximal femur fractures is observed in females over 75 years (over 75,000 from 2001 and 2007). The incidence doubles from 65 years in the women and 75 years in men, also caused by post-menopausal and senile osteoporosis (Figure 3)[13].

The duration of hospital stay due to falls varies; however, it is much longer than other injuries. It ranges from 4 to 15 days in Switzerland, Sweden, USA, Western Australia, Province of British Columbia, and Quebec in Canada. In the case of hip fractures, hospital stays extend to 20 days[9]. In Italy, the average hospital stay for femur fractures in 2002 was about 16 days[14].

Figure 3. The annual incidence of proximal femur fractures between 2000 al 2007 in Italy (from Quaderni del Ministero della Salute, 2010[13])



NOTES: the figure shows the distribution of proximal femur fractures in male (blue) and female (pink) populations with 45 years and over (data from Schede di Dimissione Ospedaliera).

These severe injuries are often associated with considerable long-term morbidity, like the impossibility of recovering the pre-facture level of independence in ambulation and daily living activities for 25-75% of community-dwelling elderly with hip fracture[4, 8].

Other serious consequences are produced by falls in this population. The elderly who fall repeatedly (recurrent fallers) have a higher risk of losing independence in daily living activities. This leads them to early admission to long-term care institutions with a probability up to 3 times superior to those with only 1 non-injurious fall[4, 15], to another recovery, to a further limitation of physical capacity and social participation, and to death[5, 9, 12]. Besides, falls may also result in a post-fall anxiety syndrome (30-73% of older people), constitute by fear of falling, which can cause, in turn, self-imposed activity restrictions, decline, depression, feelings of helplessness, and social isolation[4, 9].

Therefore, it is evident that a fall represents a fearsome event for an elderly, both for possible traumatic consequences and psychological impact that can have a devastating impact on his/her quality of life.

1.1.2.2 Costs of falls

According to the World Health Organization (WHO), falls' economic impact is critical to family, community, and society. Healthcare impacts and costs of falls in older age are significantly increasing all over the world. Falls costs are classified into two types:

- *Direct costs*, including healthcare costs such as medications and adequate services (e.g., surgery, pharmacological treatment, rehabilitation).
- *Indirect costs*, including societal productivity losses of activities in which individuals or family caregivers would have involved if he/she had not sustained fall-related injuries (e.g., lost income), or welfare economic support (e.g., disability pension)[9].

Regarding direct costs, the average health system cost per one fall injury episode for people 65 year and older in Finland and Australia was US$3,611 (€3,240) in 2001-2002 and US$1049 (€944) in 1999, respectively[9]. Besides, the average cost of hospitalization for fall-related injury for people 65 year and older ranges from US$6,646 (€5,965) in Ireland to US$17,483 (€15,692) in the USA. These costs are projected to increase to US$240 billion (€273 billion) by the year 2040[9]. In Italy, in 2002, the health system spent for 80,800 femoral fracture hospitalizations in people over 65 €806 million (€394

million direct costs for recovery + €412 for rehabilitation), determining an average direct cost of hospitalization of €10,000 (Figure 4)[14].

Figure 4. The estimate of total costs of femur fractures in people over 65 years in Italy in 2002[14]

Numero di ricoveri per frattura femorale	80.800
Costi diretti relativi ai ricoveri (euro)	394.000.000
Costi di 1 mese di riabilitazione postoperatoria (escluso 5% di mortalità acuta) (euro)	412.000.000
Costi sociali (pensioni d'invalidità ed accompagnamento per gli stimati 18000 pazienti disabili all'anno) (euro)	108.000.000
Costi indiretti (20% dei costi diretti totali) (euro)	183.000.000
Stima dei costi totali delle fratture femorali (euro)	1.097.000.000

NOTES: in the table total costs of femur fractures in people over 65 years in Italy in 2002, divided into direct and indirect costs, are presented.

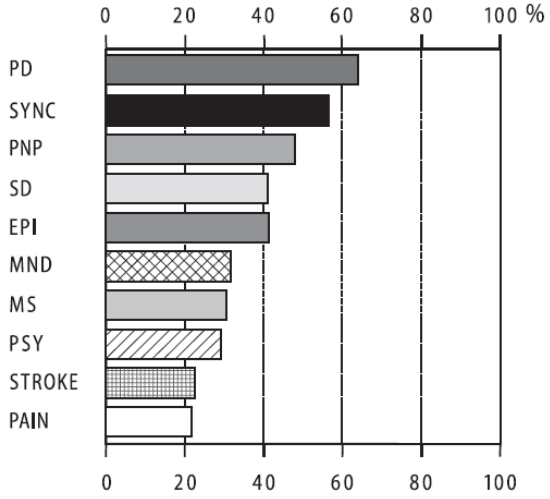
Regarding indirect costs, like the loss of productivity of family caregivers that are critical to family, in the United Kingdom in 2000, the average lost earnings could approximate US$40,000 (€36,000) per annum[9]. In Italy in 2002, the already cited work of Rossini and colleagues[14] reported that indirect costs for 80,800 hospitalizations, comprising also social costs linked to the disability pension of estimated 18,000 people with residual disability, amounted to €291 million (€108 million for disability pensions + €183 million for other indirect costs). The average indirect costs for a single recovery episode were €3,600; considering the total costs for a single recovery episode, the amount was about €14,000 (Figure 4)[14].

Given that with the progressive aging of the population, the human, health, and material costs of the falls risk of becoming unsustainable, their prevention in elderly subjects at risk is a current and priority public health objective[2].

1.1.3 Older adults with associated neurological diseases and fall risk

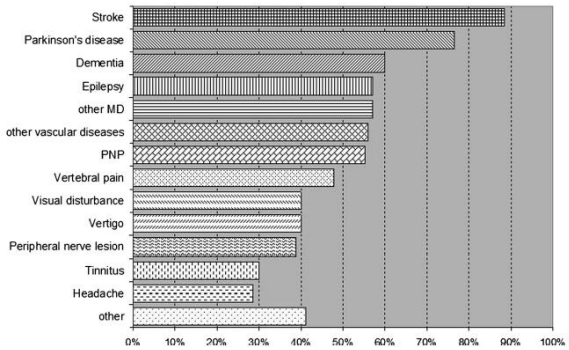
Owing to the typical sequelae of neurological diseases affecting integrative motor functions and activities like balance regulation and gait, a high prevalence of falls in neurological patients is expected[16]. In neurological elderly inpatients[16] and outpatients[17], the impact of one or more neurological diseases on top of an already increased propensity for falls, given by the aging, is substantial. Patients with specific diseases like Parkinson's Disease (PD) or stroke sequelae are particularly at risk[17]. According to the two cited studies, as shown in Figure 5 and Figure 6, PD and stroke are two of the neurological diagnosis in elderly with the highest frequency of falls in the last twelve months (Stolze, 2014, inpatients[16]: PD 62% and stroke 22% vs. Homann, 2013, outpatients[17]: PD 77% and stroke 89%).

Figure 5. Neurological diagnosis in elderly inpatients and fall risk in the last twelve months (from Stolze 2004[16])



NOTES: in the figure the ten neurological diagnoses with the highest frequency of falls in elderly neurological inpatients in the last twelve months of the cited study are displayed.
Abbreviations: PD, Parkinson's Disease; SYNC, SYNCope; PNP, PolyNeuroPathy; SD, Spinal Disorders; MND, Motor Neuron Disease; MS, Multiple Sclerosis; EPIL, EPILepsy; SAE, Subcortical Arteriosclerotic Encephalopathy; PAIN, PAIN syndrome.

Figure 6. Neurological diagnosis in elderly outpatients and fall risk in the last twelve months (from Homann, 2013[17])



NOTES: in the figure the available neurological diagnoses, in order of frequency of falls, in elderly neurological outpatients in the last twelve months of the cited study are displayed.

1.1.3.1 Parkinson's Disease and fall risk

Older adults with PD constitute 2% of the population over 65 years, being the most common neurodegenerative movement disorder[18, 19]. The clinical definition of PD is given by the presence of dopamine responsive motor impairment, including the triad of resting tremor, rigidity, and bradykinesia. These signs, when added to aging, contribute to increasing the fall risk dramatically. The reported average rate of falls of patients with PD in the more frequent follow-up period of twelve months is about 60.5%, with an average rate of 70% of fallers who are classifiable as recurrent fallers (who falls more than once in a specific period). The average rate of falls per recurrent fallers per year is 20.8%[6, 20]. These fall rates are double those reported for the general older population, and although the risk of falls increases with disease duration, falls are common even early in the disease[21-23].

Also the consequences of falls in PD are considerable. 50% of people with PD require medical care for fall-related injuries[24], and the incidence of hip fractures is 3-4 times that for older people of the same age[23, 25], conducting to a high impact on the utilization and cost of healthcare services[26].

Finally, falls are a significant determinant of poor quality of life, reduced mobility, and reduced life expectancy in people affected by PD.[20] With the number of people affected by PD expected to almost double between 2005 and 2030[27], fall prevention is set to become a great health challenge also in this population[23, 24].

1.1.3.2 Stroke sequelae and fall risk

Falls are considered as the number one medical complications after stroke[28-30]. It is also reported that the incidence and prevalence of stroke increase due to aging. As previously described for PD, stroke sequelae and age combination determines an augmented fall risk in this population[28]. Furthermore, the reduced physical activity level of individuals with chronic stroke compared to older adults (~2,800 steps a day vs. 6,565 steps a day) contributes itself to increasing the risk[28]. Regarding community-dwelling elderly who suffered from a stroke, from 40% to 73% are considered at fall risk in a one-year follow-up[28-30], with a proportion between 21% to 57% classifiable as recurrent fallers in the same time interval.

In literature, falls in individuals with stroke are lead to injuries in variable proportion (8-69%), but these injuries are usually mild (e.g., bruises). The incidence of all fractures (0.6-8.5) is similar to that of the general population. However, for these patients, many fractures (45%–59%) involve the hip, usually on the paretic side (76%–82%)[28]. The reported odds ratios for these specific fractures in the general population are 3.8 for people over 70 years and 2.1 for those over 80 years. The causes of this increased risk are linked to the augmented fall risk and probably to the loss of bone mineral density, which can be a common post-stroke complication[28].

In conclusion, individuals with stroke are much more likely to sustain a hip fracture due to a fall. They often lose independence mobility (62% vs. 31% in the general population) or even die. Also, the fear of falling (88%), decreased physical activity (44%), and the depression and social deprivation after a fall are described as severe consequences[28, 30]. Given these numbers, considerable clinical attention has to be directed toward fall prevention, even in older adults with stroke sequelae.

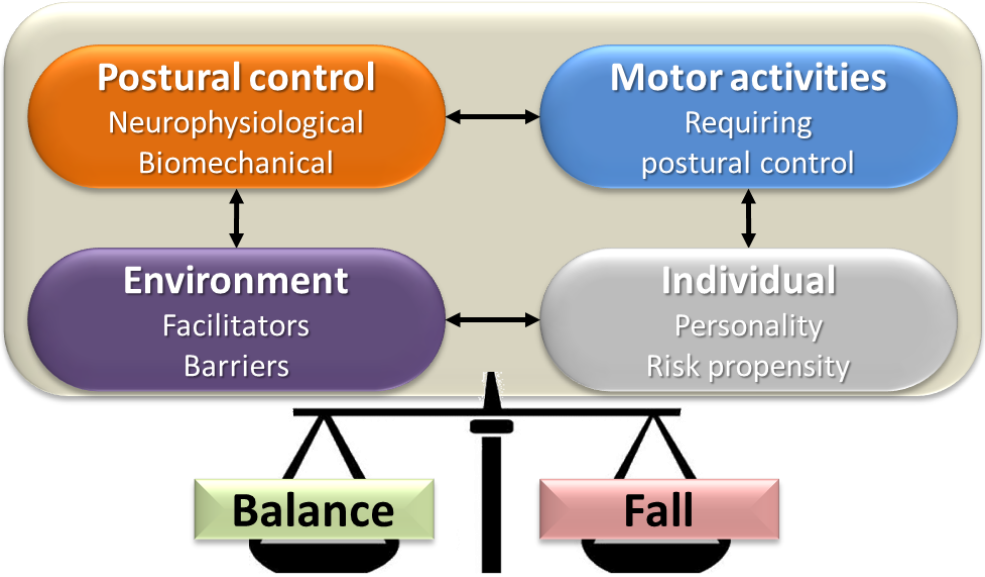
1.1.4 Definition, mechanisms and fall risk factors in older adults

1.1.4.1 Definitions of fall

Despite early attempts to achieve a consensus operational definition of ‘fall’, many definitions still exist in the literature, and, at now, there is not a universal consensus on that[2, 9, 31]. Lamb et al. in 2005 proposed a consensus statement that defined a fall as ‘an unexpected event in which the participant comes to rest on the ground, floor, or lower level, excluding intentional change in position to rest in furniture, wall or other objects’[2, 32]. The World Health Organization also shared this definition in its Global Report on Falls Prevention in Older Age[9].

From the causal point of view, the fall can be defined as an undesired outcome of a *motor activity* of an individual requiring postural control, with his own personality and risk propensity (*individual functioning element*), in which the force of gravity (*environmental element*) was not sufficiently contrasted by the *neurophysiological and biomechanical mechanisms of his postural control* [12, 33](Figure 7). On the contrary, the positive outcome of the same interaction is ‘balance’.

Figure 7. The causal definition of fall vs. balance



NOTES: in the figure, the graphical representation of the fall vs. balance's causal definition is presented.

1.1.4.2 Mechanisms of fall

1.1.4.2.1 Older adults

According to several authors, falls in the elderly were more likely to occur at home, and about half of them are caused by an environmental factor (objects tripped over, stairs, snow, or ice)[7, 34, 35]. Inside the house, rooms at the highest risk are the kitchen, the bedroom, the indoor and outdoor stairs, and the bathroom[7]. Trips and slips were the most prevalent causes, followed by misplaced steps (e.g., stepping into a hole) and losing balance[34].

According to Berg and colleagues[34], most falls occur (52%) in the afternoon, followed by morning, evening, and night (respectively 30, 14 and 4%), and the largest proportion of them (34%) takes place during the winter months of December, January, and February.

From the biomechanical point of view, over 50% of falls occur during activities that mildly displaced the subject's center of mass (COM), like standing still, basic daily living activities (ADL), and walking. Then, one third is caused by a moderate displacement of the elderly's COM (postural changes, bending over or reaching up, stepping down and up), and, finally, a little percentage by those activities which

markedly displace the COM (sports and climbing on ladders, chairs or other objects)[34, 35].

Finally, regarding reported reasons for falls, elderly refer that about one-third of the falls are attributed to hurrying too much. Approximately 20% of falls were considered the result of not looking where one was going, slipping on a wet or slippery surface, or tripping over something[34].

1.1.4.2.2 Older adults with Parkinson's Disease

In elderly with PD, generally, authors found that most falls happen at the faller's home, in particular bedroom, living area, kitchen, and garden, in line with observations in the general older population [36, 37]. Persons who fall indoors are significantly older and have more prominent non-motor symptoms of PD, whereas those who fall outdoors tend to be more likely younger, physically active, and healthy.

Ashburn in 2008 evidenced that about half the falls happen when the individual is ambulant, one-third when he is standing, and about 20% during transfers[36]. Trips and a loss of balance are the major suspected causes of falls during ambulation (i.e., walking, turning, stepping up or down, or carrying something from one place to another[36-38]). They are followed by a leg giving way, freezing or festinating, a leg not moving as expected, turning too quickly and slipping. The cause of most falls from standing (i.e., bending toward or reaching for an object, washing or dressing, or completing another everyday task such as preparing a drink, gardening, or making the bed) is a loss of balance, followed by stepping backward. When falls happen during transfers (i.e., to or from a seat or car seat, bed or toilet), the major specified cause is a loss of balance, followed by setting off too quickly after standing.

Finally, regarding reported reasons for falls, in the already cited study[36], the majority of individuals attributes falls to a misjudgment on their part, followed by having been distracted or having failed to concentrate adequately, or feeling dizzy, giddy, or tired at the time. In another conspicuous part, the individuals note that they fall without warning or that the cause remains unknown while recalling the event.

1.1.4.2.3 Older adults with stroke sequelae

Even in elderly with stroke sequelae, over 90% reported that falls had occurred in a familiar environment and 80% in patients' homes[30, 39]. In community-dwelling stroke survivors walking is the most frequently mentioned activity leading to a fall, but, differently from the elderly's general population, indoor walking is more involved than outdoor with a negligible involvement of extrinsic factors like obstacles[28, 39]. Other mentioned activities at higher risk are turning and rising from sitting to standing[39]. Falls during transfers are still a problem, but more frequently during the first phase of stroke recovery in the hospital setting[28, 30, 39]. A particularity that affects most patients with stroke sequelae is the presence of a weaker side of the body, which is reported to be the preferred direction of fall, together with their hands or knees[39].

Finally, regarding reasons for falls, the misjudgment, the lack of concentration, and the loss of balance seem the most common reported causes[39]; other already cited authors have reported the latter cause also for general older adults[34, 35].

1.1.4.3 Fall risk factors and their classification

As several authors claimed in the last decades, fall risk factors have been identified by epidemiological studies of varying quality[2, 40]. Although the studies' comparison and synthesis to identify the most important independent risk factors are difficult for some methodological reasons, more than one attempt in this direction has been made in a recent systematic review[41, 42]. Here, two among the most widespread and accredited classifications of fall risk factors are presented.

1.1.4.3.1 Effective Health Care Bulletin classification

The Effective Health Care Bulletin, in its edition of April 1996, made a classification proposal of fall risk factors from the perspective of the previous cited 'causal' definition of fall[40]:

- Risk factors depending on the *environment* (e.g., carpets, poor lighting, unsafe stairways, ill-fitting shoes, etc.).
- Risk factors which modify or act at the level of *individual functioning*:
    1. Pharmacological (e.g., antidepressants, sedatives, hypnotics, polypharmacy, etc.);

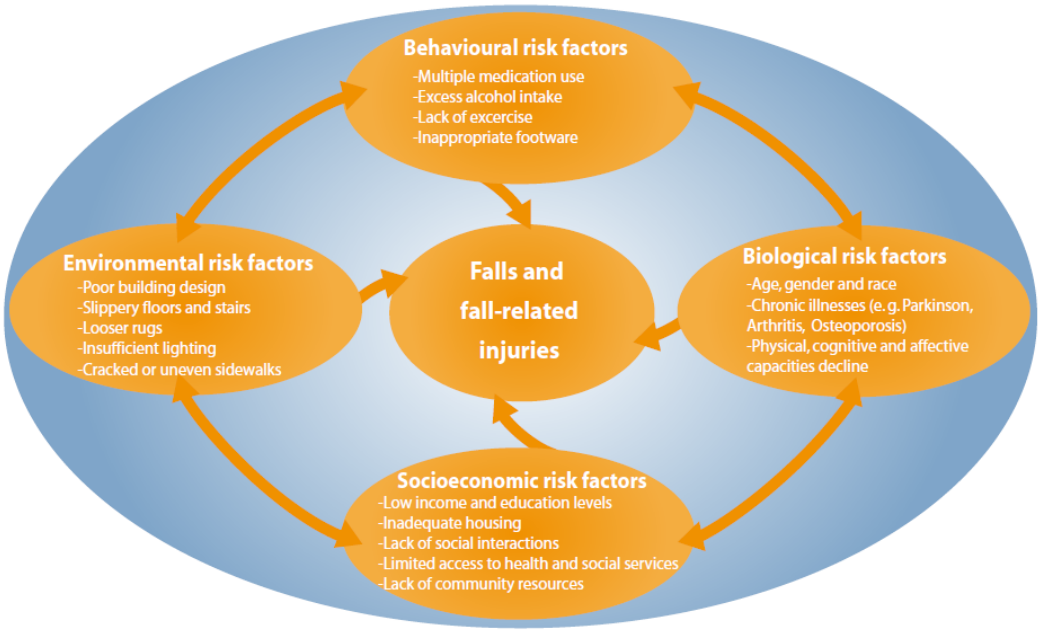
2. Medical conditions and aging changes (e.g., deterioration of vision, cognitive impairments, weakness of the lower limbs, sensitivity alterations, cardiovascular diseases, etc.);
3. Nutritional (e.g., vitamin D and calcium deficits);
4. Lack of physical exercise.

1.1.4.3.2 WHO classification

According to the WHO, in its Global Report on Falls Prevention in Older Age[9], falls occur due to a complex interaction of risk factors, reflecting the multitude of health determinants that directly or indirectly affect well-being. Those are categorized into four dimensions (Figure 8):

1. *Biological* risk factors: include individual characteristics about the human body (e.g., non-modifiable factors like age, gender, and race), also associated with changes due to aging (e.g., physical, cognitive, and affective capacities decline) and co-morbidity (e.g., chronic illnesses).

Figure 8. Risk factor model for falls in older age (from the WHO Global Report on Falls Prevention in Older Age[9])



Notes: in the figure the graphical representation of the interactions between the four dimensions of the fall risk factors and examples for each of the dimensions are displayed.

2. *Behavioral* risk factors: embrace those concerning human actions, emotions, and daily choices and are modifiable (e.g., risky behavior like the intake of multiple medications, the excess of alcohol use, a sedentary behavior).
3. *Environmental* risk factors: include the interaction between the individual's physical conditions and the surrounding environment (e.g., home hazard like narrow steps, slippery surfaces of stairs, looser rigs and insufficient light, and hazardous features in a public environment, like poor building design, slippery floor, cracked or uneven sidewalks and poor lightening in public places).
4. *Socioeconomic* risk factors: those are about the influence of social conditions and economic status of individuals and the capacity of the community to challenge them (e.g., low income, inadequate housing, lack of social interaction, limited access to health and social care, and lack of community resources).

The WHO highlights how the interactions between those four dimensions increase the fall risk. For example, muscle strength loss leads to a loss of function and a higher level of frailty, which intensifies the fall risk due to environmental hazards[9]. In their Cochrane review in 2012[2], also Gillespie and colleagues stated that only about 15% of falls result from an external event that would cause most people to fall; a similar proportion has a single identifiable cause such as syncope. However, the remainder results from multiple interacting factors.

Due to this genesis of most elderly falls, almost all guidelines recommend a multifactorial removal approach of the modifiable fall risk factors to prevent falls[43]. Examples of the application of this kind of approach are the behavioral and environmental modifications, like those related to a healthy lifestyle (non-smoking, moderate alcohol consumption, weight control, avoid sedentary) and modifications regarding home (provision of lighting and handrails) and public places (age-friendly design)[9].

1.1.5 Effective interventions to prevent falls

1.1.5.1 Effective interventions in community-dwelling older adults

The literature on the efficacy of interventions to prevent falls in community-dwelling older adults was summarized in a Cochrane systematic review in 2012[2] and,

subsequently, in an evidence report in 2018 for the US Preventive Services Task Force[44]. These reviews considered several effective interventions to prevent falls: single interventions, multiple interventions, and multifactorial interventions.

- *Single interventions*: according to Gillespie[2], among these type, the multicomponent physical exercise (both in group and individual), the Tai-Chi, home modifications, the vitamin D supplementation in people with known vitamin deficiency, the pacemaker installation in elderly with carotid sinus hypersensitivity, the gradual withdrawal of psychotropic drugs, a prescribing modification program for primary care physicians, the cataract-removal surgery, and the use of non-slip soles during winter are all effective interventions in reducing the rate and/or the fall risk. The evidence report of 2018[44], differently, showed that exercise trials demonstrate the exercise efficacy only on the fall risk, but not on the rate of falls. It also highlighted that results concerning the vitamin D supplementation interventions are mixed, but with a high dose associated with higher rates of fall-related outcomes.
- *Multiple interventions*: among them, it has been demonstrated the efficacy, in combination with other interventions, for the visual assessment, the educational interventions to minimize risk behaviors, the whole body vibration, the revision of footwear together with podiatric intervention in subjects with foot pain, and the eased access to the geriatric assessment[2].
- *Multifactorial interventions*: those interventions, which include individual risk assessment, showed extremely heterogeneous results in demonstrating their efficacy in reducing the rate of falls[2, 44].

1.1.5.2 Effective interventions in community-dwelling older adults with associated neurological diseases

In randomized control trials summarized by the cited Cochrane review[2], elderly with associated neurological diseases like PD or stroke sequelae were excluded. However, their fall risk is notoriously higher than that of the elderly general population, as we have extensively discussed.

Current literature highlights that physical exercise in elderly with PD can improve significantly balance and mobility[20, 45, 46], whose lack is one of the most important risk

factors, with a significant reduction of fall rate in at least one of the trials[20, 47]. The exercise trials' findings suggest that exercise interventions delivered early through a stage, group, and/or minimally supervised programs with a focus on challenging balance and attention are recommended. It prevents falls and maintains optimum mobility, and enhances any possible neuroprotective effect of exercise[20, 48]. As disease severity and risk increase, and where there is a history of multiple falls, injurious falls, or falls associated with dizziness or fainting, then more highly supervised exercise is recommended[20]. At present, several randomized control trials (RCTs) are in progress to test the efficacy of exercise protocols to reduce fall rate and fall risk in this population[49-51].

Regarding community-dwelling elderly with stroke sequelae, a 2013 Cochrane systematic review[29] with meta-analysis, reviewed at the end of 2019[52], affirmed that at present, there exists very little evidence about interventions other than exercises to reduce falling post-stroke. Besides, low to very low quality evidence exists that this population benefits from exercises to prevent falls, but not to reduce fallers. They concluded that fall research does not generally follow methodological gold standards, especially concerning fall definition and time post-stroke. They hoped that more well-reported, adequately-powered research could further establish the value of exercises in reducing falling, particularly per phase, post stroke[52].

The current lack of evidence in favor of fall preventing interventions in PD and stroke could have been determined by a combination of factors, among which: 1) the selection for ‘disease’ rather than for the estimate of the real fall risk; 2) the use of ‘standard’ exercises to improve balance, albeit personalized, which do not consider the more complex and specific motor impairments typical of these neurological diseases; 3) low compliance to the proposed physical exercise; 4) the absence of a multifactorial and individualized approach to the fall risk factors other than balance[50, 53]; 5) methodological limitations in study summaries due to heterogeneity between studies and high risk of bias[52].

1.2 Fall risk screening

According to WHO, fall risk screening is the first of several components of effective fall prevention programs. They aim to reduce the number of people who fall, the rate of

falls and the severity of injury should a fall occur, especially in adults older than 65, who suffer the highest number of fatal falls[54]. Fall prevention strategies should be comprehensive and multifaceted and, in particular:

1. They should prioritize research and public health initiatives to further define the burden, explore variable risk factors, and utilize effective prevention strategies.
2. They should support policies that create safer environments and reduce risk factors.
3. They should promote engineering to remove the potential for falls, healthcare providers' training on evidence-based prevention strategies, and individuals' and communities' education to build risk awareness[54].

1.2.1 Definitions of screening and its history

The practice of screening in healthcare has grown rapidly during the twentieth century and now has wide acceptance in our society[55]. According to WHO, 'screening is defined as the presumptive identification of unrecognized disease in an apparently healthy, asymptomatic population by means of tests, examinations, or other procedures that can be applied rapidly and easily to the target population'[56].

Over the years, there have been various definitions of screening starting from the fifties. One of the first definitions was by the US Commission on Chronic Illness in 1957[57]. It is similar to the latter: 'screening is the presumptive identification of an unrecognized disease or defect by applying tests, examinations or other procedures, which can be applied rapidly. Screening tests sort out apparently well persons who apparently have a disease from those who probably do not'. McKeown, in 1968[58], highlighted in his definition a further aspect of the screening, which is that the 'screening is a medical investigation which does not arise from a patient's request for advice for a specific complaint'. In the same year, WHO commissioned a report on screening from James Wilson and Gunner Jungner, who defined the screening criteria for the first time and adopted the screening definition by the Conference on Preventive Aspects of Chronic Disease (CCI), held in 1951: 'the screening is the presumptive identification of an unrecognized disease or defect by the application of tests, examinations, or other procedures which can be applied rapidly. Screening tests sort out apparently well

persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic. Persons with positive or suspicious findings must be referred to their physicians for diagnosis and necessary treatment'. Finally, the current definitions of the United Kingdom National Health Service (UK NHS) and that of the Ministero della Salute in Italia are, respectively, 'screening is the process of identifying healthy people who may have an increased chance of a disease or condition. The screening provider then offers information, further tests, and treatment. This is to reduce associated problems or complications. Screening should always be a personal choice'[59] and 'screening is a group of organized activities, aimed at a large part of the population, to early identify the presence of a disease in people who are asymptomatic'[60].

In summary, from all of these definitions, the screening may have at least two aims: it can be used for the protection of public health to identify a source of infection (e.g., in the search for the source of an outbreak of food poisoning), and for a direct contribution to the individuals' health.

Screening finds its most logical and widespread place in the oncology field. This is true because cancer shows its natural suitability for screening developing in different stages and sites. Criteria that enhance screening suitability include the potential for serious complications, physiological and financial burden, and overall a high rate of mortality (the second leading cause of death globally, about 1 in 6 deaths is due to cancer according to WHO), a prolonged preclinical phase, and an existing therapy that is simpler and more effective in reducing the mortality rate when applied to preclinical disease than to clinically evident cancer[61]. The five common cancers where screening is employed are breast, cervical, colorectal, lung, and prostate cancers[62].

The benefits of screening were first demonstrated using mass miniature radiography to identify individuals with tuberculosis in the late 1950s and early 1960s in Scotland. With the reduction in the burden of tuberculosis, the concept of screening began to be considered equally applicable to the control of other chronic diseases. This was shown particularly in the USA, where a law on controlling chronic diseases was passed in the late 1950s. The initial push for screening was particularly evident in North America, with one of the first examples: the introduction of screening for cervical cancer in British Columbia and California. In Europe, and especially in the UK, possibly because of fewer financial resources for health, screening lagged behind. However, during the

1960s, screening was progressively recognized as an important possible method of delivering preventive healthcare[55]. In Italy, screening, mainly addressed to the oncological field, began to take hold as organized programs on a large scale in the first decades of the 1990s. Until then, it existed only some isolated experience on a local and voluntary basis, which was expected to invite the population to carry out a series of prevention tests without tools capable of providing measurable feedback on the realized actions. Thanks to public health's decisive contribution, especially to epidemiology, the paradigm changed: epidemiologists reconsidered and implemented screenings as organized population programs. They also potentiated and developed theoretical and practical assessment elements to ensure the system efficacy: the path and the informative steps, the variable definitions, the indicators, the standards. Then, some years later, there was a further introduction of the impact indicators, which allowed and still allow to conduct studies to estimate the outcome of the taken measures and the made choices at the population level[63].

1.2.2 Characteristics of screening

1.2.2.1 Criteria for screening

The basic criteria to be fulfilled before introducing a specific condition screening have been well rehearsed over the years. On commission by the WHO, Wilson and Jungner in 1968[64] attempted to define screening criteria (Table 1) to guide the selection of conditions that would be suitable for screening, based, among other factors, on the capacity to detect the condition at an early stage and the availability of an acceptable treatment[65]. They wrote in a period of many technological advances in medicine, which made screening a topic of growing importance but also controversy. They pointed out that ‘the central idea of early disease detection and treatment is essentially simple. However, the path to its successful achievement (on the one hand, bringing to treatment those with previously undetected disease and, on the other, avoiding harm to those persons not in need of treatment) is far from simple though sometimes it may appear deceptively easy’[64], mainly when ‘case-finding is carried out by a public health agency, where the pitfalls may be more numerous than when a personal physician performs screening’[64].

Table 1. Wilson and Jungner screening criteria

1. The condition sought should be an important health problem.
2. There should be an accepted treatment for patients with a recognized disease.
3. Facilities for diagnosis and treatment should be available.
4. There should be a recognizable latent or early symptomatic stage.
5. There should be a suitable test or examination.
6. The test should be acceptable to the population.
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood.
8. There should be an agreed policy on whom to treat as patients.
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole.
10. Case-finding should be a continuing process and not a 'once and for all' project.

Regarding the condition, both Wilson and Jungner[64] and the UK National Screening Committee (UK NSC)[66], who summarized other fundamental indications by Cochrane[67] and Sackett[68], over the years have underlined that the importance of the health problem needs to be judged by its frequency and/or severity. Also, the epidemiology, the incidence, the prevalence, and the natural history of the condition should be understood, including the development of the disease, and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease. Furthermore, all cost-effective primary prevention interventions should have been implemented as far as practicable.

Regarding the test, it has to be simple, safe, precise and validated. Its distribution in the target population should be known, and a suitable cutoff level defined and agreed. The test, in all its steps from the sample collection to the delivery of results, should be acceptable to the target population. Also, there should be an agreed policy on the further

diagnostic investigation of individuals with a positive test result and the choices available to those individuals[64, 66].

Concerning treatment and according to both Wilson and Jungner[64] and UK NSC[66], as well as there should be an effective intervention for patients identified through screening, there should also highlight evidence that intervention at a pre-symptomatic phase leads to better outcomes for the screened individual compared with usual care. Off all the criteria, Wilson and Jungner conceived that the ability to treat the condition adequately, when discovered, is probably the most important[64]. Besides, evidence relating to the wider benefits of screening, for example, those relating to family members, should be taken into account where available. However, where there is no prospect of benefit for the individual screened, then the screening program should not be further considered.

Even the principles regarding the screening program have assumed a constantly increasing importance over the years. The effectiveness of the screening program in reducing mortality or morbidity should be evident from high quality randomized controlled trials. The same level of evidence is required to support that the test accurately measures risk if the screening aims to let people make an 'informed choice' (e.g., Down's syndrome screening). The test information and its outcome must be significant and readily understood by the individual being screened. Not only the test, but the complete screening program should demonstrate to be clinically, socially, and ethically acceptable to health professionals and the public. The benefit that the screening program procures to the screened individuals should outweigh any harm (e.g., overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings, and complications). Finally, the complete screening program cost should be economically balanced, concerning expenditure on medical care as a whole, in terms of value for money. This evaluation should be based on evidence from cost-benefit and/or cost-effectiveness analyses, also considering the resource availability[64, 66].

Then, regarding the program implementation, the authors pointed out that, before participating in the screening program, all healthcare providers should optimize the clinical management of the condition. All other options, such as improving the treatment of providing other services, should have been considered. Also, the availability of adequate staffing and facilities to perform all the screening steps should

be verified. Besides, the screening program should be continuously managed and monitored, with an agreed set of quality assurance standards[64, 66].

Finally, concerning the interface with individuals and the population, it is fundamental that the potential participants are assisted in making an informed choice, giving them evidence-based information explaining the purpose and potential consequences of screening, investigation, and preventative intervention or treatment. No less important is that health policy decisions, e.g., widening the eligibility criteria, and/or reducing the screening interval, and/or increasing the sensitivity of the testing process, should be guided and justifiable on a scientific basis[64, 66].

In 2008 and 2011, Andermann and colleagues reviewed the screening criteria over the past 40 years[65, 69], mostly linked to the recent discoveries in the sequencing of the entire human genome. Acknowledging that the Wilson and Jungner criteria have long been considered the gold standard in making decisions for implementing the screening program, they stated that a growing number of genetic screening policymaking approaches are in use. They are based on an even greater number of different sets of criteria, frequently a variation of the classic criteria of Wilson and Jungner. This situation leads to inconsistencies and difficulties in applying criteria considered too vague or theoretical to assess these criteria. Their review, which systematically identified and synthesized over 50 lists of screening criteria that have been proposed over the past 40 years, resulted in a modified version of Wilson and Jungner criteria and newly emerging criteria, which were transformed into a more elaborate decision support guide[69]. A synthesis of emerging screening criteria proposed over the past 40 years is presented in Table 2.

Table 2. Synthesis of emerging screening criteria proposed over the past 40 years[65]

1. The screening program should respond to a recognized need.
2. The objectives of screening should be defined at the outset.
3. There should be a defined target population.
4. There should be scientific evidence of screening program effectiveness.
5. The program should integrate education, testing, clinical services and program management.
6. There should be quality assurance, with mechanisms to minimize potential risks of screening.
7. The program should ensure informed choice, confidentiality and respect for autonomy.
8. The program should promote equity and access to screening for the entire target population.
9. Program evaluation should be planned from the outset.
10. The overall benefits of screening should outweigh the harm.

1.2.2.2 Organizational type of screening

From the organizational perspective, screening can be delivered according to two different modalities: *program screening* versus *opportunistic screening*[55, 62]. The program screening, also known as organized screening or population-based screening, consists of inviting groups of people who are thought to be at risk of attending screening[55]. Generally, there are personal invitations for eligible individuals and an organized process regarding invitation, intervention, management of screen-detected disease, clinical surveillance, and quality assurance[62]. Examples of this kind of screening are the public cancer screening programs developed in Europe and Australia. On the contrary, the opportunist screening or case-finding happens when individuals, in consequence to a medical contact for some reason, are addressed, by the healthcare professional, to perform other tests appropriate to the subject's age or sex (e.g., measurement of blood pressure or cholesterol)[55]. For example, in the USA, general

recommendations are given to the population, and the screening practice depends on the individual's medical insurance[62].

The opportunistic screening is considered relatively inefficient and cost-ineffective compared to the program one. A strong political commitment is required by the European Commission in 2017 regarding the cancer field to adequate and sustain logistic and fiscal support in the conversion of this kind of screening in population-based organized programs[70]. Like Getz and colleagues in 2003, other authors have already pointed out that it may be time to look critically at the question of opportunistic screening, saying, 'medical resources are increasingly shifting from making patients better to prevent them from becoming ill. Genetic testing is likely to extend the list of conditions that can be screened for - is it time to stop and consider whom we screen and how we approach it?'[71]. In that sense, they suggested that 'doctors should maintain a clear focus on each patient's reasons for seeking help rather than be distracted by an increasing list of standardized preventive measures with unpredictable relevance to the individuals. Also, they affirmed to reconsider the justification of opportunistic initiatives to prevent disease aimed at asymptomatic individuals from an ethical point of view[71].

1.2.2.3 Screening operative modes

There are two distinctively different operative modes to approach disease screening, e.g., *preventive* versus *early detection*. The first consists of preventing the disease by finding and removing precursors of the disease, whereas the second of detecting the disease as early as possible to treat and cure the patient[62].

In cancer screening, preventive methods highlight the disease before it becomes malignant by detecting and removing precancerous lesions (e.g., in cervical cancer screening detecting pre-invasive cervical neoplasia). This approach aims to prevent invasive disease and reduce the incidence of the disease, consequently reducing disease-mortality. Actually, all currently used screening prevention tools can also detect invasive tumors[62]. Examples of cancer screening tools with related operative modes are outlined in Figure 9.

Figure 9. Examples of cancer screening tools with the related operative modes (from Bretthauer et al., 2012)[62]

	Prevention	Early detection
Breast cancer		
Mammography	No	Yes
Ultrasonography	No	Yes
MRI	No	Yes
Cervical cancer		
Cervical smear testing	Yes	Yes
Colorectal cancer		
FOBT	No	Yes
Colonoscopy	Yes	Yes
Flexible sigmoidoscopy	Yes (distal colon only)	Yes (distal colon only)
Prostate cancer		
PSA	No	Yes
Lung cancer		
Chest X-ray	No	Yes
Computed tomography	No	Yes

Abbreviations: MRI, Magnetic Resonance Imaging; FOBT, Fecal Occult Blood Test; PSA, Prostate-Specific Antigen.

Otherwise, the goal of early detection tools is the reduction of disease mortality, e.g., mammography to reduce breast cancer mortality. However, they are not able to reduce the incidence of the disease. The risk linked to the early detection screening is, actually, the increase of incidence due to overdiagnosis[62].

1.2.3 Further aspects of screening

1.2.3.1 Economic evaluation of screening

In the last decades, the consequences of screening in terms of economic impact have been considered for two reasons: first, for advances in the application of economic principals in health services, and, second, for the comprehension that some screening procedures have a large consumption of money compared to the limited benefit to the population[55]. With the economic theory entrance in this field, 'it has been increasingly recognized that screening is not a universal panacea and that it may also do harm'[55], adding this consideration to others from the ethical point of view. The most important economic consequence of many individuals to test in screening programs is that these programs' actual costs are not negligible given these high numbers of screened people, and particular types of tests, e.g., body scanning, are expensive. Besides, the further exam for those found to be positive that eventually will be negative is also expensive.

For these reasons, since economic resources for health are finite, the costs of screening services have been justified for other services (e.g., better care service for elderly and chronic pathologies or other patients) [55].

Sassi developed in 2000[72] his considerations about the need for an economic evaluation of the diagnostic process in screening, highlighting the different stages of the complex economic evaluation[55]:

1. Examination of the test result's production result, i.e., sensitivity and specificity in relation to the population tested.
2. Use of the diagnostic output of screening as part of a diagnostic strategy and choice of treatment, 'which entails the estimation of post-test probabilities of disease using a Bayesian rule and the choice of treatment on the basis of disease probability thresholds balancing alternative possible outcomes'[72].
3. The production of outcomes conditional upon the choice of treatment.

The complexity of the evaluation is linked to several factors that have to be considered in terms of costs and benefits (Figure 10), e.g., the health effect of the screening; the effect of the test itself; the reassurance or distress of individuals undergoing screening; clinician's behaviors affected by the need of further diagnostic tests; psychological variables consisting in the likelihood of a patient's or clinician's willingness to take risks; the empathy and emotional relationship between the clinician and the patient; the limitations in our knowledge of the consequences of the tests or treatment the behavioral/attitudinal/psychological implications for the patient; incentives for specific actions of the practitioners; the costs of advertising, hiring and accommodating new staff and providing additional space for new machinery in the case of the introduction of a new test[55, 72].

Figure 10. The complexity of economic evaluation: factors and impact (from Holland, 2005[55])

Factor	*Impact*
Health outcomes	True results (positive and negative) False results (positive and negative) 'Psychologically mediated' health effects Short-term health outcomes
Reassurance and distress	For the patient (e.g. prenatal screening, screening for breast cancer) For the clinician (e.g. ordering of laboratory tests)
Psychological variables	Cognitive response style (risk aversion) – patient Cognitive response style (risk aversion) – clinician Emotional response style (sympathy/empathy) Cognitive limitations
Incentives	Financial (e.g. response to changes in relative fees) Non-financial (technological imperative)

Using these approaches in a later work in 2001[73], Sassi and colleagues have also raised the 'dilemma' of equity versus efficiency. In one of their examples regarding cervical cancer screening, the NHS policy aims to maximize population coverage by giving general practitioners economic incentives. However, less affluent women (known as at higher risk) in England have a lower participation rate in this program than their more affluent counterparts. As reported, the number of invasive cancer cases avoided in 1997 was likely to be 60-85% of the number of cases that might have been avoided if screening rates had increased uniformly in different social groups after the introduction of target payments to general practitioners. Equivalent cost-effectiveness ratios for cervical cancer screening could be achieved with less frequent but more even coverage. They concluded by signaling the NHS system's failure to balance equity and efficiency when they are in conflict[73].

This brief description of some of the economic considerations used in the evaluation of screening procedures demonstrates, in addition to the mentioned complexity of the problem, the need to develop appropriate economic models and perspectives and to avoid the somewhat simplistic methods that are still commonly used[55]. As Mushlin and colleagues pointed out in 2001[74], cost-effectiveness analyses regarding therapeutic interventions and diagnostic tests and procedures help make decisions to avoid unnecessary expenditure and contain cost based on the best available information. Without effort in doing these analyses, the most important consequences are measured

in 'lost opportunity to buy as many lives, cures, or extra life-years as possible with the funds allocated to healthcare' [74], and not only in financial terms. An essential role in this direction is required to physicians and medical scientists to obtain conclusions more scientifically robust and clinically valid, which constitute the basis for these analyses that can inform and direct government or medical system decision-makers in the best possible way.

1.2.3.2 Information about screening

In their already cited book, Holland and Stewart in 2003 affirmed that clear and evidence-based information about benefits and harms of any screening program should be given to all individuals invited to participate in any program[55, 75]. Besides, Skrabanek in 1988 already affirmed that 'screening healthy people without informing them about the magnitude of inherent risks of screening is ethically unjustifiable'[76]. Even Law[77] suggested that the same rigorous set of experimental data presented for the licensed use of new drugs would also be applied to medical screening. He reported that breast self-examination is an example of the spurious assumption that it must be beneficial and cannot harm. Actually, it has been not recommended and resulted in increased harm in terms of increased numbers of benign lesions identified, of biopsies performed, and considerable women's anxiety[78].

For healthcare professionals, providing the described information is not easy given the complex and rapidly changing health fields. However, it remains fundamental because individuals need help in the quality assessment of the vast amount of information through media and the internet to which they have access and the possible risks of tests and treatments. Other aspects can also increase this difficulty: consultations timed too short for an adequate explanation, the lack of knowledge on treatment options and effects by the healthcare professionals, the underestimation of the patients' wish or ability to cope with information[55]. For this reason, Shepperd in 1999[79] proposed some steps to help health professionals advise patients in the information evaluation. Besides, decision aids aimed at people facing health treatment or screening decisions have been developed as adjuncts to practitioners' counseling. People exposed to these aids feel more knowledgeable, better informed, and more precise about their values. They probably have a more active role in the decision-making and more accurate risk perceptions either within or in preparation for the consultation. There is also growing

evidence that decision aids may improve values-congruent choices and no adverse effects on health outcomes or satisfaction[79]. It has signaled the need for further research to study the effects on adherence with the chosen option, cost-effectiveness, and use with lower literacy populations[80].

Regarding decision aids for screening, Barratt and colleagues in 2004[81] proposed some considerations in developing decision aids for screening intervention, together with suggestions to developers (Table 3).

Table 3. Considerations in developing screening decision aids and suggestions to developers (modified from Barrat, 2003[81])

Considerations in developing screening decision aids	**Suggestions to developers**
Screening leads to overdetection and overtreatment	Present the chances of having pseudo-disease as well as clinically important disease detected by screening
Screening may include invasive follow-up investigations and treatments	Give information about the whole of the early detection and treatment process
Benefits of screening are delayed, whereas harms are immediate	Present balanced information about the cumulative chance of benefits and harms over equivalent time frames
Few people experience benefits from screening compared with the number who would be expected to benefit from most treatments	Present very small numbers by using large and consistent denominators, for example, outcomes per 1000 or 10 000 people screened
Individual values and preferences are critical to screening decision-making	Screening decision aids need to accommodate flexibility in labeling the outcomes of screening as benefits or harms
Evidence for screening decision aids is often limited	Explicitly declare where high-quality evidence is lacking; use ranges or some other method to convey uncertainty in numerical estimates
Public attitude is that early detection and/or prevention must be good	Explain that there is a choice and the reasons why people might decide to decline screening
Little regulation is in place to protect consumers from aggressive marketing, and there may be strong financial incentives to get people to participate in screening	Information about financial gains to the organization offering the screening test may need to be included in decision aids

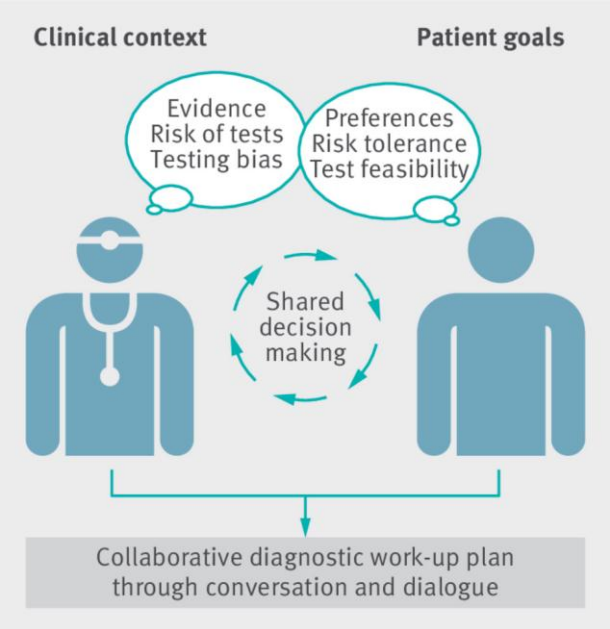
As detailed in Table 3, developers of screening decision aids must include information about the whole screening process, including follow-up tests and treatments, along with

the benefits and harms of the screening, to ensure that the individual can make an informed choice having considered all the elements[81].

A recent work by Berger and colleagues[82] highlighted the concept of patient-centered diagnosis and pointed out the difference between the shared (health professional and individual/patient) decision-making for diagnostic decisions (e.g., participation in a screening program) compared to the treatment decision-making. Three key points[82] are reported:

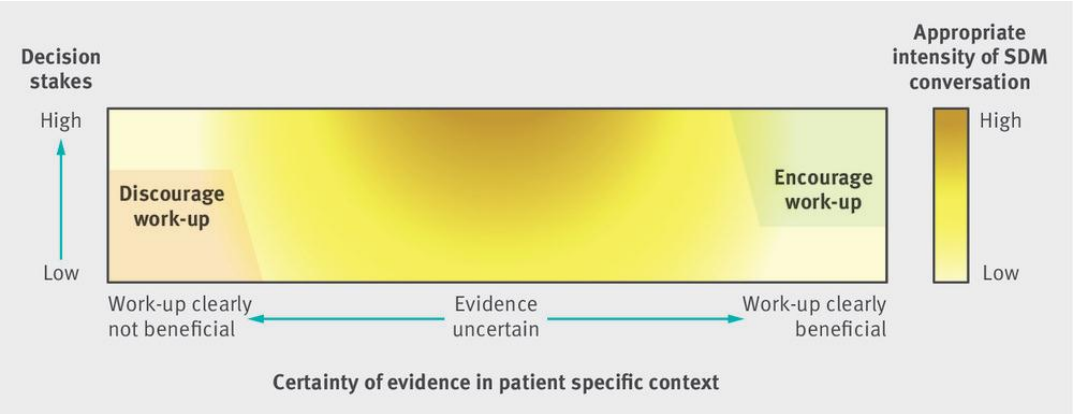
1. Patient-centered diagnosis is best practiced through shared decision-making: an iterative dialogue between doctor and patient, respecting the patient's needs, values, preferences, and circumstances (Figure 11).
2. Shared decision-making for diagnostic situations differs fundamentally from that for treatment decisions. This has important implications when considering its practical application (Figure 12).
3. The nature of dialogue should be tailored to the specific diagnostic decision; scenarios with higher stakes or uncertainty usually require more detailed conversations.

Figure 11. Conceptual model of patient-centered diagnosis from Berger, 2017[82]



NOTES: The clinical context includes the presenting patient concern, the clinician's role, and the clinical setting. Considering this and the patient's (or family's) overall goals of care, a dialogue between clinician and patient or family should be used to agree on a patient-centered diagnostic plan.

Figure 12. Framework for adjusting the intensity of shared decision-making in patient-centered diagnosis based on uncertainty and stakes from Berger, 2017[82]



NOTES: When the potential benefit to patients is uncertain, and when diagnostic decisions have high stakes, clinicians should engage in robust, detailed shared decision-making with patients. In other words, the highest is the uncertainty of evidence, the highest should be the intensity of the SDM.
Abbreviations: SDM, Shared Decision-Making.

In summary, then, information is another central concept in screening. The aim to achieve truly informed consent through shared decisions, based on a balanced and understandable picture of the options and the outcomes, should be followed.

1.2.3.3 Ethics of screening

Ethics of screening has an unbreakable bond with the just treated information. As we have already declared, screening differs significantly from the traditional medical practice. In the latter, it is the individual who seeks help from a health professional to obtain reassurance or a diagnosis and a treatment for his/her signs and symptoms. Conversely, in screening is the health service that invites individuals, who consider themselves healthy, to be tested to verify the presence of a particular condition at an early stage, either before or very soon after symptoms present. Early detection allows the treatment of the condition with a reversible or at least a containable outcome. Besides, in screening, any abnormality that is identified must be treatable, and the investigation itself must not do any harm[55]. According to the difference compared to traditional medical practice, the screening criteria presented in section 1.2.2.1 were born precisely to give references to screening implementation. This is because very considerable ethical responsibilities follow the screening application, as it potentially transforms individuals who are supposed to be ‘healthy’ to a state with some disorder or

a potential one. For the same ethical reasons, it should not be used to identify insignificant of untreatable conditions[55].

Several additional issues require consideration. First, between the cited criteria, the presence of an effective treatment for the diagnosed condition is crucial because the 'shift' of the individual from 'healthy' to 'sick' should occur when the likelihood of effective treatment and better long-term outcome (e.g., survival or better quality of life) is increased[55]. Besides, Rose's preventive paradox (or prevention paradox) is present in screening, especially regarding population prevention strategies. Rose explained that 'a preventive measure that brings considerable benefits to the community offers little to each participating individual'[83]. In such situations, the number of individuals found to have a positive test result is likely to be small, but the benefit is assessed in terms of the improvement in health (or survival) of the population investigated and not in tangible and brief-term benefit for the single person. Even when individual benefits are proven, they are enjoyed by only a few, usually many years in the future (e.g., a decreased rate of coronary artery events as a result of taking cholesterol-lowering drugs). At the same time, everyone participating in the program is at risk of harm, which often manifests itself immediately (e.g., anxiety resulting from an abnormal mammogram or impotence from a radical prostatectomy). Although harms, like benefits, affect only a few, most affected receive no compensatory gain[84]. Here appears the contraposition between 'individual good' (e.g., individuals identified at risk at an early stage who have been treated successfully) and 'community good' (e.g., screening policies not advocated universally for the same condition)[55]. The described 'lucky' individuals may find difficult to understand this kind of screening policies, but 'the difficulty in discussions on screening and the reason why stringent principles must be followed is that screening may also cause harm'[55].

However, in all instances, some members of the population screened may experiment disadvantages: the false positive subjects, who screen positive without suffering of the disease, and the false negatives, who screen negative but actually have the abnormality. Unfortunately, even in screening tests as in diagnostic processes, human and technical errors and variations can lead to mistakes. Hence, an assessment of the screening harm-benefit ratio must be studied[55, 62].

A need for balance in careful and rigorous consideration of all screening practices and proposals in the screening debate is present, between the extremes of enthusiasm and doubt. Any preventive measure or investment in health should be driven by whether the gains in healthcare are or not a reasonable return for the risks and costs involved[55], mainly because the thought that screening could be harmful is counterintuitive for many. Scientific evidence is needed underlying the harm-benefit ratio assessment, and it takes at least 10-15 years; only when findings that benefits are more than harms the screening test can be suggested to be implemented and communicated to politicians and the general public[62].

In this optics, the analysis of screening benefits and harms (disadvantages) that Chamberlain[85] did thirty-six years ago is still actual and considerable (Figure 13).

Figure 13. Benefits and disadvantages of screening (from Chamberlain 1984[85])

Benefits	*Disadvantages*
Improved prognosis for some cases detected by screening	Longer morbidity for cases whose prognosis is unaltered
Less radical treatment which cures some early cases	Overtreatment of questionable abnormalities
Resource savings	Resource costs
Reassurance for those with negative test results	False reassurance for those with false negative results
	Anxiety and sometimes morbidity for those with false positive results
	Hazard of screening test

The first and most crucial benefit for screened people is the improved prognosis for some cases to whom the diagnosis has been made at an early stage, not all. Conversely, for cases whose prognosis stays unaltered screening, screening can cause longer morbidity, longer awareness of their condition, and the submission to unpleasant and ineffective treatment[85].

The second benefit is the need for a less radical effective treatment for those screening identifies the disease in an early stage. Even in this case, there are two of the most important disadvantages, which are the ‘overdiagnosis’ and the following

'overtreatment' of similar lesions. If not detected, they would not have progressed to overt disease and would not have been identified clinically in someone's remaining lifetime[62, 85]. There are no markers in cancer screening, which can differentiate between lethal and non-lethal cancer, and all people detected as diseased are treated. In turn, the overtreatment does not benefit these subjects, but it leads to the risk of complications and adverse effects, together with the psychological distress diagnosis-related[62]. An example is the overdiagnosis in prostate cancer (PSA) screening, which has been estimated to be nearly 50%, which, together with its modest effect on mortality, has led to recent recommendations against PSA screening (D recommendation)[62, 86].

The third benefit regards resource savings related to the avoidance of more expensive services needed when the diagnosis is made at an early stage compared to later. On the counterpart, there are the resource costs of the screening program, the additional investigation for positive cases generated by screening, and the overtreatment of non-progressive disease[85]. The reassurance for those with negative test results by screening regards the great majority of screened people.

As pointed out by Chamberlain[85], the last three disadvantages are linked to the screening properties (i.e., sensitivity and specificity), which are not perfect by definition. The results falsely reassure the false-negative subjects, and, as a consequence of that, they may not consider symptoms and feel negative emotions if or when the diagnosis is made. On the contrary, false-positive patients experiment psychological and physical adverse effects from additional investigation to establish if they are really affected by the disease. Finally, the tests themselves may cause harm (e.g., amniocentesis). Included in this latter disadvantage, some authors also consider the anxiety a subject can prove in the time intervals between he/she is submitted to a screening test, receives the results, then, sometimes, performs the diagnostic test, and receives the final results[55, 87-89].

Despite the presented complex ethical analysis of screening, its public perception tends to remain positive. The majority of people continue to believe in a better treatment and improvement of diagnosis thanks to early staging, especially in cancer and heart disease[55]. This perspective's attraction leads individuals to request the availability of screening programs beyond the demonstration that diagnosis guarantees or not an improved outcome and assuming that, generally, more tests and treatments indicate

superior care[55, 90]. In her systematic review in 2015, Hoffmann showed that screening participants rarely had accurate expectations of benefits and harms, and they tend to overestimate benefits and underestimate harms for many interventions[90]. These optimistic intervention expectations by patients and the public contribute to enlarge the already cited problems of overdiagnosis and overtreatment and influence clinicians' decisions to provide interventions even with limited or no benefits. At the base of this belief, there is the false thought that 'identifying the presence of a condition equates with the ability to alter its natural history' [55]. Hoffmann[90] concluded that many participants reported in several studies that they would have stopped or not commenced screening if they had known that screening harms were high or outweighed the benefits.

In summary, thus, the ethics of screening appears as an immensely complex subject and does require further debate and clarification given ever-advancing technology.

1.2.3.4 Audit, evaluation, and quality control of screening

In any screening program, as with any other service program, adequate steps must be taken to ensure that the original objectives are being met and that the methodology meets appropriate standards. The general objective of screening is to reduce mortality and morbidity - that is, to identify a condition at a stage when it is reversible or at least containable[55]. Randomized controlled trials are the 'gold standard' for evaluating screening tools and methods, as they are for clinical medicine[55, 62]. This design enabled the investigators to control some of the common biases that can affect screening evaluation. In some cases, an even higher level of evidence studies is available, such as systematic reviews of evaluations of tests, which have the same aims as those of treatment interventions: to produce estimates of test performance and impact based on all available evidence, to evaluate the quality of published studies, and to account for variation in findings between studies[91].

In 1971, Cochrane and Holland[67] suggested seven criteria for the screening evaluation, which are still valid today (Table 4).

Table 4. Summary of criteria for evaluation of screening (from Holland, 2005[55] and Cochrane, 1971[67])

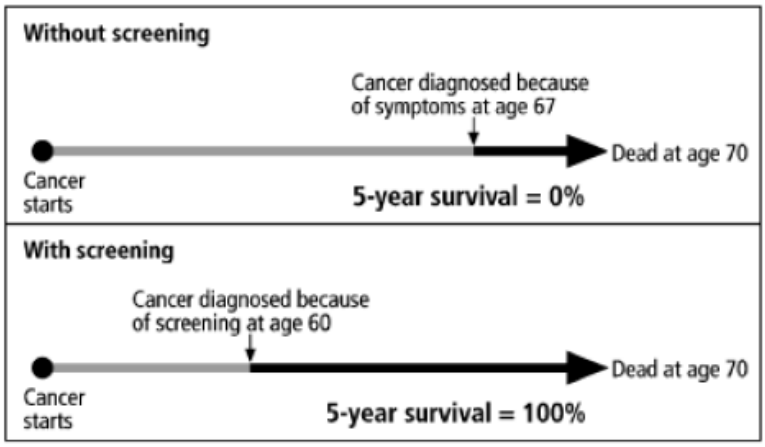
Factor	Criteria
Simplicity	The test should be simple to perform, easy to interpret, and, where possible, capable of use by paramedical and other personnel
Acceptability	Since participation in screening is voluntary, the test must be acceptable to those undergoing it
Accuracy	The test must give a true measurement of the condition or symptom under investigation
Cost	The expense of the test must be considered in relation to the benefits of early detection of the disease
Repeatability	The test should give consistent results in repeated trials
Sensitivity	The test should be capable of giving a positive finding when the individual being screened has the condition being sought
Specificity	The test should be capable of giving a negative finding when the individual being screened does not have the condition being sought

Cochrane and Holland also pointed out that consideration should be given to each of these criteria. In the case of a test that fulfills more one condition than another (e.g., tests with greater accuracy but, at the same time, more expensive and time-consuming), the test's choice must therefore often be based on compromise[67].

Regarding biases that can affect screening, at least four can be identified:

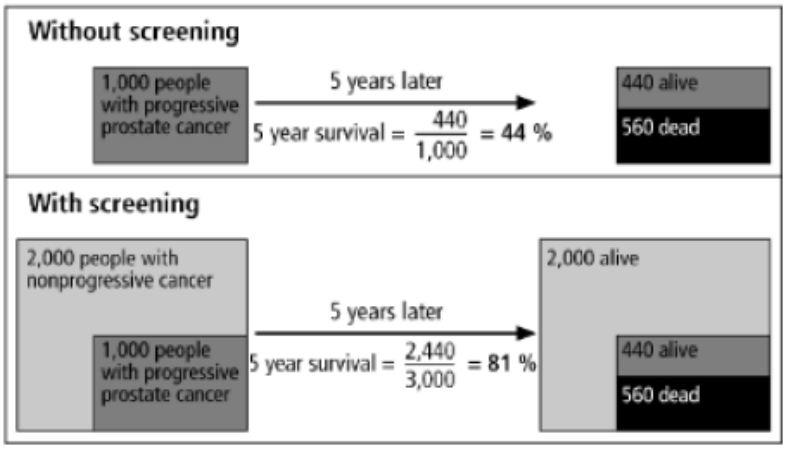
1. *Lead-time bias*[55, 62, 92](Figure 14). It is defined as the time interval from the diagnosis of a screening-detected disease (e.g., cancer) to the time point at which the disease (e.g., cancer) would have been detected clinically[62].

Figure 14. Lead-time bias (from Gigerenzer, 2008[92])



NOTES: even if the time of death is not changed by screening—and thus no life is saved or prolonged—advancing the time of diagnosis in this way can result in increased 5-year survival rates, causing such statistics to be misleading.

Figure 15. Overdiagnosis bias (from Gigerenzer, 2008[92])



NOTES: even if the number of people who die is not changed by screening—and thus no life is saved or prolonged—screening detected non-progressive cancers could inflate the 5-year survival rates, causing such statistics to be misleading.

Gigerenzer and colleagues[92] report a good example. A group of men diagnosed with prostate cancer at age 67: they all die at 70, so their survival at 5 years is 0% (Figure 14). What would have happened if cancer had been diagnosed earlier at age 60? They would have died the same at 70, but in this case, the 5-year-survival rate would have been 100% (Figure 14). The authors concluded that 'survival rates can be increased by setting the diagnosis time earlier, even if no life is prolonged or saved'. Hence, in this case, the screening does not produce a real extension of survival, but it is an 'artifact' of an earlier diagnosis.

2. *Length-biased sampling*[55]. It happens when individuals who have rapidly progressing disease will develop symptoms that cause them to consult a health professional immediately. Thus, only less rapidly progressive cases are likely to be detected by screening. Of course, the former have a poorer prognosis than the latter, and therefore the results will suggest that screening is more effective than is really the case.

3. *Selection bias*[55, 62]. Since any participant in a screening study is a volunteer, those who are most health-conscious will likely participate, and they are likely to survive longer in any case, whether they are screened or not.

4. *Overdiagnosis bias*[55, 62, 92](Figure 15). It is present when some of the identified lesions, which are counted as disease, may not present clinically during their lifetime[55]. In the upper part of Figure 15, there is the example of a group of 1,000 men with progressive cancer not submitted to a screening; the 5-year-survival for these men is 44%. Differently, we can consider another population submitted to a PSA screening. The test identifies 1,000 men with progressive cancer and 2,000 men with non-progressive cancer (who, by definition, will not die in the following 5 years at least for this diagnosed cancer). In this way, the 5-year-survival rate is inflated to 81%, but actually, the number of people who die has not changed at all (Figure 15)[92].

Regarding screening endpoints, the incidence, the mortality, and the long-term adverse effects of the intervention can be used to evaluate screening[62]. All current cancer screening tools aim to reduce cancer mortality through early detection of cancer or indirectly by reducing incidence. As Bretthauer highlights, none of the currently available cancer screening tools has been shown to have life-saving or life-prolonging effects. If screening works, an individual can exclude the disease he/she wants to die from by participating in screening, as screening reduces cancer-specific death, but not overall death. It is unrealistic that cancer screening has the overall-mortality reduction as a goal, given the magnitude of the effect of the currently used screening tools and the small proportion of each cancer to all-cause mortality. Conversely, all-cause mortality should be monitored carefully in every screening program because it is essential that screening does not increase it[62]. Due to lead-time, selection, and overdiagnosis biases,

the survival rate is not a valid endpoint in cancer screening because it has no reliable relationship to changes in mortality[62, 92].

Randomized controlled trials, although they are considered as the 'gold standard' for the evaluation of screening tools and methods, are particularly challenging to perform in screening for several reasons: 1) a long time period may pass from the screening intervention to the outcome of interest (death or cancer diagnosis), which often exceeds 10 years[62]; 2) the low event rates of cancer and cancer death in the target group (for example, the lifetime risk for colorectal cancer is about 5 per cent in Europe, so that 95 per cent of the population will never get the disease independent of any screening), which require very large study sample sizes, often several tens of thousands[62]; 3) the high costs of intervention and follow-up[55, 62]; 4) this kind of study can sometimes be ethically unacceptable, as an experience with screening for cervical cancer in the UK in 1964 suggests (i.e., a national cervical cancer screening program was established without sufficient evidence that the identification of abnormalities with the possibility of further investigation and treatment would reduce mortality from the condition, but at this point women would not have wanted to be included in the control group of a possible randomized control trial - RCT)[55]. Because of the explained reasons, the performance of observational studies such as case-control studies, cheaper and quicker than the RCTs, can be intriguing. However, it is necessary to use information from this kind of easier studies considering their specific limitations. Often screening participants have a different background risk for the disease than those who do not comply with the screening; the first can be more healthy and less prone to cancer. Therefore only high-quality observational studies, in addition to RCTs, were considered to provide robust data to evaluate the effectiveness of screening, as the International Agency for Research on Cancer working group concluded in their last report on breast cancer in 2015[93]. In particular, they 'gave the greatest weight to cohort studies with long follow-up periods and the most robust designs, which included those that accounted for lead time, minimized temporal and geographic differences between screened and unscreened participants, and controlled for individual differences that may have been related to the primary outcome'[93]. As the UK National Screening Committee, other institutions continue to declare that screening evidence in reducing mortality or morbidity should be derived from high quality randomized controlled trials[66].

Also, the regular scrutiny of all screening programs to control their performance and effectiveness is essential[55]. A widely accepted definition of 'effectively organized screening program' was proposed by Hakama and colleagues in 1985[55, 94, 95]. According to these authors (Hakama[94] and then modified by Holland[55]), organized screening programs should meet the following criteria:

a. the target population has been identified;

b. individual people to be screened are identified;

c. mechanisms are implemented to guarantee high coverage and attendance (e.g., a personal letter of invitation, suitable timing of screening examinations to suit the needs of those involved);

d. there are adequate field facilities (premises, equipment, and staff) for performing the screening tests in pleasant and acceptable conditions for individuals;

e. there is a defined quality control program concerning how the tests are performed and interpreted to ensure the maintenance of the best standards of the test:

   - initial and continuing training of the personnel conducting the test(s);
   - the demonstration, by appropriate records, of the maintenance standards of equipment used in the examination (e.g., calibration of X-ray machines in mammography);
   - routine checks of the validity of the tests performed (e.g., random duplicate measurements for biochemistry, cytology, and reading of X-rays).

f. adequate facilities exist for diagnosis and the appropriate treatment of confirmed abnormalities.

g. there is a carefully designed and agreed upon referral system, an agreed link between the participant, the screening center, and the clinical facility for diagnosis of an abnormal screening test, for management of any abnormality found, and for providing information about routine screening tests;

h. regular checks on feelings of satisfaction of those who have undergone the screening process are present - including those investigated, screen-negatives, and those who were invited and have not participated;

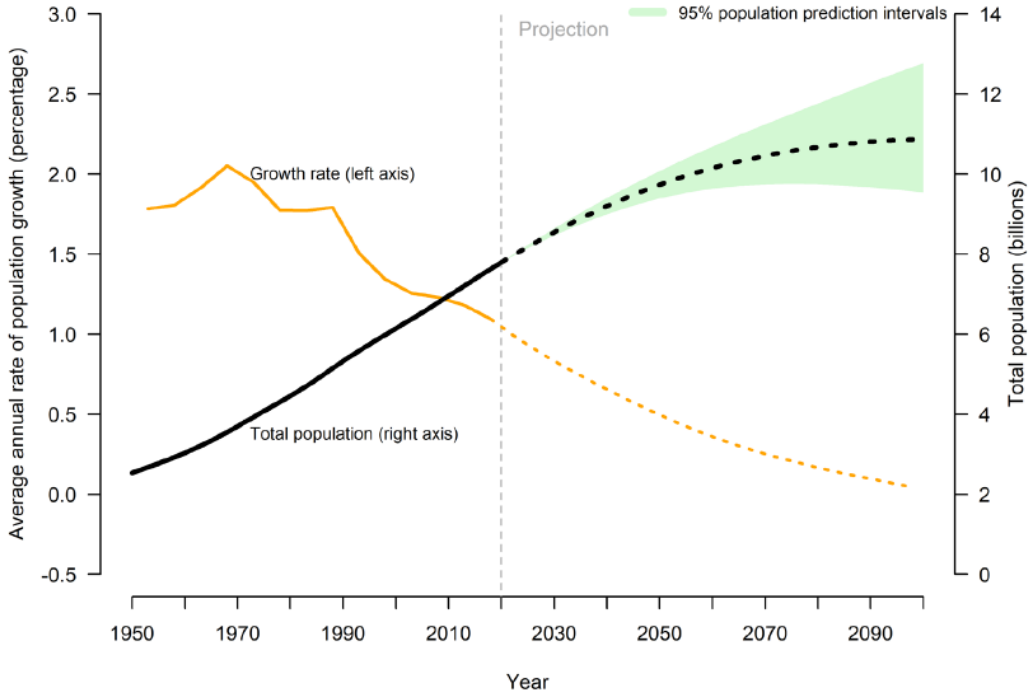
i. evaluation and monitoring of the total program are organized in terms of incidence and mortality rates among those attending and not at the total target population level. Quality control of the epidemiological data should be established.

In his discussion of the prerequisites for successful screening programs, Miller affirms that quality assurance in screening affects all program components, and, to ensure this, a quality organization is indispensable. Furthermore, an organized program is more effective and less consuming than a program that is not organized[96]. Hence, 'an effective organized program is one that is adequately planned, on the basis of the needs and resources of a country or region, one that is adequately funded, and one that is efficiently managed'[97]. In summary, quality must exist throughout the screening pathway.

1.2.4 Screening in the elderly

As all we already know, the world population grows, albeit at a slower pace than at any time since 1950, owing to reduced levels of fertility. From an estimated 7.7 billion people worldwide in 2019, the medium-variant projection indicates that the global population could grow to around 8.5 billion in 2030, 9.7 billion in 2050, and 10.9 billion in 2100. Europe and Northern America are projected to reach a peak population size and to begin to decline before the end of this century[98] (Figure 16).

Figure 16. Population size and annual growth rate for the world: estimates, 1950-2020, and medium-variant projection with 95% prediction intervals, 2020-2100 (from DESA,2019[98])

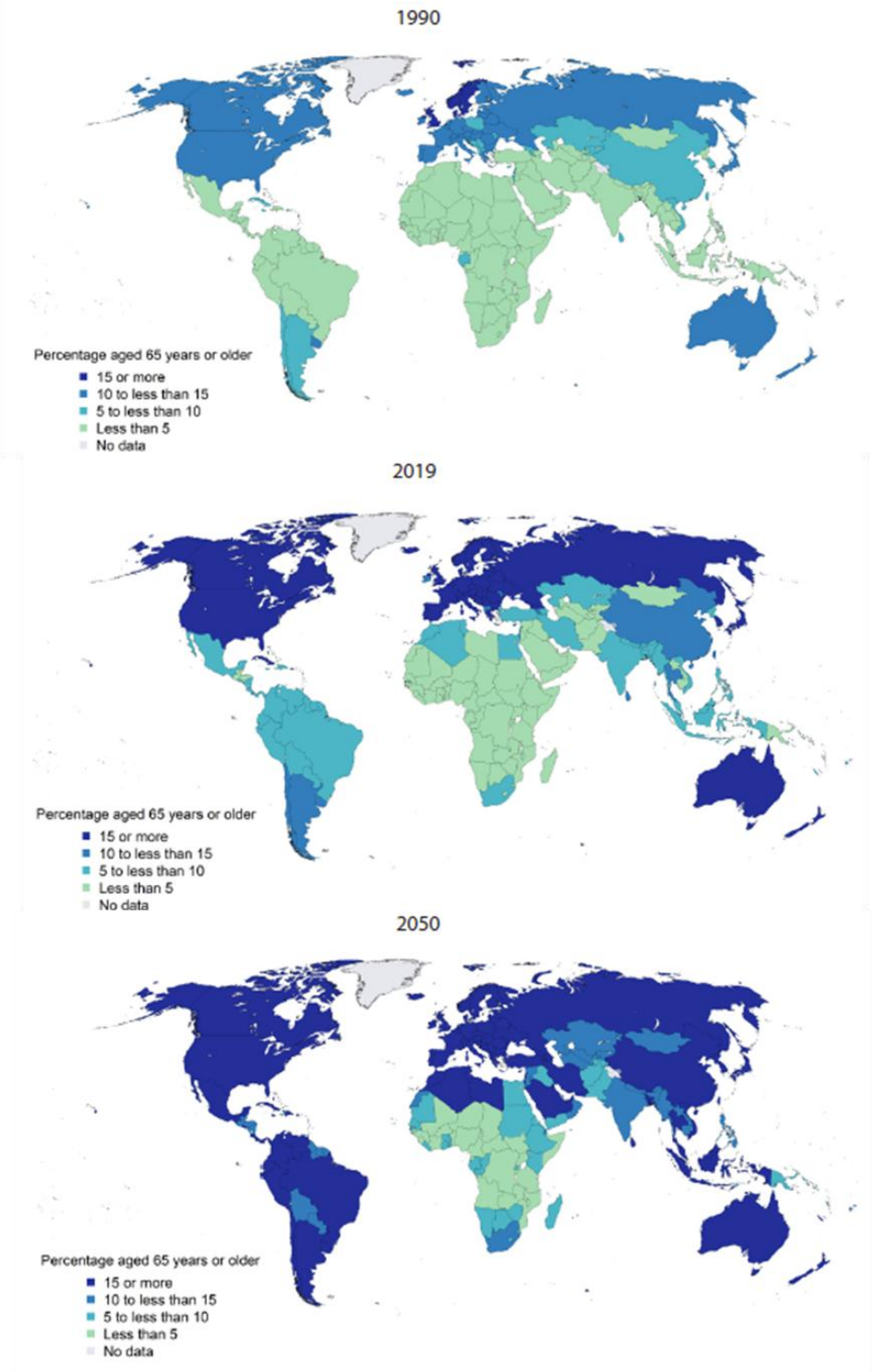


Data source: United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019.*

Notes: population growth continues at the global level, but the rate of increase is slowing, and the world population could cease to grow around the end of the century.

In 2018, for the first time in history, persons aged 65 years or over worldwide outnumbered children under age five. Projections indicate that by 2050 there will be more than twice as many persons above 65 as children under five. By 2050, the number of persons aged 65 years or over globally will also surpass the number of adolescents and youth aged 15 to 24 years[98] (Figure 17).

Figure 17. Percentage of population aged 65 years or over in 1990, 2019, and 2050, according to the medium-variant projection (from DESA,2019[98])



Data source: United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Population Prospects 2019*.

NOTES: virtually all countries and areas are experiencing population aging.

With increased longevity, more people are at risk of developing health problems primarily related to older age, so screening for these problems that would emerge in

later life is deemed as essential[99]. At present, there are screening or preventive care guidelines for older people available as a reference to guide primary care physicians[100-104]. Most of these guidelines focus on older people's functional ability and preventive activities in older age, including screening for physical function, vision, hearing, cognition, osteoporosis, fall prevention, and immunization.

However, because of the relative lack of reputable studies addressing in particular screening processes with beneficial health outcomes among older adults, there is little scientific evidence to support the benefits of screening in this age group, representing the key challenge[55, 100]. Besides, the effectiveness of preventive strategies in the presence of geriatric syndromes and comorbidities has not been addressed yet. Therefore, it is suggested that older adults' screening should focus on activities and health-related quality of life as the vital outcomes[99]. Identifying disabilities, such as hearing loss or reduced mobility, it can make a significant difference to the elderly's quality and enjoyment of life[55].

Hence, as pointed out by Holland[55], 'although screening in this age group will always have its opponents and its advocates, this is in our view a complex and important stage of the life cycle for screening and healthcare, since many of the difficulties encountered (e.g., impairments of sight, hearing, mental health or mobility) can with proper care and surveillance be helped if not cured'. An example of this is the comprehensive geriatric assessment (CGA), one of the procedures designed to improve older people's health[105-107]. The assessment regards the older adult's medical, psychosocial, functional, and environmental resources and needs and produces an overall plan for treatment and follow-up[106, 107]. A systematic review with meta-analysis of randomized control trials by Stuck and colleagues[107] in 1993 indicated that CGA programs linked to geriatric evaluation with long-term management effectively improve survival and physical and cognitive functional status in older people.

Another work by Spalding and colleagues[108], considering leading and actual causes of death, proposed as key recommendations for practice, based on an A evidence rating (consistent, good-quality patient-oriented evidence), the following screening evaluations in older adults:

- tobacco use in all adults;

- discuss aspirin therapy in all patients at increased risk of coronary heart disease;
- abdominal aortic aneurysm by ultrasonography in all men 65 to 75 years of age who have ever smoked;
- continue mammography screening in women older than 65 years;
- screen for colorectal cancer beginning at 50 years of age;
- perform cholesterol screening in men 35 years and older and women 45 years and older who are at increased risk of coronary heart disease;
- screen for diabetes in persons with hypertension or hyperlipidemia.

Also Bulpitt and colleagues[109], in 1990, advocated screening for both disease and consequences of them, such as disabilities, for common and potentially treatable conditions. They looked at thirteen possible screening tests, assessed based on specific criteria, and they provided evidence that screening may be worthwhile for the need for chiropody, varicose veins/ulcer, hearing loss, obesity, visual impairment, hypothyroidism, hypertension, anemia, and diabetes mellitus, even if these assessments needed to be tested prospectively in randomized controlled trials.

Considering the latter Bulpitt's work[109] and others, Holland recommended in 2005[55] the following assessment of the elderly in primary care:

- *Physical assessment*: hypertension, early heart failure, hearing loss, vision loss, incontinence, lack of physical activity, foot problems, review of medication.
- *Mental assessment*: depression, alcohol use.
- *Social assessment*: falls, under-nutrition, isolation.

According to Holland, the most appropriate form of delivery of this kind of screening in this age group seems to be regular surveillance and case finding in primary case, with an important role played by general practitioners, even if the resource implications of this for general practice must be confronted. These actions should always aim to improve quality of life and preserve function and independence, rather than on providing treatments to prevent mortality or prolong suffering[55].

1.2.4.1 Screening in the elderly: the Italian situation

Also in Italy, the definition of the health question in older people is a crucial issue. It is essential in such a country that experiments progressive aging and allocates more than half of its health resources to the elderly. However, it seems unable to guarantee equal care throughout the country. To answer these questions, Profea, the applied epidemiology program of the Istituto Superiore di Sanità, promoted the studio Argento[110] in 2003, conducted on eleven Italian regions. Together with a further multi-purpose investigation by Istat, the Studio ILSA (Italian Longitudinal Study on Aging)[111], which for over ten years has been following older population groups in an important cohort study, and the Rete Argento Università Cattolica S. Cuore, which in recent years has studied in depth the health needs of the elderly, they give birth to a contemporary production of relevant scientific investigations on older people's health questions[112].

In the Italian federalist health system, given the strategic guidelines defined by the Piano Sanitario Nazionale and the implementation lines agreed in the Livelli Essenziali di Assistenza (LEA), the single Region, the Azienda Sanitaria, and the District have autonomous managerial and operational responsibilities that involve decisive choices on the offer of services for the elderly. For this reason, it is urgent and necessary to adopt political decision-making that includes the citizen, providing the scientifically validated cognitive basis to allow health decision-makers to make their choices on scientifically valid data. This is not only for the programming of services offered but also for evaluating the efficiency of the services provided[112].

In summary, from these studies emerged some main results to consider given the implementation of screening programs throughout the national territory[112]:

- 3% of women and 2% of men aged 65 to 69 need daily assistance: this percentage rises to 25% in women over 80 and 18% in men. At 65, a woman has a life expectancy of 20 years, half of which will pass them in good health; the data are similar for men (Studio ILSA[111]).
- The perceived quality of life is very different in the country: 1% of the elderly in Bolzano judges himself in poor health against 24% of the elderly in Napoli of the same age (Studio Argento ISS[110]).

- The prevalence of the elderly's chronic diseases increases over time by about 50% in the last 10 years (Studio ILSA Firenze[111]).
- The cost of assistance increases over time, and according to age: as the number of older people as ‘users’ increases, the health cost of ‘users’ increases (Studio ILSA Firenze[111]).
- Despite the free offer of flu vaccine, only 50% of the elderly are vaccinated, mainly at the general medical doctor. Very little practiced is the vaccination against pneumococcus (Studio Argento ISS[110]).
- Substantial geographical differences: the elderly have different health, both objectively and subjectively. The average disability is 12% of the 65-year-olds in the center of Italy and 11% in the South and the North, while the severe disability in males is around 13% (Studio ILSA[111]).
- Subjected to screening for cognitive disorders, the elderly of Bolzano were positive to 11%, but the elderly of Basilicata were positive for cognitive deficits three times more (Studio Argento ISS[110]).
- Older people are large drug users and 92% of the elderly of Veneto take drugs in a continuous or cyclical against 79% of the elderly in Sardegna and 95% in Puglia (Studio Argento ISS[110]).

1.3 Fall risk screening guidelines and tools

1.3.1 Fall risk screening guidelines

Several medical societies and national health agencies proposed fall risk screening guidelines in community-dwelling older adults, containing recommendations, algorithms, and instruments. Three of the most important international guidelines by the following authors will be analyzed in details, together with one Italian guideline:

1. American Geriatrics Society/British Geriatrics Society (ABS/BGS) (last version 2010)[43];
2. UK National Institute for Health and Care Excellence (NICE) (last version 2013)[113];

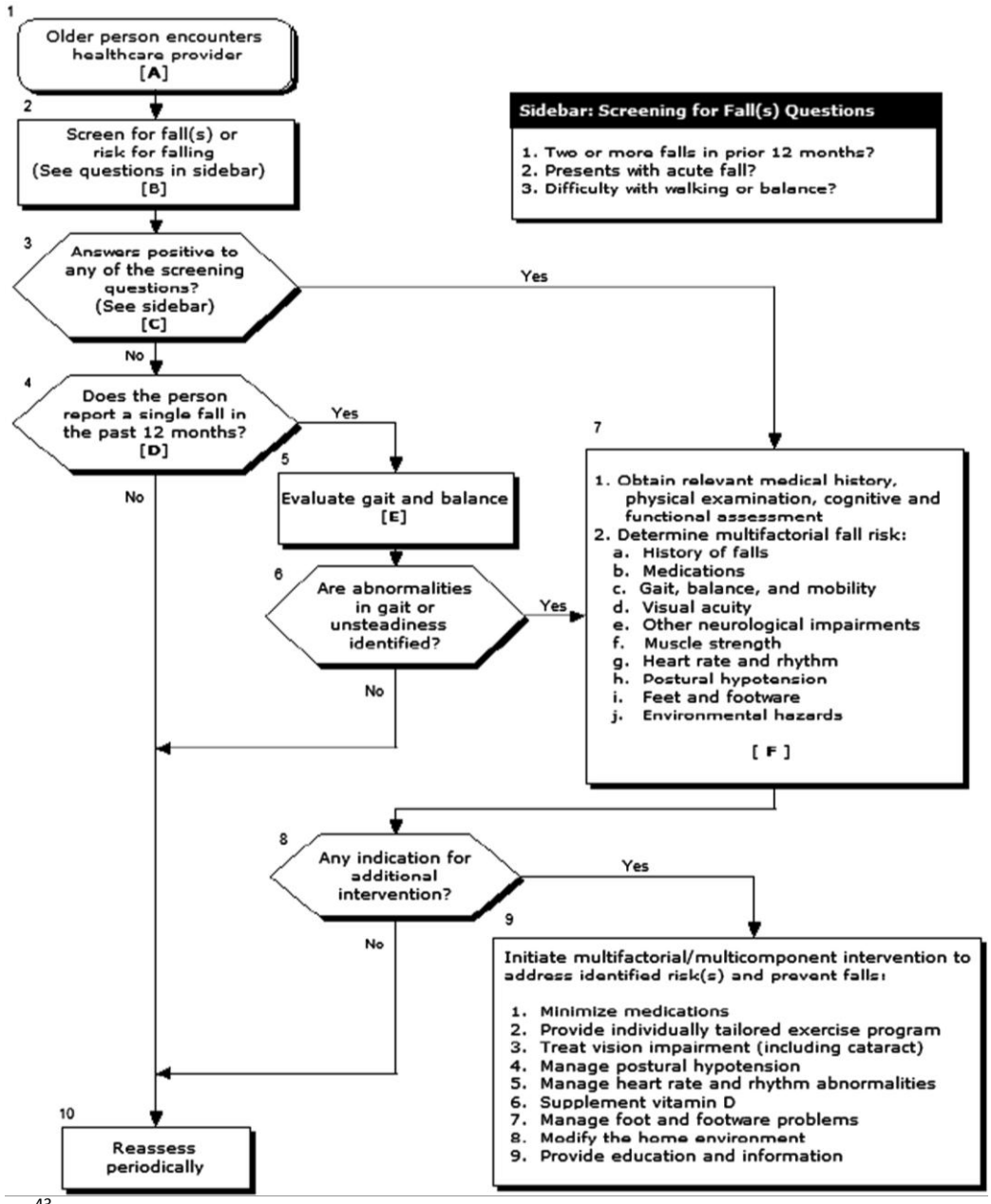
3. Australian Commission on Safety and Quality in HealthCare (ACSQHC) (last version 2009)[114];
4. Prevenzione delle cadute da incidente domestico negli anziani, Programma Nazionale delle Linee Guida, Ministero della Salute (last version 2009)[12].

1.3.1.1 American Geriatrics Society/British Geriatrics Society guideline

The 'Clinical Practice Guideline for the Prevention of Falls in Older Persons' by the American Geriatrics Society and the British Geriatrics Society (ABS/BGS) was published first in May 2001 and updated in 2010[43]. The latter was realized by evaluating evidence and analyses since 2001 and providing revised recommendations based on these evaluations.

Authors provide a clinical algorithm (Figure 18), which describes the systematic process of decision-making and intervention that should occur in the management of older persons who present in a clinical setting with recurrent falls or difficulty walking, or in the emergency department after an acute fall. When outcome data was insufficient to allow evidence-based recommendations or the existing literature was ambiguous or conflicting, the recommendations were based on panel consensus after intensive discussion.

Figure 18. Clinical algorithm for the prevention of falls in older persons in the community by the AGS/BGS[43]



NOTES[43]:

- *Annotation A*: Older Adult Encounters with Healthcare Provider. This guideline algorithm is to be used in the clinical setting for assessment and intervention to reduce falls in community-residing older persons (≥65). The guideline algorithm is not intended to address fall injuries per se or falls that occur in the hospital.
- *Annotation B*: Screen for Falls or Risk for Falling. The screening for falls and risk for falling is aimed at preventing or reducing fall risk. Any positive answer to the screening questions puts the person screened in a high-risk group that warrants further evaluation. All older adults under the care of a health professional (or their caregivers) should be asked at least once a year about falls, frequency of falling, and gait or balance difficulties.
- *Annotation C*: Screen Positive for Falls or Risk for Falling. Persons at higher fall risk,

identified by screening, should be assessed for known risk factors. A multifactorial fall risk assessment should be performed for community-dwelling older persons who report recurrent (≥2) falls, report difficulties with gait or balance, or seek medical attention or present to the emergency department because of a fall.

- *Annotation D*: Report of a Single Fall in the Past 12 Months. A (first) single fall may indicate difficulties or unsteadiness in walking or standing. In older individuals, a fall may be a sign of problems in gait or balance that was not present in the past.
- *Annotation E*: Evaluation of Gait and Balance. Gait and balance deficits should be evaluated in older individuals reporting a single fall as a screen for identifying individuals who may benefit from a multifactorial fall risk assessment. For persons who screen positive for falls or fall risk, evaluation of balance and gait should be part of the multifactorial fall risk assessment. Frequently used gait or balance tests include the Get Up and Go Test, the Timed Up&Go Test (TUG), the Berg Balance Scale (BBS), and the Performance-Oriented Mobility Assessment.
- *Annotation F*: Determination of Multifactorial Fall Risk. A multifactorial fall risk assessment can reveal the factors that put an older adult at fall risk and help identify the most appropriate interventions. A multifactorial fall risk assessment followed by intervention to modify any identified risk is a highly effective strategy to reduce falls and fall risk in older persons.

1. The algorithm starts with the encounter between a health provider and an older adult. As reported in annotation A (notes, Figure 18), the present guideline algorithm is addressed to older people (≥ 65 years) who live in the community.

2. The health provider screens for fall(s) or fall risk the older person through three questions (sidebar, Figure 18). According to annotation B (notes, Figure 18), any positive answer to the screening questions puts the person screened in a high-risk group that warrants further evaluation. Questions like these have to be made each year to all older adults under the care of a health professional (or their caregivers).

3. In the case of any positive answer to the screening questions (sidebar, Figure 18), the person at fall risk, according to annotation C (notes, Figure 18), should be assessed for known risk factors (point 7).

4. In the case of no positive answers to the screening questions (sidebar, Figure 18), the health provider asks the older person for a single fall in the past twelve months. As pointed out in annotation D (notes, Figure 18), the report of a single fall may indicate difficulties or unsteadiness in standing and/or walking or be a sign of problems in gait or balance not present before.

5. The health provider assesses gait and balance of the older adult. According to annotation E (notes, Figure 18) for people with a single fall in the past twelve

months, this assessment helps identify individuals who can benefit from a multifactorial fall risk assessment. Otherwise, for people screened positive for fall or fall risk (sidebar questions, Figure 18), this evaluation should be a part of the multifactorial fall risk assessment.

6. Those persons who present gait abnormalities or difficulties at point 5 should be addressed to a multifactorial fall risk assessment.
7. The health provider does a multifactorial fall risk assessment of the older adult. The assessment expects information about medical history, physical examination, and cognitive and functional assessment and determines the presence of fall risk factors (history of falls; medications; gait, balance, and mobility; visual acuity; other neurological impairments; muscle strength; heart rate and rhythm; postural hypotension; feet and footwear; environmental hazards). As pointed out in annotation F (notes, Figure 18), the multifactorial assessment can help identify the most appropriate interventions to modify any identified risk and reduce falls and fall risk in older people.
8. The health provider, given the multifactorial assessment, evaluates indications for additional interventions.
9. In the case of the need for additional interventions, the health provider proposes to the older person to initiate a multifactorial/multicomponent intervention to address the identified risk(s) and prevent falls (for interventions see algorithm, point 9, Figure 18).
10. If there is no need for additional interventions, the health provider will reassess periodically (each year) the older person's fall risk through this clinical algorithm.

Finally, from this guideline, it is interesting to report two recommendations regarding the screening and assessment section:

- Older persons reporting only a single fall in the past year and reporting or demonstrating no difficulty or unsteadiness during the evaluation do not require a fall risk assessment.

- A clinician (or clinicians) with appropriate skills and training should perform the multifactorial fall risk assessment.

1.3.1.2 UK National Institute for Health and Care Excellence guideline

The 'Falls in older people: assessing risk and prevention – clinical guideline' by the UK National Institute for Health and Care Excellence (NICE) was published in June 2013[113], updating those published in 2004 by the same institute. It provides recommendations for the assessment and prevention of falls in older people, and it addresses healthcare and other professionals and staff who care for older people who are at fall risk. In the 2013 version, new recommendations for older people in hospital are added to the original from the 2004 guideline. All guideline recommendations cover all people aged 65 or older because they have the highest fall risk.

Recommendations regard preventing falls in older people in the community and during a hospital stay. Focusing on the first, the guideline covers the following sections: case/risk identification, multifactorial fall risk assessment, and multifactorial interventions (Table 5).

Table 5. Recommendations for preventing falls in older people by the UK NICE guidelines[113]

1.1.1 Case/risk identification

1.1.1.1 Older people in contact with healthcare professionals should be asked routinely whether they have fallen in the past year and asked about the frequency, context, and characteristics of the fall/s. [2004]

1.1.1.2 Older people reporting a fall or considered at fall risk should be observed for balance and gait deficits and considered for their ability to benefit from interventions to improve strength and balance. (Tests of balance and gait commonly used in the UK are detailed in section 3.3 of the full guideline). [2004]

1.1.2 Multifactorial falls risk assessment

1.1.2.1 Older people who present for medical attention because of a fall, or report recurrent falls in the past year, or demonstrate abnormalities of gait and/or balance should be offered a multifactorial falls risk assessment. This assessment should be performed by a healthcare professional with appropriate skills and experience, normally in the setting of a specialist falls service. This assessment should be part of an individualized, multifactorial intervention. [2004]

1.1.2.2 Multifactorial assessment may include the following:

- identification of falls history;
- assessment of gait, balance, and mobility, and muscle weakness;
- assessment of osteoporosis risk;
- assessment of the older person's perceived functional ability and fear relating to falling;
- assessment of visual impairment;
- assessment of cognitive impairment and neurological examination;
- assessment of urinary incontinence;
- assessment of home hazards;
- cardiovascular examination and medication review. [2004]

1.1.3 Multifactorial interventions

1.1.3.1 All older people with recurrent falls or assessed as being at increased fall risk should be considered for an individualised multifactorial intervention. [2004]

In successful multifactorial intervention programmes the following specific components are common (against a background of the general diagnosis and management of causes and recognised risk factors):

- strength and balance training;
- home hazard assessment and intervention;
- vision assessment and referral;
- medication review with modification/withdrawal. [2004]

1.1.3.2 Following treatment for an injurious fall, older people should be offered a multidisciplinary assessment to identify and address future risk and individualised intervention aimed at promoting independence and improving physical and psychological function. [2004]

NOTES: in the table, recommendations from UK NIC guidelines taken from the three sections in bold are reported.

The recommendations regarding the case/risk identification section (Table 5, first section) agree with those by AGS/BGS[43], providing that periodically, health professionals, when in contact with older people, asked them information about fall(s) in the last year. If fallen or considered at fall risk, they advise submitting the person to a balance and gait assessment to understand the need for a strength and balance program.

Then, concerning the multifactorial fall risk assessment section (Table 5, second section), NICE recommendations suggest that this kind of assessment is indicated in case of fall or recurrent fallers in the past year or demonstrated abnormalities of gate and balance. Hence, it is possible to identify a difference compared to AGS/BGS guideline[43]: for the latter, the single fall in the past year without a contextual situation of balance and/or gait deficits is not followed by a multifactorial assessment, but only by a periodically reassessing through the fall risk screening. Otherwise, they agree that a healthcare professional should perform the multifactorial fall risk assessment with appropriate skills and experience and evaluate the fall risk factors.

Furthermore, regarding multifactorial interventions (Table 5, third section), the NICE guideline indicates an individualized multifactorial intervention for those at high fall risk or recurrent falls, as AGS/BGS does, with an extended agreement also to the successful included specific components[43]. Besides, NICE provides a further recommendation addressed to people submitted to treatment for an injurious fall. It suggests a multifactorial fall risk assessment and an individualized intervention to promote independence and physical and psychological functions.

Finally, in 2019 a 'Surveillance report' of falls in older people[115] checked this guideline and established the necessity to update it due to new evidence in several areas:

- *Fall risk assessment tools*: new evidence suggests that tools based on clinical characteristics lack sensitivity, specificity, or both. However, the addition of new technologies to measure gait may improve on assessments based on clinical risk factors. Additionally, topic experts indicated a need to include frailty and previous fragility factors as risk factors for falls because people with these characteristics may have more severe fall consequences[115].
- *Multifactorial interventions*: new evidence indicates they may not be effective. This finding contradicts current recommendations to offer multifactorial

interventions, indicating that offering interventions based on individual risk factors may not be effective, but offering all interventions irrespective of individual risk factors may be effective[115].

- *Exercise interventions*: they appear to be effective using a wider range of exercise types than are currently recommended. The update should consider how to safely encourage people at risk of falls to undertake the recommended physical activity levels and maintain benefits after the prescribed exercise program ends[115].
- *Vitamin D*: evidence suggests that it may be associated with fewer falls, although conversely, high doses of vitamin D may increase falls risk. The update should consider the role of vitamin D in falls prevention[115].

1.3.1.3 Australian Commission on Safety and Quality in HealthCare guideline

The Australian Commission on Safety and Quality in HealthCare (ACSQHC) in 2009, to lead and coordinate improvements in the safety and quality of Australian healthcare, produced the 'Preventing Falls and Harm From Falls in Older People - Best Practice Guidelines for Australian Community Care'[114]. Here, the focus is on the guideline for fall prevention in the community, which is the specific one for that setting. Of course, it has common elements with the other two considered settings (hospitals and residential aged care facilities), but more specific information and recommendations for the community. So, this document 'aims to reduce the number of falls and the harm from falls experienced by older people in the community'[114]. All guideline recommendations cover all people aged 65 or older and younger people at increased fall risk, and they have been developed for all those involved in the care of older people[114].

The guideline covers several parts of the screening, assessment, and intervention for preventing falls in community older adults. We will now consider the two following chapters and relative sections: 4. Falls prevention interventions, and 5. Fall risk screening and assessment (Table 6).

Table 6. Recommendations for preventing falls in older people (community care) by the Australian Commission on Safety and Quality in HealthCare (ACSQHC)

Part B - Standard falls prevention strategies

Chapter 4 Falls prevention interventions

Recommendations

Intervention

- Use effective interventions to reduce falls in the community, for example certain exercise programs, assessment followed by multifactorial treatment, home safety interventions in high-risk groups, and academic detailing for general practitioners by a pharmacist. (Level I)

Single interventions

- Older people should be encouraged to exercise to prevent falls. Certain programs have been shown to be effective and largely focus on balance training. (Level I)
- Older people with visual impairment primarily related to cataracts should undergo cataract surgery as soon as practicable. (Level II)
- When conducted as a single intervention, home environment interventions are effective for reducing falls in high-risk older people. (Level I)
- For individual older people, gradual and supervised withdrawal of psychoactive medications should be considered to prevent falls. (Level II)
- People with severe visual impairment should receive a home safety assessment and modification program specifically designed to prevent falls. (Level II)
- Use cardiac pacing in older people who live in the community and who have carotid sinus hypersensitivity and a history of syncope or falls to reduce the rate of falls. (Level II)
- Collaborative review and modification of medication by general practitioners and pharmacists, in conjunction with individual patients, is recommended to prevent falls. (Level II)
- Vitamin D and calcium supplementation should be recommended as an intervention strategy to prevent falls in older people who live in the community, particularly if they are not exposed to the minimum recommended levels of sunlight. Benefits from supplementation are most likely to be seen in people who have vitamin D insufficiency (25(OH)D <50 nmol/L) or deficiency (25(OH)D <25 nmol/L). (Level I)7 (Level I-*)

Multiple interventions

- The combination of exercise targeting strength and balance, education, and home safety intervention (the Stepping On Program) is recommended to reduce the rate of falls in older people who live in the community. (Level I)

Multifactorial interventions

- In older people at risk of falls, individualised assessment leading directly to tailored interventions is recommended. (Level I)

Good practice points

- The general practitioner can ‘prescribe’ verbal or written instructions for falls prevention interventions (e.g., exercise programs) for the older person to improve or maintain independence, and encourage adherence.
- Managing many of the risk factors for falls (e.g., balance problems, medication) will have wider benefits beyond falls prevention.

Part C Management strategies for common falls risk factors

Chapter 5 Falls risk screening and assessment

Recommendations

Screening and assessment

- Older people should be asked about falls at least once every year by their general practitioner or other healthcare providers.
- Older people with a history of one or more falls in the past year should be assessed using a simple, validated balance test or falls risk screen.
- Older people who perform poorly on a simple test of balance or gait, or on a falls risk screening tool, should undergo a detailed assessment to identify contributory risk factors.
- Falls risk screening and assessment tools used should be evidence-based (meaning that they have demonstrated good predictive accuracy and have been evaluated in the relevant setting in more than one site).
- Falls prevention interventions may need to be modified to make sure they are suitable for the individual, and often the carer(s) and family members will also play important roles in implementing fall prevention actions.

Good practice points

Falls risk screening

- Falls risk screening should be used to guide more detailed assessment and intervention, and the outcomes of the screen should be documented and discussed with the older person and their carer(s).
- When the threshold score of a screening tool is exceeded, a falls risk assessment should be conducted as soon as practicable. If the score is not exceeded, standard falls prevention strategies apply.

Falls risk assessment

- To develop an individualised plan for preventing falls, healthcare professionals need to identify systematically and comprehensively the factors contributing to the older person's increased fall risk.
- Interventions delivered as a result of the assessment provide benefit rather than the assessment itself; therefore, it is essential that interventions systematically address the risk factors identified.
- Identifying the presence of cognitive impairment should form part of the falls risk assessment process.

NOTES: in the table, recommendations from ACSQHC guideline (community care) taken from the two chapters in bold are reported. For several recommendations, the level of evidence (I, II, III-1, III-2, III-3, IV) based on the National Health and Medical Research Council levels of evidence are reported too[116].

Regarding the fall prevention interventions chapter (Table 6, first chapter), this guideline is in line with those by ABS/BGS[43] and by NICE[113]. It reports similar indications regarding the types of interventions recommended to reduce falls in older people. Besides, differently from the others, it highlights the collaborative role of general practitioners and pharmacists in the modification of medication and that of the

firsts in the prescription of instructions for falls prevention interventions to improve or maintain independence and encourage adherence[114].

Then, even concerning the fall risk screening and assessment chapter (Table 6, second chapter), the ACSQHC guideline reports similar recommendations to ABS/BGS[43] and NICE[113]. In particular, it indicates that healthcare providers should ask about falls to older people at least once every year. In the case of one or more falls, the fall risk screening is required, and in the case of poor performance on the screening, a detailed multifactorial assessment is provided. Besides, the ACSQHC guideline confirms the need for a tailored and fall risk-oriented intervention for those who required it, and it highlights that cognitive impairment should be a part of the fall risk assessment process[114].

1.3.1.4 Italian Health Ministry guideline

In 2007, with the following update in 2009, the Italian Health Ministry, together with the Istituto Superiore di Sanità, in the Programma Nazionale delle Linee Guida framework, drafted a guideline 'Prevenzione delle cadute da incidente domestico negli anziani'[12]. The aim was to summarize the current scientific knowledge on the argument, obtaining an evidence-based guideline, and to outline information and intervention policies aimed at preventing as much as possible the phenomenon, also considering the continuous increase in longevity of the Italian population. Every year in Italy, 3-4 million domestic accidents happen, principally in older people with consequent disability, recoveries, and mortality. Falls represent the most critical voice between these accidents, with extraordinary human, social, and material costs[12]. The document took as its starting point the already cited guideline published by the National Institute for Clinical Excellence (NICE)[113]. The recommendations address people aged 65 or older who live in the community, or resident in long-term or extended care facilities at fall risk or already fallen, or older people who, following a fall, need basic or secondary assistance, but that maintain totally or partially their self-sufficiency[12].

The guideline covers several parts of the screening, assessment, and intervention for preventing falls in community older adults, in particular: 1) fall risk factors and the useful interventions to identify persons at risk; 2) instruments to measure the fear of falling; 3) efficient tools to identify the modifiable fall risk factors; 4) safe interventions and strategies to prevent the fall risk[12]. We will focus on the first and fourth parts.

Table 7. Recommendations for the preventions of falls from domestic accidents in older adults by the Italian Ministero della Salute[12]

1) Fall risk factors and useful interventions to identify persons at risk

History of previous falls

- Considered the high level of falls relapse, it is recommended to periodically interview older adults to understand if they had fallen in the last year and to eventually know the frequency, the characteristics, and the context (level III/A).

The role of drugs

- Interrogate older people on the taken therapies and keep under strict control the administered drugs (level III/A).
- Review the pharmacological prescriptions periodically, with particular attention to the intake of drugs such as benzodiazepines, antidepressants o to the simultaneous intake of 3 drugs. Where possible, review those prescriptions to reduce exposure to risk (level III/A).

Motor disorders

- Keep under control balance and gait problems in older adults who refer to a previous fall or who are considered at risk. The tests used to assess the fall risk should be simple to execute, of short duration, repeatable, to allow a follow-up (level III/A).

Vision impairment

- Medical doctors, who are responsible for the clinical handling of the patient, have to carry out or have carried out an accurate eye examination, to verify eventual problems or alterations (level III/A).

Domestic risks

- Following a fall, it is recommended a domestic risk analysis in order to verify the presence of modifiable obstacles or dangers (level III/A).

Social isolation, financial difficulties

- In the event of serious situations of social hardship, support involving different services is also recommended for prevention or recurrence of falls (level III/A).

Cognitive deterioration

- There is not enough evidence to consider the cognitive deterioration as a fall risk factor in older adults. There are no certain elements to recommend or discourage cognitive examination to prevent falls (level III/C).

Urinary incontinence

- There is not enough evidence to consider urinary incontinence as a fall risk factor in older adults. There are no certain elements to recommend or discourage urinary incontinence to prevent falls (level III/C).

Fear of falling

- Interrogate older adults to understand if they feel fear of falling; in that event, the level and the

reason of this fear should be assessed by a health professional (level III/A).

4) Safe interventions and strategies to prevent the fall risk

Multifactorial, multidisciplinary, personalized interventions

- A multidisciplinary, multidimensional approach to the management of elderly subjects, victims of traumatic falls that have been treated as part of basic care or acute assistance, has been described as an effective intervention package in two studies. Important components of the treatment must be the assessment of the patient's general health, interventions on risk factors, and a detailed discharge plan. The first objective to be set in taking care of the older adult is an accurate clinical evaluation. Attention should be paid to identifying and possibly to treat the pathologies or conditions that could favor the onset of falls (level I/A).
- Older persons who come to medical attention due to a fall or who report recurrent falls in the past year or have gait abnormalities and/or of balance should be subjected to a multifactorial risk assessment.
  The assessment should be carried out by one or more health professionals that have specific skills and experience in the field of falls. This assessment should be part of an individualized and multidimensional intervention that includes:
  - identification of a history of falls;
  - gait, balance, and possible muscle weakness assessment;
  - osteoporotic risk assessment;
  - evaluation of the functional ability subjectively perceived by the elderly and fear related to falling;
  - vision impairment assessment;
  - cognitive deterioration assessment and neurological examination;
  - domestic risk assessment;
  - cardiovascular examination and possible revision of pharmacological treatment.

  Also, if present, arthritis, diabetes, dementia, vestibular system disorders, and cognitive deficits should be considered (level I/A).

Physical exercise to improve strength and balance

- Personalized intervention programs to improve strength, walking, and balance are recommended (level I/A).

Domestic risk and safety measures

- Give information and educational intervention to enhance risk awareness (level II/A).
- Control of the environmental situation and of the present dangers, through the execution of house visits, should be offered to people who are discharged from hospital or emergency room due to a fall (level III/A).
- Train health and social workers (e.g., prevention department and health district workers) to the acquisition of basic skills for the safety verification of the domestic environment and the related social-welfare aspects (level III/B).
- Advise the elderly on the installation of devices (fire extinguishing lights, anti-slip stripes, handles, etc.) that can make safer the home environment (level I/A).

NOTES: in the table, recommendations from Italian Ministero della Salute (community care)

taken from the two parts in bold are reported. For each recommendation, the level of evidence (I, II, III, IV, V, IV) and the force of recommendations (A, B, C, D, E) based on the guideline classification are also reported.

Concerning the fall risk factors and useful interventions to identify persons at risk (Table 7, part 1), Italian recommendations are in line with all the previously cited guidelines. They highlighted the assessment of the same fall risk factors in older people already reported previously, in particular the periodical interrogation on falls of the previous year, the review of pharmacological treatment, the balance and walking assessment in older adults who refer a previous fall or who are considered at risk, the vision evaluation, the domestic risk assessment and that of the fear of falling. Attention is also given to the evaluation of eventual social isolation and financial difficulties[12].

Then, regarding the safe interventions and strategies to prevent the fall risk (Table 7, part 4), once again in agreement with the previous guidelines, this one pointed out the indication for a multidisciplinary multidimensional approach to the management of the elderly who already experienced a traumatic fall, paying particular attention to identify and, possibly, to treat the pathologies or conditions that could favor the recurrence of fall. For those who had a fall or recurrent fallers or demonstrated balance and/or gait abnormalities, it was still recommended a multifactorial risk assessment, realized by health professionals with specific skills and experience in the field of falls. Finally, physical exercise to improve balance and strength and the domestic risk assessment with suggested modifications to improve the home environment safeness are again reported[12].

1.3.2 Fall risk screening tools

To intervene promptly and effectively in reducing the fall risk, it is essential to identify and quantify this risk in the population of interest. It is generally detected through the use of clinical screening tests, which have, as their primary objective, to distinguish between high and low fall risk, thus requiring high sensitivity and specificity[117].

Even regarding the fall risk screening instruments, the literature offers a wide availability of tools, which differ in content, spread, properties, and setting. Three systematic reviews performed in the last twenty years, together with the presentation of a further instrument, will be presented in details:

1. Perell, 2001[118];
2. Gates, 2008[119];
3. Park, 2018[117]
4. FRAT-Up (Fall-Risk Assessment Tool), 2016[120].

1.3.2.1 Fall Risk Assessment Measures: An Analytic Review (Perell, 2001)

In 2001 Perell and colleagues performed a systematic review of the literature[118] to summarize information regarding existing fall assessment scales to help clinicians make more informed choices given their patients. In fact, they claimed that clinicians are not often completely aware of the numerous available fall risk assessment tools, and they tend to develop their own assessment scales, sometimes lacking adequate psychometric properties. The authors' search yielded 21 articles from 1984 to 2001 that described 20 fall risk assessment scales, reviewed independently by five reviewers using a standardized review form, and guaranteeing a high interrater reliability[118] (Table 8).

Table 8. Reviewed articles of the three considered systematic review by Perell 2001[118], Gates 2008[119] and Park 2018[117]

Systematic review (Perell, 2001)[118]	Systematic review (Gates, 2008)[119]	Systematic review and meta-analysis (Park, 2018)[117]
Berg Balance Scale (by Berg)	Berg Balance Scale (by Bogle Thorbahn)	Berg Balance Scale (by Zur, Santos, Muir, Lajoie, and Bogle Thorbahn)
Timed Up&Go (by Shumway-Cook)	Timed Up&Go (by Morris, Okumiya, and Trueblood)	Timed Up&Go (by Martinez, Alexandre, Möller, Wrisley, and Whitney)
Dynamic Gait Index (by Whitney)	-	Dynamic Gait Index (by Wrisley)
-	Downton Fall Risk Index (by Rosendahl)	Downton Fall Risk Index (by Rosendahl and Möller)
Elderly Fall Screening Test (by Cwikel)	Elderly Fall Screening Test (by Cwikel)	-
-	Fall Risk Assessment Tool (by Nandy)	Falls Risk Assessment Tool (by Nandy)
-	Mobility Interaction Fall Chart (by Lundin-Olsson)	Mobility Interaction Fall Chart (by Ivziku and Lundin-Olsson)
STRATIFY (by Oliver)	-	STRATIFY (by Oliver, Webster, and Papaioannou)
-	Timed Walk/Distance Walked (by Murphy and Verghese)	Timed Gait (Walk) (by Verghese)
-	Tinetti Balance Scale (by Faber, Murphy, Tinetti, Trueblood, and Verghese)	Tinetti Balance scale (by Verghese and Raîche)
Tinetti Performance Oriented Mobility (by Tinetti)	Tinetti Performance Oriented Mobility (by Faber, Hale, Raiche, and Tinetti)	-
-	Walking While Talking Complex (by Verghese)	Walking While Talking Complex (by Verghese)
-	Walking While Talking Simple (by Verghese)	Walking While Talking Simple (by Verghese)
Assessment for High Risk to Fall	5-Step-Test	Activities-specific Balance Confidence scale
Fall Assessment Questionnaire	9-Test Battery	Demura's Fall Risk Assessment chart
Fall Prediction Index	Coalition for Community Fall Prevention Screen	Entry Fall Status Criterion
Fall Risk Assessment Tool (by MacAvoy)	Dynamic Posturography	FROP-Com tool
Fall Risk Assessment Tool (by Schmid)	Fall Risk Assessment (FRA) (by Flemming)	Fullerton Advanced Balance Scale
Fife	Floor Transfer	Functional Gait Assessment
Hendrich Fall Risk Model	Functional Reach	Hendrich II Fall Risk Model (by Zhang, Caldevilla, and Ivziku)
High Risk for Falls Assessment	Getting Up From Lying on Floor	LASA fall risk profile
Modified Gait Abnormality Rating scale	75% Limits of Stability	Modified Johns Hopkins-fall risk assessment tool
Morse Fall scale (by Morse and McCollum)	Mobility Screen	Performance-based fall risk assessment tool
Patient Fall Questionnaire	Modified Clinical Test of Sensory Interaction on Balance	Posturographic Fall Risk Index
Reassessment Is Safe Kare	One-Leg Balance	Risk Model for Recurrent Falls
Resident Assessment Instrument	Stops Walking When Talking	Test battery
Royal Melbourne Hospital Risk Assessment Tool	Tandem Stance	Thai Falls Risk Assessment Test
	Tinetti Gait Scale (by Faber, Tinetti, and Trueblood)	TIMG fall risk assessment chart
		Zur Balance scale

NOTES: in green, the two instruments reported by all three systematic reviews, and in yellow those reported by at least two of the systematic reviews; the following are the other tools reported by each study.

Among the 20 described papers, 14 were nursing assessment tools developed within the hospital or nursing home setting, and 6 functional assessment scales. The majority were addressed to older people over 60 years[118].

The following intrinsic risk factors appeared most often in the 14 nursing assessment tools within this review, in particular mental status (13 tools), mobility (10 tools), history of previous fall (10 tools), secondary or specific diagnoses (8 tools), incontinence or toileting issues (8 tools), medications (7 tools), and sensory deficits (e.g., vision, hearing, sensation) (7 tools) [118].

Data on interrater reliability was provided for 7 nursing (from 79% to 100%) and 4 functional assessment tools (from 58% to 98%)[118].

Information on predictive validity was included for 7 nursing and 5 functional assessment studies, identifying a threshold or cutoff score to define people at high fall risk for 10 nursing and 5 functional assessments. Regarding sensitivity and specificity, 5 functional and 8 nursing assessments reported them. Sensitivity varied from 43% (Fall Risk Assessment tool by MacAvoy) to 100% (Fall Prediction Index by Nyberg). Specificity varied from 38% (Dynamic Gait Index by Whitney) to 88% (STRATIFY by Oliver and Fall Assessment Questionnaire by Rapport) across all assessment tools combined (Table 8)[118].

Finally, they concluded by recommending the following criteria to choose the most appropriate assessment tool for a specific setting: high sensitivity, specificity, and interrater reliability; similarity of the patient population to ones in which the instrument was developed or studied; written procedures explicitly outlining appropriate use of the form; reasonable time required to administer the scale; and established thresholds identifying when to initiate interventions[118]. Besides, they gave indications about tools according to the setting: nursing assessment scales (in particular STRATIFY by Oliver and Fall Risk Assessment tool by Schmid) in acute care setting; functional assessment tools (in particular Elderly Fall Risk Screening Test by Cwikel and the Timed Up&Go by Shumway-Cook) in the outpatient setting; implementation of an overall fall prevention program in extended care settings (e.g., nursing homes and rehabilitation units)[118].

1.3.2.2 Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults (Gates, 2008)

In 2008 Gates and colleagues realized a systematic review of the literature[119] to assess and summarize the evidence for the accuracy of screening tests at predicting fallers in community-dwelling populations and indicate where more research was needed. Indeed, they noticed that many fall risk screening tools had been introduced into clinical practice in recent years as components of clinics intended to reduce falls in community-living older people, but not based on evidence-based usefulness in discriminating between fallers and non-fallers[119]. At the end of the systematic search, 25 studies were eligible and included in the review, reporting results for 29 different screening tests (Table 8). Data were extracted by two reviewers independently, with a third reviewer that resolved discrepancies. Most tests were assessed by only one included study, the main exceptions being the Tinetti Gait, Balance, and Mobility scales (8 studies) and Timed Up&Go test (4 studies) (Table 8). Where several studies evaluated the same test, differences in the conduction of the test, scoring, cutoff points, or outcome measures meant that the results could not be compared or combined in meta-analyses[119].

Screening tests had, in general, higher specificity (at least 80% for 22 tools) than sensitivity (at least 80% for 8 tools), indicating that a higher proportion of non-fallers than fallers were correctly identified. Only two tests, the 5-Step test and the Mobility Interaction Fall Chart reported sensitivity and specificity, both exceeding 80%[119].

Besides, the authors highlighted that some evidence exists that simple screening questions may perform as well as more complex screening tests in predicting who will fall, like a history of falls and reported abnormalities of gait or balance which are consistently found to be the best predictors of future falls[119].

Finally, they claimed that, at present, recommending any screening test for routine clinical use is not possible given the lack of evidence that any of them is useful to identify fallers and the presence of heterogeneity between studies that do not allow to perform a quantitative synthesis (meta-analysis). So, for the use in clinical practice, future studies with a sufficiently large sample size to estimate sensitivity and specificity with high precision, conducted in a clinically relevant population, with a sufficient duration of follow-up, and with reliable methods of the recording of falls are needed[119].

1.3.2.3 Tools for assessing fall risk in the elderly: a systematic review and meta-analysis (Park, 2018)

In 2018 Park and colleagues did a systematic review with meta-analysis[117] of the literature to compare the diagnostic accuracy of several currently available fall risk assessment tools developed for the elderly, to identify the assessment tools most frequently used to discriminate fallers and non-fallers and those having the highest predictive validity, and to provide scientific evidence for selecting the best tool to use in practice[117]. They pointed out that some studies have attempted to provide an overview of which assessment tool has the highest validity for the elderly at fall risk. However, they presented results based on a specific tool or different setting or did not perform a quantitative analysis of the predictive validity, making it challenging to conclude which tools are really effective for assessing the fall risk in the elderly population[117].

They kept thirty-three articles regarding people over 60 years, after two reviewers' selection process independently with a third reviewer that resolved discrepancies. Article quality was assessed by two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2). To obtain a quantitative synthesis, they performed a diagnostic meta-analysis with a random-effects model to reflect inter-study heterogeneity of pooled sensitivity, specificity, positive- and negative-likelihood ratios, and diagnostic odds ratios. The area under the curve (AUC) and index Q* were used to measure diagnostic accuracy, whereas inter-study heterogeneity was assessed with Higgins's I^2 homogeneity test (I^2 test) [117].

Twenty-six assessment tools for fall risk were used in the selected articles (Table 8), and they tended to vary based on the setting[117]. In the acute care setting, where typically the risk is assessed by medical staff (e.g., nurses), two tools were used, the Hendrich II Fall Risk Model and the STRATIFY, making decisions on fall risk involving items such as conscious state, urinary function, and drug-taking. Differently, in a community-dwelling setting, 23 fall risk instruments were found, and the only 4 of them used in two or more studies were Berg Balance Scale (BBS), Timed Up&Go (TUG) test, Downton Fall Risk Index (DFRI), and Tinetti Balance scale. In the latter case, except in the DFRI, which is similar to the acute care setting tools, the fall risk is assessed directly measuring balancing ability in everyday activities, such as walking up and downstairs and mobility and the speed[117].

Regarding diagnostic accuracy, they highlighted that all fall risk assessment tools used for the elderly, except for the Mobility Interaction Fall Chart and the Tinetti Balance scale, showed a sensitivity of ≥0.7 and low or no inter-study heterogeneity. In most assessment tools except for BBS, specificity was lower than sensitivity (i.e., under 0.6), and inter-study heterogeneity was high (i.e., over 90%). Therefore, they concluded that the predictive validity of the fall risk assessment tools currently used for the elderly is not sufficient, despite using a large variety of fall risk assessment tools in this population[117].

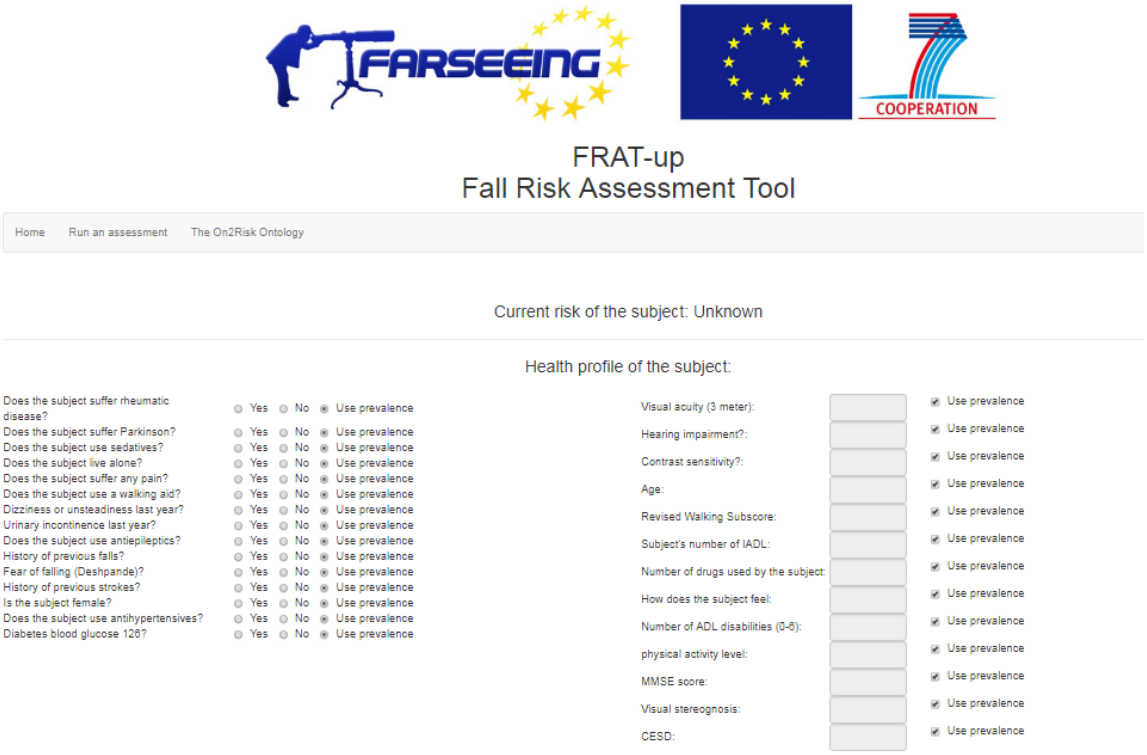
Finally, they suggested the combined use of two assessment tools, one with a stable sensitivity and the other with a stable specificity (e.g., the combined use of BBS and TUG in a community-dwelling setting), to obtain an adequate predictive validity, maximizing the characteristics and the predictability of each test[117].

1.3.2.4 FRAT-up (Fall-Risk Assessment Tool)

A further fall risk screening test is available in the literature, called FRAT-up (Fall-Risk Assessment Tool)[121]. It is a web-based free tool (http://ffrat.farseeingresearch.eu/), which was developed to provide the fall risk assessment of subjects aged 65 or up and living in a community-dwelling (Figure 19). Indeed, Cattelani and colleagues reviewed the actual literature situation extensively, appreciating that they have been insufficient for predicting falls despite numerous existing assessment tools for fall risk. Moreover, several reviews and meta-analyses of the statistical correlation between the exposure to risk factors and the fall risk in terms of odds ratios are available, thus providing a solid scientific base about fall risk factors[121].

This instrument, which is an example of how applied computer technology can serve the clinical decision, has been generated in the context of the European project ‘Farseeing’, coordinated by Prof. Lorenzo Chiari of the Department of Ingegneria dell'Energia Elettrica e dell'Informazione ‘Guglielmo Marconi’ of the University of Bologna. For the tool's discriminative performance and calibration, they used the InCHIANTI dataset, where 1453 persons have been initially enrolled (1150 subjects aged 65 or more) and have undergone four consecutive visits globally covering a 9-year follow-up from 1998[121].

Figure 19. The web-based Fall-risk Assessment Tool (FRAT-up)[121]



NOTES: in the figure, a preview of the web-based FRAT-up, available at http://ffrat.farseeingresearch.eu/.

The authors tried to represent information in a structured way, using formal approaches from computer science, like ontologies and the semantic Web, and basing on the hypothesis that FRAT-up considers the fall risk directly related to the subject's exposure to known risk factors[121]. They followed three steps:

1. Starting from the literature, they realized an ontological classification of fall risk factors (by kind, reversibility, and setting), based on a well-established meta-analysis on known risk factors for falls in community-dwelling older people by Deandrea et al. [41], which also provided the odds ratio for each risk factor.
2. They extracted probabilities from odds ratios using a few mathematical steps. The probability contribution is at the base of the FRAT-up tool.
3. From a data structure containing probabilities representation, they compiled Logic Programming with Annotated Disjunctions (LPAD) rules, which allow representing the contribution of each risk factor in terms of probabilistic rules and probabilistic reasoning[121].

According to the authors, this instrument is addressed to two different health professionals: 1) general practitioners (GPs) delivering primary care provisions, with no specific knowledge about falls, who need an assessment tool to evaluate patients' fall risk and possible early interventions, and 2) professionals involved in fall prevention and treatment, who need a tool to assess the fall risk in a reliable and quantitative manner[121].

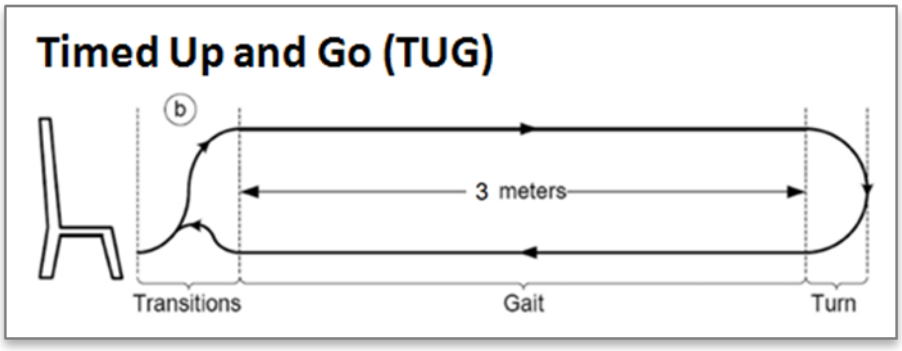
Cattelani and colleagues studied the diagnostic accuracy of the FRAT-up in several studies. Regarding the calibration paper[121], on the InCHIANTI dataset, the tool showed an Area Under Curve (AUC) value of 0.642 (95% Confidence Interval 0.614-0.669), classifiable as sufficient diagnostic accuracy[122], and at the same time, a miscalibration (Hosmer-Lemeshow test P-value <.001) due to overestimation of the fall risk that, however, they considered of less concern. A further study[123], which used a subsample of the InCHIANTI dataset, demonstrated a sufficient[122] AUC of 0.638 (95% Confidence Interval 0.610–0.666) in predicting single falls and a good[122] AUC of 0.713 (0.675–0.752) for multiple falls. These values were higher than those provided by the number of previous falls, the gait speed, and the Short Physical Performance Battery score. Finally, in a third work by the same authors[120], based on 4 European cohorts of older people data, the FRAT-up predicted sufficiently[122] the fall risk with a mean AUC of 0.646 (95% CI 0.584–0.708).

1.3.3 Inaccuracy of the actual screening tests

To summarize the diagnostic accuracy conclusions of the three systematic reviews[117-119] and the works on the further tool FRAT-up[120, 121, 123] in the prediction of the fall risk in older adults, we can observe that:

- Two instruments have been reported by all three systematic reviews (in green, Table 8) as between the most popular in the literature, such as the Berg Balance Scale and the Timed Up&Go (graphical representation in Figure 20). In all these secondary studies, the BBS showed higher specificity (0.86[118]; 0.92[119]; 0.90[117]) compared to sensitivity (0.77[118]; 0.53[119]; 0.73[117]). Regarding, instead, the TUG, the previous works highlighted a higher specificity (0.77[118]; >0.73[119]) in two of them and higher sensitivity (0.76[117]) in the remaining.

Figure 20. The Timed Up&Go test



NOTES: in the figure one among the most known clinical fall risk screening tests used in geriatric and rehabilitation is represented, in which the operator times the subject in getting up from a chair, walking for three meters, backing-to-front, walking another three meters, and sitting down. The time taken to perform this test defines the subject's fall risk based on the literature cutoffs.

- Other several tools, also well known, have been reported by at least two of the systematic reviews, like Dynamic Gait Index, Downton Fall Risk Index, Elderly Fall Screening Test, Fall Risk Assessment Tool (by Nandy), Mobility Interaction Fall Chart (MIFC), STRATIFY, Timed Walk, Tinetti Balance Scale, Tinetti Performance Oriented Mobility, Walking While Talking Simple and Complex (in yellow, Table 8).

- Globally the selected screening tests presented higher values of sensitivity (0.43-1.00[118]; ≥0.70[117]) in two of the systematic reviews, but the contrary in the remaining (specificity ≥0.80[119]).

- The three systematic reviews drew different conclusions. The first of them recommended, for the outpatient setting, the Best Elderly Fall Screening test and the TUG, as they have shown sensitivity and specificity higher than the median value[118]. The second highlighted that tests with the best diagnostic accuracy had been the 5-step-tests and the MIFC, with both values higher than 0.80[119]. However, they concluded that, at present, it was not possible to recommend any screening test for routine clinical use because there was no strong evidence that any of them was useful for identifying fallers. The last presented review stated, like the latter, that the predictive validity of the fall risk assessment tools currently used for the elderly was insufficient and suggested the combined use

of two tools with complementary properties in terms of sensitivity and specificity (e.g., BBS and TUG).

- A further screening test, the FRAT-up, which based its fall risk prediction on the presence of evidence-based risk factors, may be an option to improve the diagnostic accuracy of this phenomenon with its sufficient/good AUC values (all fallers 0.642[121], 0.638[123], 0.646[120]; recurrent fallers 0.713[123]).

From all these findings, the global conclusion is that, despite the presence and the use of a large variety of fall risk assessment tools in the elderly, at now, it was not possible to predict older adult fallers with optimal accuracy. This evidence is in line with previous studies, in which the most reliable criterion for identifying those at fall risk, and for this reason used in most randomized control trials, is the history of a previous fall. Deandrea et al. showed this. in 2010[41]: between the most recognized fall risk factors, the best predictor of fall risk, both for all fallers and recurrent fallers, is right the history of a previous fall (odds ratio 2.8 and 3.5). A similar approach prevents the primary prevention of the fall events, excluding from the intervention all the subjects that, because of a not diagnosed balance impairment associated with other fall risk factors, are at high risk even if they have not yet had the opportunity to fall.

A solution suggested by Park et al. in their systematic review[117] is to carry out multiple fall risk assessments using more than one screening test with different characteristics, to improve the accuracy of the diagnostic algorithm. Since the aim of a fall risk test is to discriminate between high-risk persons and low-risk persons, a method which combines high sensitivity and specificity should be chosen to reduce both false positives (patients 'positive to the test' who are not at fall risk) and false negatives (patients 'negative to the test' who are at high risk). The combined use of two tools, for example, TUG and BBS, the first having a relatively high sensitivity (0.76) and the second a very high specificity (0.90), could be useful to increase the diagnostic accuracy of the fall risk[117].

1.4 The World Health Organization International Classifications

1.4.1 The WHO Family of International Classifications (WHO-FIC)

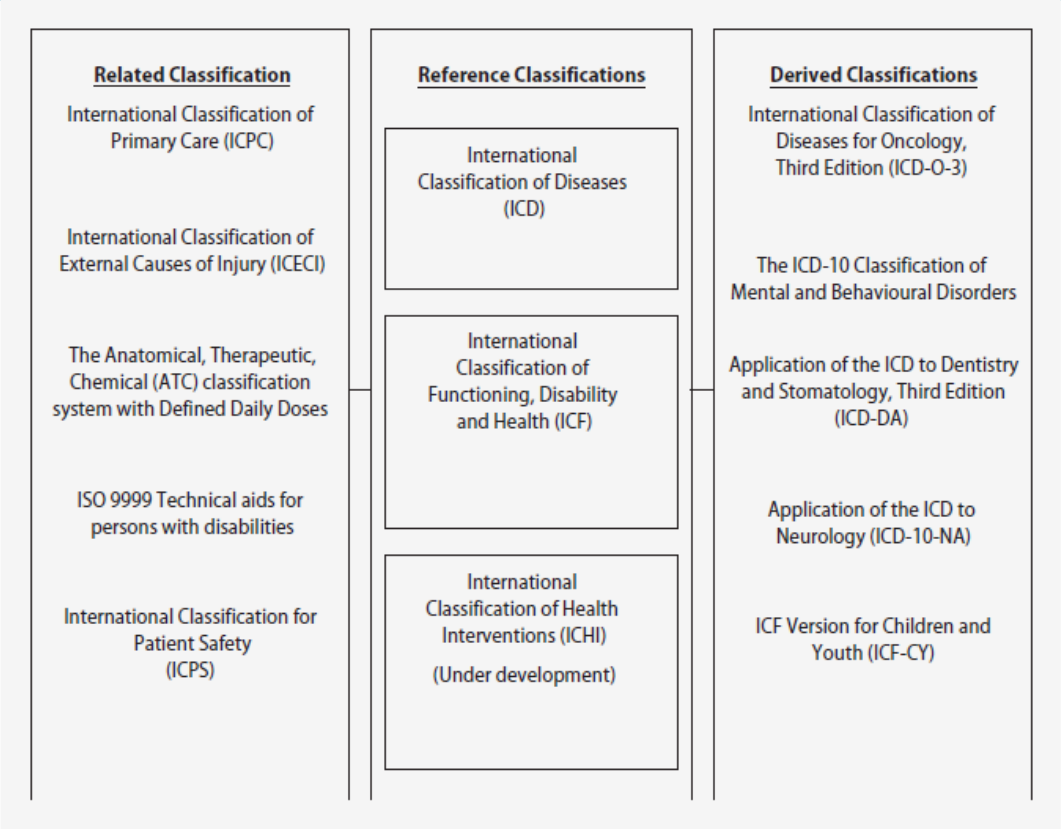
The WHO Family of International Classifications (WHO-FIC) is a suite of classification products that may be used in an 'integrated fashion to compile and compare health information nationally and internationally'[124]. This set of classifications aims to provide sound health information to support the decision-making at all levels and a conceptual framework of information domains for which classifications are, or are likely to be, required for purposes related to health and health services. Joint use of these classifications renders better health information, identifying associations between diseases, disability, and interventions[124].

The WHO maintains WHO-FIC through an international Network, which consists of an international range of collaborating centers and experts. Decisions reflect a worldwide consensus aimed at consistent, high-quality health data for both national and international purposes. To achieve its purpose, members of the WHO Family are based on sound scientific and taxonomic principles; the WHO-FIC must be culturally appropriate and internationally applicable. Most importantly, it has to address the different dimensions of deaths, diseases, disability, and health interventions in a joint framework[124].

The WHO-FIC is designed to provide a framework to code a wide range of health information (e.g., diagnosis, functioning, and disability, reasons for contact with health services) and uses a standardized common language, permitting comparison and communication about health and healthcare across the world in various disciplines and sciences. They also provide a valuable tool to describe and compare the health of populations in an international context. The information on mortality (provided by ICD) and on health outcomes (provided by ICF) may be combined in summary measures of population health to monitor the health of populations and its distribution and assess the contributions of different causes of mortality morbidity[124].

The WHO-FIC classifications and the broader United Nations family of economic and social classifications are of three major types (Figure 21): reference, derived, and related classifications[124].

Figure 21. Types of the WHO Family of International Classifications[124]



NOTES: in the figure, the three major types of the WHO-FIC are represented.

We will focus on the reference classifications, particularly on the International Classification of Diseases (ICD) and the International of Functioning, Disability, and Health (ICF).

The reference classifications cover the health system's main parameters, such as death, disease, functioning, disability, health, and health interventions. They are a product of international agreements, and they have achieved broad acceptance and official agreement for use. Also, they are approved and recommended as guidelines for international reporting on health. They may be used as models for the development or revision of other classifications, concerning both the structure, the character, and the definition of the categories. To this end, WHO has developed two reference classifications that can describe a person's health state at a particular point in time: the already cited ICD and ICF[124].

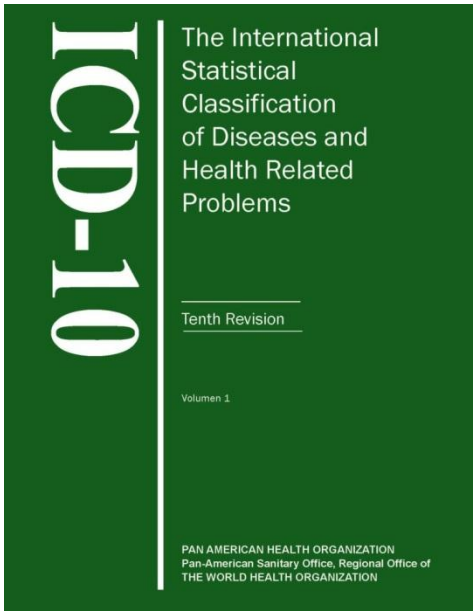
1.4.1.1 The International Classification of Diseases (ICD)

The International Classification of Diseases (ICD) is the standard diagnostic tool for epidemiology, health management, and clinical purposes. It includes the analysis of the general health situation of population groups[125], and it is the foundation for the identification of health trends and statistics globally and the international standard for reporting diseases and health conditions. It also constitutes the standard diagnostic classification for all clinical and research purposes. ICD defines the universe of diseases, disorders, injuries and other related health conditions, listed in a comprehensive, hierarchical fashion that allows for:

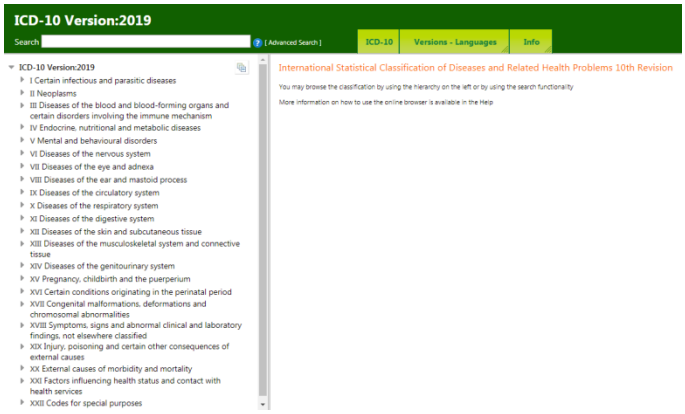
- Easy storage, retrieval, and analysis of health information for evidence-based decision-making;
- Sharing and comparing health information between hospitals, regions, settings, and countries;
- Data comparisons in the same location across different periods.

Uses include monitoring the incidence and prevalence of diseases, observing reimbursements and resource allocation trends, and keeping track of safety and quality guidelines. They also include the counting of deaths and diseases, injuries, symptoms, reasons for encounter, factors that influence health status, and external causes of disease[126, 127]. The ICD is used in systematic full mortality registration in more than 117 countries and translated into 43 languages[124, 127].

Figure 22. The International Classification of Diseases (10th version) – ICD-10 (on the left) and the English online version of the International Classification of Diseases (10th version) – ICD-10 (on the right)



NOTES: in the figure, the cover of the ICD-10 is presented.



NOTES: in the figure, the English online version of the ICD-10 is presented. It is available at https://icd.who.int/browse10/2019/en.

According to WHO, it is essential because it provides a common language for reporting and monitoring diseases, which can be used by physicians, nurses, other providers, researchers, health information managers and coders, health information technology workers, policy-makers, insurers, and patient organizations[128]. It is articulated in 3 volumes, which are all necessary to the correct coding of diseases[124, 127]:

- *Volume 1 – Analytical Classification:* it contains the tabular list of diseases, definitions, and nomenclature regulations.
- *Volume 2 – Instructions Manual:* it includes the manual with an extensive description of the classification and methods for use in mortality and morbidity.
- *Volume 3 – Alphabetical Index:* it includes separate indices for diseases, external causes, and drugs/substances.

ICD-10 was endorsed in May 1990 by the Forty-third World Health Assembly (Figure 22). It is cited in more than 20,000 scientific articles and is used by more than 100 countries worldwide. A version of ICD-11 was released on 18 June 2018 to allow the Member States to prepare for implementation, including translating ICD into their

national languages and starting reporting using ICD-11 on 1 January 2022[126, 127]. Italy still uses, at now, the International Classification of Diseases, 9th revision – Clinical Modification (ICD-9-CM)[129] for the codification of information in the Scheda di Dimissione Ospedaliera flow, but the version of the Italian translation and adaptation of ICD-10 (10th revision, 5th edition, 2016) is available.

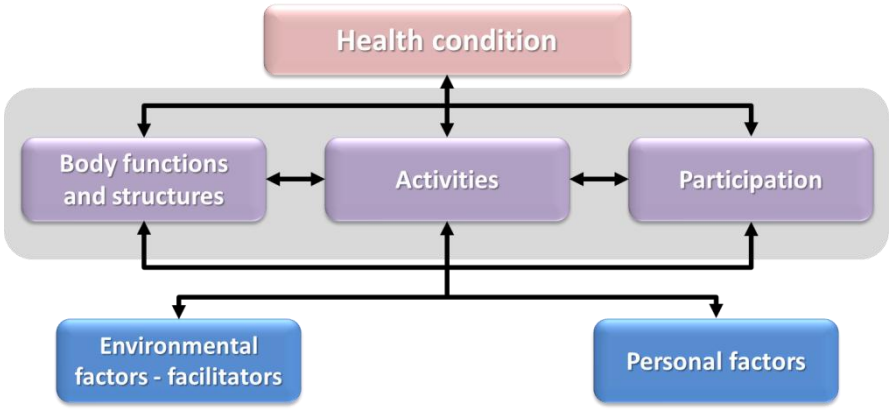
The classification can also be consulted online in English and French in the WHO website https://www.who.int/classifications/en/ (Figure 22), and in national languages through the relative national institutions (for Italy in the Portale Italiano delle Classificazioni Sanitarie https://www.reteclassificazioni.it/classificationbrowser/#/node_tree/ICD-10/2009).

1.4.1.2 The International Classification of Functioning, Disability, and Health (ICF)

1.4.1.2.1 ICF overview

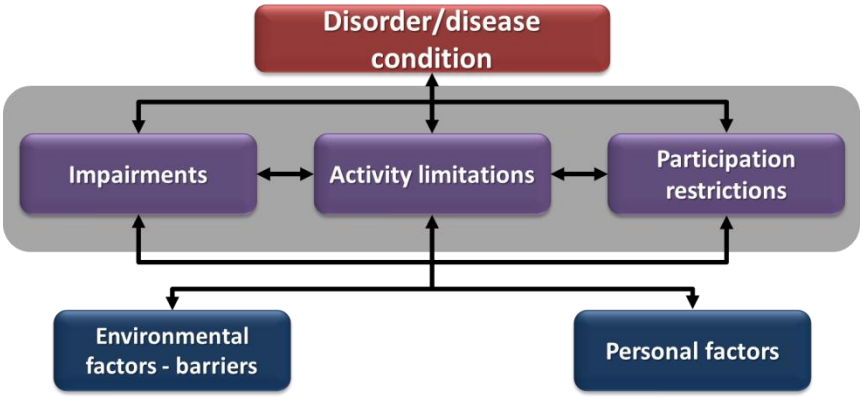
The International Classification of Functioning, Disability, and Health, known more commonly as ICF, is a classification of health and health-related domains based on a bio-psycho-social model. As the individual's functioning and disability occur in a context, ICF also includes a list of contextual (environmental and personal) factors (Figure 23 and Figure 24)[130].

Figure 23. The ICF bio-psycho-social model of health



NOTES: in the figure, the ICF bio-psycho-social model of health is represented. Three domains constitute it: body functions, body structures, and activity and participation. They interact with each other, with the health condition and with facilitator environmental factors and personal factors. The light grey area defines the *functioning*, an umbrella term encompassing body functions and structures, activities, and participation.

Figure 24. The ICF bio-psycho-social model of disability

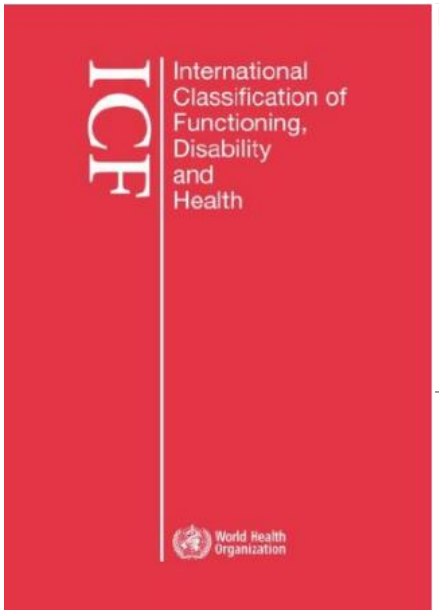


NOTES: in the figure, the ICF bio-psycho-social model of disability is represented. It is constituted by the alterations in one or more of the three domains: impairments (body functions and/or structures), activity limitation, and participation restriction. They interact with each other, with the disorder/disease condition and with barrier environmental factors and personal factors. The dark grey area defines the *disability*, an umbrella term encompassing impairments, activity limitations, and participation restrictions.

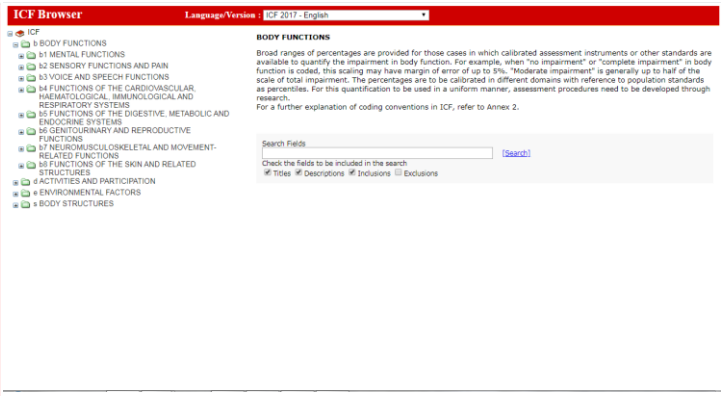
ICF replaced the International Classification of Impairments, Disabilities, and Handicaps (ICIDH), which was first published by the World Health Organization for trial purposes in 1980, after developing systematic field trials and international consultation over five years[124, 131]. The new classification has moved away from being 'consequences of disease' classification (1980 version) to become 'components of health' classification. 'Components of health' identifies the constituents of health, whereas 'consequences' focuses on the impacts of diseases or other health conditions that may follow as a result. Thus, ICF takes a neutral stand about etiology so that researchers can draw causal inferences using appropriate scientific methods[131].

ICF was officially endorsed by all 191 WHO Member States in the Fifty-fourth World Health Assembly on 22 May 2001 as the international standard to describe and measure health and disability (Figure 25)[130, 131]. ICF is translated into over 38 languages and can be accessed online in the 6 official languages at the WHO website https://www.who.int/classifications/en/ (Figure 25) (for Italy in the Portale Italiano delle Classificazioni Sanitarie https://www.reteclassificazioni.it/classificationbrowser/#/node_tree/ICF/2001).

Figure 25. The International Classification of Functioning, Disability, and Health (ICF) (on the left) and the English online version of the International Classification of Functioning, Disability, and Health (ICF) (on the right)



NOTES: in the figure, the cover of the ICF is presented.



NOTES: in the figure, the English online version of the ICF is presented. It is available at https://apps.who.int/classifications/icfbrowser/.

ICF is the WHO framework for measuring health and disability at both individual and population levels. The ICF's overall aim is to provide a unified and standard language and framework for the description of health and health-related states of an individual. It defines components of health and some health-related components of well-being (such as education and labor)[124, 131]. These domains are described from the perspective of the body, the individual, and the society in two basic lists:

1. Body Functions and Structures;
2. Activities and Participation.

As a classification, ICF systematically groups different person domains in a given health condition. *Functioning* is an umbrella term encompassing all body functions, activities, and participation (Figure 23); similarly, *disability* serves as an umbrella term for impairments, activity limitations, or participation restrictions (Figure 24)[124, 131].

ICF also lists contextual (environmental and personal) factors (Figure 23 and Figure 24), interacting with all these constructs. The basic construct of the environmental factors component is the *facilitating or hindering* impact of features of the physical,

social, and attitudinal world (Figure 23: facilitators; Figure 24: barriers). In this way, it enables the user to record useful profiles of individuals' functioning, disability, and health in various domains, e.g., pre-treatment and post-treatment assessment (e.g., for outcome measurement)[124, 131].

1.4.1.2.2 ICF aims and structure

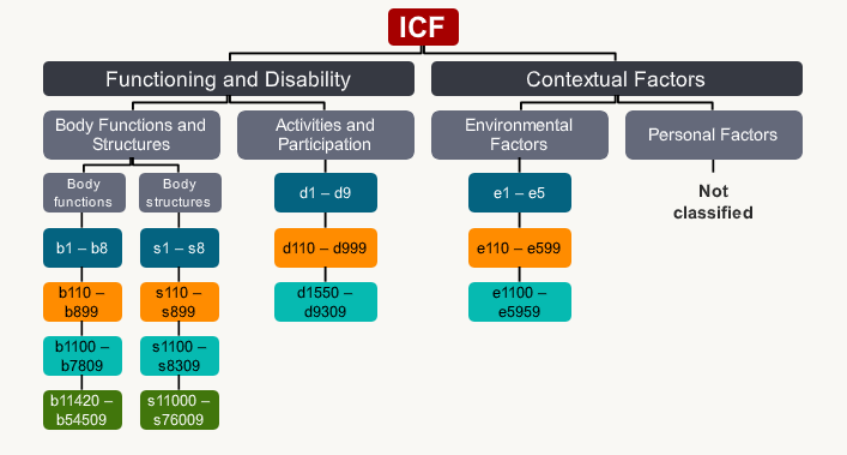
ICF is a multipurpose classification designed to serve various disciplines and different sectors. Its specific aims can be summarized as follows:

- To provide a scientific basis for understanding and studying health and health-related states, outcomes, and determinants.
- To establish a common language for describing health and health-related states to improve communication between different users, such as healthcare workers, researchers, policy-makers, and the public, including people with disabilities.
- To permit a comparison of data across countries, healthcare disciplines, services, and time; to provide a systematic coding scheme for health information systems.

These aims are interrelated since the need for, and the use of ICF requires constructing a meaningful and practical system that can be used by various consumers for health policy, quality assurance, and outcome evaluation in different cultures[131].

ICF describes situations concerning human functioning and its restrictions and serves as a framework to organize this information. It structures the information in a meaningful, interrelated, and easily accessible way (Figure 26).

Figure 26. The International Classification of Functioning, Disability and Health structure[132]



NOTES: in the figure, the ICF structure is represented (2 parts: Functioning and Disability, Contextual Factors; 4 components, 2 for each part: Body Functions and Structures, Activities and Participation; Environmental Factors and Personal Factors). For each component, the related categories indicated by a letter and followed by 1 to 5 numbers are the ICF units of classification.

ICF organizes information in two parts. Part 1 deals with *Functioning and Disability*, while Part 2 covers *Contextual Factors*. Each part has two components:

1. Components of Functioning and Disability

The *Body* component comprises two classifications, one for *functions* of body systems and one for body *structures*. The chapters in both classifications are organized according to the body systems. The *Activities and Participation* component covers the complete range of domains denoting aspects of functioning from both an individual and a societal perspective (Figure 26)[131].

2. Components of Contextual Factors

A list of *Environmental Factors* is the first component of Contextual Factors. Environmental factors impact all functioning and disability components and are organized in sequence from the individual's most immediate environment to the general environment. *Personal Factors* is also a component of Contextual Factors, but they are not classified in ICF because of the enormous social and cultural variance associated with them (Figure 26)[131].

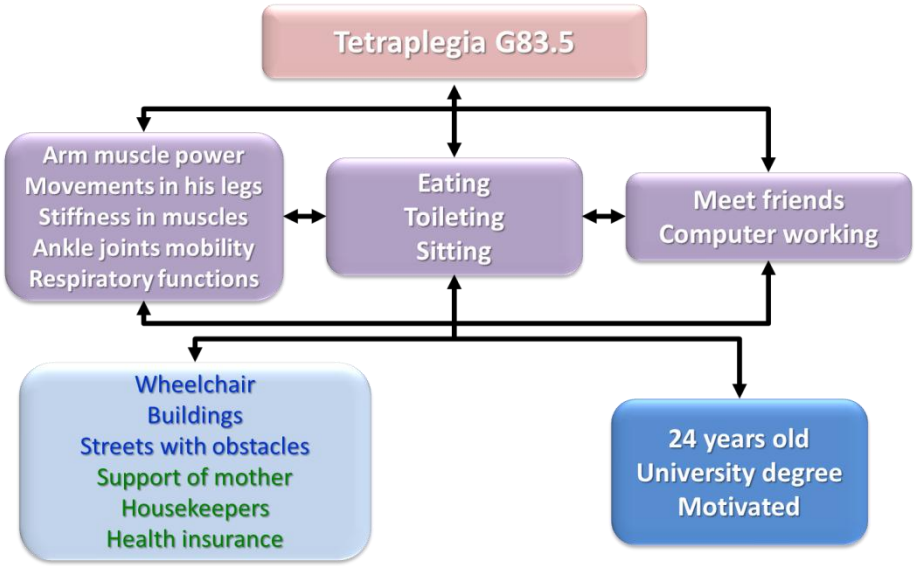
Therefore, the ICF unit of classification is the *category* within the health and health-related domains, indicated by a letter followed by 1 to 5 numbers (figure 23). It is important to note that, in ICF, persons are not the units of classification; that is, ICF

does not classify people but describes each person's situation within an array of health or health-related domains. Moreover, the description is always made within the context of environmental and personal factors[131].

1.4.1.3 The combined use of ICD and ICF

Joint use of the WHO classifications renders better health information, identifying associations between diseases, disability, and interventions[124]. Indeed, health conditions (diseases, disorders, injuries, etc.) are classified primarily in ICD, which provides an etiological framework, whereas functioning and disability associated with health conditions are classified in ICF. ICD and ICF are therefore complementary, and users are encouraged to utilize these two members of the WHO-FIC together. ICD provides a 'diagnosis' of diseases, disorders, or other health conditions, and this information is enriched by the additional information given by ICF on functioning (Figure 27).

Figure 27. Example of the combined use of ICD and ICF (modified by ICF e-learning tool, ICF Research Branch[132])



NOTES: in the figure, as an example of the combined use of ICD and ICF, health/disability information about Wun, a young boy with tetraplegia, is presented[132]. This is possible given the ICF bio-psycho-social model, which conceives health and disability on the same continuum.
Wun has a condition of tetraplegia due to a car accident (disease classify in ICD with code G83.5). Linked bi-directionally to this condition, his functioning (disability) is described in terms of:

- Body functions and structure (impairments): arm muscle power, movements in his legs, stiffness in muscles, ankle joints mobility, respiratory functions.
- Activities (activities limitations): eating, toileting, sitting.
- Participation (participation restrictions): meet friends, computer working.

The integrity/alteration of these components determines/is caused by his tetraplegia condition. Contextual factors influence this interaction:

- Environmental factors: wheelchair, buildings, and streets with obstacles as *barriers* (in blue); support of the mother, housekeepers, and health insurance as *facilitators* (in green).
- Personal factors: 24 years old, Engineering university degree, motivated.

The blue environmental factors act as barriers for their specific characteristics and worsen his functioning and tetraplegia condition. On the other hand, the green ones improve his condition. Also, personal factors influence other elements.

It is fundamental to state that two persons with the same disease can have different functioning levels, and two persons with the same level of functioning do not necessarily have the same health condition. Hence, joint use enhances data quality for medical purposes. Together, information on diagnosis plus functioning provides a broader and more meaningful picture of people or populations' health, which can then be used for decision-making purposes.

1.4.2 Lack of a common classification framework for fall risk factors

As we have already seen in section 1.1.4.3 Fall risk factors and their classification and subsections, several epidemiologic studies have investigated risk factors for falls in recent decades. However, investigators have not used consistent classifications, so that the lack of a common classification framework for fall risk factors is still present[41].

Previously we have already reported two of the major most widespread and accredited classifications of fall risk factors: the first by the Effective Health Care Bulletin[40] and the second by the World Health Organization[9].

A further general categorization of fall risk factors is between *intrinsic* and *extrinsic* factors. Intrinsic factors are individual-specific and include, for example, advanced age, chronic disease, muscle weakness, gait and balance disorders, and cognitive impairment. Extrinsic factors generally include external elements that can increase fall risk, like medication use, environmental hazards, and hazardous activities[41].

Lord et al.[133] proposed a more *analytic classification*, dividing risk factors into socio-demographic factors, balance and mobility factors, sensory and neuromuscular factors, psychologic factors, medical factors, medication use, and environmental factors.

Finally, an *ontological classification* is suggested by Catellani et al. in their validation paper of the FRAT-up tool[121]. They classified the fall risk factors by kind (behavioral, endogenous, environmental), by setting (acute care risk factor, long term care risk factor, supporting housing risk factor, community risk factor), and by reversibility (surely reversible, subject-specific reversible, or irreversible).

Despite the face validity of the listed classifications, none of them is accepted and used as a universal reference framework to classify the fall risk factors.

1.4.3 Practical application of the WHO-FIC

As previously described in detail, WHO-FIC was born with a practical nature, specifically in its applications for describing health and health-related states of an individual[124, 131].

We will analyze their use as a conceptual framework to classify fall risk factors in our context. In particular, it is conceivable that the fall risk factors are substantially

independent of the disease that predisposes to fall. In other words, referring to the conceptual framework of the ICF, it can be said that it is not the disease itself to expose the subject to a fall (e.g., Parkinson's Disease or stroke), but rather the consequences of the disease, expressed in terms of impairment, activity limitation and reduced participation, which are not pathological-specific. For example, a fall may occur for inadequate footwear, visual impairment, intake of 4 medications, and balance impairment due to fragility, Parkinson's Disease, or stroke.

The use of the WHO-FIC as a conceptual framework for the classification of these risk factors can be implemented through the adherence to the ICF linking rules and the creation of adapted sets of descriptors, i.e., the core sets.

1.4.3.1 ICF linking rules

To cover the lack of a universal reference framework to classify the fall risk factors, the WHO-FIC can be the more natural and logical solution. We already reported that these classifications are designed to provide a framework to code a wide range of health information (e.g., diagnosis, functioning, and disability) and use a unified and standardized common language, permitting comparison and communication about health and healthcare across the world in various disciplines and sciences. They are also approved and recommended as guidelines for international reporting on health, and they may be used as models for the development or revision of other classifications[124, 131]. In particular, ICF has become the generally accepted framework to describe functioning in rehabilitation. In the clinical context, it is intended for use in needs assessment, matching interventions to specific health states, rehabilitation, and outcome evaluation[134].

The use of the WHO-FIC, particularly ICF, as a reference framework to describe individual health and health-related states requires the availability of a sound, standardized procedure that enables health information to be linked to the ICF[135-137]. Cieza and colleagues, in three papers in 2002, 2005, and 2019, have progressively defined the linking rules to link technical and clinical measures, health-status measures, and interventions to the ICF categories with the primary aims to ensure comparability, strengthen the overall transparency, and the reliability of the linking process[135-137]. In 2002, the focus was on the method for linking questionnaire items to the classification[135]; then, in the update of 2005, the rules were extended to other modes of

application, including specific clinical outcome measures (including technical measures for laboratory, imaging, and other clinical examinations), data on intervention targets and qualitative data[136]. After this definition, authors in 2019 claimed that ‘these rules were used by researchers across the globe in more than 100 articles, published in more than 55 peer-reviewed journals across more than 50 topic areas’[137]. In Figure 28, the 10 ICF linking rules updated to 2019 with further explanations are presented.

Figure 28. The 10 ICF linking rules (updated to 2019[137])

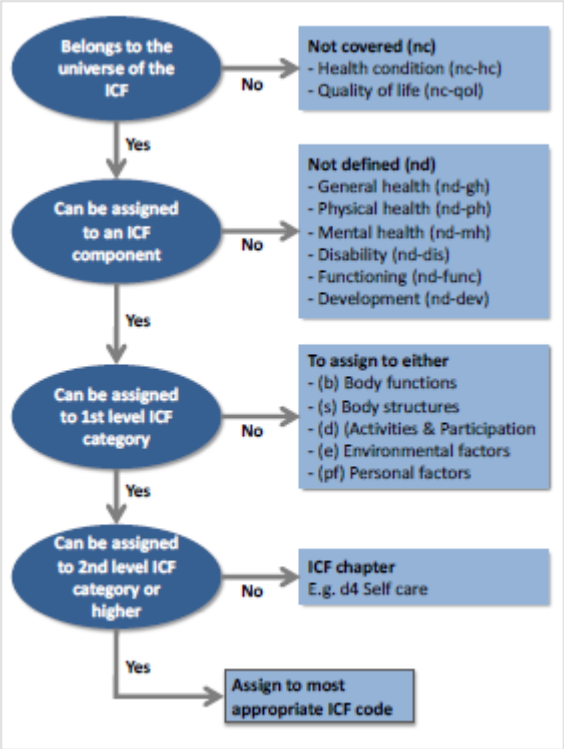
	Rule	Examples	
1	Acquire good knowledge of the conceptual and taxonomical fundamentals of the ICF, as well as of the chapters, domains and categories of the detailed classification, including definitions before starting to link meaningful concepts to the ICF categories		
2	Identify the purpose of the information to be linked by answering the question *What is this piece of information about?* or *What is this item about?* The answer to these questions will help to identify the main concept(s) most relevant to be linked to the ICF	"Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses." (Health Utility Index III (HUI-3), Vision Item 1)	*Main concept:* seeing *Additional concepts:* reading, recognize a friend on the other side of the street, glasses or contact lenses
3	Identify any additional concepts contained in the piece of information in addition to the main concept(s) already identified in the previous step		
4	Identify and document the perspective taken on within a certain piece of information when linking it to the ICF	"How satisfied are you with your sleep?" (WHOQoL-BREF, Item F3.3)	*Perspective:* Appraisal
	Take into consideration the most frequently used perspectives when collecting health or health-related information as listed in Table 2	"What was your level of need for help with not sleeping?" (Supportive Care Needs Survey, Item 3)	*Perspective:* Needs or dependency
5	Identify and document the categorization of the response options	"During the past month could you easily comb or brush your hair?"	*Categorization of response option:* Frequency
	Take into consideration the most frequently used approaches as listed in Table 3 Note: this rule applies only to instruments, questionnaires, assessments or tests that contain response options	Response options: All days, most days, some days, few days, no days (Arthritis Impact Measurement Scale (AIMS2), Item 18)	
6	Link all meaningful concepts, the most relevant and additional ones, to the most precise ICF category	"Play cards and other games." (West Haven-Yale Multidimensional Pain Inventory, Item C4)	*Main concept*: Play cards and other games *Most precise ICF category:* d9200 Play
7	Use "other specified [8]" or "unspecified [9]" ICF categories as appropriate	"Are you dependent on other persons in order to get out of or into your home?" (Nordic Mobility-related Participation Outcome Evaluation of Assistive Device Intervention (NOMO), Item 2a)	*Main concept:* get out of or into your home *Most precise ICF category:* d4608 Moving around in different locations, other specified – get out of or into home
	At the end of the chapter, and at the end of each embedded set of third- or fourth-level ICF categories, there are categories with the final code number 8 for "other specified" and 9 for "unspecified"		
	"8" is to be used when the concept is not contained within any of the other specific categories at the respective level of a chapter. The additional information is documented after the ICF code	"Does it every happen to you, that people invite you to a party or a dinner?" (Social Support List of Interactions (SSL-12-I), Item 1)	*Meaningful concept:* being invited to a party or dinner by people *Most precise ICF category:* d799 Interpersonal interactions and relationships, unspecified – being invited to dinner or party
	"9" is used when the concept to be linked fits within a given chapter but there is not sufficient information at hand to assign it to a specific ICF category		
8	If the information provided by the meaningful concept is not sufficient for making a decision about the most precise ICF category, assign the concept to nd (not definable)	"I have unpleasant side effects from my medication." (St. George's Hospital Respiratory Questionnaire Section 5)	*Main concept:* side effects *Most precise ICF category:* nd
	Concepts referring to health in general, physical health or mental (emotional) health in general, are assigned nd-gh, nd-ph or nd-mh (not definable-general health, not definable-physical health, not definable-mental health), respectively, as well as to disability in general (nd-dis), functioning (nd-func), or a child's development (nd-dev)	"In general, would you say your health is . . .?" (SF-36, Item 1)	*Main concept:* health *Most precise ICF category:* nd-gh
9	If the meaningful concept is not contained in the ICF, but is clearly a personal factor as defined in the ICF, assign the meaningful concept to pf (personal factors)	". . . Your faith in God?" (Quality of Life Index – cardiac version IV, Item 29)	*Main concept:* faith in God *Most precise ICF category:* pf
10	If the meaningful concept is not contained in the ICF, assign this meaningful concept to nc (not covered)	". . . attempts at suicides." (Hamilton Rating Scale for Depression, Item 3)	*Main concept:* attempts at suicides *Most precise ICF category:* nc
	Further specifications: Meaningful concepts referring to a diagnosis or health condition are assigned to nc-hc (not covered-health condition).	"Has the patient had a myocardial infarction? (MI)" (Charlson Comorbidity Index, Item 1)	*Main concept:* myocardial infarction *Most precise ICF category:* nc-hc
	Meaningful concepts referring to quality of life or life in general are assigned nc-qol (not covered-quality of life).	"How would you rate your quality of life?" (WHOQoL-BREF, Item 1)	*Main concept:* quality of life *Most precise ICF category:* nc-qol

NOTES: The 10 ICF linking rules updated to 2019 are presented. Specifically:

- Rule 1 highlights the importance of '*acquire a good knowledge*' of the ICF structure and composition before linking any kind of information.
- Rules 2 and 3 regards the identification of the *main and additional concept(s)* to be linked (e.g., main concept 'seeing' and additionals concepts 'reading, recognize a friend on the other side of the street, glasses or contact lenses of the Item 1 Vision of the Health Utility Index III).
- Rule 4 indicates identifying the *perspective* of the piece of information when linked to the ICF between descriptive perspective (performance or capacity), appraisal, need, or dependency.
- Rule 5 suggests to identify the *categorization of the response options* between (applies to instruments or similars that contain response option): intensity, frequency, duration, confirmation or agreement, qualitative attributes.
- Rule 6 regards linking *the identified main and additional concept(s)* to the most precise ICF category.
- Rule 7 defines the appropriate use of *'other specified (8)' and 'unspecified (9)'* ICF categories. 8 is used 'when the concept is not contained within any of the other specified categories at the respective level', otherwise 9 'when the concept fits within a given chapter', but sufficient information is not available to assign it to a specific ICF category.
- Rule 8 suggests assigning *'nd'(not definable)* to a concept when the available information is insufficient to link to a specific ICF category.
- Rule 9 identifies a *person factor* as a meaningful concept not classified in ICF but definable as a person factor.
- Rule 10 indicates to assign *'nc' (not covered)* to a concept that is not classified in the ICF (e.g., myocardial infarction: not covered-health condition. It is covered by ICD-10 (I21.9)).

In order to help new users in the linking of concepts to the ICF, as well as expert linkers, Cieza et al. constructed a 'linking decision tree', which is reported in Figure 29.

Figure 29. The ICF linking decision tree (from Cieza e al., 2019[137])



Finally, an example of the ICF linking extraction table, as the result of the extraction process, is available in Figure 30.

Figure 30. Example of an ICF linking extraction table (from Cieza e al., 2019[137])

1. Name of instrument or other identifier	2. Verbatim health information (e.g., wording of item or instruction)	3. Perspective adopted in information	4. Response options	5. Classification of response options	6. Main concept: What is this information about?	7. Additional concepts contained in information	8. ICF category of main concept	9. ICF category of other concepts	10. Annotation
WHODAS 2.0, Item D4.3	How much difficulty did you have in getting along with people who are close to you?	Descriptive; Performance	None, Mild, Moderate, Severe, Extreme/cannot do	Intensity	Getting along with people who are close to you		d799 Interpersonal interactions and relationships, unspecified		People who are close to you
WHOQOL-100, Item 21	How satisfied are you with your sex life?	Appraisal	Very dissatisfied, Dissatisfied, Neither satisfied nor dissatisfied, Satisfied, Very satisfied	Intensity	Sexual life		d7702 Sexual relationships		
LIFE-H, Item 2	Preparing your meals (including using electric kitchen appliances)	Need or dependency	Type of assistance (Check 1 or more, as required): No assistance Assistive device Adaptation Human assistance	Qualitative attributes	Preparing meals	(using electric kitchen appliances)	d630 Preparing meals	(d6403 Using household appliances)	
Manual Muscle Test	Say: "I am going to push down and I want you to resist me. Keep your arm up as I push down" (while patient seated and glenofemoral joint abducted to 90 degrees and elbow flexed to 90 degrees)	Descriptive; Capacity	0 (no mm contraction can be seen or felt) to 5 (part moves through complete range of motion against gravity and maximal resistance)	Intensity	Shoulder (Middle deltoid) strength against gravity		b7300 Muscle power functions of isolated muscles & muscle groups		Deltoid muscle(s) Against gravity

NOTES: in the figure, an example of the ICF linking extraction table, as the result of the extraction process, is represented. Specifically:

- Three columns (1,2 and 4) contain information to extract when linking health information to the ICF (column 4 only in case of instruments or similar linking).
- Two columns (6 and 7) contain main and additional concept(s)of the piece of information to be linked to the ICF.
- Four columns (3,5,8 and 9) contain the actual linking process (column 5 only in case of instruments or similar linking).

1.4.3.2 The ICF core sets

As we already described in detail, the International Classification of Functioning, Disability, and Health (ICF) offers a comprehensive and universally-accepted framework to describe functioning, disability, and health in persons with all kinds of diseases or conditions. The ICF, an exhaustive classification by its very nature, is quite complicated for use in daily practice. In daily practice, clinicians and other professionals need only a fraction of the categories found in the ICF[138]. A common base of categories is needed to compare with other health conditions and interventions. However, on the other hand, 'variability' is also required to capture the detail to describe the profile of a unique group for specialized clinical settings. The latter requirement was the primary motivation for WHO, in collaboration with the ICF Research Branch, to develop Comprehensive and Brief ICF Core Sets[138].

According to Selb et al.[139], 'an ICF Core Set (ICF-CS) is a selection of essential categories from the full ICF classification that are considered most relevant for describing the person's functioning with a specific health condition or in a specific healthcare context'. They are the minimum standard set of categories for assessing and reporting functioning and health in clinical practice and research[139].

For every ICF-CS, there are the Brief and the Comprehensive versions. The *Brief ICF-CS* for a specific condition (e.g., stroke) includes a list of ICF categories as few as possible to gain usability, but sufficient to guarantee the description of the typical spectrum of problems in the functioning of patients with stroke in clinical studies and practice. This list's shortness also assures to be used as a minimum data set to describe the disease's burden in a comparable way across studies[140]. Otherwise, the *Comprehensive ICF-CS* for a specific condition includes a list of ICF categories sufficiently exhaustive to describe in a comprehensive and multidisciplinary assessment the typical spectrum of problems in patients' functioning with a specific condition. The latter will be longer than the brief one[140]. An example of the Brief ICF-CS for Stroke[141] is available in Table 9.

Table 9. The Brief ICF core set for stroke (from Geyh et al., 2004[141])

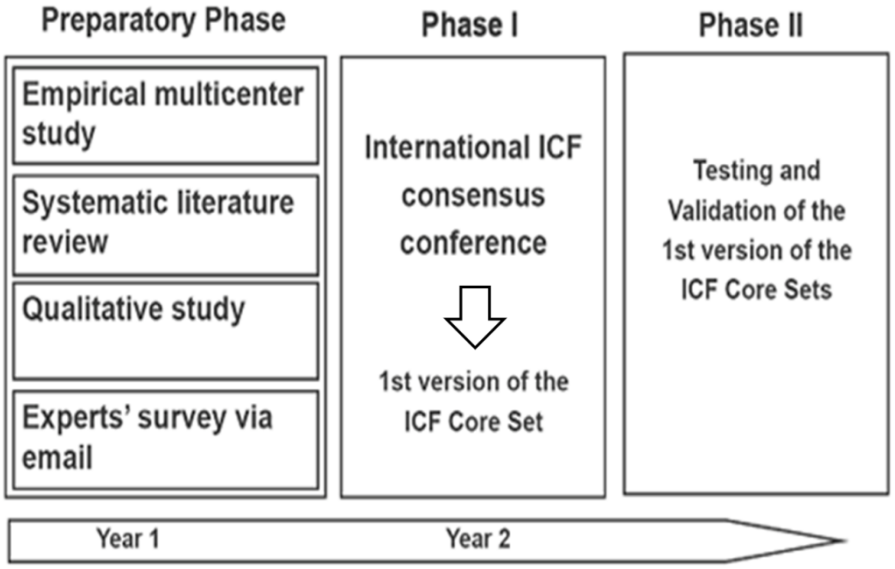
Body functions
b110 Consciousness functions
b114 Orientation functions
b730 Muscle power functions
b167 Mental functions of language
b140 Attention functions
b144 Memory functions
Body structures
100 s110 Structure of brain
s730 Structure of upper extremity
Activities and participation
d450 Walking
d330 Speaking
d530 Toileting
d550 Eating
d510 Washing oneself
d540 Dressing
d310 Communicating with – receiving – spoken messages
Environmental factors
e310 Immediate family
e355 Health professionals
e580 Health services, systems and policies

NOTES: in the figure, the categories related to body functions, structures, activities and participation, and environmental factors, which constitute the Brief ICF core set for stroke, are reported.

Over 60 ICF-CSs have been developed until now (Table 10), following a defined process[138, 139], that we will discuss in the following lines. They can be differentiated in setting (acute, early post-acute, and long-term) and for macro-groups of conditions, such as neurological, musculoskeletal and cardiopulmonary conditions, spinal cord

injury, and vocational rehabilitation[142]. Other authors, which follow different scientific process ways, made other ICF core sets (e.g., ICF core set for fall risks in acute rehabilitation settings[143], ICF core set for patients with systemic sclerosis[144], ICF core set for the rehabilitation of respiratory diseases[145]).

Figure 31. Process for developing an ICF core set (from ICF Research Branch[138] and Selb, 2015[139]



NOTES: in the figure, the detailed scientifically-based process for developing core sets of ICF categories for specific purposes is represented.

The process to develop core sets of ICF categories for specific purposes is standardized and scientifically-based[138, 139]. According to Selb and colleagues[139], which described this process in 2015, the core sets adhere to *three principles*: 1) the evidence-based process integrates evidence from preparatory studies; 2) ICF-CSs reflect the health professionals' and other experts' perspectives, but also that of persons with the specific health condition; 3) health professionals and experts come from various discipline fields and the 6 WHO world regions, allowing the application of the core sets in multidisciplinary settings and all over the world.

Then, the entire process is decomposable in *three phases*, as shown in Figure 31. The *preparatory phase*, in which the objective is to collect evidence, consists of four preparatory studies, each capturing a different perspective: the clinical perspective (multicenter empirical study), the health professionals' one (expert survey), the researchers' one (systematic literature review with the ICF linking process of identified

concepts according to the already mentioned rules[137]), and the perspective of a person with the specific health condition (qualitative study). The *phase I* foresees the organization of an International ICF consensus conference. Experts and health professionals evaluate the set of 'candidate' ICF categories derived from the preparatory phase and decide which categories will constitute the related first version of ICF-CS. Finally, in *phase II*, the testing and validation of this first version are implemented[139].

Table 10. Currently available ICF Core Sets (ICF-CSs) (from ICF Research Branch[138])

Neurological conditions

- ICF Core Set for adults with cerebral palsy (CP) (in process)
- ICF Core Set for children and youth (CY) with cerebral palsy (CP) (5)
- ICF Core Set for Multiple Sclerosis (MS) (2)
- ICF Core Set for Spinal Cord Injury (SCI) (4)
- ICF Core Set for Traumatic Brain Injury (TBI) (2)

Cardiovascular and respiratory conditions

- ICF Core Set for Chronic Obstructive Pulmonary Diseases (2)
- ICF Core Set for Obesity (2)
- ICF Core Set for Diabetes Mellitus (2)
- ICF Core Set for Stroke (2)
- ICF Core Set for Chronic Ischaemic Heart Disease (2)

Cancer

- ICF Core Set for patients with Head and Neck Cancer (2)
- ICF Core Set for Breast Cancer (2)

Mental health

- ICF Core Set for schizophrenia (2)
- ICF Core Set for Depression (2)
- ICF Core Set for Bipolar Disorders (2)

Musculoskeletal conditions

- ICF Core Sets for Ankylosing Spondylitis (2)
- ICF Core Set for Chronic Widespread Pain (2)
- ICF Core Set for Osteoporosis (2)
- ICF Core Set for Osteoarthritis (2)
- ICF Core Set for Low Back Pain (2)
- ICF Core Set for Rheumatoid Arthritis (2)

Diverse situations
- ICF Core Sets for acute and post-acute settings (cardiopulmonary, neurological, and musculoskeletal conditions) (12)
- ICF Core Set for Geriatric patients (2)
- ICF Generic and Rehabilitation Sets (2)
- ICF Core Set for Vocational Rehabilitation (2)
Other health conditions
- ICF Core Set for Autism Spectrum (ASD) (5)
- ICF Core Set for Attention Deficit Hyperactivity Disorder (ADHD) (5)
- ICF Core Set for Hearing Loss (2)
- ICF Core Set for Vertigo (2)
- ICF Core Set for Inflammatory Bowel Diseases (2)
- ICF Core Set for Sleep (2)
- ICF Core Set for persons following an amputation (in process)
- ICF Core Set for Hand Conditions (2)

NOTES: in the table, the current available ICF Core Sets from the ICF Research Branch are listed. In brackets, for each core set, the number of versions is available.

From the ICF core set webpage provided by the ICF Research Branch (https://www.icf-core-sets.org/) is possible to have access to the ICF-based Documentation Tool, which allows creating an ICF-based Documentation form, starting from the selection of one or more ICF core sets. The tool consents to add optionally further relevant categories from the whole ICF to fill the created form by rating the ICF categories, specifying the source of information and the description of the assessed person's problem, and obtaining an Individual Functioning Profile (Figure 32).

Figure 32. Example of a part of the ICF-based Documentation Form of the ICF Generic Set (from ICF Research Branch https://www.icf-core-sets.org/)

ICF-based Documentation Form

Reminder: The categories of the Generic Set are indicated by the letter (G).

PATIENT INFORMATION

BODY FUNCTIONS Physiological functions of body systems (including psychological functions) ***How much impairment does the person have in ...***		No impairment	Mild impairment	Moderate impairment	Severe impairment	Complete impairment	Not specified	Not applicable
		0	1	2	3	4	8	9
b130	**Energy and drive functions (G)**	☐	☐	☐	☐	☐	☐	☐
	General mental functions of physiological and psychological mechanisms that cause the individual to move towards satisfying specific needs and general goals in a persistent manner. Inclusions: functions of energy level, motivation, appetite, craving (including craving for substances that can be abused) and impulse control Exclusions: consciousness functions (b110); temperament and personality functions (b126); sleep functions (b134); psychomotor functions (b147); emotional functions (b152)							
	Sources of information: ☐ Case history ☐ Patient reported questionnaire ☐ Clinical examination ☐ Technical investigation							
	Description of the problem:							

NOTES: in the figure, the first body functions category of the ICF Generic Set is represented within the ICF-based Documentation Form. For each core set, it is possible to fill the created form by rating the ICF categories, specifying the source of information, and the description of the assessed person's problem. In the end, the Individual Functioning Profile is printable.

1.5 Summary of the current knowledge and aims of the thesis

In summary, the current knowledge presented in the introduction of this thesis can be summarized as follows.

1.5.1 Summary of the current knowledge

1.5.1.1 Fall risk

- The fall risk and of a related injury increase with age.

- Falls constitute two-thirds of deaths for unintentional injuries, which are the fifth leading cause of death in older adults (after cardiovascular disease, cancer, stroke, and pulmonary disorders).
- A fall represents a fearsome event for an elderly, both for possible traumatic consequences and psychological impact that can have a devastating impact on his quality of life.
- The human, health, and material costs of the fall risk are becoming unsustainable, and fall prevention in elderly subjects at risk is a current and priority public health objective.
- In the elderly with one or more associated neurological diseases, the impact of these conditions leads to increased fall risk, considering the already increased propensity for falls due to aging.
- The fall can be defined as an undesired outcome of an individual's motor activity requiring postural control, with his own personality and risk propensity (individual functioning element), in which the force of gravity (environmental element) was not sufficiently contrasted by the neurophysiological and biomechanical mechanisms of his postural control. On the contrary, the positive outcome of the same interaction is 'balance'.
- Falls in the elderly were more likely to occur at home. About half of them are caused by an environmental factor, happen during activities that mildly displaced the subject's center of mass, and are due to hurrying too much and misjudgment.
- In epidemiological studies of varying quality, fall risk factors have been identified, particularly by the Effective Health Care Bulletin (risk factors depending on environment and individual functioning) and World Health Organization (biological, behavioral, environmental, and socioeconomic risk Factors).
- Almost all international guidelines recommend a multifactorial removal approach of the modifiable fall risk factors to prevent falls.

- Effective interventions to prevent falls in community-dwelling older adults, also in those with an associated neurological disease, have been demonstrated.

1.5.1.2 Fall risk screening

- Fall risk screening is the first of several components of effective fall prevention programs. They aim to reduce the number of people who fall, the rate of falls, and the severity of injury should a fall occur, especially in adults older than 65, who suffer the highest number of fatal falls.
- Screening is defined as the presumptive identification of unrecognized disease in an apparently healthy, asymptomatic population by means of tests, examinations, or other procedures that can be applied rapidly and easily to the target population.
- Screening criteria were defined in 1968 by Wilson and Jungner to guide the selection of conditions that would be suitable for screening, based, among other factors, on the capacity to detect the condition at an early stage and the availability of an acceptable treatment. In the last decade, a review of the screening criteria over the past 40 years was performed, and they were transformed into a more elaborate decision-support guide.
- Screening can be delivered according to two different modalities: program screening versus opportunistic screening. The latter is considered relatively inefficient and cost-ineffective compared to the program one.
- Screening operates according to two different modalities: preventive versus early detection. The first consists of preventing the disease by finding and removing precursors of the disease, whereas the second of detecting the disease as early as possible to treat and cure the patient.
- The assessment of the screening economic consequences is possible for the advances in applying economic principles in health services. This evaluation's complexity is linked to several factors, which have to be considered in terms of costs and benefits.
- Clear and evidence-based information about the benefits and harms of any screening program should be given to all individuals invited to participate in any

program. The information aims to achieve truly informed consent through shared decisions based on a balanced and understandable picture of the options and the outcomes.

- Ethical screening responsibilities follow the screening application, as it potentially transforms individuals who are supposed to be 'healthy' to a state with some disorder or a potential one, and it should not be used to identify insignificant of untreatable conditions. Effective treatments for the diagnosed condition is crucial to increase the likelihood of effective treatment and better long-term outcomes.
- The regular scrutiny of all screening programs to control their performance and effectiveness is essential. Cochrane and Holland, in 1971, suggested seven criteria for the screening evaluation, which are still valid today. At least four biases can affect screening: lead-time bias, length-biased sampling, selection bias, and overdiagnosis bias.
- By 2050, projection indicates that the number of persons aged 65 years or over globally will surpass the number of adolescents and youth, so screening for health problems primarily related to older age is deemed as essential. However, there is little scientific evidence to support the benefits of screening in this age group, representing the key challenge.
- The most appropriate form of screening delivery in the elderly seems to be the regular surveillance and the case finding in primary case, with an important role played by general practitioners, to improve quality of life and preserve function and independence.

1.5.1.3 Fall risk screening guidelines and tools

- Several medical societies and national health agencies proposed fall risk screening guidelines in community-dwelling older adults, containing recommendations, algorithms, and instruments.
- To intervene promptly and effectively in reducing the fall risk, it is essential to identify and quantify this risk in the population of interest. Nevertheless, despite

the presence and use of various fall risk assessment tools in the elderly, it was not possible to predict older adult fallers with optimal accuracy.

- The most reliable criterion for identifying those at fall risk remains the history of a previous fall, which prevents the primary prevention of the fall events. It excludes from the intervention all the subjects that, because of undiagnosed balance impairment associated with other fall risk factors, are at high risk even if they have not yet had the opportunity to fall.

1.5.1.4 The WHO International Classifications

- The WHO Family of International Classifications is designed to provide a framework to code a wide range of health information and uses a standardized common language, permitting comparison and communication about health and healthcare across the world in various disciplines and sciences.
- The International Classification of Diseases is the standard diagnostic tool for epidemiology, health management, and clinical purposes, while the International Classification of Functioning, Disability, and Health is a classification of health and health-related domains, based on a bio-psycho-social model.
- In recent decades, several epidemiologic studies have investigated risk factors for falls. However, investigators have not used consistent classifications, so that the lack of a common classification framework for fall risk factors is still present.
- To cover the lack of a universal reference framework to classify the fall risk factors, the WHO-FIC can be the more natural and logical solution. In particular, it is conceivable that these factors are substantially independent of the disease that predisposes to fall. In other words, referring to the conceptual framework of the ICF, it can be said that it is not the disease itself to expose the subject to a fall (e.g., Parkinson's Disease or stroke), but rather the consequences of the disease, expressed in terms of impairment, activity limitation and reduced participation, which are not pathological-specific.

- Some authors have progressively defined the linking rules to link technical and clinical measures, health-status measures, and interventions to the ICF categories and have created and propose the ICF Core Sets.

1.5.2 Aims of the current thesis

1.5.2.1 Gap of knowledge

Considering the already available knowledge, it would be useful to know whether:

1. It is possible to improve the diagnostic accuracy of current screening systems in the older adult population, particularly having an instrument with high sensitivity and high specificity.
2. The current fall risk screening systems, including the FRAT-up, can demonstrate a good diagnostic accuracy even in a mixed population of older people at fall risk, including also subjects with Parkinson's Disease and stroke sequelae.
3. It is conceivable that the fall risk factors are substantially independent of the disease that predisposes to fall: it is not the disease itself to expose the subject to a fall (e.g., PD or stroke), but rather the consequences of that disease, which represent the fall risk factors. They can be expressed in terms of ICF impairment, limitation of activities, and reduction of participation concerning the subject's functioning.

1.5.2.2 Aims

Thus, specifically, the following aims were set for this thesis:

1. *to validate a fall risk serial screening algorithm with a high level of diagnostic accuracy in a sample of community-dwelling older people, also with associated neurological diseases (Parkinson's Disease and stroke sequelae).* This algorithm could be constituted by a first step based on anamnestic indicator(s) administered by phone and followed by a second step based on objective test(s) administered by a healthcare professional (e.g., physiotherapist). The two groups of indicators, anamnestic and objective, could be assembled in measurement scales, on which to identify the fall risk prediction cutoffs that maximize sensitivity and specificity with consequent better diagnostic accuracy.

2a. *to assess the effect of the neurological diseases on the fall risk screening tests in a sample of community-dwelling older people, also with associated neurological diseases (Parkinson's Disease and stroke sequelae).*

2b. *to validate an ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases (Parkinson's Disease and stroke sequelae).*

2 Methods

2.1 Study design and population

In order to achieve the planned aims, data of participants of the PRE.C.I.S.A. (PREvenzione Cadute e promozione Invecchiamento Sano e Attivo) randomized control trial[146] were used (Figure 33).

2.1.1 The PRE.C.I.S.A. study

The PRE.C.I.S.A. study is a randomized controlled trial with blind assessments, conducted from 2014 to 2016 at Modena and Reggio Emilia hospitals[146]. This study aimed to assess the effectiveness of a multicomponent therapeutic intervention combined with a multifactorial and personalized intervention to reduce individual fall risk factors in the prevention of falls in the following twelve months in community-dwelling older adults. The distinctive feature of the study was to propose a unified and interdisciplinary model of intervention for the prevention of falls, which was independent of the underlying disease, through the extension of the intervention also to elderly at very high fall risk, such as those with Parkinson's Disease or stroke sequelae. Specifically, patients were recruited according to the following inclusion and exclusion criteria:

Inclusion criteria

1. Community-dwelling older people at fall risk (≥ 65 years);
2. High to medium fall risk, associated with age and/or neurological diseases, like Parkinson's Disease and/or stroke sequelae;
3. Ability to walk for ten meters without assistance (possible the use of an aid);
4. Availability to give informed consent to participation.

Exclusion criteria

1. Any medical condition which constitutes, to the General Practitioner's and/or study physicians' judgment, a contraindication to the physical exercise;

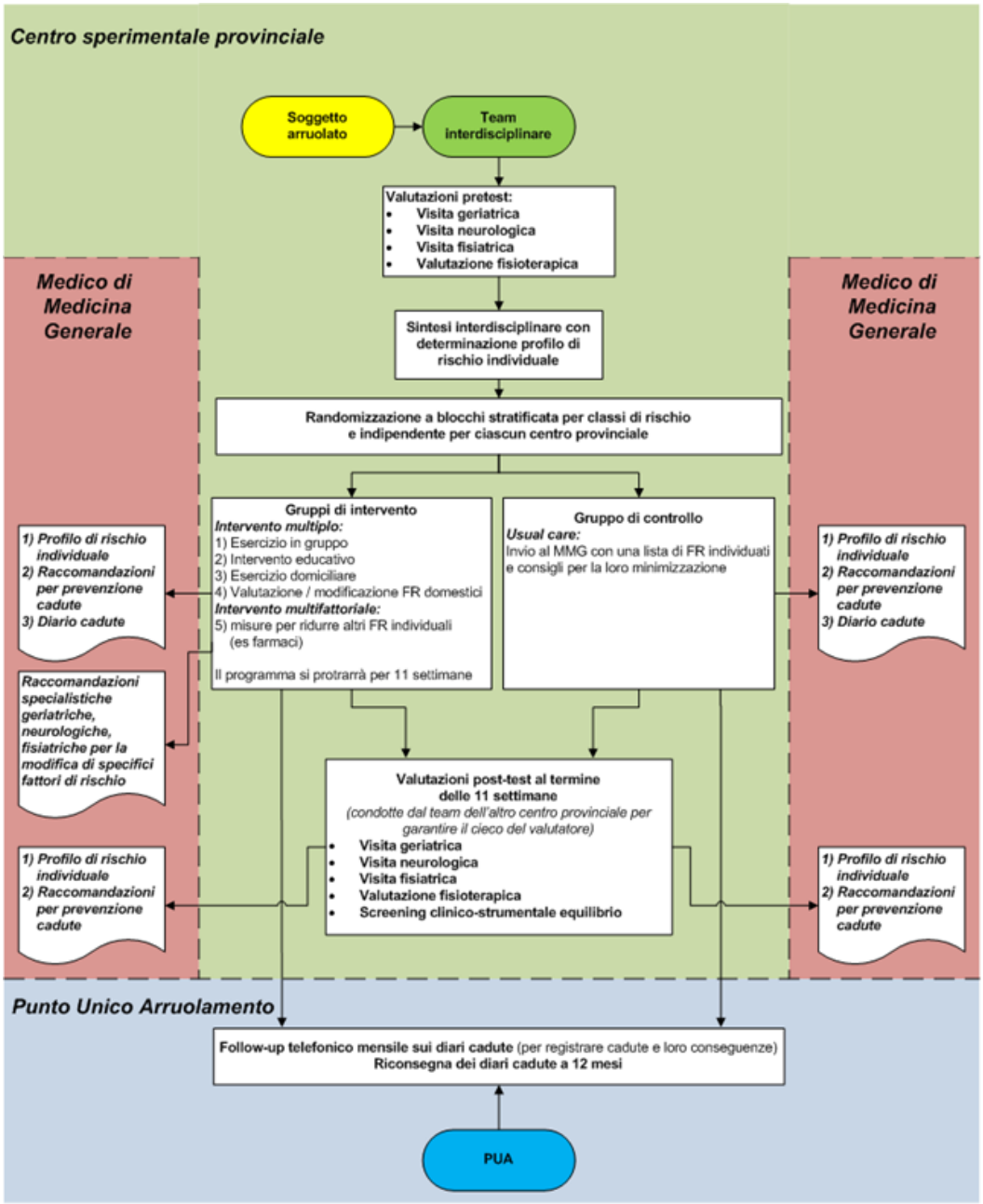
2. Known cognitive impairment or dementia (Mini-Mental total score <24 or a sufficiently severe cognitive impairment to compromise the comprehension of simple instruction or guidelines and/or to collaborate);
3. Severe deafness (he/she can understand less than 80% of the ordinary conversation despite the use of a hearing aid);
4. Severe low vision (visual function deficit not correctable with lenses, which limits the patient in the execution of at least one of the activities of daily living, for example, required assistance for moving caused by visual function deficit);
5. Severe aphasia or visuospatial disorders (sufficiently serious linguistic or visuospatial disorders to compromise the visual or auditory comprehension of simple instructions or guidelines);
6. Vertigo arose less than 3 months ago (vertigo must be objective, i.e., visual perception of object/room rotation, or subjective, i.e., a sensation of head and/or body rotation). Sensations of 'empty head' or 'instability' are not vertigo;
7. At the time of recruitment, participation in physical therapy which can influence target variables (balance, walking, etc.).

In the PRE.C.I.S.A. trial, the following treatments were administered (Figure 33):

- Experimental group (N=203): multicomponent therapeutic intervention (reducing the environmental fall risk factors at home, home exercise program in synergy with a group exercise program, educational intervention), combined with a multifactorial and personalized intervention to reduce individual fall risk factors.
- Control group (N=200): usual care (structured information about the individual fall risk factors for the general practitioner, together with an information booklet containing advice to reduce the fall risk).

The primary endpoint of the study was represented by the total number of falls that occurred in each group over 12 months, monitored monthly by telephone follow-up, and verified at the end of the follow-up by the return of the falls diaries[146].

Figure 33. PRE.C.I.S.A. study summary flowchart (from La Porta, 2018)[146]



2.1.2 Current study design, sample and collected data

For the current study project, the two groups of randomized subjects of the PRE.C.I.S.A. study (N=386, missing data 17) were considered as a unique cohort. It was justified given that the PRE.C.I.S.A. primary endpoint analysis did not show a significant difference between the two groups. Besides, data of further 365 subjects excluded from the PRE.C.I.S.A. study because 'at low fall risk' or 'not satisfied criteria'

were added. Hence, a sample of 768 subjects was available to achieve the primary and secondary aims of the current project (validation of a fall risk serial screening algorithm, evaluation of the effect of the neurological diseases on the fall risk screening tests, and validation of an ICD&ICF core set for the fall risk) with cross-sectional study design. As described in detail in the following 'statistical analyses' section, it was composed of three main parts: the psychometric validation of scales, the study of diagnostic accuracy of scales and algorithms, and the core set validation.

For each participant, the variables collected at different study selection steps for the recruitment to the PRE.C.I.S.A. trial were used:

1. *Anamnestic assessment of eligibility (VAE).* This assessment constituted the PRE.C.I.S.A. first selection step, and it was administered by a trained nurse ('Punto Unico di Arruolamento – PUA'), through a telephone call, to older adults who had been signaled as 'at fall risk' by another health professional (medical specialist or general practitioner) or who had self-reported as 'at fall risk'. The aim was to confirm the inclusion/exclusion criteria for the study recruitment and evaluate the most important fall risk factors. People who resulted in 'medium-high fall risk' after the combination of the assessment results accessed the successive selection step (Objective assessment of eligibility - VOE). In contrast, those resulted at 'low fall risk' or 'not satisfying study criteria' were excluded. The latter were contacted one year later by the cited trained nurse to collect the number of falls (study outcome). The following variables were evaluated:
    - A question related to one of the inclusion criteria: could the patient be at fall risk?
    - 5 items from the Fall risk Assessment Tool (Nandy, 2004)[147]:
        a. Is there a history of any fall in the previous year?
        b. Is the patient/client on four or more medications per day?
        c. Does the patient/client have a diagnosis of stroke or Parkinson's Disease?
        d. Does the patient/client report any problem with his/her balance?

e. Is the patient/client unable to rise from a chair of knee height?

- Further question: does the patient feel fear of falling?

2. *Objective assessment of eligibility (VOE).* This assessment constituted the PRE.C.I.S.A. second selection step, and it was administered by a trained physiotherapist, during an outpatient visit, to older adults who had resulted at 'medium-high fall risk' at the previous selection step (VAE). The aim was to evaluate in detail all fall risk factors described in the literature and, hence, to confirm the eligibility for the study (be at 'medium-high fall risk' after the combination of the assessment results). Even at this step, those resulted at 'low fall risk' were excluded but contacted one year later by the cited trained nurse to collect the number of falls (study outcome). The following variables were evaluated:

   - 8 items from the FROP-Com[148]:

     a. Number of falls in the past 12 months

     b. Was an injury sustained in any of the fall/s in the past 12 months?

     c. Prior to this fall, how much assistance was the individual requiring for personal care activities of daily living (e.g., dressing, grooming, toileting)?

     d. Has this (assistance for personal care activities of daily living) changed since the most recent fall?

     e. Prior to this fall, how much assistance was the individual requiring for instrumental activities of daily living (e.g., shopping, housework, laundry)?

     f. Has this (assistance for instrumental activities of daily living) changed since the most recent fall?

     g. Does the individual, upon observation of walking and turning, appear unsteady or at risk of losing their balance?

     h. Has the individual's level of physical activation changed since the most recent fall?

- 3 items from the Fall risk Assessment Tool (Stapleton, 2009)[149]:
    a. Medications (sedatives, anti-depressants, anti-Parkinson's, diuretics, anti-hypertensives, hypnotics)
    b. Psychological status (anxiety, depression, loss of cooperation, of insight or judgment)
    c. Cognitive status (according to the Abbreviated Mental Test Score by Hodkinson, 1972[150])
- Balance and mobility tests:
    a. 10 meters walking test[151]
    b. Use of aid in the 10 meters walking test[151]
    c. Timed Up&Go test[152]
    d. Use of aid in the Timed Up&Go test[152]
    e. Standing balance, taken from the 4 Stage Balance Test [153]
    f. 30-second Chair stand test[154, 155]
    g. Short Physical Performance Battery[156]
    h. Functional Reach test[157]
- Adequate vision (Snellen Chart).

The PRE.C.I.S.A. outcome variable, available for the entire sample (N=768), was the number of falls one year after enrollment, and it was used as the 'gold standard' test for the diagnostic part of the study. For the randomized subjects, this data was collected through monthly telephone monitoring by the cited trained nurse and the return of the falls diaries at the end of the follow-up. Differently, as already mentioned, only a single telephonic follow-up, at 12 months after the exclusion, has been done by the same nurse for the elderly excluded from the study because 'at low fall risk' or for 'not satisfied criteria' to collect the number of falls of this period.

2.2 Statistical analyses

2.2.1 Validation of a fall risk serial screening algorithm with a high level of diagnostic accuracy

Regarding the statistical analyses about the primary objective of *validation of a fall risk serial screening algorithm with a high level of diagnostic accuracy*, the following analyses were conducted.

2.2.1.1 Calibration of measurement scales from VAE and VOE variables

2.2.1.1.1 Extraction of the variables of interest from the PRE.C.I.S.A. dataset

Regarding the VAE variables of the PRE.C.I.S.A. study dataset, the following variables of interest were extracted for a total sample of 768 older adults with complete observations. These variables were the 7 candidates to compose the items of the VAE scale:

- One item from the inclusion criteria:

  VAE00.Could the patient be at fall risk? (0 = no; 1 = yes)

- Five items from the Fall risk Assessment Tool (Nandy, 2004)[147]:

Item code	Item description	Item scoring
VAE01	Is there a history of any fall in the previous year?	0 = no 1 = yes
VAE02	Is the patient/client on four or more medications per day?	Number of medications (med): 1 = ≤1 med 2 = 2 med 3 = 3 med 4 = 4 med 5 = 5 med 6 = 6 med 7 = ≥7 med
VAE03	Does the patient/client have a diagnosis of stroke or Parkinson's Disease?	0 = no 1 = yes
VAE04	Does the patient/client report any problem with his/ her balance?	0 = no 1 = yes
VAE05	Is the patient/client unable to rise from a chair of knee height?	0 = no 1 = yes

- Additional item:

  VAE06.Does the patient feel fear of falling? (0 = no; 1 = yes)

Regarding the VOE variables of the PRE.C.I.S.A. study dataset, the following variables of interest were extracted for a total sample of 574 older adults with complete observations. These variables were the 22 candidates to compose the items of the VOE scale:

- Ten items from balance and mobility tests (10 Metres Walking test[151]; Timed Up&Go test[152]; Standing balance, taken from the 4 Stage Balance Test[153]; 30-second Chair stand test[154, 155]; Short Physical Performance Battery[156]; Functional Reach test[157]):

Item code	Item description	Item scoring
VOE01	10 Metres Walking test	Seconds to perform the test transformed into an item with 3 score categories according to Bowden's cutoff[158]: 0 = >0.8 m/s: community ambulators 1 = 0.4-0.8 m/s: limited community ambulators 2 = <0.4 m/s: household ambulators)
VOE02	Does the patient need a walking aid to perform the 10 meters walking test?	0 = no 1 = yes
VOE03	Timed Up&Go test	Seconds to perform the test transformed into an item with 2 score categories according to Shumway-Cook's cutoff[159]: 0 = <13.5 seconds: low fall risk 1 = ≥13.5 seconds: high fall risk
VOE04	Does the patient need a walking aid to perform the Timed Up&Go test?	0 = no 1 = yes
VOE05	Standing balance (ability to maintain balance in tandem stance for more than 10 seconds)	0 = no 1 = yes
VOE06	30-second chair stand test	Number of lifts from a chair in 30 seconds transformed in an item with 2 score categories according to the STEADI's cutoff[160]: 0 = not at fall risk; 1 = at fall risk)
VOE07	Short Physical Performance Battery 1 – Balance assessment in 3 positions	0 = tandem stand for 10 seconds 1 = tandem stand for 3-9 seconds 2 = tandem stand for 0-2 seconds 3 = semi-tandem stand for 0-9 seconds 4 = side-by-side stand for 0-10 seconds
VOE08	Short Physical Performance Battery 2 - Gait speed assessment for 4 metres	0 = <4.1 seconds 1 = 4.2-5.3 seconds 2 = 5.4-7.5 seconds 3 = >7.5 seconds 4 = unable to walk for 4 metres
VOE09	Short Physical Performance Battery 3 - Repeated chair-stand assessment for 5 times	0 = <11,2 seconds 1 = 11.2-13.6 seconds 2 = 13.7-16.6 seconds 3 = ≥16.7 seconds 4 = unable to complete 5 chair-stands
VOE10	Functional Reach test	Mean of three trials in centimeters transformed in an item of 2 score categories according to Thomas' cutoff[161]:

	0 = ≥ 18,5 cm: not at fall risk 1 = <18,5 cm: at fall risk)

- Eight items from the FROP-Com[148]:

Item code	Item description	Item scoring
VOE11	Number of falls in the past 12 months?	0 = no falls 1 = 1 fall 2 = 2 falls 3 = 3 or more falls
VOE12	Was an injury sustained in any of the fall/s in the past 12 months? *(Rate most severe injury due to a fall in the past 12 months)*	0 = no 1 = minor injury, did not require medical attention 2 = minor injury, did require medical attention 3 = severe injury (fracture, etc.)
VOE13	Prior to this fall, how much assistance was the individual requiring for personal care activities of daily living (e.g., dressing, grooming, toileting)? *(NOTE: If no fall in last 12 months, rate current function)*	0 = none (completely independent) 1 = supervision 2 = some assistance required 3 = completely dependent
VOE14	Has this (assistance for personal care activities of daily living) changed since the most recent fall? *(leave blank if no falls in 12 months)*	0 = no 1 = yes
VOE15	Prior to this fall, how much assistance was the individual requiring for instrumental activities of daily living (e.g., shopping, housework, laundry)? *(NOTE: If no fall in last 12 months, rate current function)*	0 = none (completely independent) 1 = supervision 2 = some assistance required 3 = completely dependent
VOE16	Has this (assistance for instrumental activities of daily living) changed since the most recent fall? *(leave blank if no falls in 12 months)*	0 = no 1 = yes
VOE17	When walking and turning, does the person appear unsteady or at risk of losing their balance? *- Observe the person standing, walking a few meters, turning, and sitting. If the person uses an aid, observe the person with the aid. Do not base on self-report.* *- If the level fluctuates,*	0 = no unsteadiness observed 1 = yes, minimally unsteady on walking or turning 2 = yes, moderately unsteady on walking or turning (needs supervision) 3 = yes, consistently and severely unsteady on walking or turning (needs constant hands on assistance)

	tick the most unsteady rating. If the person is unable to walk due to injury, score as 3.	
VOE18	Has the individual's level of physical activation changed since the most recent fall?	0 = no 1 = yes

- Three items from the Fall risk Assessment Tool (Stapleton, 2009)[149]:

Item code	Item description	Item scoring
VOE19	Medications (sedatives, anti-depressants, anti-Parkinson's, diuretics, anti-hypertensives, hypnotics)	1 = not taking any of these 2 = taking one 3 = taking two 4 = taking more than two
VOE20	Psychological status (anxiety, depression, loss of cooperation, of insight or judgment)	1 = does not appear to have any of these 2 = appears mildly affected by one or more 3 = appears moderately affected by one or more 4 = appears severely affected by one or more
VOE21	Cognitive status (according to the Abbreviated Mental Test Score by Hodkinson, 1972[150])	1 = AMTS 9-10 or intact 2 = AMTS 7-8 mildly impaired 3 = AMTS 5-6 moderately impaired 4 = AMTS 4 or less severely impaired

- Additional item:

  VOE22.Visual acuity (Snellen Chart) (visual acuity of at least 1/10: 0 = no impairment; 1 = one-side impairment; 2 = bilateral impairment).

2.2.1.1.2 Psychometric analyses of the candidate variables

We performed the psychometric analyses of the candidate variables to calibrate at least one 'anamnestic scale', which can be administrated by telephone, and one or more 'objective scales', administered to the subject by a health professional. Specifically, descriptive statistics (sample and items), the preliminary assessment of the dimensionality of the scales using the Confirmatory Factor Analysis (CFA) and the Mokken analysis (MA), and the Rasch analysis (RA) were performed.

- *Descriptive statistics* were performed for the main demographic and clinical variables of persons (age, gender, neurological diseases, access sources to the recruitment, fall risk, falls at twelve months), along with analyses of the frequency of score categories and their distribution, of missing data, of inter-item and item-total score correlations, and internal consistency reliability of the VAE and VOE variables.

- *Confirmatory Factor analysis*: to assess the dimensionality of the VAE and VOE scales, we performed a non-parametric CFA for ordinal data based on polychoric correlations [162]. Within the CFA, described in detail elsewhere[163, 164], we assessed the model fit under the assumption of unidimensionality of the two item sets using the following indicators:
    - the Root Mean Square Error of Approximation (RMSEA), where values ≤0.08 are generally indicative of 'mediocre fit' but deemed to be sufficient for a preliminary assessment of dimensionality before the Rasch analysis[165] and ≤0.06 of 'good fit'[166];
    - the Standardized Root Mean square Residual (SRMR), where values ≤0.08 were considered indicative of 'adequate fit'[166] and ≤0.05 were of 'well-fitting' solution[167];
    - the Comparative Fit Index (CFI) and the non-normed fit index (Tucker-Lewis Index – TLI), for which values >0.95 [0, 1] were considered of a well-fitting solution[167].

  Within the CFA, for each item set, we first tested a one-factor model. Should this model fail to fit, we would attempt to improve fitness to the model by allowing the correlation of error terms between pair of items displaying high modification indices (MI)[162, 167], which indicate local dependence[168-172]. Should this modified model fail to fit, we would consider this information of preliminary insufficient unidimensionality in the next steps of analysis.

- *Mokken analysis*: to obtain further evidence of the dimensionality of the scale, we performed a Mokken Analysis of the VAE and VOE variables. This scaling procedure for both ordinal dichotomous and polytomous items, based on the Monotone Homogeneity Model (MHM), is the most general nonparametric IRT model[173]. MHM assumes the unidimensionality of the latent trait, the monotonicity, and the local independence of responses; it can be used to partition a set of items into Mokken scales using an automated item selection procedure (AISP)[173-175]. In the MA, items belonging to the same Mokken scale should have an item scalability coefficient Hj (calculated as the normed covariance between the item score and the rest score) greater than a positive

lower bound c, which measures the consistency of the item j with the scale[175], and have to be specified (recommended default value c = 0.3)[176] together with the *nominal significance level α*[177]. The latter is the nominal significance level of the inequality tests used in the automated item selection procedure, and its recommended default value is 0.05[178]. AISP also considers the item-pair scalability coefficients Hij, calculated as the normed covariance between the item scores: items belonging to the same Mokken scale should have positive item-pair scalability coefficients[177, 178]. Finally, the scalability coefficient H indicates the overall quality of a scale (i.e., the degree to the test data follow a perfect Guttman scalogram)[177, 178].

At the end of the procedure, the analysis shows the number of scales needed for scaling all items. Should the automated procedure estimate the need for more than one scale to accommodate all the items, we would consider this information of preliminary insufficient unidimensionality in the next analysis steps.

- *Rasch analysis*: following the above analyses, the VAE and VOE variables were fitted separately to the Rasch model[179]. It is a unidimensional mathematical model, which postulates that a subject with a specific ability on the latent variable (i.e., fall risk) is expected to affirm (pass) items associated with less risk and not to affirm (fail) items representing a higher risk[180]. If data are conformed to this pattern, together with the demonstration of the based assumptions of local independence and unidimensionality, it is possible to affirm that they satisfy the requirements of the model and that raw total scores of these scales can be transformed into interval-level measures[181-183]. The process of iteratively testing whether the data meet the requirements of the Rasch model is widely known as Rasch analysis[165, 179, 182, 184-187]. Within this analytical framework, here based upon the partial credit parameterization of the model, which does not place constraints on the item threshold parameters[188], we tested the following measurement indicators[171, 189]:

    - Internal construct validity[183], which included the assessment of requirements specific to the Rasch model for model fit (i.e., item homogeneity or invariance[179, 182, 184]), and adherence to a probabilistic Guttman pattern[182, 184, 186]. Furthermore, other general requirements of

item response theory models [190, 191], such as monotonicity[182, 190], local independence[190, 192], unidimensionality [182, 190, 193], and absence of differential item functioning (DIF) or item bias for a subgroup of persons in the sample (age, gender, neurological diseases, days since lesion, center)[182, 185, 186, 190].

- Separation reliability, here represented both by the Person Separation Index (PSI), the Cronbach's α [182, 184, 194, 195], the number of statistically Distinct Levels of Performance Ability (DLPA)[196], and the distribution-independent person separation index (DI-PSI) [196].
- Targeting, indicating how well the measurement range of the scale matches the distribution of the calibrating sample[182, 184, 197], here expressed as floor and ceiling effects[197] and targeting index[197].

Where these assumptions failed, an iterative phase involving item modifications was undertaken, aiming at finding a solution that satisfied both the model requirements and the model expectations, as well as the theoretical expectations of the measured construct. These post-hoc item modification strategies included:

- *Item rescoring* (modification of the item scoring structure), to manage the violation of monotonicity[182, 184, 189, 198];
- *Item grouping* or 'testlet' creation (grouping of items), to handle the violation of local independence[165, 199, 200];
- *Item splitting* (separation of the item), to manage the presence of uniform DIF[180, 182, 185, 201];
- *Item deleting* (elimination of the item), an adopted strategy when other strategies failed, and the item still showed a model non-adherence, or it was affected by non-uniform DIF[200].

This phase was followed by a reassessment of the internal construct validity after each modification cycle, described in detail elsewhere [179, 185, 186, 202-204]. To assess the unidimensionality of the subscale structure after the creation of testlets, the following indicators were evaluated:

- c, which is the 'unique variance' for each subscale[205];

- r, which is the 'latent correlation' between the subscales[205, 206];
- A, which is the proportion of non-error variance common to all subscales, can be considered equivalent to the 'explained common variance' of the bifactor model[206-208].

In the case of the multidimensionality of the subscale structure, c will be high and r and A low, whereas in the case of unidimensionality of the full item set despite the subscale structure, c will be low and r and A high. Particularly, values of A>0.90 will indicate 'essential unidimensionality', as the subscales do not share only 10% of the common variance. In the case of a two-testlet solution, also the conditional total item-trait interaction chi-squares were evaluated because the unconditional ones are not reliable for a sample size of 200 or more. In contrast, the conditional fit statistics remain reliable for sample sizes ≤2,000[209].

Differential Item Functioning was examined through the response standardized residuals between the observed and the expected responses of each person to an item. Each person was assigned to a factor group (e.g., neurological diseases) and classified by the ability measure on the latent trait into one of the class intervals (groups of approximately equivalent size across the sample to approximate ability groups). Then, for each item, the residuals are analyzed with a two-way analysis of variance (ANOVA) according to the two factors: person factor and class interval. The presence of DIF is indicated by statistically significant inter-person-group variance[210]. Should DIF be detected, the influence of the item/testlet splitting on the person estimates would be tested using the procedure presented by Maritz and colleagues[207]. Specifically, after the item/testlet splitting, we would anchor the 'splitted' solution on the 'un-splitted' one, using an item or testlet free from DIF. We would then compare the person estimates of the two solutions, calculating an effect size (Cohen's d) of the paired t-test of the difference. Should Cohen's d be <0.2, it would be considered negligible, and, thus, the DIF would not be adjusted for[207]; otherwise, the 'splitted' solution would be the final one[207].

Should a final solution fitting the model following the above modifications be found, its total score could be transformed into interval-level measurements, whose unit is the logit[179, 182, 184].

2.2.1.2 Diagnostic accuracy of the available screening tools

2.2.1.2.1 Compilation of the two versions of the FRAT-up

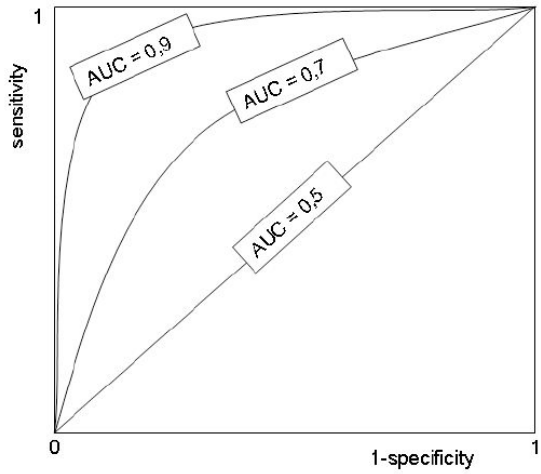
We decided to use the FRAT-up tool (available at http://ffrat.farseeingresearch.eu/) as a comparator for the calibrated VAE, VOE1, and VOE2 scales in terms of diagnostic accuracy in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months, with a purpose of 'external validation', being already validated and published[120, 121]. Hence, before proceeding to the diagnostic accuracy study, we compiled the two versions of the FRAT-up, the first based on the VAE variables and the second on the VOE variables. The adaptation and the conversion of the information given by the variables to the different FRAT-up items were realized through two designed tables, one for the VAE and one for the VOE variables (Appendix A). In case of missing data for a risk factor, corresponding prevalence data taken from several literature sources was used, as expected in the original validation of the instrument[121].

2.2.1.2.2 Definition of the optimal cutoffs of the available screening tools

The optimal cutoffs of the calibrated scales (separately on the VAE and VOE variables) and of the two calculated FRAT-up (separately on the VAE and VOE variables) in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months were defined. The cutoffs showed different units of measure according to the type of assessment tool (logit for the VOE and VAE scales, probability for the two calculated FRAT-up).

One of the most commonly used methods to analyze the effectiveness of a diagnostic test is the *receiver operating characteristic curve (ROC) analysis*. It offers a graphical illustration of the trade-off mentioned above between a test sensitivity and specificity and depicts true positive rate against false positive rate for each cutoff value[211, 212] (Figure 34).

Figure 34. ROC curve, the area under the curve and diagnostic accuracy (from Šimundić 2009)[122]



NOTES: in the figure a ROC curve is represented, with three examples of curve shapes designed by a test defining an area under the curve (AUC), with range 0-1. It represents an indicator of the goodness of the test. In the case of AUC=0.5, the test has a 'null' diagnostic accuracy, between 0.5 and 0.7 the accuracy passes from 'null' to 'sufficient', then between 0.7 and 0.9 it goes from 'good' to 'very good'; over 0.9 it becomes 'excellent'.

The shape of a ROC curve and the area under the curve (AUC) help us estimate how high is the discriminative power of a test. The closer the curve is located to the upper-left-hand corner and the larger the area under the curve, the better the test is at discriminating between diseased and non-diseased. The area under the curve can have any value between 0 and 1, and it is a good indicator of the goodness of the test. A perfect diagnostic test has an AUC of 1.0, whereas a non-discriminating test has an area of 0.5. AUC is a global measure of diagnostic accuracy, and it is useful for general assessment and comparison of two or more diagnostic tests. However, it does not give information about individual parameters, such as sensitivity and specificity, predictive values, or the test's contribution to ruling in or ruling out a diagnosis. The relationship between the area under the ROC curve and the diagnostic accuracy were interpreted as follows: 0.9-1 excellent, 0.8-0.9 very good, 0.7-0.8 good, 0.6-0.7 sufficient, 0.5-0.6 bad, <0.5 test not useful[122] (Figure 34).

AUC is particularly useful when two or more diagnostic tests are compared. Having a higher AUC, a test with a ROC curve that lies entirely above another curve is clearly a better one. The methods for calculating the AUC are mainly based on a nonparametric statistical test, the Wilcoxon rank-sum test, proposed by DeLong and Hanley[212-215]. In

this thesis, we performed the calculation and the comparison of the ROC curves and the AUCs for all the scales by testing the equality of two or more ROC areas (chi-square test)[213]. In the case of significant statistics, we made a pairwise comparison between scales to identify those with a real different discriminative power.

In tests with multiple values, choosing an appropriate threshold (cutoff) value is of paramount importance in using a test effectively, to distinguish which subjects are at risk and not regarding the studied condition[211, 216]. Several criteria, mostly based on ROC analysis, have been proposed for choosing the most appropriate cutoff value[212, 216, 217]. Each point on a ROC curve corresponds to a cutoff value associated with a test sensitivity and specificity. Thus, locating the cutoff point requires a compromise between the two properties. According to the choice of the cutoff level, it is privileged a certain value of sensitivity at the expense of specificity, because a high level of sensitivity implies a loss of specificity and viceversa[211], with an impact on the classification of false positive and false negative subjects (Table 11 and Table 12). Between the cited existing methods, we remember:

- *Youden's index*[218]. It is calculated by deducting 1 from the sum of the sensitivity and specificity of the test, expressed not as a percentage but as a part of a whole number: (sensitivity + specificity) – 1[122] (Table 12). This index minimizes the chance to find false positive and false negative[211] (Table 11). It is one of the oldest measures for diagnostic accuracy and a global measure of test performance used to evaluate the overall discriminative power of a diagnostic procedure and the comparison of one test with other tests. For a test with poor diagnostic accuracy, Youden's index equals 0, and in a perfect test, it equals 1. It is not sensitive to differences in the sensitivity and specificity of the test, which is its main disadvantage: a test with sensitivity 0.9 and specificity 0.4 has the same Youden's index (0.3) as a test with sensitivity 0.6 and specificity 0.7. It is clear that those tests are not comparable in terms of diagnostic accuracy and that it is not correct to discriminate the power of a test solely based on this index[122].
- *The point closest-to-(0, 1) corner in the ROC plane* approach[219]. It selects the optimal cut-point to minimize the Euclidean distance between the ROC curve and the (0, 1) point[216]. The (0, 1) corner in the ROC curve represents the ideal situation with maximum sensitivity (true positives = 1) and specificity (1 – false

positives = 0). In its calculation, this method minimizes a quadratic term that it is not clinically interpretable[216, 219].

- *Concordance probability criterion* approach[220]. It is based on the cut-point choice that achieves the maximum of the product of sensitivity (true positives) and specificity (1 – false positives)[216, 219].
- *Minimum p-value* approach[221]. It identifies the optimal threshold that best separates the two risk groups (i.e., diseased and diseased-free) according to the maximum achievable value of the chi-square statistic on the association between the test result and the outcome variable (disease vs. not disease)[216].
- *Misclassification-Cost Term*[212, 217]. It is based on the minimization of a term that measures the cost of incorrect classifications, and it considers only the cost ratio of a FP to a FN result because the costs of true decisions are assumed to be null[217]. Besides, it defines the same cut-point of the analytical method to maximize the Number Needed to Misdiagnose (NNM = 1/FN+FP)[212].
- *Other methods,* based on the maximization of sensitivity, specificity, predictive values, diagnostic likelihood ratios, and prevalence[217].

According to several authors' demonstrations, we performed multiple attempts, using the cited methods, to establish the optimal cutoff. For each scale, we reported the cutoff determined through the Youden Index method, which is one of the most common methods, and an 'ad hoc' cutoff that allows obtaining the optimal trade-off, from a clinical perspective, between false positives and false negatives in the following steps of the generation of screening algorithms.

2.2.1.2.3 Analyses with standard techniques of diagnostic accuracy of the available screening tools

Based on the cutoffs defined with the Youden Index and the 'ad hoc' clinical methods, the analyses with standard techniques of diagnostic accuracy of the calibrated scales and the two calculated FRAT-up (separately on the VAE and VOE variables) in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months were performed.

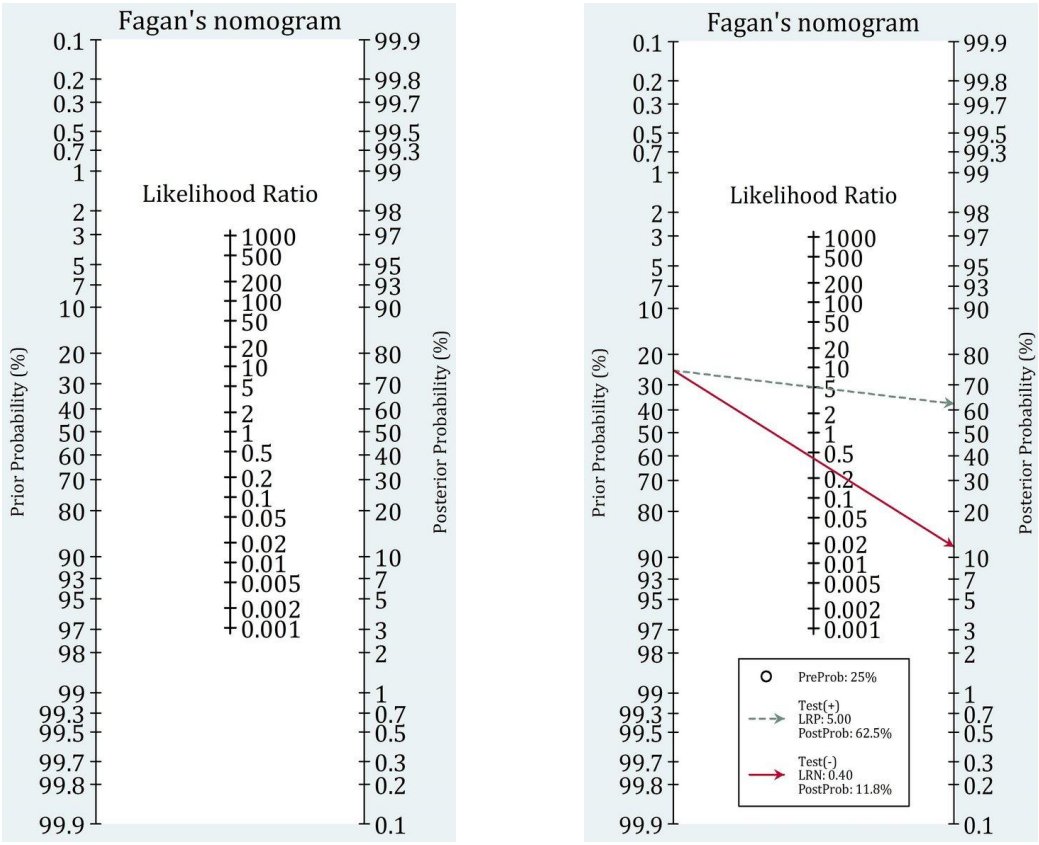
Specifically, the validity of the 'test' (i.e., scales and, later, the fall risk screening algorithms) was described in terms of:

- The classification of the population of examined subjects with and without disease in subgroups, defined by the calculated cutoff[122]: *true positives* (subjects with the disease with the value of the parameter of interest above the cutoff), *false positives* (subjects without the disease with the value above the cutoff), *true negatives* (subjects without the disease with the value below the cutoff), *false negatives* (subjects with the disease with the value below the cutoff). The first step in the calculation of the following parameters (sensitivity, specificity, etc.) was to create a 2x2 table with groups of subjects divided into columns according to a gold standard (i.e., 'fall' outcome), and categories in rows according to the test (Table 11)[122].
- The degree to which persons with and without the condition under study (be at fall risk) were correctly categorized[222]: *sensitivity* (proportion of persons with the condition who tested positive) and *specificity* (proportion of persons without the condition who were correctly categorized as negative by the test)[122, 223] (Table 12). Given that these properties were expressed in terms of proportion, the desired values were as much as possible equal to 1 (or 100%)[211]. The disease prevalence influences neither sensitivity nor specificity; hence, results from one study could easily be transferred to some other setting with a different prevalence of the population's disease. Nonetheless, sensitivity and specificity can vary greatly depending on the disease's spectrum in the studied group[122].
- The proportion of correctly classified subjects (True positives + True Negatives) among all subjects, called *diagnostic accuracy (effectiveness)*[122] (Table 12). This measure is affected by the disease prevalence: with the same sensitivity and specificity, it increases as the disease prevalence decreases. However, this does not mean that the test is better if we apply it in a population with low disease prevalence, but that, in absolute number, the test gives more correctly classified subjects. This percentage of correctly classified subjects should always be weighed, considering other diagnostic accuracy measures, especially predictive values[122].

- The extent to which being categorized as positive or negative actually predicted the presence of the condition[222]: *positive predictive value* (PPV - the proportion of those with a positive test who have the condition) and *negative predictive value* (NPV - the proportion of those with a negative test who do not have the condition)[223] (Table 12). Similar to sensitivity and specificity, the desired values were as much as possible equal to 1 (or 100%)[211]. Besides, these measures are not invariant characteristics of the tests and depend significantly on the disease's prevalence in the population tested[122, 224, 225]. PPV is increasing, while NPV decreases with the increase of the disease's prevalence in a population. Whereas the change in PPV is more substantial, NPV is somewhat weaker influenced by the disease prevalence[122].

- The ratio of the specific test probability result in people who do have the disease to the probability in people who do not, called *likelihood ratio*[211, 225]. Specifically, the probability that a positive to the test has actually to have the examined condition is the *positive likelihood ratio (*the ratio between the sensitivity and the reciprocal of the specificity). The probability that a negative to the test actually has to have the condition is the *negative likelihood ratio* (the ratio between the reciprocal of the sensitivity and the specificity)[211](Table 12). These two indexes are useful to identify the condition under exam, so if these ratios are equal to 1, the test is not useful, and they are independent of the prevalence of the condition[122, 211, 224]. Likelihood ratios above 10 and below 0.1 are considered to provide strong evidence to rule in or rule out diagnoses respectively in most circumstances[122, 225]. The *post-test odds* that the patient has the disease can be estimated by multiplying the *pre-test odds* by the likelihood ratio (Table 12). In order to avoid the conversion of odds in probabilities, the Fagan nomogram (Figure 35) is available to directly and graphically obtain the *post-test probability* of disease starting from the *pre-test probability*. In our sample, the pre-test probability (disease prevalence) for at least one, two, and three falls were calculated on the results of the PRE.C.I.S.A. sample data, considering thus the ‘fall’ outcome collected in the study.

Figure 35. Fagan's nomogram with examples (from Safari and colleagues, 2016)[226]



NOTES: on the left, the Fagan's nomogram, and on the right, an example of the use of the Fagan's nomogram with a test. The pre-test probability of the disease is 25%; the positive likelihood ratio of this test is 5.0; the negative likelihood ratio is 0.4. Starting from the left axis (heading: prior probability %), we need to localize the 25%. Then, if our patient is positive to the test, we draw a straight line between the left axis and the right one (heading: posterior probability %), passing from the central axis (heading: likelihood ratio) in correspondence to the 5.0 (positive likelihood ratio). In this way, we can know the post-test probability, which is 62.5%. In other terms, using that test, if our patient is positive to the test, his probability of disease increases from 25% to 62.5%.
We can apply the same procedure to calculate the post-test probability in case of a negative test. We only need to connect with a straight line the left and the right axis passing from the central one in correspondence to 0.4: if our patient is negative to test, his probability of disease decreases from 25% to 11.8%.

- The ratio between the odds of positivity in subjects with disease and the odds in subjects without disease, called *Diagnostic Odds Ratio (DOR)*[122] (Table 12). It constitutes another global measure for diagnostic accuracy, used for the general estimation of the discriminative power of diagnostic procedures and the comparison of diagnostic accuracies between two or more diagnostic tests. DOR depends significantly on the sensitivity and specificity of a test: a test with high

specificity and sensitivity with a low rate of false positives and false negatives has high DOR[122].

Table 11. Standard 2x2 table for a diagnostic test

		Disease status		
		Present (+)	Absent (-)	
Diagnostic test	Positive (+)	a - True positive (TP)	b - False positive (FP)	Total positive tests (a+b)
	Negative (-)	c - False negative (FN)	d - True negative (TN)	Total negative tests (c+d)
		Total diseased (a+c)	Total healthy (b+d)	Total sample (a+b+c+d)

Table 12. Test parameters with their respective formulas

Test parameter	Formula
Sensitivity (Sn)	a/(a+c) or TP/(TP+FN)
Specificity (Sp)	d/(d+b) or TN/(TN+FP)
Diagnostic effectiveness or accuracy (DA)	(a+d)/(a+c+b+d) or (TP+TN)/(TP+TN+FP+FN)
Positive Predictive Value (PPV)	a/(a+b) or TP/(TP+FP)
Negative Predictive Value (NPV)	d/(d+c) or TN/(TN+FN)
Positive Likelihood Ratio (LR+)	Sensitivity/(1-Specificity)
Negative Likelihood Ratio (LR-)	(1-Sensitivity)/Specificity
Pre-test probability or prevalence of disease (PrTP)	(a+c)/(a+b+c+d) or taken from literature
Pre-test odds (PrOdds)	PreProb/(1-PreProb)
Post-test odds (positive) (PoOdds+)	PreOdds x LR^+
Post-test odds (negative) (PoOdds-)	PreOdds x LR^-
Post-test probability (positive) (PoTP+)	(PreOdds x LR^+)/(PreOdds x LR^+ + 1)
Post-test probability (negative) (PoTP-)	(PreOdds x LR^-)/(PreOdds x LR^- + 1)
Diagnostic Odds Ratio (DOR)	(TP/FN)/(FP/TN)

NOTES
Abbreviations: a or TP, True Positive; d or TN, True Negative; b or FP, False Positive; c or FN, False Negative; LR^+, Positive Likelihood Ratio; LR^-, Negative Likelihood Ratio; PreProb, Pre-test Probability; PreOdds, Pre-test Odds; PostOdds, Post-test Odds; PostProb, Post-test Probability; DOR, Diagnostic Odds Ratio.

2.2.1.3 Generation of the screening algorithms and their diagnostic accuracy

2.2.1.3.1 Screening algorithms with calibrated scales and their diagnostic accuracy

The generation of the screening algorithms, obtained with serial combinations of the calibrated scales, and the analyses with standard techniques of their diagnostic accuracy in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months were realized. Specifically, we did:

- *Analysis of the strength of the association* between the calibrated scales (Spearman rho). In the case of an absent or weak/moderate correlation (≤0.5) between the described instruments, we used all these scales to generate the screening algorithms. Should a strong correlation between the variables be found, it would be necessary to choose fewer scales for the successive algorithms.

- *Construction of multistep algorithms with serial combinations of the calibrated scales* in predicting at least one, two, and three (recurrent fallers) falls in the following twelve months. To obtain algorithms with higher diagnostic accuracy compared to that of the single tools, the scales were combined using the 'serial' method with the 'AND rule': this procedure implies that the first test is administered to all patients, the second one only to those patients positive to the first test, and the third test only to those positive to the second one. The diagnosis of disease (i.e., fall risk) is formulated when the result is positive to all three tests. This method is particularly cost-efficient, has the advantage of avoiding unnecessary tests and making decisions with fewer tests, but the disadvantage of potentially delaying treatment for diseased patients by lengthening the diagnostic testing period[227-229]. Besides, the 'serial method with AND rule' leads to higher overall specificity and positive predictive value, based on a decrease in false positives, than either test by itself. At the same time, it implies an increase in false negatives compared to the administration of only one of the tests. This suggests that this procedure should be applied in situations where priority should be given to the specificity of diagnosis[227-230].

  In addition to the choice of the described method, it was analyzed the 'cost' ratio between false negative and false positive rates and decided to limit the overall false negative rate to the following percentages: 30-35% for the prediction of ≥1 fall, 20-25% for ≥2 falls, and 10-15% for ≥3 falls. The definition of the optimal cutoff for each scale was based on the satisfaction of this overall result ('ad hoc method'). Globally, six different serial algorithms with the calibrated scales were devised:

  1. Algorithm with scale cutoffs based on the 'Youden Index method', prediction of at least one fall in the following twelve months;

2. Algorithm with scale cutoffs based on the 'Youden Index method', prediction of at least two falls in the following twelve months;
3. Algorithm with scale cutoffs based on the 'Youden Index method', prediction of at least three (recurrent fallers) falls in the following twelve months;
4. Algorithm with scale cutoffs based on the 'ad hoc method', prediction of at least one fall in the following twelve months;
5. Algorithm with scale cutoffs based on the 'ad hoc method', prediction of at least two falls in the following twelve months;
6. Algorithm with scale cutoffs based on the 'ad hoc method', prediction of at three (recurrent fallers) falls in the following twelve months.

- *Analyses with standard techniques of the diagnostic accuracy of the screening algorithms.* The calculation of the diagnostic accuracy measures of the combined tests using the described method (i.e., the algorithm) was realized based on a combined 2x2 table for each predicted outcome (Table 13)[227-230].

Table 13. Combined 2x2 table for serial algorithms with 'AND' rule

		Disease status		
		Present (+)	Absent (-)	
Diagnostic tests	Positive (+) to all three tests	a - True positives (TP) of the algorithm	b - False positives (FP) of the algorithm	Total positive tests (a+b) of the algorithm
	All other combinations	c - False negatives (FN) of the algorithm	d - True negatives (TN) of the algorithm	Total negative tests (c+d) of the algorithm
		Total diseased (a+c) of the algorithm	Total healthy (b+d) of the algorithm	Total sample (a+b+c+d)

These calculations were formulated under the assumption of test independence subject to the actual state of disease. It is assumed that the sensitivity and specificity of the second test do not depend on the result obtained in the first test, and the same for the third test[229]. However, in most situations, this assumption is violated. Under these conditions, the calculation of the positive predictive value of the series, as described in this paragraph, systematically leads to its overestimation[229].

Besides, we performed a supplementary analysis to permit a further interpretation of the diagnostic accuracy of the constructed algorithms. Given

that most single screening tests have not a post-test probability (PoTP) large enough to cross the intervention threshold, each scale's results for all the algorithms were combined to calculate a cumulative PoTP value. In effect, the PoTP of one test became the pre-test probability for the next test. If both pre-test probability and a test/measures likelihood ratio values are low or moderate, the cumulative PoTP can be thought of as increasing surety, and the presence of two or more positive tests with a high cumulative PoTP value (above the baseline pre-test probability) suggests the individual is at high risk of experiencing falls and supports the need for intervention[231, 232].

2.2.1.3.2 Screening algorithms with the two FRAT-up and their diagnostic accuracy

The generation of the screening algorithms, obtained with serial combinations of the FRAT-up VAE and FRAT-up VOE, and the analyses with standard techniques of their diagnostic accuracy in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months were realized, as described for the calibrated scales. The FRAT-up, being already validated and published[120, 121], was used as a comparator for 'external validation'. Specifically, we did:

- *Analysis of the strength of the association* between the two calculated FRAT-up VAE and VOE. In the case of an absent or weak/moderate correlation (≤0.5) between the described instruments, we used the two tools to generate the screening algorithms. Should a strong correlation between the variables be found, it would be necessary to choose only one instrument for the successive algorithms.
- *Construction of multistep algorithms with serial combinations of the two calculated FRAT-up VAE and VOE*, in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months. The methods were those used for the calibrated scales. Globally, six different serial algorithms with FRAT-up VAE and VOE were devised:
    1. Algorithm with FRAT-up cutoffs based on the 'Youden Index method', prediction of at least one fall in the following twelve months;
    2. Algorithm with FRAT-up cutoffs based on the 'Youden Index method', prediction of at least two falls in the following twelve months;

3. Algorithm with FRAT-up cutoffs based on the 'Youden Index' method, prediction of at least three (recurrent fallers) falls in the following twelve months;
4. Algorithm with FRAT-up cutoffs based on the 'ad hoc method', prediction of at least one fall in the following twelve months;
5. Algorithm with FRAT-up cutoffs based on the 'ad hoc method', prediction of at least two falls in the following twelve months;
6. Algorithm with FRAT-up cutoffs based on the 'ad hoc method', prediction of at three (recurrent fallers) falls in the following twelve months.

- *Analyses with standard techniques of the diagnostic accuracy of the screening algorithms.* The methods were those used for the calibrated scales.

2.2.1.3.3 Comparison of the diagnostic accuracy of the generated algorithms

After their generation, we compared the diagnostic accuracy of the serial screening algorithms composed by the calibrated scales vs. those composed by the two calculated FRAT-up in the prediction of the outcomes mentioned above. The comparison concerned the indexes described in the previous paragraphs (Table 12, Table 13, and cumulative PoTP).

2.2.1.4 Construction of additional screening algorithms based on a logistic regression model

Finally, we realized the construction of additional screening algorithms based on a logistic regression model for the same predicted outcomes, integrating the calibrated scales and other available covariates excluded from the scales. The calculated FRAT-up VAE and VOE were used, even in this case, as comparators for 'external validation'. In particular, for each predicted outcome, we did:

- *Univariable logistic regression model*[233-235] for each tool (independent variable) for the prediction of the specific outcome (dependent variable).
- *Generation of the predicted probabilities* of each fitted model.

- *Comparison of the ROC curves and their relative AUCs* of the predicted probabilities of the scales by testing the equality of two or more ROC areas (chi-square test)[213] for each outcome.
- *Multivariable logistic regression models*[233-235]: starting from the scale with the most discriminant ROC curve as a covariate, we added a new scale (the second with the highest AUC) as a new covariate for a second attempt. Then, we added the third scale in terms of AUC, as a further covariate, for the third attempt. Each time we compared the new multivariable model with the previous model, which demonstrated the highest discriminant power (highest AUC of the predicted probabilities). Finally, we chose the best discriminant model between those subjects who experienced the outcome of interest versus those who did not for each predicted outcome to be submitted to the following analytical steps.
- *Goodness-of-fit test of the best discriminant model* (i.e., screening algorithm) *to the sample data* for each predicted outcome (Hosmer-Lemeshow goodness-of-fit test[235, 236]).
- Realization of a *graph with the sensitivity and specificity* versus the *probability cutoff* of the best discriminant model (i.e., screening algorithm)[235] for each predicted outcome. This graph could represent an additional way to choose the optimal cutoff for classification purposes, which may vary according to the desired trade-off between false positives and false negatives, as already discussed in the previous section[235].
- *Comparison of the ROC curves and their relative AUCs* of *the best discriminant model* (i.e., screening algorithm) composed by the calibrated scales vs. that composed by the two calculated FRAT-up for the specific outcome in terms of discriminant power, by testing the equality of two or more ROC areas (chi-square test)[213].

The described analysis and the successive reporting of the results have been conducted according to the STARD 2015 guidelines for reporting diagnostic accuracy studies[237, 238].

2.2.2 Assessment of the effect of neurological diseases on the fall risk screening tests

Regarding the statistical analyses about the secondary objective (a) of the *assessment of the effect of neurological diseases on the fall risk screening tests,* the following analyses were realized:

- Analysis of the *presence of differential item functioning (DIF) or item bias* [182, 185, 186, 190] for two subgroups of neurological diseases (elderly and elderly with associated neurological diseases like PD or stroke), in the context of Rasch analysis, for each of the item of the calibrated measurement scales (VAE and VOE scales) on the primary objective, according to the methods described in the previous paragraph on Rasch analysis.
- *An independent-samples t-test* to compare two subgroups of neurological diseases (elderly and elderly with associated neurological diseases like PD or stroke) for the two versions of the FRAT-up calculated on the VAE and VOE variables. In the case of the detection of a significant difference between the estimates of the two subgroups, an effect size (Cohen's d) will be calculated to determine the magnitude of this effect.

2.2.3 Validation of an ICD&ICF core set for the fall risk in community-dwelling older adults (also with associated neurological diseases)

Regarding the secondary objective (b) of the *validation of an ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases,* the following steps were realized:

- Searching for the most complete and recent systematic reviews with meta-analysis on the fall risk factors in the elderly, also with associated neurological diseases. The search was conducted on the Pubmed website (https://pubmed.ncbi.nlm.nih.gov/) in the 'free search field', indicating 'meta-analysis' in the 'article type filter', and using the following search strings:
  - Older people: 'risk factors falls community-dwelling older people';
  - Parkinson's Disease: 'risk factors falls Parkinson's Disease community-dwelling older people';

  - Stroke: ‘risk factors falls community-dwelling stroke older people’.

- Extraction from each of the systematic reviews of the following data:
  - Fall risk factors;
  - Measures of the effect of each extracted risk factor (odds ratio, risk ratio, or hazard ratio) in the comparison between fallers vs. non-fallers in the following twelve months and between recurrent fallers (>3 falls per year) vs. non-fallers in the same period;
  - Bibliographic reference(s) for each fall risk factor and related measure of effect.
- Linking of each risk factor with an ICD category or an ICF category, according to the coded linking rules published by Cieza and colleagues[135-137].
- Unification of all risk factors into a new unique core set.
- Comparison of the new core set with the other available ‘ICF core set for falls in acute rehabilitation settings’ by Yen and colleagues published in 2014[143]. This core set was developed in three steps (a systematic review to identify risk factors, linking of the factors to the ICF categories, three Delphi surveys to reach a consensus on the selection of categories), and it contains 88 fall risk factors linked to 66 ICF categories (33 belonged to ‘Body functions’, 3 to ‘Body structures’, 19 to ‘Activities and participation’, and 11 to ‘Environmental factors’), and 5 personal factors. Besides, it refers to the description of the fall risk in a hospital setting, particularly ‘acute rehabilitation’. According to the authors, it may also be a useful tool for applying the ICF model in preventing falls in other clinical settings.
- Integration of the two above-mentioned core sets to create a new comprehensive ICD&ICF core set for falls in community-dwelling older adults, also with associated neurological diseases.

2.3 Statistical notes, software and sample size issues

All descriptive statistics and correlations and the logistic regression were performed using STATA software (STATA/IC 13.1 for Windows).

CFA for ordinal data was undertaken using the Mplus software (Mplus version 6.0. Muthen & Muthen, Los Angeles, CA; 1998–2010; www.statmodel.com). It was estimated that 768 observations for the VAE variables (ratio of subjects towards 7 items 109.7:1) and 574 observations for the VOE variables (ratio of subjects towards 22 items 26.1:1) would constitute adequate samples for this analysis[239].

Mokken analysis was performed using R (R version 3.5.0 (2018-04-23)). Sample sizes of 768 observations for the VAE variables and 574 for observations for the VOE variables were considered sufficient to perform this kind of analysis[240].

Rasch analysis was carried out using the RUMM2030 software (version 5.52 for Windows. RUMM Laboratory Pty Ltd, Perth, Australia; 1997–2016; www.rummlab.com). We estimated that the two sample sizes of 768 (VAE variables) and 574 observations (VOE1 and VOE2 variables) would be sufficient to estimate item difficulty, with α of .01 to <±.5 logits, irrespective of the targeting of persons to the items [241]. A significance value of 0.05 was used throughout and corrected for the number of tests by Bonferroni correction [242]. For all reliability analyses, cutoffs of >0.70 and >0.90 were considered adequate for group and individual person measurements, respectively[168, 243]. To facilitate the interpretation of the results of each Rasch analysis, an ad hoc Excel 2007™ application (RUMM logbook) (La Porta F, on behalf of the ERRTG (European Rasch Research & Teaching Group). RUMM Logbook v1.9.5. Bologna, Italy; 2018), developed using Microsoft Visual Basic™ macros (Microsoft Office Excel for Windows, version 12.0. Microsoft Corporation, Redmond, WA, USA; www.microsoft.com)[171], was used.

The analysis of the diagnostic accuracy of the scales and the algorithms were performed using an on-line software 'easyROC: a web-tool for ROC curve analysis' (ver. 1.3.1)[244], and an ad hoc Excel 2007™ spreadsheet with formulas.

To facilitate the interpretation of the absolute values of correlation coefficients, a modified version[245] of the cutoff criteria provided by Pallant[246] was adopted: negligible: 0–0.09; weak: 0.10–0.29; moderate: 0.30–0.49; strong: 0.50–0.79; very strong: ≥ 0.80.

Effect sizes were interpreted according to the criteria provided by Cohen[247]: small: 0.20–0.49; medium: 0.50–0.79; large: ≥ 0.80.

3 Results

3.1 Validation of a fall risk serial screening algorithm with a high level of diagnostic accuracy

3.1.1 Calibration of the VAE scale

3.1.1.1 Extraction of the variables of interest from the PRE.C.I.S.A. study dataset

As described in the methods chapter, the variables of interest regarding the VAE assessment were extracted from the PRE.C.I.S.A. study dataset.

3.1.1.2 Descriptive statistics of the sample and items

The sample was constituted of 768 community-dwelling elderly, also with associated neurological diseases, and derived from the PRE.C.I.S.A. study sample as described in the methods chapter. The principal demographic and clinical characteristics of these patients were reported in Table 14.

The average age of the subjects was 76.3 years (Standard Deviation - SD 6.6), of which females were 65.6% of the total sample. Concerning the presence of neurological diseases, 29.7% was also affected by an associated neurological disease on the whole sample of older adults. In detail, most part were elderly with 65-80 years (24.6%), followed by older adults with 80 years (10.3%), those with associated PD (10.4%), and finally those with associated stroke sequelae (8.2%). Unfortunately, 53.5% of this detailed information was not available because it was recorded in a successive step of assessment (VOE). Almost half of the sample was signaled by the Rehabilitation Units (48.7%) regarding the recruitment sources. A further 22.4% was self-reported, whereas General Practitioners signaled only 10.7% of the subjects.

Almost 53% of the sample was constituted by those randomized in the PRE.C.I.S.A. experimental and control groups, whereas the remaining part by subjects ‘not recruitable’ because at low fall risk at the VAE and/or VOE selection steps. Regarding VAE, 72% of the sample was at high risk, 22% at moderate risk, while only one subject was ‘not recruitable’ at this step for ‘not satisfied criteria’. Regarding VOE, instead, 452 persons have been judged as ‘recruitable’ and then ‘randomizable’, whereas the

remaining 122 ‘not recruitable’ because at low risk. This decision was based on combining the risks from two tests, the FROP-COM screen[148] and the FRAT[149]. As it is notable in Table 14, the instruments did not always define a specific subject as a member of the same ‘risk class’. Indeed, the FRAT provided a systematic lower assessment of the fall risk than the FROP-COM screen (low risk 53.4% vs. 16.4%; medium risk 23.7% vs. 43%; high risk 1.0 vs. 18.8).

Finally, regarding the number of falls at twelve months (the primary endpoint of the PRE.C.I.S.A. study), the mean for this sample was 1 (SD 3), with 37.4% of subjects who fell at least one time. If we consider this variable distinctly between the randomized subjects (R) and non-randomized (nR), it is notable that the average number of falls at twelve months in the R group has been higher compared to the nR group (2 vs. 0, t=12.6, Degrees of Freedom (DF)=475, p-value <0.000), as well as the number of subjects who fell at least one time (55.8% vs. 17%, chi-square=123.5, p-value <0.000).

Table 14. Principal characteristics of the used sample (N=768)

	N	%	Mean	SD	Media	Range
Age (years)	768	100	76.3	6.6	76	[65, 100]
Gender	768	100				
Male	264	34.4				
Female	504	65.6				
Presence of neurological diseases	768	100				
Older adults	540	70.3				
Older adults with neurological diseases	228	29.7				
Presence of neurological diseases (detailed)	768	100				
Older adults 65-80 years	189	24.6				
Older adults >80 years	79	10.3				
Older adults with stroke sequelae	63	8.2				
Older adults with PD	80	10.4				
Not available	357	53.5				
Recruitment sources	768	100				
Self-reporting	172	22.4				
General Practitioner	82	10.7				
Rehabilitation Medicine	374	48.7				
Neurology	73	9.5				
Geriatrics	34	4.4				
Other hospital units	10	1.3				
Other	1	0.1				
Not available	22	2.9				
PRE.C.I.S.A. study randomization	768	100				
Yes	403	52.5				
No	365	47.5				
Global risk at VAE	768	100				
Non recruitable (not satisfied criteria)	42	5.5				
Non recruitable (low risk)	1	0.1				
Moderate risk	170	22.1				

High risk	555	72.3				
Risk according to FROP-COM screen	768	100				
Low risk	126	16.4				
Medium risk	330	43.0				
High risk	144	18.8				
Not available	168	21.9				
Risk according to FRAT	768	100				
Low risk	410	53.4				
Medium risk	182	23.7				
High risk	8	1.0				
Not available	168	21.9				
Falls at 12 months	768	100	1	3	0	[0, 33]
Falls at 12 months (R vs nR)	768	100				
Randomized	403		2	3	1	[0, 33]
Non-randomized	365		0	1	0	[0, 3]
Subjects with at least 1 fall	768	100				
Yes	287	37.4				
No	481	62.6				
Subjects with at least 1 fall (R)	403	100				
Yes	225	55.8				
No	178	44.2				
Subjects with at least 1 fall (nR)	365	100				
Yes	62	17.0				
No	303	83				

NOTES
Abbreviations: N, Number of subjects; SD, Standard Deviation; PD, Parkinson's Disease; VAE, Valutazione Anamnestica dell'Eleggibilità; VOE, Valutazione Oggettiva dell'Eleggibilità; FRAT, Fall Risk Assessment Tool; R, Randomized; nR, non-Randomized.

All score categories of the seven items were available, and no missing data were present. The average inter-item correlation (AIC) was 0.161 (range: [-0.019,0.320]). The item-to-total correlations were moderate (median value: 0.413), ranging from 0.158 (VAE00) to 0.917 (VAE02). All these correlations were lower than 0.460, except for VAE02, which had a significantly higher correlation (0.917). Besides, the Cronbach's alpha was 0.381, indicating an unacceptable internal consistency upon the assumption of unidimensionality. Only if deleting VAE02, this value increased to 0.468, indicating the lowest correlation values with the other items.

3.1.1.3 Preliminary assessment of dimensionality

3.1.1.3.1 Confirmatory Factor Analysis

The baseline Confirmatory Factor Analysis (CFA), undertaken on the whole sample (N=768), failed to support the unidimensionality of the scale (RMSEA= 0.078; SRMR= 0.082; CFI= 0.874; TLI= 0.810). However, two pairs of items showed large

modification indices (MI) (VAE02–03=57.4; VOE06-03=27.9). After allowing the highest correlation of the errors within the VAE02-03 pair, it was possible to well-fit a final model for the VAE scale, reaching the intended cutoff for a preliminary unidimensionality before Rasch analysis (RMSEA = 0.029; SRMR= 0.057; CFI = 0.983; TLI = 0.973).

3.1.1.3.2 Mokken analysis

The automated item selection procedure (AISP) within the Mokken Analysis (MA) showed the scalability of 5 items (VAE00-01-04-05-06) on scale 1 and the remaining 2 items (VAE02-03) on scale 2. All item-pair scalability coefficients Hijs inside the respective scales were positive, satisfying the first criterion of a Mokken scale. VAE01 item scale coefficient Hj was less than the lower bound c = 0.3 and violated the second criterion of a Mokken scale. Finally, the scalability coefficient for the entire scale, H, was equal to 0.29, which was too low even for the qualification 'weak scale'[178].

3.1.1.4 Rasch analysis

Despite the preliminary evidence of the multidimensionality of the 7-item set highlighted by the Mokken analysis, we decided to submit the entire set of items to the Rasch analysis.

The base Rasch analysis showed that the scale, as a whole, did not fit the Rasch model (Table 15, analysis 1), failing the item homogeneity or invariance requirement (χ^2_{49}=124.3; p=0.0000). One item (VAE02) showed fit residuals <-2.5, thus suggesting that the responses to this item were too predictable (model overfit). One item (VAE03), instead, had fit residuals >2.5, thus highlighting that responses to this item were too unpredictable (model underfit). Beyond this, there were highly significant chi-squares for three items (VAE01-02-04), suggesting a violation of the requirement of stochastic invariance of the item hierarchy. The scale satisfied the unidimensionality requirement, as the Proportion of Significant T-test (PST) and the Lower Bound of Binomial Confidence Interval for proportions (LBBCI) were 0%. Furthermore, there were disordered thresholds for VAE02 (violation of the monotonicity requirement), and one pair of items (VAE04-06) had residual correlations above the local dependence relative cutoff (LDRC, here set at 0.112, indicative of a violation of the local independence requirement).

Table 15. Rasch analysis details for the VAE scale

No	Description of analysis	N	Fitness to the Rasch Model						Targeting					Separation reliability					Unidimensionality	
			FitRes Items		FitRes Persons		Item-trait interaction		Person location		Floor effect (%)	SEM	Targeting index	DDR			DIR			
			Mean	SD	Mean	SD	χ^2_{df}	P^	Mean	SD				PSI	α	Strata	DLPA	DI-PSI	PST	BCI
1	Base analysis	768	0.051	3.076	-0.157	0.301	124.3_{49}	0.000	1.372	1.136	0.4	0.893	1.536	0.637	0.381	2.1	3	0.900	0%	0%
2	After rescoring	768	-0.187	1.980	-0.157	0.302	78.3_{35}	0.000	1.433	1.319	0.4	0.916	1.564	0.520	0.517	1.7	2	0.800	4.1%	3.4%
3	After subtesting	768	0.305	5.337	-0.384	0.736	36.0_{10}	0.000	0.934	1.340	0.4	0.945	0.989	0.517	0.502	1.7	2	0.800	0.1%	0%
4	After splitting	768	-1.056	5.321	-0.228	0.614	43.4_{13}	0.000*	0.424	3.411	0.4	n/a	0.420**	0.912	n/a	4.6	n/a	n/a	n/a	n/a

NOTES

Abbreviations: FitRes, Fit Residual; DDR, Distribution-Dependent Reliability; DIR, Distribution-Independent Reliability; SD, Standard Deviation; $\chi 2_{df}$, chi-square and its degrees of freedom; p, Bonferroni-corrected χ2 probability value; SEM, Standard Error of Measurement of the person locations; PSI, Person Separation Index; α, Cronbach's alpha; DLPA, Distinct Levels of Performance Ability; DI-PSI, Distribution-Independent Person Separation Index (based on DLPA); LD, Local Dependence; PST, Proportion of Significant T-test carried out on the estimates that, within a principal component analysis of residuals, loaded positively and negatively (factor loading >±.3) on the first component; BCI, Binomial Confidence Interval for PST; n/a, not available.

Values are mean (SD) or as otherwise indicated.

^P-value is considered significant when <0.05

*P=0.080 for a sample size of 350 to handle type I error due to a larger sample size

**Targeting index based on the PSI

In the next steps of analysis, we performed several item modifications to achieve a final fitting solution of the scale:

- One item was rescored (VAE02), as it showed disordered thresholds (Table 16);
- 'Testlets' (or super-items) were created according to the 'alternative 2-testlet approach'[207], thus obtaining a two-testlet solution (testlet 1: VAE01: history of falling, VAE03: diagnosis of stroke or PD and VAE05: inability to rise from a chair; testlet 2: VAE00: at fall risk, VAE02: medications, VAE04: balance problems and VAE06: fear of falling).
- The testlet 1 was splitted to resolve a uniform DIF for neurological diseases. For further details on the analysis, see results section '3.2 Assessment of the effect of neurological diseases on the fall risk screening tests'.

Table 16. Item parameters and fit statistics for the VAE scale (N=768, analysis no. 2)

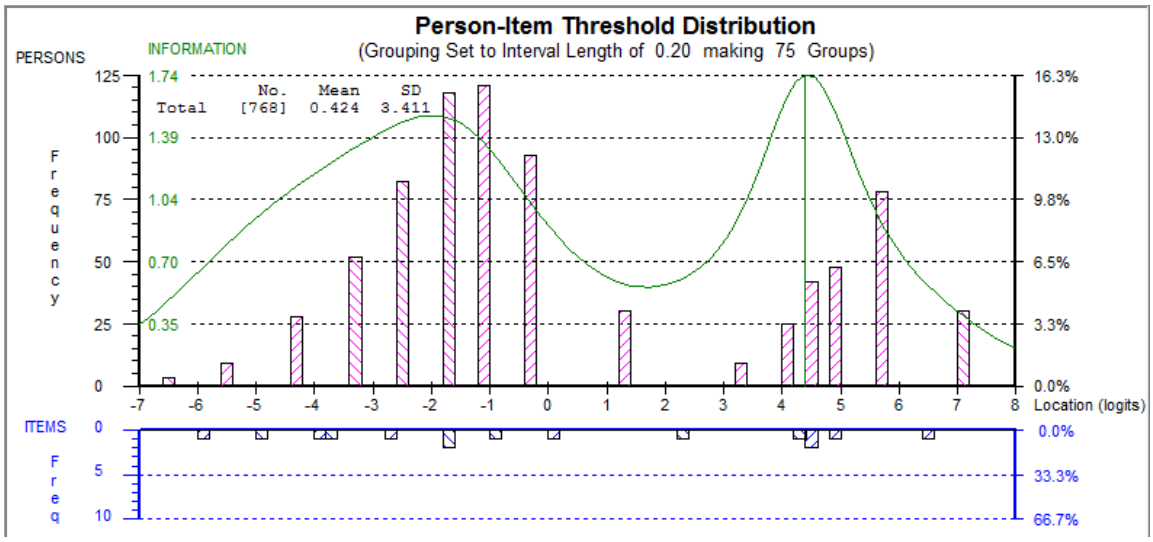
	Item parameters and fit statistics					Scoring model						
VAE items	**Loc**	**SE**	**FR**	**χ^2**	**p***	**0**	**1**	**2**	**3**	**4**	**5**	**6**
VAE00 – At fall risk	-4.182	0.382	-0.745	1.865	0.867	0	1					
VAE05 – Inability to rise from a chair	-2.288	0.189	-0.368	6.324	0.276	0	1					
VAE04 – Balance problems	0.499	0.088	-1.729	21.558	0.000	0	1					
VAE06 – Fear of falling	0.813	0.085	0.076	13.413	0.020	0	1					
VAE01 – History of falling	1.293	0.083	1.152	14.080	0.015	0	1					
VAE02 – Medications	1.466	0.049	-2.903	8.810	0.117	0	1	1	2	2	2	3
VAE03 – Diagnosis of stroke or PD	2.399	0.087	3.211	12.237	0.032	0	1					

NOTES: VAE items are ordered by progressively increasing difficulty from top to bottom. The location is expressed in logits. The degrees of freedom for each χ2 were 5 for all items.
*The Bonferroni-corrected p-value indicating statistical significance at the 0.05 level was 0.007.
Abbreviations: VAE, Valutazione Anamnestica dell'Eleggibilità; Loc, Location; SE, Standard Error; FR, Fit Residual; χ2, chi-square; P, χ2 Probability.

After these modifications, the final 7-item solution for the VAE scale showed a satisfying fit to the Rasch model (Table 15, analysis 4). The scale was strictly unidimensional (overall PST 0.1%, LBBCI 0%, c=0.137, r=0.981, A=0.972), and also satisfied all the other Rasch model requirements in terms of invariance (unconditional χ^2_{13}=20.7; p=0.080 for a sample size of 350 to handle type I error due to larger sample size, conditional class-interval based not available for analysis with splitting), local independence (no pairs of residual correlations of items above the LDRC) and monotonicity (no disordered thresholds). All the subjects' responses fitted the model.

The targeting graph of the VAE scale (Figure 36) highlighted that subjects were spread across fourteen logits, with negligible floor (0%) and ceiling effects (4%). The mean person ability of 0.424 logits and a targeting index of 0.420 (calculated on the PSI) indicated, on average, a 'very good' matching between person ability and item difficulty (set by default at 0 logits)[197].

Figure 36. Targeting of the VAE scale (n=768)



NOTES: in the picture, persons and items were displayed, respectively, in the upper and the lower part of the graph, separated by the logit scale. Grouping set to interval length of 0.20, making 75 groups.
Abbreviation: Freq, Frequency.

The separation reliability expressed as PSI was 0.912, indicating precision of measurement at the individual level[248]. Given the PSI, persons could be separated by the scale in 4.6 strata, i.e., the statistically distinct levels of ability that the VAE could reliably distinguish in this sample[194]. The calculation of the Cronbach's Alpha, of the distinct levels of performance ability, and the related distribution-independent person separation index were not possible due to missing data following the splitting analysis.

The item hierarchy (Table 15, analysis 2 and Table 16) was consistent with the theoretical expectations about the hierarchy levels of the latent variable. Notably, it suggested that the earlier aspects which highlight the presence of fall risk were the health professionals' or the own subject's perception to be at fall risk (VAE00), the inability to rise from a chair (VAE05), and the presence of balance problems (VAE04). On the other hand, the latest aspects linked to the fall risk were the number of

medications taken by the subject (VAE02) and the associated diagnosis of stroke or PD (VAE03).

Based on the item calibration, it was possible to construct two tables for older adults with or without associated neurological disease, to convert the scale raw scores into interval-level estimates of the fall risk (Table 17).

Table 17. Raw-score-to-measure-estimates conversion tables for the VAE scale

Older adults				
Raw score	**Logit scale**	**±95% CI**	**0-9 scale**	**±95% CI**
0	-6.577	3.048	0.0	1.0
1	-5.437	2.356	0.8	0.8
2	-4.270	2.066	1.5	0.7
3	-3.246	1.868	2.2	0.6
4	-2.434	1.723	2.7	0.6
5	-1.756	1.674	3.2	0.6
6	-1.095	1.748	3.6	0.6
7	-0.306	2.029	4.1	0.7
8	1.239	3.375	5.2	1.1
9	5.466	-	8.0	-
Older adults with associated neurological disease				
Raw score	**Logit scale**	**±95% CI**	**0-9 scale**	**±95% CI**
0	-5.55	-	0.7	-
1	-4.253	2.689	1.5	0.9
2	-2.646	-	2.6	-
3	-0.206	3.847	4.2	1.3
4	3.349	2.001	6.6	0.7
5	4.032	1.568	7.0	0.5
6	4.474	1.490	7.3	0.5
7	4.933	1.597	7.6	0.5
8	5.726	2.080	8.1	0.7
9	7.03	3.069	9.0	1.0

NOTES: two raw-score-to-measure-estimates conversion tables are presented, the upper for older adults without associated neurological disease and the lower for older adults with associated neurological diseases (PD or stroke). The creation of two different tables was needed to account for detected DIF by neurological diseases. Where ±95%CI calculated based on the standard error was not available, this was attributable to missing data generated by the splitting analysis.
Person estimates were expressed in logits and into a 0 to 9 scale (VAE original scaling).
Abbreviations: 95%CI, 95% Confidence Interval (equal to 1.96 standard errors of measurement).

3.1.2 Calibration of the VOE scale

3.1.2.1 Extraction of the variables of interest from the PRE.C.I.S.A. study dataset

As described in the methods chapter, the variables of interest regarding the VOE assessment were extracted from the PRE.C.I.S.A. study dataset.

3.1.2.2 Descriptive statistics of the sample and items

The sample was constituted of 574 community-dwelling elderly, also with associated neurological diseases, and derived from the PRE.C.I.S.A. study sample as described in the methods chapter. The principal demographic and clinical characteristics of these patients were reported in Table 18.

The average age of the subjects was 75.8 years (SD 6.4), of which females were 65.3% of the total sample. Concerning the presence of neurological diseases, 29.1% was also affected by an associated neurological disease on the whole sample of older adults. In detail, most part were elderly with 65-80 years (32.2%), followed by older adults with 80 years (13.6%), those with associated PD (12.4%), and finally those with associated stroke sequelae (10.5%). Unfortunately, 31.3% of this detailed information was not available. Almost half of the sample was signaled by the Rehabilitation Units (45.1%) regarding the recruitment sources. A further 27.5% was self-reported, whereas General Practitioners signaled only 11% of the subjects.

Almost 70% of the sample was constituted by those who were randomized in the PRE.C.I.S.A. experimental and control groups, whereas the remaining 30% by subjects ‘not recruitable’ because at low fall risk at the VAE and/or VOE selection steps. Regarding VAE, 80% of the sample was at high risk, 19.9% at moderate risk, while only one subject was ‘not recruitable’ at this step for ‘not satisfied criteria’. Regarding VOE, instead, 452 persons have been judged as ‘recruitable’ and then ‘randomizable’, whereas the remaining 122 ‘not recruitable’ because at low risk. This decision was based on combining the risks from two tests, the FROP-COM screen[148] and the FRAT[149]. As it is notable in Table 18, the instruments did not always define a specific subject as a member of the same ‘risk class’. Indeed, the FRAT provided a systematic lower assessment of the fall risk than the FROP-COM screen (low risk 22% vs. 70%; medium risk 28.7% vs. 56.4%; high risk 1.26 vs. 21.6).

Finally, regarding the number of falls at twelve months (the primary endpoint of the PRE.C.I.S.A. study), the mean for this sample was 1 (SD 3), with 43.5% of subjects who fell at least one time. If we consider this variable distinctly between the randomized subjects (R) and non-randomized (nR), it is notable that the average number of falls at twelve months in the R group has been significantly higher compared to the nR group (2 vs. 0, test statistic t=9.2, DF=476, p-value <0.000), as well as the number of subjects who fell at least one time (55.9% vs. 18%, chi-square=73.8, p-value <0.000).

Table 18. Principal characteristics of the used sample (N=574)

	N	%	Mean	SD	Media	Range
Age (years)	574	100	75.8	6.4	75.6	[65, 99]
Gender	574	100				
Male	199	34.7				
Female	375	65.3				
Presence of neurological diseases	574	100				
Older adults	407	70.9				
Older adults with neurological diseases	167	29.1				
Presence of neurological diseases (detailed)	574	100				
Older adults 65-80 years	185	32.2				
Older adults >80 years	78	13.6				
Older adults with stroke sequelae	60	10.5				
Older adults with PD	71	12.4				
Not available	180	31.3				
Recruitment sources	574	100				
Self-reporting	158	27.5				
General Practitioner	63	11.0				
Rehabilitation Medicine	259	45.1				
Neurology	56	9.8				
Geriatrics	15	2.6				
Other hospital units	7	1.2				
Other	1	0.2				
Not available	15	2.6				
PRE.C.I.S.A. study randomization	574	100				
Yes	386	67.2				
No	188	32.8				
Global risk at VAE	574	100				
Non recruitable (not satisfied criteria)	1	0.2				
Non recruitable (low risk)	0	0.0				
Moderate risk	114	19.9				
High risk	459	80.0				
Global risk at VOE	574	100				
Non recruitable	122	21.3				
Recruitable	452	78.7				
Risk according to FROP-COM screen	574	100				
Low risk	126	22.0				
Medium risk	324	56.4				
High risk	124	21.6				
Risk according to FRAT	574	100				
Low risk	402	70.0				

Medium risk	165	28.7				
High risk	7	1.2				
Falls at 12 months	574	100	1	3	0	[0, 33]
Falls at 12 months (R vs nR)	574	100				
Randomized	386		2	4	1	[0, 33]
Non-randomized	188		0	1	0	[0, 3]
Subjects with at least 1 fall	574	100				
Yes	250	43.5				
No	324	56.5				
Subjects with at least 1 fall (R)	386	100				
Yes	216	55.9				
No	170	44.1				
Subjects with at least 1 fall (nR)	188	100				
Yes	34	18.0				
No	154	82.0				

NOTES
Abbreviations: N, Number of subjects; SD, Standard Deviation; PD, Parkinson's Disease; VAE, Valutazione Anamnestica dell'Eleggibilità; VOE, Valutazione Oggettiva dell'Eleggibilità; FRAT, Fall Risk Assessment Tool; R, Randomized; nR, non-Randomized.

All score categories of the twenty-two items were represented, and no missing data were present. The average inter-item correlation was 0.246 (range: [-0.077,0.911]). The item-to-total correlations were strong[245, 246] (median value: 0.544), ranging from 0.006 (VOE22) to 0.748 (VOE08). All these correlations were higher than 0.300, except for VOE22, which had a significantly lower correlation (0.006). Besides, the Cronbach's alpha was 0.860, indicating good internal consistency upon the assumption of unidimensionality. If deleting VOE12 and VOE22, this value increased to 0.864 and 0.863, indicating the lowest correlation values with the other items. On the other hand, if deleting VOE08 and VOE17, the Cronbach's alpha decreased to 0.844 and 0.846, indicating the highest correlation values with the other items.

3.1.2.3 Preliminary assessment of dimensionality

3.1.2.3.1 Confirmatory Factor Analysis

The baseline CFA, undertaken on the whole sample (N=574), failed to support the unidimensionality of the scale (RMSEA= 0.155; SRMR= 0.180; CFI= 0.824; TLI= 0.805). However, twenty-one pairs of items showed large modification indices (MI), between 10.2 and 484.7. After allowing correlation of the errors within the dependent pairs, it was possible to well-fit a final model for the VOE scale, reaching the intended cutoff for a preliminary unidimensionality before Rasch analysis (RMSEA = 0.036; SRMR= 0.059; CFI = 0.991; TLI = 0.989).

3.1.2.3.2 Mokken analysis

The automated item selection procedure within the MA showed the scalability of 13 items (VOE01-02-03-04-05-06-07-08-09-10-13-15-17) on scale 1, of 5 items (VOE11-12-14-16-18) on scale 2, and 2 items (VOE19-20) on scale 3. Finally, two items (VOE21: cognitive state and VOE22: visual acuity) were not scalable. All item-pair scalability coefficients Hijs inside the respective scales were positive, satisfying the first criterion of a Mokken scale. Eight out of twenty-two item scale coefficient Hjs were less than the lower bound c = 0.3 and violated the second criterion of a Mokken scale; especially, the item-scalability coefficient for VOE22 was even less than zero (Hj = -.02). Finally, the scalability coefficient for the entire scale, H, was equal to 0.351, which allowed the qualify as 'weak scale'[178].

3.1.2.4 Rasch analysis

According to the evidence of the multidimensionality of the 22-item set highlighted by the Mokken analysis, the Rasch analysis was conducted separately on the scale 1 (13 items) and the scale 2 (5 items). The remaining 4 items (VOE19-20 scalable on scale 3 and VOE21-22 not scalable) were not considered for the calibration of the VOE scale through RA.

3.1.2.4.1 Rasch analysis on VOE items scalable on Mokken scale 1 (VOE1)

The first RA was conducted on the 13-item set scalable by the MA on scale 1: VOE01-02-03-04-05-06-07-08-09-10-13-15-17.

The base Rasch analysis showed that the scale, as a whole, did not fit the Rasch model (Table 19, analysis 1), failing the item homogeneity or invariance requirement (χ^2_{78}=322.5; p=0.0000). Three items (VOE01-03-04) showed fit residuals <-2.5, thus suggesting that the responses to these items were too predictable (model overfit). One item (VOE07), instead, had fit residuals >2.5, thus highlighting that responses to this item were too unpredictable (model underfit). Beyond this, there were highly significant chi-squares for eight items, suggesting a violation of the requirement of stochastic invariance of the item hierarchy. Also, the scale failed the unidimensionality requirement, as the Proportion of Significant T-test (PST) was 11.1%, and the Lower Bound of Binomial Confidence Interval for proportions (LBBCI) was 10.2%. Furthermore, there were disordered thresholds for three items (VOE07-13-15),

indicating a violation of the monotonicity requirement. Nine pairs of items had residual correlations above the local dependence relative cutoff (LDRC, here set at 0.150), indicating a violation of the local independence requirement.

Table 19. Rasch analysis details for the VOE items scalable on Mokken scale 1 (VOE1)

No	Description of analysis	N	Fitness to the Rasch Model						Targeting					Separation reliability					Unidimensionality	
			FitRes Items		FitRes Persons		Item-trait interaction		Person location		Floor effect (%)	SEM	Targeting index	DDR			DIR			
			Mean	SD	Mean	SD	χ^2_{df}	P^	Mean	SD				PSI	α	Strata	DLPA	DI-PSI	PST	BCI
1	Base analysis	574	-0.662	2.385	-0.246	0.884	322.5_{78}	0.000	-1.123	1.779	1	0.610	-1.842	0.890	0.889	4.1	6	0.973	11.1%	10.2%
2	After rescoring	574	-0.873	2.033	-0.314	0.832	206.0_{78}	0.000	-1.303	2.011	1	0.653	-1.998	0.894	0.893	4.2	5	0.962	7.2%	6.3%
3	After subtesting	574	-0.447	2.649	-0.486	0.781	17.5_{12}	0.130	-0.909	1.261	1	0.514	-1.768	0.817	0.834	3.1	3	0.900	2.1%	1.2%

NOTES

Abbreviations: FitRes, Fit Residual; DDR, Distribution-Dependent Reliability; DIR, Distribution-Independent Reliability; SD, Standard Deviation; $\chi 2_{df}$, chi-square and its degrees of freedom; p, Bonferroni-corrected χ2 probability value; SEM, Standard Error of Measurement of the person locations; PSI, Person Separation Index; α, Cronbach's alpha; DLPA, Distinct Levels of Performance Ability; DI-PSI, Distribution-Independent Person Separation Index (based on DLPA); LD, Local Dependence; PST, Proportion of Significant T-test carried out on the estimates that, within a principal component analysis of residuals, loaded positively and negatively (factor loading >±.3) on the first component; BCI, Binomial Confidence Interval for PST.

Values are mean (SD) or as otherwise indicated.

^P-value is considered significant when <0.05

In the next steps of analysis, we performed several item modifications to achieve a final fitting solution of the scale:

- Three items were rescored (VOE07-13-15), as they showed disordered thresholds;
- 'Testlets' (or super-items) were created according to the mixed application of the 'traditional testlet approach'[207], which creates testlets oriented at conceptually associated items and based on their residual correlations, and the 'factor loading 2-testlet approach', which creates two testlets with items with similar Principal Component Analysis loading factor. Thus, we obtained a two-testlet solution (testlet 1: VOE01: 10MWT, VOE02: 10MWT_aid, VOE03: TUG, VOE04: TUG_aid, VOE08: SPPB_Walking; testlet 2: VOE05: standing balance, VOE06: chair stand test, VOE07: SPPB_Balance, VOE09: SPPB_Chair stand, VOE13: ADL pre-fall, VOE15: IADL pre-fall, VOE17: balance stability).

Table 20. Item parameters and fit statistics for the VOE1 scale (N=574, analysis no. 2)

	Item parameters and fit statistics					**Scoring model**				
VOE1 items	**Loc**	**SE**	**FR**	**χ^2**	**p***	**0**	**1**	**2**	**3**	**4**
VOE06 – Chair stand test	-2.411	0.111	-0.197	10.408	0.108	0	1			
VOE09 – SPPB_Chair stand	-1.452	0.059	2.293	11.202	0.082	0	1	2	3	4
VOE05 – Standing balance	-1.407	0.109	-0.162	12.187	0.058	0	1			
VOE03 – Timed Up&Go test	-1.386	0.109	-4.748	44.434	0.000	0	1			
VOE17 – Balance stability	-0.592	0.082	-1.172	2.945	0.816	0	1	2	3	
VOE07 – SPPB_Balance	-0.396	0.085	1.180	8.325	0.215	0	1	1	2	
VOE08 – SPPB_Walking	-0.029	0.069	-1.836	17.710	0.007	0	1	2	3	4
VOE02 – 10 meters Walking test_aid	0.496	0.125	-1.479	12.402	0.053	0	1			
VOE15 – IADL pre-fall	0.735	0.094	0.368	10.866	0.092	0	1	1	2	
VOE04 - Timed Up&Go test_aid	0.855	0.130	-3.159	22.520	0.000	0	1			
VOE01 – 10 meters Walking test	1.281	0.101	-3.467	18.151	0.006	0	1	2		
VOE10 – Functional reach test	1.924	0.156	1.081	28.669	0.000	0	1			
VOE13 – ADL pre-fall	2.381	0.116	-0.054	6.175	0.404	0	1	1	2	

NOTES: VOE1 items were ordered by progressively increasing difficulty from top to bottom. The location was expressed in logits. The degrees of freedom for each χ2 were 6 for all items.

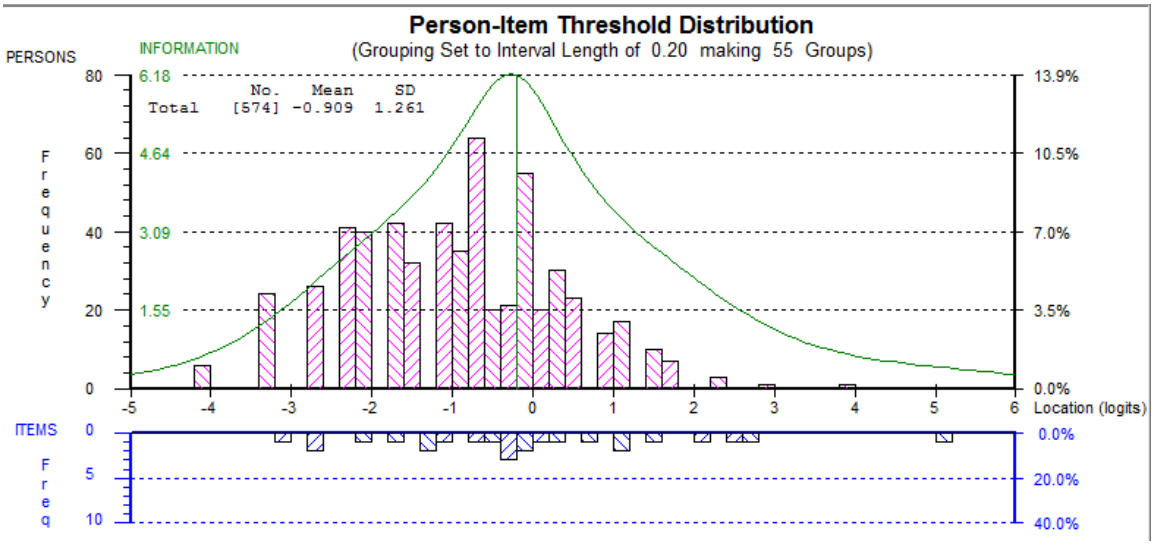
*The Bonferroni-corrected p-value indicating statistical significance at the .05 level was 0.004.

Abbreviations: VOE, Valutazione Oggettiva dell'Eleggibilità; Loc, Location; SE, Standard Error; FR, Fit Residual; χ2, chi-square; P, χ2 Probability.

After these modifications, the final 13-item solution for the VOE1 scale showed a satisfying fit to the Rasch model (Table 19, analysis 3). The scale was strictly unidimensional (overall PST 2.1%, LBBCI 1.2%, c=0.339, r=0.874, A=0.932), and also satisfied all the other Rasch model requirements in terms of invariance (unconditional χ^2_{12}=17.5; p=0.130, conditional class-interval based χ^2=26.1; p=0.072), local independence (no pairs of residual correlations of items above the LDRC), and monotonicity (no disordered thresholds). No DIF was detected for any of the tested person factors. All the subjects' responses fitted the model.

The targeting graph of the VOE1 scale (Figure 37) highlighted that subjects were spread across eight logits, with negligible floor (1%) and ceiling effects (0%). The mean person ability of -0.909 logits and a targeting index of -1.768 indicated, on average, a 'fair' matching between person ability and item difficulty (set by default at 0 logits)[197].

Figure 37. Targeting of the VOE1 scale (n=574)



NOTES: in the picture, persons and items were displayed, respectively, in the upper and the lower part of the graph, separated by the logit scale. Grouping set to interval length of 0.20, making 55 groups.
Abbreviation: Freq, Frequency.

The separation reliability expressed as PSI and Cronbach's Alpha was, respectively, 0.817 and 0.815, both indicating precision of measurement at the group level[248]. Given the PSI, persons could be separated by the scale in 3.2 strata, i.e., the statistically distinct levels of ability that the VOE1 was able to reliably distinguish in this sample[194]. Furthermore, persons could be separated into 3 Distinct Levels of Performance Ability

(DLPA), with a Distribution Independent PSI of 0.900, indicating the precision of measurement at the individual level.

The item hierarchy (Table 19, analysis 2 and Table 20) was consistent with the theoretical expectations about the hierarchy levels of the latent variable. Notably, it suggested that the earlier aspects that highlight the fall risk were the difficulty of rising from a chair (VOE06 and VOE09) and maintaining tandem standing (VOE05). On the other hand, the latest aspects which are linked to the fall risk were walking speediness (VOE01), the reduction of the standing limits of stability (VOE10), and the level of required assistance in activities of daily living before the fall (VOE13).

Based on the item calibration, it was possible to construct a table to convert the scale raw scores into interval-level estimates of the fall risk (Table 21).

Table 21. Raw-score-to-measure-estimates conversion table for the VOE1 scale

Raw score	Logit scale	±95% CI	0-25 scale	±95% CI
0	-4.049	2.389	0.0	2.9
1	-3.302	1.701	1.8	2.1
2	-2.753	1.374	3.2	1.7
3	-2.351	1.219	4.1	1.5
4	-2.012	1.123	5.0	1.4
5	-1.709	1.053	5.7	1.3
6	-1.435	0.992	6.4	1.2
7	-1.188	0.939	7.0	1.1
8	-0.967	0.888	7.5	1.1
9	-0.773	0.847	8.0	1.0
10	-0.602	0.815	8.4	1.0
11	-0.447	0.796	8.8	1.0
12	-0.3	0.788	9.1	1.0
13	-0.155	0.792	9.5	1.0
14	-0.005	0.808	9.8	1.0
15	0.157	0.835	10.2	1.0
16	0.341	0.878	10.7	1.1
17	0.555	0.933	11.2	1.1
18	0.805	1.000	11.8	1.2
19	1.091	1.070	12.5	1.3
20	1.413	1.151	13.3	1.4
21	1.778	1.252	14.2	1.5
22	2.214	1.409	15.2	1.7
23	2.811	1.705	16.7	2.1
24	3.971	2.425	19.5	2.9
25	6.225	4.167	25.0	5.1

NOTES: person estimates are expressed both in logits and into a 0 to 25 scale (VOE1 original scaling).
Abbreviations: 95%CI, 95% Confidence Interval (equal to 1.96 standard errors of measurement).

3.1.2.4.2 Rasch analysis on VOE items scalable on Mokken scale 2 (VOE2)

The second RA was conducted on the 5-item set scalable by the MA on scale 2: VOE11-12-14-16-18.

The base Rasch analysis showed that the scale, as a whole, did not fit the Rasch model (Table 22, analysis 1), failing the item homogeneity or invariance requirement (χ_{20}=163.6; p=0.0000). No items showed fit residuals <-2.5, suggesting that the responses to these items were too predictable (model overfit). One item (VOE011), instead, had fit residuals >2.5, thus highlighting that responses to this item were too unpredictable (model underfit). Beyond this, there were highly significant chi-squares for four out of five items, suggesting a violation of the requirement of stochastic invariance of the item hierarchy. The scale satisfied the unidimensionality requirement, as the Proportion of Significant T-test (PST) was 1.3%, and the Lower Bound of Binomial Confidence Interval for proportions (LBBCI) was 0%. Furthermore, there were disordered thresholds for two items (VOE11-12), indicative of a violation of the monotonicity requirement, and three pairs of items had residual correlations above the local dependence relative cutoff (LDRC, here set at 0.024, indicative of a violation of the local independence requirement).

Table 22. Rasch analysis details for the VOE items scalable on Mokken scale 2 (VOE2)

No	Description of analysis	N	Fitness to the Rasch Model						Targeting					Separation reliability					Unidimensionality	
			FitRes Items		FitRes Persons		Item-trait interaction		Person location		Floor effect (%)	SEM	Targeting index	DDR			DIR			
			Mean	SD	Mean	SD	χ^2_{df}	P^	Mean	SD				PSI	α	Strata	DLPA	DI-PSI	PST	BCI
1	Base analysis	574	0.390	2.005	-0.059	0.772	163.6_{20}	0.000	-0.947	2.105	31.5	1.090	-0.869	n/a	0.732	n/a	n/a	n/a	1.3%	0.0%
2	After rescoring	574	0.146	2.286	-0.032	0.708	298.2_{15}	0.000	-0.889	1.965	31.5	0.881	-1.009	0.487	0.799	1.6	n/a	n/a	1.3%	0.0%
3	After subtesting	574	0.202	1.353	-0.355	0.797	32.9_{6}	0.000*	-1.384	3.149	31.5	1.703	-0.998	0.805	0.708	3.0	2	0.800	1.6%	0.3%

NOTES

Abbreviations: FitRes, Fit Residual; DDR, Distribution-Dependent Reliability; DIR, Distribution-Independent Reliability; SD, Standard Deviation; $\chi 2_{df}$, chi-square and its degrees of freedom; p, Bonferroni-corrected χ2 probability value; SEM, Standard Error of Measurement of the person locations; PSI, Person Separation Index; α, Cronbach's alpha; DLPA, Distinct Levels of Performance Ability; DI-PSI, Distribution-Independent Person Separation Index (based on DLPA); LD, Local Dependence; PST, Proportion of Significant T-test carried out on the estimates that, within a principal component analysis of residuals, loaded positively and negatively (factor loading >±.3) on the first component; BCI, Binomial Confidence Interval for PST; n/a, not available.

Values are mean (SD) or as otherwise indicated.

^P-value is considered significant when <0.05

*P=0.040 for a sample size of 150 to handle type I error due to a larger sample size

In the next steps of analysis, we performed several item modifications to achieve a final fitting solution of the scale:

- Two items were rescored (VOE11-12), as they showed disordered thresholds;
- 'Testlets' (or super-items) were created according to the 'factor loading 2-testlet approach', which creates two testlets with items with similar Principal Component Analysis loading, thus obtaining a two-testlet solution (testlet 1: VOE11: history of falls, VOE02: lesions post-fall; testlet 2: VOE14: ADL post-fall, VOE16: IADL post-fall, VOE18: physical activity post-fall).

Table 23. Item parameters and fit statistics for the VOE2 scale (N=574, analysis no. 2)

	Item parameters and fit statistics					Scoring model			
VOE2 items	**Loc**	**SE**	**FR**	**χ^2**	**p***	**0**	**1**	**2**	**3**
VOE11 – History of falls	-3.605	0.126	4.119	64.863	0.000	0	1	1	2
VOE12 – Lesions post-fall	-0.068	0.090	-0.099	29.120	0.000	0	1	1	2
VOE18 – Physical activity post-fall	0.023	0.113	-1.596	56.985	0.000	0	1		
VOE16 – IADL post-fall	1.606	0.134	-0.691	61.392	0.000	0	1		
VOE14 – ADL post-fall	2.044	0.149	-1.005	85.831	0.000	0	1		

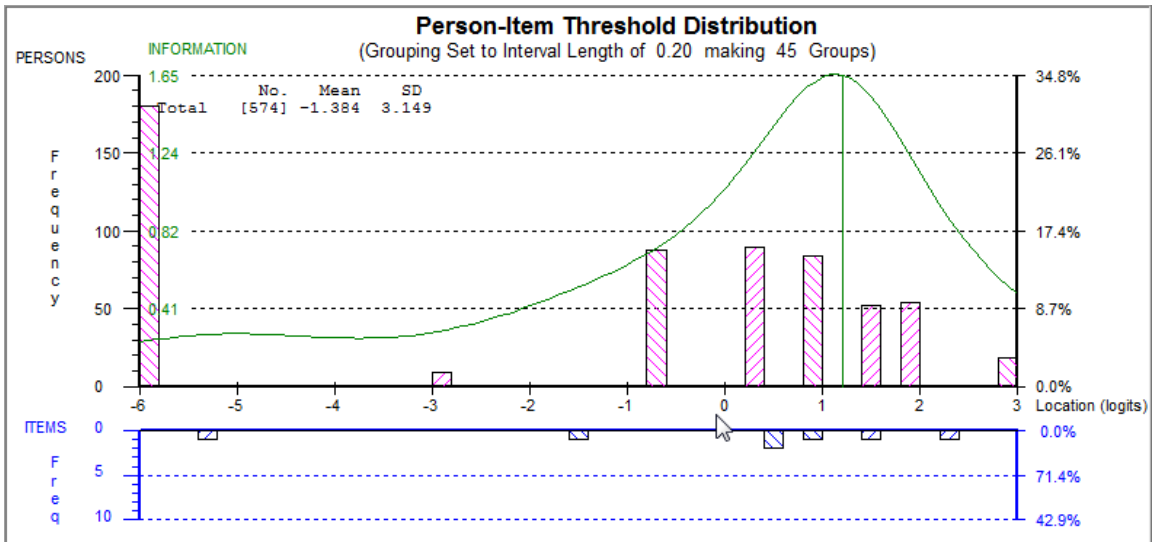
NOTES: VOE2 items were ordered by progressively increasing difficulty from top to bottom. The location was expressed in logits. The degrees of freedom for each χ2 were 3 for all items.
*The Bonferroni-corrected p-value indicating statistical significance at the 0.05 level was 0.010.
Abbreviations: VOE, Valutazione Oggettiva dell'Eleggibilità; Loc, Location; SE, Standard Error; FR, Fit Residual; χ2, chi-square; P, χ2 Probability.

After these modifications, the final 5-item solution for the VOE2 scale showed a satisfying fit to the Rasch model (Table 22, analysis 3). The scale was strictly unidimensional (overall PST 1.6%, LBBCI 0.3%, c=0.454, r=0.829, A=0.887), and also satisfied all the other Rasch model requirements in terms of invariance (unconditional χ^2_6=13.1; p=0.040 for a sample size of 150 to handle type I error due to larger sample size, conditional item based χ^2=9.7; p=0.139), local independence (no pairs of residual correlations of items above the LDRC), and monotonicity (no disordered thresholds). No DIF was detected for any of the tested person factors. All the subjects' responses fitted the model, except 27 subjects out of 574, who showed a model overfit (too predictable responses).

The targeting graph of the VOE2 scale (Figure 38) highlighted that subjects were spread across nine logits, with remarkable floor (31.5%) and negligible ceiling effect (3%). The

mean person ability of -1.384 logits and a targeting index of -0.998 indicated, on average, a 'fair' matching between person ability and item difficulty (set by default at 0 logits)[197].

Figure 38.Targeting of the VOE2 scale (n=574)



NOTES: in the picture, persons and items were displayed, respectively, in the upper and the lower part of the graph, separated by the logit scale. Grouping set to interval length of 0.20, making 45 groups.
Abbreviation: Freq, Frequency.

The separation reliability expressed as PSI and Cronbach's Alpha was, respectively, 0.805 and 0.708, both indicating precision of measurement at the group level[248]. Given the PSI, persons could be separated in 3.0 strata, i.e., the statistically distinct levels of ability that the VOE2 was able to reliably distinguish in this sample[194]. Furthermore, persons could be separated into two Distinct Levels of Performance Ability (DLPA) with a Distribution Independent PSI of 0.800, indicating even in this case, the precision of measurement at the group level.

The item hierarchy (Table 22, analysis 2 and Table 23) was consistent with the theoretical expectations about the hierarchy levels of the latent variable. Notably, it suggested that the earlier aspects that highlight the fall risk were the history of falls (VOE11) and the severity of the post-fall lesions (VOE12). On the other hand, the latest aspects linked to the fall risk were the modification of the assistance required in IADL and ADL post-fall compared to the pre-fall (VOE16 and VOE14).

Based on the item calibration, it was possible to construct a table to convert the scale raw scores into interval-level estimates of the fall risk (Table 24).

Table 24. Raw-score-to-measure-estimates conversion table for the VOE2 scale

Raw score	Logit scale	±95% CI	0-7 scale	±95% CI
0	-5.835	3.940	0.0	1.6
1	-2.854	3.579	2.4	1.4
2	-0.693	2.287	4.2	0.9
3	0.378	1.725	5.0	0.7
4	0.941	1.541	5.5	0.6
5	1.400	1.554	5.8	0.6
6	1.959	1.813	6.3	0.7
7	2.833	2.621	7.0	1.1

NOTES: person estimates are expressed both in logits and into a 0 to 7 scale (VOE2 original scaling).
Abbreviations: 95%CI, 95% Confidence Interval (equal to 1.96 standard errors of measurement).

3.1.3 Diagnostic accuracy of the available screening tools

3.1.3.1 Calculation of the FRAT-up on the VAE and VOE variables

To use the FRAT-up tool as an external comparator for the calibrated VAE, VOE1, and VOE2 scales in terms of diagnostic accuracy, we calculated the FRAT-up probabilities using the information provided by the VAE and VOE variables on the same samples, as described in the methods chapter.

3.1.3.2 Definition of the optimal cutoffs of the available screening tools

As planned in the methods, the definition of the optimal cutoff of the calibrated scales (VAE, VOE1, and VOE2) and the two calculated FRAT-up (VAE and VOE) in the prediction of at least one, two, and three (recurrent fallers) falls in the following twelve months was realized. We performed the calculation and comparison of the ROC curves and the AUCs for all the scales (Table 25, column AUC), and we defined the optimal cutoffs using the Youden's index method (Table 25, column Cutoff (YI)), and an 'ad hoc' method that allowed to obtain the optimal trade-off, from a clinical perspective, between false positives and false negatives in the following steps of the generation of screening algorithms (Table 26, column Cutoff (Cm)). The defined cutoffs showed different units of measure according to the type of assessment tool (logit for the VOE and VAE scales, probability for the calculated FRAT-up). Based on these cutoffs, we calculated the further measures of DA for each scale: sensitivity and specificity (with

false positive and negative rates), diagnostic effectiveness or accuracy, positive and negative predictive values, pre-test probability or disease prevalence, positive and negative likelihood ratios, post-test probability, and diagnostic odds ratio.

Regarding the AUCs for the ROC curves, the mean value was 0.678, indicating an overall sufficient DA of the instruments[122]. The FRAT-up VOE demonstrated the highest values of AUC in the prediction of at least one fall (0.689, CI 95% [0.645; 0.732]) and three falls (0.721, CI 95% [0.668; 0.775]), expressing respectively a sufficient and a good DA[122], whereas, in the prediction of at least two falls, the FRAT-up VAE showed the highest value (0.699, CI 95% [0.651; 0.747]), qualify as good DA[122]. Besides, for all the predictions, the VAE scale expressed the lowest value of AUC in the prediction of at least one fall (0.638, CI 95% [0.593; 0.683]), and VOE1 in that of at least two and three falls (respectively 0.628, CI 95% [0.576; 0.679]; 0.656 [0.594; 0.718], expressing a sufficient DA[122](Table 25, column AUC and Figure 39). Through the test of equality of two or more ROC areas[213], we found a significant chi-square statistics in the comparison of all scales for the prediction of all three outcomes (≥1 fall: chi2(4)=12.46, p-value 0.014; ≥2 falls: chi2(4)=14.66, p-value 0.006; ≥3 falls: chi2(4)=15.55, p-value 0.004), so we performed a pairwise comparison between scales. According to the test, we obtained evidence of a significant different discriminant power between VAE and FRAT-up VAE (chi2(1)=3.77, p-value 0.052) and between VAE and FRAT-up VOE (chi2(1)=4.36, p-value 0.037) in favor of the two FRAT-up in the prediction of at least one fall. Besides, VOE1 showed a significantly lower discriminatory capacity in the prediction of at least two falls compared to FRAT-up VAE and VOE (respectively chi2(1)=6.14, p-value 0.0132 and chi2(1)=13.38, p-value 0.000). Finally, VOE1 demonstrated again a significant lower discriminant power in the prediction of at least three falls compared to FRAT-up VOE (chi2(1)=11.41, p-value 0.001) (Table 25, column AUC and Figure 39).

3.1.3.3 Analyses with standard techniques of diagnostic accuracy of the available screening tools

The further measures of DA based on the scale cutoffs calculated with the Youden Index and the 'ad hoc' methods were reported in Table 25 and Table 26.

In detail, regarding the *DA based on the Youden Index cutoff*, for the prediction of at least one fall, the mean Youden Index was 0.277, and the mean sensitivity and

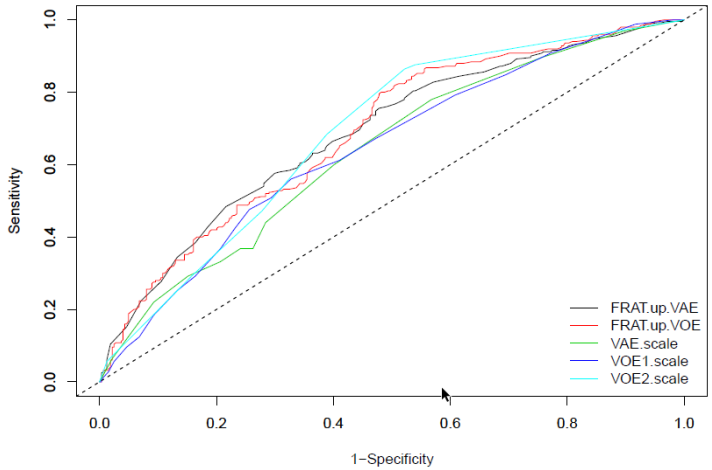
specificity were, respectively, 75.2% and 52.5%. The best sensitivity was shown by the VOE2 scale (86.4%, false positive rate of 52.2%, and false negative rate of 13.6%). The highest value of specificity was 67.3%, demonstrated by the VOE1 scale (false positive rate of 32.7% and false negative rate of 44%). VOE1 scale showed also the highest values of PPV (56.9%) and of LR+ (1.71), increasing the disease probability from 43.6% to 56.9% in case of a positive test result. Finally, the VOE2 highlighted the highest values of NPV (82.0%), of LR- (0.28), decreasing the disease probability from 43.6% to 18% in case of a negative test result, of DA (0.646), and of DOR (5.83, CI 95% [3.82; 8.89]) (Table 25).

Table 25. Diagnostic accuracy of the calibrated scales and the two FRAT-up based on the cutoff defined by the Youden Index method and areas under the curve

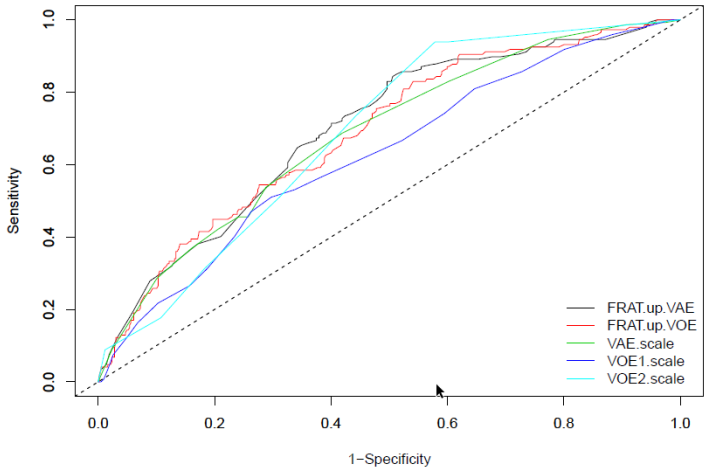
State	Unit of measure	Cutoff (YI)	N	YI	Sens (%)	Spec (%)	DA	FP (%)	FN (%)	PPV (%)	NPV (%)	Pre-test prob (%)	LR +	LR -	Post-test prob (+) (%)	Post-test prob (-) (%)	DOR (CI 95%)	AUC (CI 95%)
VAE scale																		
≥1 fall	Logit	-1.095	574	0.212	78.0	43.2	0.584	56.8	22.0	51.5	71.8	43.6	1.37	0.51	51.5	28.2	2.70 (1.86; 3.91)	0.638 (0.593; 0.683)
≥2 falls	Logit	-0.306	574	0.268	68.7	58.1	0.608	41.9	31.3	36.1	84.4	25.6	1.64	0.54	36.1	15.6	3.04 (2.04; 4.53)	0.683 (0.635; 0.732)
≥3 falls	Logit	-0.306	574	0.319	75.9	56.1	0.591	43.9	24.1	23.6	92.9	15.2	1.73	0.43	23.6	7.1	4.01 (2.38; 6.76)	0.707 (0.652; 0.761)
VOE1 Scale																		
≥1 fall	Logit	-0.602	574	0.233	56.0	67.3	0.624	32.7	44.0	56.9	66.5	43.6	1.71	0.65	56.9	33.5	2.62 (1.86; 3.68)	0.644 (0.600; 0.689)
≥2 falls	Logit	-0.300	574	0.213	51.0	70.3	0.653	29.7	49.0	37.1	80.6	25.6	1.72	0.70	37.1	19.4	2.46 (1.68; 3.61)	0.628 (0.576; 0.679)
≥3 falls	Logit	-0.300	574	0.317	62.1	69.6	0.685	30.4	37.9	26.7	91.1	15.2	2.04	0.55	26.7	8.9	3.75 (2.33; 6.02)	0.656 (0.594; 0.718)
VOE2 scale																		
≥1 fall	Logit	-0.693	574	0.342	86.4	47.8	0.646	52.2	13.6	56.1	82.0	43.6	1.66	0.28	56.1	18.0	5.83 (3.82; 8.89)	0.686 (0.643; 0.728)
≥2 falls	Logit	-0.693	574	0.360	93.9	42.2	0.554	57.8	6.1	35.8	95.2	25.6	1.62	0.15	35.8	4.8	11.17 (5.54; 22.53)	0.688 (0.643; 0.732)
≥3 falls	Logit	-0.693	574	0.347	96.6	38.2	0.470	61.8	3.4	21.8	98.4	15.2	1.56	0.09	21.8	1.6	17.30 (5.39; 55.53)	0.671 (0.622; 0.72)
FRAT-up VAE																		
≥1 fall	Probability	0.378	574	0.277	75.6	52.2	0.624	47.8	24.4	54.9	73.5	43.6	1.58	0.47	54.9	26.5	3.38 (2.35; 4.85)	0.687 (0.643; 0.730)
≥2 falls	Probability	0.379	574	0.337	84.4	49.4	0.584	50.6	15.6	36.5	90.2	25.6	1.67	0.32	36.5	9.8	5.27 (3.25; 8.54)	0.699 (0.651; 0.747)
≥3 falls	Probability	0.378	574	0.337	88.5	45.2	0.517	54.8	11.5	22.4	95.7	15.2	1.61	0.25	22.4	4.3	6.34 (3.21; 12.56)	0.685 (0.629; 0.741)
FRAT-up VOE																		
1 fall	Probability	0.347	574	0.322	80.0	52.2	0.643	47.8	20.0	56.3	77.2	43.6	1.67	0.38	56.3	22.8	4.36 (2.99;6.37)	0.689 (0.645; 0.732)
2 falls	Probability	0.350	574	0.289	83.0	45.9	0.554	54.1	17.0	34.6	88.7	25.6	1.53	0.37	34.6	11.3	4.14 (2.59;6.63)	0.688 (0.639; 0.736)
≥3 falls	Probability	0.355	574	0.352	89.7	45.6	0.523	54.4	10.3	22.7	96.1	15.2	1.65	0.23	22.7	3.9	7.26 (3.56;14.81)	0.721 (0.668; 0.775)

NOTES: the unit of measure of the defined cutoffs is the logit for the Rasch calibrated scales and the probability for the two calculated FRAT-up.
Abbreviations: Cutoff (YI), Cutoff defined using the Youden Index method; N, sample size; YI, Youden Index; Sens, Sensitivity; Spec, Specificity; DA, Diagnostic effectiveness or Accuracy; FP, False Positives; FN, False Negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value; pre-test prob, pre-test probability; LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio; post-test prob (+), post-test probability if the test result is positive; post-test prob (-), post-test probability if the test result is negative; DOR, Diagnostic Odds Ratio; AUC, Area Under the Curve; CI 95%, 95% Confidence Interval.

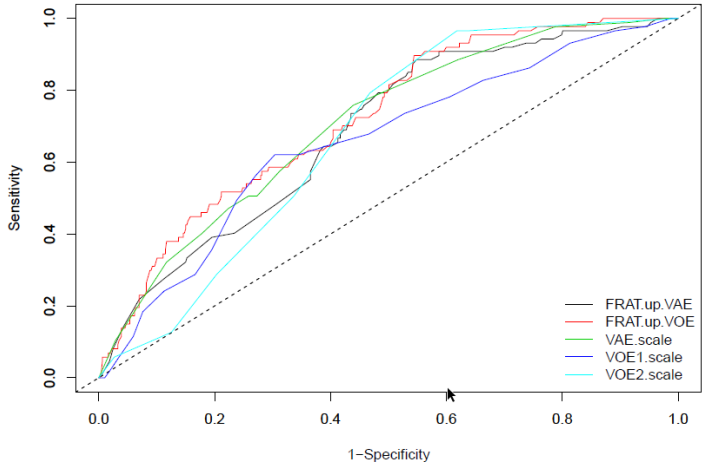
Figure 39. ROC curves of the five instruments in the prediction of at least one, two, and three falls



(a – prediction of at least one fall)



(b – prediction of at least two falls)



(c prediction of at least three falls)

NOTES: in the figure, ROC curves of the five instruments in the prediction of at least one (a), two (b), and three (c) fall(s) are presented. The FRAT-up VAE demonstrated the highest values of AUC in the prediction of at least one and two falls (black curves), whereas the highest value of AUC in the prediction of at least three falls was shown by FRAT-up VOE (red curve). Regarding AUC comparison, in (a) we note the significant different discriminatory capacity between VAE (0.638) and FRAT-up VAE and VOE (0.688 and 0.689), in (b) between VOE1 (0.628) and FRAT-up VAE and VOE (0.701 and 0.688), in (c) between VOE1 (0.656) and FRAT-up VOE (0.721).

Then, for the prediction of at least two falls, the mean Youden Index was 0.293, and the mean sensitivity and specificity were, respectively, 76.2% and 53.2%. The best sensitivity was shown again by the VOE2 scale (93.9%, false positive rate of 57.8%, and false negative rate of 6.1%). The highest value of specificity was 70.3%, demonstrated by the VOE1 scale (false positive rate of 29.7% and false negative rate of 49%). VOE1 scale showed also the highest values of PPV (37.1%), of DA (0.653), and of LR+ (1.71), increasing the disease probability from 25.6% to 37.1% in case of a positive test result. Finally, the VOE2 scale showed the highest values of NPV (95.2%), of LR- (0.15), decreasing the disease probability from 25.6% to 4.8% in case of a negative test result, and of DOR (11.17, CI 95% [5.54; 22.53]) (Table 25).

Finally, for the prediction of at least three falls, the mean Youden Index was 0.334, and the mean sensitivity and specificity were, respectively, 82.6% and 50.9%. The best sensitivity was shown once more by the VOE2 scale (96.6%, false positive rate of 61.8%, and false negative rate of 3.4%). The highest value of specificity was 69.6%, demonstrated again by the VOE1 scale (false positive rate of 30.4% and false negative rate of 37.9%). VOE1 scale showed also the highest values of PPV (26.7%), of DA (0.685), and of LR+ (2.04), increasing the disease probability from 15.2% to 26.7% in case of a positive test result. Finally, the VOE2 scale showed the highest values of NPV (98.4%), of LR- (0.09), decreasing the disease probability from 15.2% to 1.6% in case of a negative test result, and of DOR (17.3, CI 95% [5.39; 55.53]) (Table 25).

In detail, *regarding the DA based on the 'ad hoc' cutoff*, for the prediction of at least one fall, the mean sensitivity and specificity were, respectively, 81% and 45.1%. The best sensitivity was shown by the VOE2 scale (86.4%, false positive rate of 52.2%, and false negative rate of 13.6%). The highest value of specificity was 52.2%, demonstrated by the FRAT-up VAE (false positive rate of 47.8% and false negative rate of 24.4%). FRAT-up VOE showed the highest values of PPV (56.2%) and of LR+ (1.66), increasing the disease probability from 43.6% to 56.2% in case of a positive test result. Finally, the VOE2 highlighted the highest values of NPV (82.0%), of LR- (0.28), decreasing the disease probability from 43.6% to 18% in case of a negative test result, of DA (0.646), and of DOR (5.83, CI 95% [3.82; 8.89]) (Table 26).

Then, for the prediction of at least two falls, the mean sensitivity and specificity were, respectively, 87.2% and 45.1%. The best sensitivity was shown again by the VOE2

scale (93.9%, false positive rate of 57.8%, and false negative rate of 6.1%). The highest value of specificity was 49.4%, demonstrated again by the FRAT-up VAE (false positive rate of 50.6% and false negative rate of 15.6%). FRAT-up VAE showed also the highest values of PPV (36.5%), of DA (0.584), and of LR+ (1.67), increasing the disease probability from 25.6% to 36.5% in case of a positive test result. Finally, the VOE2 scale showed the highest values of NPV (95.2%), of LR- (0.15), decreasing the disease probability from 25.6% to 4.8% in case of a negative test result, and of DOR (11.17, CI 95% [5.54; 22.53]) (Table 26).

Finally, for the prediction of at least three falls, the mean sensitivity and specificity were, respectively, 91.3% and 37.1%. The best sensitivity was shown once more by the VOE2 scale (96.6%, false positive rate of 61.8%, and false negative rate of 3.4%). The highest value of specificity was 45.6%, demonstrated again by the FRAT-up VOE (false positive rate of 54.4% and false negative rate of 10.3%). FRAT-up VOE showed also the highest values of PPV (22.7%), of DA (0.523), and of LR+ (1.65), increasing the disease probability from 15.2% to 22.7% in case of a positive test result. Finally, the VOE2 scale showed the highest values of NPV (98.4%), of LR- (0.09), decreasing the disease probability from 15.2% to 1.6% in case of a negative test result, and of DOR (17.3, CI 95% [5.39; 55.53]) (Table 26).

Table 26. Diagnostic accuracy of the calibrated scales and the two FRAT-up based on the cutoff defined by an 'ad hoc' clinical method

State	Unit of measure	Cutoff (Cm)	N	Sens (%)	Spec (%)	DA	FP (%)	FN (%)	PPV (%)	NPV (%)	Pre-test prob (%)	LR +	LR -	Post-test prob (+) (%)	Post-test prob (-) (%)	DOR (CI 95%)
VAE scale																
≥1 fall	Logit	-1.095	574	78.0	43.2	0.584	56.8	22.0	51.5	71.8	43.6	1.37	0.51	51.5	28.2	2.70 (1.86; 3.91)
≥2 falls	Logit	-1.095	574	83.0	39.8	0.509	60.2	17.0	32.2	87.2	25.6	1.38	0.43	32.3	12.8	3.23 (2.01; 5.17)
≥3 falls	Logit	-1.095	574	88.5	38.0	0.456	62.0	11.5	20.3	94.9	15.2	1.43	0.30	20.3	5.1	4.72 (2.38; 9.35)
VOE1 Scale																
≥1 fall	Logit	-1.709	574	84.8	30.6	0.542	69.4	15.2	48.5	72.3	43.6	1.22	0.50	48.5	27.7	2.45 (1.62; 3.73)
≥2 falls	Logit	-2.012	574	91.8	19.9	0.383	80.1	8.2	28.3	87.6	25.6	1.15	0.41	28.3	12.4	2.80 (1.48; 5.28)
≥3 falls	Logit	-2.012	574	93.1	18.7	0.3	81.3	6.9	17.0	93.8	15.2	1.14	0.37	17.0	6.2	3.10 (1.31; 7.33)
VOE2 scale																
≥1 fall	Logit	-0.693	574	86.4	47.8	0.646	52.2	13.6	56.1	82.0	43.6	1.66	0.28	56.1	18.0	5.83 (3.82; 8.89)
≥2 falls	Logit	-0.693	574	93.9	42.2	0.554	57.8	6.1	35.8	95.2	25.6	1.62	0.15	35.8	4.8	11.17 (5.54; 22.53)
≥3 falls	Logit	-0.693	574	96.6	38.2	0.470	61.8	3.4	21.8	98.4	15.2	1.56	0.09	21.8	1.6	17.30 (5.39; 55.53)
FRAT-up VAE																
≥1 fall	Probability	0.378	574	75.6	52.2	0.624	47.8	24.4	54.9	73.5	43.6	1.58	0.47	54.9	26.5	3.38 (2.35; 4.85)
≥2 falls	Probability	0.379	574	84.4	49.4	0.584	50.6	15.6	36.5	90.2	25.6	1.67	0.32	36.5	9.8	5.27 (3.25; 8.54)
≥3 falls	Probability	0.378	574	88.5	45.2	0.517	54.8	11.5	22.4	95.7	15.2	1.61	0.25	22.4	4.3	6.34 (3.21; 12.56)
FRAT-up VOE																
1 fall	Probability	0.347	574	80.0	51.9	0.641	48.1	20.0	56.2	77.1	43.6	1.66	0.39	56.2	22.9	4.31 (2.95; 6.29)
2 falls	Probability	0.350	574	83.0	45.9	0.554	54.1	17.0	34.6	88.7	25.6	1.53	0.37	34.6	11.3	4.14 (2.59; 6.63)
≥3 falls	Probability	0.355	574	89.7	45.6	0.523	54.4	10.3	22.7	96.1	15.2	1.65	0.23	22.7	3.9	7.26 (3.56; 14.81)

NOTES: the unit of measure of the defined cutoffs is the logit for the Rasch calibrated scales and the probability for the two calculated FRAT-up.
Abbreviations: Cutoff (Cm), Cutoff defined using an 'ad hoc' clinical method; N, sample size; YI, Youden Index; Sens, Sensitivity; Spec, Specificity; DA, Diagnostic effectiveness or Accuracy; FP, False Positives; FN, False Negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value; pre-test prob, pre-test probability; LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio; post-test prob (+), post-test probability if the test result is positive; post-test prob (-), post-test probability if the test result is negative; DOR, Diagnostic Odds Ratio; AUC, Area Under the Curve; CI 95%, 95% Confidence Interval.

3.1.4 Generation of the screening algorithms and their diagnostic accuracy

3.1.4.1 Screening algorithms with calibrated scales and their diagnostic accuracy

Before the generation and the following analyses of multistep algorithms with serial combinations of the calibrated scales VAE, VOE1, and VOE2, we performed an analysis of the strength of the association between the calibrated scales (Spearman rho). It showed the presence of a weak correlation between the described instruments (≤0.433), allowing the use of all these scales to generate the following screening algorithms.

Then, we generated six different serial algorithms with the calibrated scales as defined in the methods and, for each algorithm, we performed the study of the diagnostic accuracy using the same measures reported above for the single instruments (Table 27), along with the cumulative post-test probability values (Table 28).

Table 27. Diagnostic accuracy of the six generated algorithms with the calibrated scales for the prediction of at least one, two, and three falls in the following twelve months (cutoffs based on Youden Index and clinical 'ad hoc' methods)

Algorithm	VAE cutoff	VOE1 cutoff	VOE2 cutoff	Sens (%)	Spec (%)	DA	FP (%)	FN (%)	PPV (%)	NPV (%)
YIm – 1 f	-1.095	-0.602	-0.693	43.6	84.3	0.666	15.7	56.4	68.1	65.9
Cm – 1 f	-1.095	-1.709	-0.693	62.4	71.0	0.672	29.0	37.6	62.4	71.0
YIm – 2 f	-0.306	-0.300	-0.693	38.1	86.2	0.739	13.8	61.9	48.7	80.2
Cm – 2 f	-1.095	-2.012	-0.693	72.8	63.2	0.657	36.8	27.2	40.5	87.1
YIm – 3 f	-0.306	-0.300	-0.693	47.1	84.8	0.791	15.2	52.9	35.7	90.0
Cm – 3 f	-1.095	-2.012	-0.693	79.3	60.0	0.629	40.0	20.7	26.1	94.2

NOTES: the unit of measure of the defined cutoffs is the logit.
Abbreviations: YIm, Youden Index method for the cutoff calculation for each scale; Cm, 'ad hoc' clinical method for the cutoff calculation for each scale; 1 f, 1 fall; 2 f, 2 falls, 3 f, 3 falls; Sens, Sensitivity; Spec, Specificity; DA, Diagnostic Accuracy; FP, False Positives; FN, False Negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Table 28. Cumulative post-test probability of the six generated algorithms with the calibrated scales for the prediction of at least one, two, and three falls in the following twelve months (cutoffs based on Youden Index and clinical 'ad hoc' methods)

Algorithm	PrTP (%)	LR+ VAE	LR- VAE-	PoTP + VAE/ PrTP + VOE 1 (%)	PoTP - VAE (%)	LR+ VOE 1	LR- VOE 1	PoTP + VOE 1/ PrTP + VOE 2 (%)	PoTP - VOE 1/ PrTP - VOE 2 (%)	LR+ VOE 2	LR- VOE 2	PoTP + VOE 2 (%)	PoTP - VOE 2 (%)
YIm – 1 f	43.6	1.37	0.51	51.4	28.2	1.71	0.65	64.4	40.7	1.66	0.28	75.0	15.3
Cm – 1 f	43.6	1.37	0.51	51.4	28.3	1.22	0.50	56.4	34.6	1.66	0.28	68.2	13.6
YIm – 2 f	25.6	1.64	0.54	36.1	15.7	1.72	0.70	49.3	28.3	1.62	0.15	61.1	6.9
Cm – 2 f	25.6	1.38	0.43	32.2	12.9	1.15	0.41	35.3	16.3	1.62	0.15	47.0	5.0
YIm – 3 f	15.2	1.73	0.43	23.6	7.1	2.04	0.55	38.7	14.5	1.56	0.09	49.6	3.4
Cm – 3 f	15.2	1.43	0.30	20.3	5.1	1.14	0.37	22.6	8.6	1.56	0.09	31.2	2.0

NOTES

Abbreviations: YIm, Youden Index method for the cutoff calculation for each scale; Cm, 'ad hoc' clinical method for the cutoff calculation for each scale; 1 f, 1 fall; 2 f, 2 falls, 3 f, 3 falls; PrTP, Pre-Test Probability (prevalence); LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio; PoTP+, positive Post-Test Probability; PrTP+, positive Pre-Test Probability; PoTP-, negative Post-Test Probability; PrPT-, negative Pre-Test Probability.

Regarding the diagnostic accuracy of the six generated algorithms (Table 27), it was possible to highlight that the algorithms defined using the Youden Index method showed high values of specificity (average 85.1%) and low values of sensitivity (average 42.9%), with false negatives equal to 50-60%, and false positives to 15% for all the predicted outcomes. In contrast, algorithms defined using as an 'ad hoc' method for the cutoff score calculation, such as the limitation of the overall false negative rates (30-35% for the prediction of ≥1 fall, 20-25% for ≥2 falls, and 10-15% for ≥3 falls), demonstrated more balanced values of sensitivity (average 71.5%) and specificity (average 64.7%) and, consequently, of false negative (20-35%) and false positive rates (30-40%). Concerning positive and negative predictive values, for the prediction of at least one fall, both algorithms evidenced balanced and equivalent PPVs and NPVs, whereas, for the other predicted outcomes, NPVs were consistently higher than PPVs for both the types of algorithms. With the increase of the disease prevalence (≥1 fall 43.6%; ≥2 falls 25.6%; ≥3 falls 15.2%), PPVs increased, while NPVs decreased. Finally, the diagnostic accuracy value was slightly higher for the 'ad hoc' clinical algorithm only in predicting at least one fall (0.672 vs. 0.666), whereas it was lower for the other predictions (Table 27).

Table 28 provided complementary information to the already presented results, showing the cumulative post-test probabilities obtained with the serial combination of the scales, in which the post-test probability of one test became the pre-test probability for the next test. Considering the diagnostic accuracy measures, which also include the obtained low positive and negative likelihood ratios of every single test presented in Table 28, the Youden Index algorithms generally tended to ensure a higher probability of disease in case of positive tests (increased specificity and higher LR+), while the 'ad hoc' algorithms a greater reduction in the probability of disease in case of negative tests (increased sensitivity and lower LR-).

3.1.4.2 Screening algorithms with the two FRAT-up and their diagnostic accuracy

To use the FRAT-up, being already validated and published[120, 121], as a comparator for 'external validation', we followed the same analytical steps described for the calibrated scales.

Before the generation and the following analyses of multistep algorithms with serial combinations of the FRAT-up VAE and VOE, we performed an analysis of the strength of the association between the two tools (Spearman rho). It showed the presence of a moderate correlation between the described instruments (0.463), allowing the use of both to generate the following screening algorithms.

Then, we generated six different serial algorithms with the two FRAT-up as defined in the methods and, for each algorithm, we performed the study of the diagnostic accuracy using the same measures reported above for the single instruments (Table 29), along with the cumulative post-test probability values (Table 30).

Table 29. Diagnostic accuracy of the six generated algorithms with the two FRAT-ups for the prediction of at least one, two, and three falls in the following twelve months (cutoffs based on Youden Index and clinical 'ad hoc' methods)

Algorithm	FRAT-up VAE cutoff	FRAT up VOE cutoff	Sens (%)	Spec (%)	DA	FP (%)	FN (%)	PPV (%)	NPV (%)
YIm – 1 f	0.378	0.347	64.8	69.1	0.672	30.9	35.2	61.8	71.8
Cm – 1 f	0.369	0.357	62.0	67.3	0.650	32.7	38.0	59.4	69.6
YIm – 2 f	0.379	0.350	74.1	65.6	0.678	34.4	25.9	42.6	88.1
Cm – 2 f	0.381	0.344	72.8	64.9	0.669	35.1	27.2	41.6	87.4
YIm – 3 f	0.378	0.355	81.6	62.8	0.657	37.2	18.4	28.2	95.0
Cm – 3 f	0.372	0.356	79.3	61.2	0.639	38.8	20.7	26.7	94.3

NOTES: the unit of measure of the defined cutoffs is the probability.
Abbreviations: YIm, Youden Index method for the cutoff calculation for each scale; Cm, 'ad hoc' clinical method for the cutoff calculation for each scale; 1 f, 1 fall; 2 f, 2 falls, 3 f, 3 falls; Sens, Sensitivity; Spec, Specificity; DA, Diagnostic Accuracy; FP, False Positives; FN, False Negatives; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

Table 30. Cumulative post-test probability of the six generated algorithms with two FRAT-up for the prediction of at least one, two, and three falls in the following twelve months (cutoffs based on Youden Index and clinical 'ad hoc' methods)

Algorithm	PrTP (%)	LR+ FRAT-up VAE	LR- FRAT-up VAE	PoTP+ FRAT-up VAE/ PrTP+ FRAT-up VOE (%)	PoTP- FRAT-up VAE (%)	LR+ FRAT-up VOE	LR- FRAT-up VOE	PoTP+ FRAT-up VOE	PoTP – FRAT-up VOE
YIm – 1 f	43.6	1.58	0.47	54.9	26.6	1.67	0.38	67.1	31.7
Cm – 1 f	43.6	1.50	0.46	53.6	26.2	1.59	0.49	64.8	36.2
YIm – 2 f	25.6	1.67	0.32	36.5	9.9	1.53	0.37	46.8	17.5
Cm – 2 f	25.6	1.65	0.34	36.2	10.5	1.48	0.39	45.7	18.1
YIm – 3 f	15.2	1.61	0.25	22.3	4.3	1.65	0.23	32.2	6.2
Cm – 3 f	15.2	1.54	0.27	21.6	4.6	1.61	0.28	30.7	7.2

NOTES
Abbreviations: YIm, Youden Index method for the cutoff calculation for each scale; Cm, 'ad hoc' clinical method for the cutoff calculation for each scale; 1 f, 1 fall; 2 f, 2 falls, 3 f, 3 falls; PrTP, Pre-Test Probability (prevalence); LR+, positive Likelihood Ratio; LR-, negative Likelihood Ratio; PoTP+, positive Post-Test Probability; PrTP+, positive Pre-Test Probability; PoTP-, negative Post-Test Probability; PrPT-, negative Pre-Test Probability.

Regarding the diagnostic accuracy of the six generated algorithms (Table 29), it was possible to highlight that both algorithms, for the prediction of at least one fall, showed balanced values of sensitivity and specificity near 65-70%, with the specificity slightly higher than the sensitivity. Differently, for the prediction of at least two and three falls, both algorithms demonstrated a high value of sensitivity (near 75% for two falls and near 80% for three falls), with values of specificity between 61 and 65%. False negative and positive rates were similar between the two methods for each predicted outcome

and followed the tendency of sensitivity and specificity, being respectively their one-complement. Concerning positive and negative predictive values, for the prediction of at least one fall, both algorithms evidenced similar high NPVs (70%) and low PPVs (60%), whereas, for the other predicted outcomes, NPVs were consistently higher than PPVs for both the types of algorithms. Even for the FRAT-up algorithms, with the increase of the disease prevalence (≥1 fall 43.6%; ≥2 falls 25.6%; ≥3 falls 15.2%), PPVs increased, while NPVs decreased. Finally, the diagnostic accuracy value was slightly higher for the Youden Index algorithms for all the predicted outcomes (Table 29).

As for the analogous table for the calibrated scales (Table 28), Table 30 provided complementary information to the already presented results, showing the cumulative post-test probabilities obtained with the serial combination of the two FRAT-up, in which the post-test probability of one test became the pre-test probability for the second test. Considering the diagnostic accuracy measures, which also include the obtained low positive and negative likelihood ratios of every single test presented in Table 30, both algorithms had similar behavior, allowing at the end of the series of tests to increase the disease probability of 15-20% in case of positive results and to decrease it of 8-10% in case of negative results, without any substantial difference.

3.1.4.3 Comparison of the diagnostic accuracy of the generated algorithms

Following the presented analysis of the diagnostic accuracy of algorithms, we compared the obtained results between those composed by the calibrated scales (CS algorithms) versus those by the two calculated FRAT-up (FU algorithms) (Table 27, Table 28, Table 29, Table 30), considering the latter as a comparator for 'external validation', being already validated and published[120, 121].

Regarding the *algorithms generated based on the Youden Index method* (Table 27 and Table 29), which tends to minimize the chance to find false positives and false negatives, for all the predicted outcome, the CS algorithms showed higher values of specificity (respectively 84.3% vs. 69.1%; 86.2% vs. 65.6%; 84.8% vs. 62.8%), but the FU algorithms demonstrated more balanced properties (average sensitivity FU 73.5%, CS 42.9%; average specificity FU 65.8%, CS 85.1%). Consequently, the application of CS algorithms determined false negative and positive rates respectively higher than 50% and lower than 16%, compared to the FU false negative and positive rates lower

than 35% and 37%. Besides, PPVs were lower than NPVs for all the algorithms for all the predicted outcomes, except for the CS algorithm in the prediction of at least one fall, which was slightly higher than the NPV (68.1% vs. 65.9%), and the two values were more balanced; the latter consideration was correct also for the FU algorithm in the same predicted outcome (PPV 61.8%, NPV 71.8%). In the prediction of at least two and three falls, for both algorithms, NPVs increased, and PPVs decreased according to the decrease of prevalence (≥2 falls 25.6%; ≥3 falls 15.2%), and NPVs of the FU algorithms were slightly higher than those by the CS algorithms, as well as the PPVs were slightly lower. Generally, except for the CS algorithm in the prediction of at least one fall, all algorithms were more effective in ruling out the diagnosis of fall risk (higher NPVs than PPVs). The diagnostic accuracy was similar between the two algorithms in the prediction of at least one fall (CS 0.666 vs. Fu 0.672), whereas was consistently in favor of the CS algorithms for the other outcomes (≥2 falls CS 0.739 vs. FU 0.678; ≥3 falls CS 0.791 vs. FU 0.657). Finally, comparing the cumulative post-test probabilities obtained with the serial combination of the tests (CS vs. FU) for each predicted outcome (Table 28 and Table 30), the CS algorithms performed more effectively in ruling in and ruling out the disease for all the predicted outcomes. In fact, they led to a higher increase of the disease probability in positive results and a higher decrease in negative results.

Regarding *the algorithms generated based on the 'ad hoc' clinical method* (Table 27 and Table 29), which limited the overall false negative rates (30-35% for the prediction of ≥1 fall, 20-25% for ≥2 falls, and 10-15% for ≥3 falls), the direct consequence was the same level of sensitivity for both algorithms in each predicted outcome (≥1 fall 62%, false negative 38%; ≥2 falls 72.8%, false negative 27.2%; ≥3 falls 79.3%, false negative 20.7%). Regarding specificity and false positive rates, they were different between algorithms only in the prediction of at least one fall: the CS algorithm showed a higher level of specificity (71% vs. 67.3%) and, so, a lower value of false positive rate compared to the FU algorithm (29% vs. 32.7%). In other cases, these values were similar, with false positive rates of 35-36% for at least two falls and of 40% for at least three falls. Also, for positive and negative predictive values, the major difference concerned the one-fall outcome: PPV and NPV were both slightly higher for CS algorithms compared to FU ones. For the other outcomes, they were almost identical

(≥2 falls PPVs 40% and NPVs 87%; ≥3 falls PPVs 26% and NPVs 94%), giving evidence as for the Youden Index method of more effective capacity in ruling out the diagnosis of fall risk compared to the ruling in. The diagnostic accuracy was around 0.6 for all the algorithms; for the one-fall outcome, the CS algorithm demonstrated a higher value (0.672 vs. 0.650), whereas, for the other outcomes, DA values were slightly in favor of the FU algorithms (two-fall outcome CS 0.657 vs. FU 0.669; three-fall outcome CS 0.629 vs. FU 0.639). Finally, comparing the cumulative post-test probabilities obtained with the serial combination of the tests (CS vs. FU) for each predicted outcome (Table 28 and Table 30), even with the 'ad hoc' clinical method, the CS algorithms performed more effectively in ruling in and ruling out the disease for all the predicted outcomes. In fact, they led to a higher increase of the disease probability in case of positive results and a higher decrease in negative results.

3.1.5 Construction of additional screening algorithms based on a logistic regression model

As a final way to study the diagnostic accuracy of the calibrated scales and their use in screening algorithms, we performed their construction using the logistic regression technique. With this analysis, we attempted to integrate the scales, also with the other four available fall risk factors that the psychometric analyses excluded from the calibration of the scales (VOE19: medications; VOE20: psychological status; VOE21: cognitive status; VOE22: visual acuity).

First of all, we conducted fifteen univariable logistic regression analyses, using each tool as a covariate (three calibrated scales and the two FRAT-up) and each predicted outcome as the regression binary outcome (≥1 fall: 0=no; 1=yes, ≥2 falls: 0=no; 1=yes, and ≥3 falls: 0=no; 1=yes). All the tools (three scales and two FRAT-up) showed a significant odds ratio (p-value ≤0.05) in the prediction of all the outcomes (Table 31).

Table 31. Univariable logistic regression analyses results

	OR	SE	z	P-value	CI 95%
Prediction ≥ 1 fall					
VAE	1.13	0.03	4.66	0.000	[1.07; 1.19]
VOE1	1.52	0.11	5.74	0.000	[1.32; 1.76]
VOE2	1.30	0.04	8.29	0.000	[1.22; 1.39]
FRAT-up VAE	1.13	0.02	7.03	0.000	[1.09; 1.17]
FRAT-up VOE	1.07	0.01	7.53	0.000	[1.05; 1.09]
Prediction ≥ 2 falls					
VAE	1.19	0.03	6.02	0.000	[1.13; 1.26]
VOE1	1.44	0.11	4.55	0.000	[1.23; 1.69]
VOE2	1.37	0.06	7.13	0.000	[1.26; 1.50]
FRAT-up VAE	1.16	0.03	6.65	0.000	[1.11; 1.21]
FRAT-up VOE	1.07	0.01	6.67	0.000	[1.05; 1.09]
Prediction ≥ 3 falls					
VAE	1.21	0.04	5.49	0.000	[1.13; 1.30]
VOE1	1.52	0.15	4.27	0.000	[1.25; 1.84]
VOE2	1.38	0.08	5.46	0.000	[1.23; 1.55]
FRAT-up VAE	1.16	0.03	5.19	0.000	[1.10; 1.22]
FRAT-up VOE	1.08	0.01	6.37	0.000	[1.06; 1.11]

NOTES: sample size = 574.
Abbreviations: OR, Odds Ratio; SE, Standard Error; z, z statistics; P-value, P-value referred to Z statistics; CI95%, Confidence Interval at 95%.

Then, for each fitted univariable model, we generated the predicted probabilities, and the comparison of the ROC curves and their relative AUCs of the predicted probabilities of the calibrated scales for each outcome was realized (Table 32). By testing the equality of the probabilities of the three scales for each outcome, we did not obtain any statistically significant difference (p-values >0.05 for AUCs and their respective 95% confidence intervals see Table 32).

Table 32. Areas under curve of ROC curves of the predicted probabilities of univariable logistic models with each calibrated scale (VAE, VOE1, and VOE2) as a covariate in the prediction of the three outcomes

	≥1 fall		≥2 falls		≥3 falls	
	AUC	CI 95%	AUC	CI 95%	AUC	CI 95%
p_VAE	0.638	0.593-0.683	0.684	0.635-0.732	0.707	0.652-0.761
p_VOE1	0.644	0.599-0.689	0.628	0.576-0.679	0.656	0.594-0.718
p_VOE2	0.686	0.643-0.728	0.688	0.644-0.732	0.671	0.622-0.720

NOTES: sample size = 574. No statistically significant differences were found between the three AUCs of the three calibrated scales for each predicted outcome (chi-square test)[213].
Abbreviations: p_VAE, predicted probabilities of a univariable model with VAE scale as a covariate; p_VOE1, predicted probabilities of a univariable model with VOE1 scale as a covariate; p_VOE2, predicted probabilities of a univariable model with VOE2 scale as a covariate; AUC, Area Under the Curve; CI 95%, 95% Confidence Interval of the AUC.

Later, we constructed multivariable logistic regression models for each predicted outcome, and we compared their discriminant capacity to choose the best one. As

defined in the methods chapter, for each outcome, we started from the scale with the most discriminant ROC curve as a covariate and, in every successive attempt, we added the following scale in terms of discriminant capacity, and we compared the new multivariable model with the previous model which demonstrated the highest discriminant power (highest AUC of the predicted probabilities). Finally, we chose the best discriminant model for each predicted outcome to be submitted to the following analytical steps.

Regarding the prediction of at least one fall in the following twelve months, the VOE2 scale was again the instrument with the highest discriminant capacity (AUC 0.686, Table 32), even if this superiority was not statistically significant as defined above. So, we did a multivariable logistic regression model, adding the VOE1 scale, which was the second scale in terms of AUC (0.644, Table 32). The latter seemed to discriminate better than the model with the only VOE2 (VOE2+VOE1 AUC 0.7314, CI 95% [0.691; 0.772], chi2(1)=13.72, p-value 0.000), so in the third multivariable model, the VAE scale was added to the two-variable model. Whit this attempt, we did not obtain a more discriminant model, as well as in the attempts with the use, as covariates, of the other four available fall risk factors not included in the scales cited above. Hence, we kept the model VOE2+VOE1 as the best discriminant model to predict at least one fall.

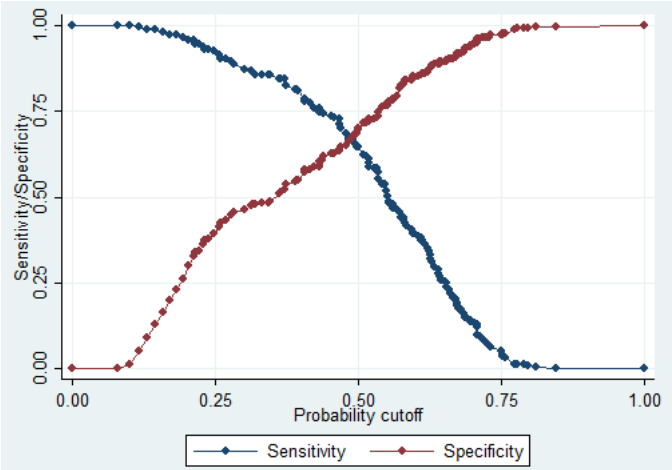
Regarding the prediction of at least two falls in the following twelve months, the VOE2 scale was the instrument with the highest discriminant capacity (AUC 0.688, Table 32), even if this superiority was not statistically significant as defined above. So, we did a multivariable logistic regression model adding the VAE scale, which was the second scale in terms of AUC (0.684, Table 32). The latter seemed to discriminate better than the model with the only VOE2 (VOE2+VAE AUC 0.752, CI 95% [0.710; 0.794], chi2(1)=16.46, p-value 0.000), so in the third multivariable model, the VOE1 scale was added to the two-variable model. As before, whit this attempt, we did not obtain a more discriminant model, as well as in the attempts with the use, as covariates, of the other four available fall risk factors not included in the scales cited above. Hence, we kept the model VOE2+VAE as the best discriminant model to predict at least two falls.

Regarding the prediction of at least three falls in the following twelve months, the VAE scale was the instrument with the highest discriminant capacity (AUC 0.707, Table 32), even if this superiority was not statistically significant, as defined above. So, we did a

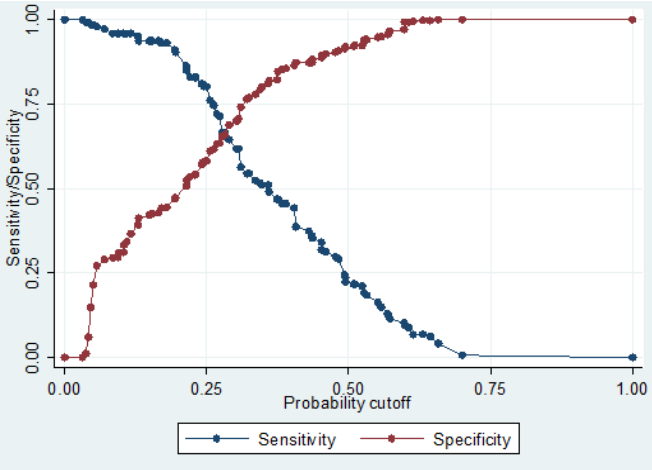
multivariable logistic regression model, adding the VOE2 scale, which was the second scale in terms of AUC (0.671, Table 32). The latter showed a higher AUC of borderline significance compared to the model with the only VAE (VAE+VOE2 AUC 0.745, CI 95% [0.697; 0.794], chi2(1)=3.75, p-value 0.053), so in the third multivariable model, the VOE1 scale was added to the two-variable model. Whit this attempt, we did not obtain a more discriminant model, as well as in the attempts with the use, as covariates, of the other four available fall risk factors not included in the scales cited above. Hence, we kept the model VAE+VOE2 as the best discriminant model to predict at least three falls.

Once the best discriminant model for each predicted outcome has been identified, the three models were submitted to the Hosmer-Lemeshow goodness-of-fit test to verify the model fit to the data sample. As expected, it showed a non-significant result for all the models (≥1 fall: chi2(8)=8.71, p-value 0.368; ≥2 falls: chi2(8)=4.87, p-value 0.772; ≥3 falls: chi2(8)=7.24, p-value 0.512), which expressed the satisfaction with the fit of our models.

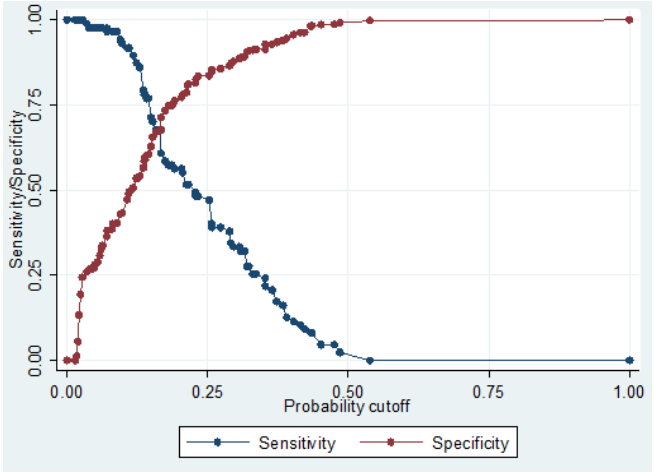
Figure 40. Sensitivity and specificity versus probability cutoff graphs of the best discriminant model for each predicted outcome



(a – prediction of at least one fall)



(b – prediction of at least two falls)



(c – prediction of at least three falls)

NOTES: in the figure, the sensitivity and specificity versus probability cutoff graphs of the best discriminant model for the prediction of at least one (a), two (b), and three (c) fall(s) are presented (x axis: predicted probabilities by the fitted model; y axis: sensitivity (blue) and specificity (red) for each predicted probability).

For the best discriminant model of each predicted outcome, we realized a sensitivity and specificity versus all possible probability cutoffs point graphs (Figure 40). Regarding graph 'a', a probability cutoff of 0.50 in predicting at least one fall in the following twelve months, defined by the model VOE2+VOE1, was associated with a sensitivity of 64.8 and a specificity of 70.1%. It corresponded almost to the point in which the two curves crossed (0.48), maximizing both properties (sensitivity 66.8%, specificity 66.4%). Differently, regarding graph 'b' about predicting at least two falls, a probability cutoff of 0.50 (VOE2+VAE) was linked to a sensitivity of 21.8% and a specificity of 92.3%. In this case, this cutpoint did not correspond to the one that maximizes the two properties, which was instead of 0.28 (sensitivity 66.7% and a specificity of 66.0%). Finally, concerning graph 'c' about predicting at least three falls, a probability cutoff of 0.50 (VAE+VOE2) was linked to a sensitivity of 0.0% and a specificity of 99.8%. Even for this prediction, this cutoff did not correspond to the one which maximizes the two properties, which was even lower than before and equal to 0.16 (sensitivity 67.8% and a specificity of 67.8%).

Finally, we constructed three multivariable logistic regression models, one for each predicted outcome, using the FRAT-up VAE and VOE as covariates. We generated, as well as for the calibrated scales, the predicted probabilities of the fitted models and the ROC curves with their relative AUCs of these predicted probabilities for each outcome (≥1 fall: AUC 0.716, CI 95% [0.674; 0.757]; ≥2 falls: AUC 0.728, CI 95% [0.683; 0.774]; ≥3 falls: AUC 0.745, CI 95% [0.694; 0.795]). Then, we compared the AUC of the model based on the two FRAT-up with the AUC of the best discriminant model based on the calibrated scales presented above for each predicted outcome. For all the three outcomes, the equality test (chi-square test) between the three pairs of AUC did not show any statistically significant difference between the models (≥1 fall: chi2(1)=1.03, p-value 0.311; ≥2 falls: chi2(1)=1.06, p-value 0.302; ≥3 falls: chi2(1)=0.18, p-value 0.669).

3.2 Assessment of the effect of neurological diseases on the fall risk screening tests

As planned in the methods, the analysis of the presence of differential item functioning (DIF) or item bias for two subgroups of neurological diseases (elderly and elderly with

associated neurological diseases like PD or stroke), in the context of Rasch analysis, for each of the item of the calibrated measurement scales (VAE, VOE1, and VOE2 scales) on the primary objective was performed.

Regarding the VAE scale, as already cited in the Rasch analysis results section, the testlet 1 including 3 items (VAE01: history of falling, VAE03: diagnosis of stroke or PD, and VAE05: inability to rise from a chair) (analysis no.3 'After subtesting', Table 15) showed a uniform DIF for neurological diseases, as elderly with associated neurological diseases (PD or stroke) found this super-item systematically more challenging to pass compared to those without this further disease. In order to solve the DIF, the testlet was splitted into two different items of different difficulty (one for elderly and one for elderly with associated neurological diseases). We compared the paired person estimates of the splitted vs. the un-splitted solution through a paired sample t-test, and the test resulted in a significant difference (t=29.2, p-value 0.000, N=768). This difference yielded a Cohen's d (effect size) of 1.437 (CI 95% [1.365; 1.6]). Given this effect size >0.2, the effect of DIF on person estimates was considered large[247], and DIF was accounted for[207]. Due to the presence of DIF, it was necessary the construction of two tables to convert the VAE scale raw scores into interval-level estimates of the fall risk (Table 17) for older adults with or without associated neurological disease.

Regarding the VOE1 and VOE2 scales, no significant DIF was found for any subgroup of neurological diseases.

Concerning the FRAT-up VAE and VOE, independent sample t-tests were conducted to explore the impact of neurological diseases on levels of fall risk, as measured by these tools (fall risk probability). Participants were divided, as before, into two groups according to their neurological diseases (elderly and elderly with associated neurological diseases like PD or stroke). There was not a statistically significant difference at the $p<0.05$ level for the FRAT-up VAE (t=-0.7260, df=766, p-value 0.468), whereas a statistically significant difference was found for the FRAT-up VOE (t=-4.705, df=572, p-value 0.000), determined by a significantly higher fall risk mean probability for the elderly with associated neurological disease group compared to the elderly group. This difference yielded a Cohen's d (effect size) of 0.432 (CI 95% [0.251; 0.614], which was considered small[247].

3.3 Validation of an ICD&ICF core set for the fall risk in community-dwelling older adults (also with associated neurological diseases)

The searching for the most recent systematic reviews with meta-analysis on the fall risk factors in the community-dwelling older people, and also in those with Parkinson's Disease and stroke sequelae, were performed according to criteria described in the methods chapter. The followed string searches produced 28 citations for older people, one citation for older people with Parkinson's Disease, and two citations for older people with stroke sequelae.

In the context of the 28 citations for older people, the systematic review chosen was that by Deandrea and colleagues (2010), with the title 'Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis'[41], because it was recognized as a methodologically rigorous review with a meta-analysis recently realized.

Regarding older people with Parkinson's Disease, the followed string search produced only one citation (the already cited review by Deandrea), which was considered inadequate for our purposes because it considered Parkinson's Disease as a fall risk factor and not as the reference population. For this reason, the string was modified to 'risk factors falls Parkinson's Disease', with article type filter 'review', and it allowed to obtain 58 citations. Between the available works, two reviews, one by Cunning et al. (2014), titled 'Prevention of falls in Parkinson's Disease: a review of fall risk factors and the role of physical interventions'[23], and the other by Fasano et al. (2017), titled 'Falls in Parkinson's Disease: a complex and evolving picture'[20], were selected because they were updated and focused adequately on fall risk factors in this population.

Finally, regarding older people with stroke sequelae, the followed string search produced two citations, which were considered inadequate for our purposes because they considered interventions or only one specific fall risk factor. Hence, we modified the string search to 'risk factors falls community-dwelling stroke' keeping 'meta-analysis' as article type filter and found six results. Between these six citations, a systematic review with meta-analysis 'Risk factors for falls in community stroke survivors: a systematic review and meta-analysis' by Xu and colleagues (2018)[30] was chosen, given its updating and pertinence to the desired topic.

From the cited reviews, a total of 86 fall risk factors were available. In particular, we extracted:

- 24 fall risk factors in older people (Deandrea, 2010)[41] (Table 33), along with significant odd ratios for all fallers and recurrent fallers (>3 falls per year) for each factor in the following twelve months.
- Respectively, 23 and 32 fall risk factors in older people with PD by Canning, 2014[23], and Fasano, 2017[20] (Table 33). For the first, a significant measure of effect (odds ratio or risk ratio or hazard ratio) for each factor was extracted (21 for all fallers and 4 for recurrent fallers in the following twelve months).
- 7 fall risk factors in older people with stroke sequelae (Xu,2018) [30], along with significant odd ratios for each factor (7 for all fallers and 2 for recurrent fallers) (Table 33).

Further details (measures of effect and other bibliographic references) were available in Appendix B.

Table 33. Fall risk factors extracted by the four reviews[20, 23, 30, 41]

N	Older adults fall risk factors (Deandrea, 2010)[41]	Older adults with PD fall risk factors (Canning, 2014)[23]	Older adults with PD fall risk factors (Fasano, 2017)[20]	Older adults with stroke sequelae fall risk factors (Xu, 2018)[30]
1	Age	Ankle dorsiflexor strength of the stronger leg	Osteoporosis	Postural instability/altered balance
2	Cognitive impairment	Ankle dorsiflexor strength of the weaker leg	Age	Cognitive impairment
3	Depression	Attention	Anxiety	Depression
4	Dizziness and vertigo	Central processing speed (cognitive reaction time)	Arthrosis	Disability in self-care
5	Fear of falling	Cognitive impairment (both for all and recurrent fallers)	Axial rigidity	History of falls (both for all and recurrent fallers)
6	Gait problems	Depression	Cardiac arrhythmia	Impaired mobility
7	Hearing impairment	Difficulty or assistance with ADL (only for recurrent fallers)	Cognitive impairment	Use of sedative and psychotropic medications (both for all and recurrent fallers)
8	History of falls	Dyskinesia	Daily use of alcohol	
9	History of stroke	Executive functions	Depression	
10	Instrumental disability	Fear of falling	Disease severity (PD)	
11	Living situation (alone vs. not alone)	Freezing of gait (both for all and recurrent fallers)	Dual tasking	
12	No. medications (for 1-drug increase)	History of falls (only for recurrent fallers)	Dyskinesia	
13	Pain	Knee extensor strength	Environmental hazards	
14	Parkinson's Disease	Knee extensor strength of the stronger leg	Female gender	
15	Physical disability	Knee extensor strength of the weaker leg	Freezing of gait	
16	Rheumatic disease	Knee flexor strength of the weaker leg	Functional neurosurgery (particularly subthalamic nucleus deep brain stimulation)	
17	Self-perceived health status	Perceived effort to complete tasks	Higher total doses of levodopa	
18	Sex	Physical activity levels	History of falls	
19	Urinary incontinence	Postural instability/altered balance	Impaired mobility	
20	Use of antiepileptics	Proprioception	Orthostatic hypotension	
21	Use of antihypertensives	Semantic fluency	Other comorbidities (peripheral neuropathy)	
22	Use of sedatives	Visual contrast	Polypharmacy (use of>3 drugs other than anti-PD)	
23	Visual impairment	Walking aid use	Postural abnormalities	
24	Walking aid use		Postural instability/altered balance	
25			Shuffling and small-scaled gait	
26			Transfers	
27			Urinary incontinence	

N	Older adults fall risk factors (Deandrea, 2010)[41]	Older adults with PD fall risk factors (Canning, 2014)[23]	Older adults with PD fall risk factors (Fasano, 2017)[20]	Older adults with stroke sequelae fall risk factors (Xu, 2018)[30]
28			Use of dopamine agonists, anticholinergic	
29			Use of sedatives	
30			Visual and oculomotor impairments	
31			Walking aid use	
32			Weakness due to inactivity	

NOTES: for each review, the extracted fall risk factors are reported in alphabetical order. In green, fall risk factors common to the four reported reviews.

Each extracted fall risk factor was linked to 2nd and 3rd level categories of the ICD or ICF classifications, according to the coded linking rules published by Cieza and colleagues, which have been updated over the years[135-137] (Figure 28 in the Introduction chapter). Each factor could be linked to one or more categories, based on the number of main concepts of the factor. In particular:

- 6 risk factors were linked to 7 ICD categories (ICD-11 version 04/2019)[249]. These factors were those related to disease states (Appendix B).
- The remaining 80 factors were linked to 51 ICF categories (ICF 2017 - English)[131]. These factors were related to the subject's functioning (body functions and structures, activities, and participation) and environmental and personal components. In detail, risk factors were linked to 24 categories of the 'Body functions' component, 20 categories of 'Activities and participation', and 7 categories of 'Environmental factors'; no factors were linked to 'Body structures' categories. Finally, 8 risk factors were linked to 7 'Personal factors', which are not coded in the classification.

Examples of the linking of the fall risk factors to ICD and ICF categories are reported in Table 34. For the complete linking, see Appendix B.

Table 34. Examples of the ICD&ICF linking of the fall risk factors

Fall risk factor	Classification category
Osteoporosis[20]	FB83.1 Osteoporosis (ICD)
Dual tasking[20]	b1402 Dividing attention (ICF Body functions)
Amputation[143]	s750 Structure of lower extremity (ICF Body structures)
Physical activity levels[23, 250]	d920 Recreation and leisure (ICF Activities and participation)
Living situation (alone vs. not alone)[41]	e310 Immediate family (ICF Environmental factors)
Age (5-year increase)[20, 41]	Not coded (ICF Personal factors)

NOTES: in the table, one example for each type of linking was reported. Codes with the initial letter 'b' indicate 'Body functions', 's' indicates 'Body structures', 'd' indicates 'Activities and participation', and 'e' indicates 'Environmental factors'.

In order to create the new ICD&ICF core set for fall risk in community-dwelling older adults, also with associated neurological diseases, we followed these steps:

- All the risk factors with the related ICD or ICF categories were assembled into a new unique core set.
- The new core set was compared to the available 'ICF core set for falls in acute rehabilitation settings' by Yen and colleagues published in 2014[143].

- The new core set was integrated with the Yen's core set to create a new comprehensive ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases (Figure 41 and Appendix B for the whole core set).

Figure 41. Example of the ICD and ICF linking for the fall risk factors

Category[1-3]	Descriptor	Risk factor	Effect measure (OR, HR, RR with CI 95%) for all fallers	Effect measure (OR, HR, RR with CI 95%) for recurrent fallers
Pathology (ICD-11 version 04/2019)[4]				
8A00.0	Parkinson disease	Parkinson disease	2.71 (1.08-6.84)[5]	2.84 (1.77-4.58)[5]
8A00.2	Secondary parkinsonism			
8C00-8C03	*Polyneuropathies and other disorders of the peripheral nervous system*	*Other comorbidities (peripheral neuropathy)[6]*	-	-
8B00-8B26	Cerebrovascular diseases	History of stroke	1.61 (1.31-1.98)[5]	1.79 (1.51-2.13)[5]
FA00-FA38	Arthropathies	Rheumatic disease	1.47 (1.28-1.70)[5]	1.57 (1.42-1.73)[5]
FA00-FA05	*Arthrosis*	*Arthrosis[6]*	-	-
FB83.1	*Osteoporosis*	*Osteoporosis[6]*	-	-
Body functions (ICF 2017 - English)[7]				
b110	Consciousness functions*	*Confusion/disorientation[8]*	-	-
b114	Orientation functions*			
b140-b189	Specific mental functions	Cognitive impairment	1.36 (1.12-1.65)[5] 3.38 (1.31-8.67)[9, 10] 0.92* (0.86-0.99)[9, 11] 0.80 (0.63-1.02)[9, 12]	1.56 (1.26-1.94)[5] 3.44 (1.05-11.5)[9, 13]

NOTES: an image of the ICD&ICF core set for the fall risk in older people, also with associated neurological diseases, is presented. As an example of the linking of a fall risk factor to an ICD category, we can look to the risk factor 'History of stroke'. It was linked to the ICD group categories '8B00-8B26 Cerebrovascular disease'. For the ICF classification, the risk factor 'Confusion/disorientation' was linked to the 2nd level ICF category b110 'Consciousness functions'. In the last two columns, effect measures with 95% confidence interval for all fallers and recurrent fallers are reported when available.

The new comprehensive core set is composed of a total amount of 103 fall risk factors, extracted from the four reviews[20, 23, 30, 41] and the already existing ICF core set for falls by Yen[143], linked to 74 categories (7 to ICD and 67 to ICF) and 7 ICF 'Personal factors' (Appendix B). In particular:

- 6 risk factors were linked to 7 ICD categories (ICD-11 version 04/2019)[249]. These factors were those related to disease states (Appendix B).
- The remaining 97 factors were linked to 67 ICF categories (ICF 2017 - English)[131]. These factors were related to the subject's functioning (body functions and structures, activities, and participation), and the environmental and personal components:
    - 45 risk factors were linked to 35 ICF categories of the 'Body functions' component.
    - 1 risk factors were linked to 2 ICF categories of the 'Body structures' component.

- 9 risk factors were linked to 22 ICF categories of the ‘Activities and Participation’ component.
- 15 risk factors were linked to 8 ICF categories of the ‘Environmental factors’ component.
- 11 risk factors were linked to 7 ICF ‘Personal factors’, which are not coded.

4 Discussion

In this thesis, we have pursued the three aims of (1) validation of a fall risk serial screening algorithm, of (2a) assessment of the effect of neurological diseases on the fall risk screening tests, and of (2b) validation of an ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases, which have been described in detail in the introduction chapter (pages 99-100). Now we will discuss them in light of the presented results.

4.1 Validation of a fall risk serial screening algorithm with a high level of diagnostic accuracy

As mentioned in the methods chapter, the sample for these analyses was constituted by 768 people enrolled in a multicenter randomized control trial, the PRE.C.I.S.A. study. It aimed to study the efficacy of a tailored interdisciplinary, multicomponent, and multifactorial intervention to reduce the number of falls at twelve months in a community-dwelling elderly population, also including subjects with associated neurological diseases (PD or stroke sequelae) for one-third of the sample number[146]. This inclusion represents quite a novelty because, in most studies on older adults, those who also suffer from neurological diseases are generally excluded or considered in separate investigations. Then, based on the inclusion and exclusion criteria, we can affirm that the validated fall risk screening algorithm is addressed to subjects with a certain degree of independence, as the ability to walk at least for 10 meters without any assistance (aid if necessary), and the integrity of the cognitive functions. This is because, as already detailed in the introduction, about half of falls is caused by an environmental factor [7, 34, 35], and happens during activities that mildly displaced the subject's center of mass (e.g., standing, basic ADLs, walking)[34, 35]. Those people who already need assistance for motor or cognitive impairments, or activity limitations, are excluded: this does not mean that they are not at fall risk but constitute a separated population who have to be assessed differently. For example, activities like transfers on the wheelchair or on the toilet (20% falls according to Ashburn in PD[36]) can be more significant in terms of risk assessment. Our sample was numerically consistent (N=768), with a good representation of the older age (median=76, range [65; 100]), and two-third of women, whose 'gender', per se, constituted a fall risk factor[20, 41]. These proportions

have been maintained in the reduced sample of 574 persons, used to calibrate the VOE scales and the screening algorithms. An important result to highlight is that the elderly's self-reporting constituted the second source of recruitment of the PRE.C.I.S.A. sample older adults (about 23%, N=768 and 28%, N=574) for the assessment of their own fall risk. It gives evidence to the need and the perception of this population to be monitored, that frequently exceed the professionals' decisions. Regarding the PRE.C.I.S.A. levels of selection to consider persons as recruitable in the RCT, they selected as at medium&high fall risk the 59% (452, N=768) of the sample, merging the risk definitions by three fall risk tools (FRAT_Nandy, FRAT_Stapleton, FROP-COM screen), of whom the 89% (403) were eventually recruited. The significantly higher number of falls and subjects who fell at least one time in the following twelve months in the recruited group could represent evidence of the appropriate selection of people at fall risk. The fact that the calibration of the scales of the present study, which constituted the algorithms, is based on the same variables permits to extend this conclusion of selection efficacy also to our sample.

4.1.1 Calibration of the VAE, VOE1, and VOE2 scales

The main strength of the validated algorithms is their composition made by the three scales calibrated with Rasch analysis (VAE, VOE1, and VOE2), which were compared to the two versions of another instrument (FRAT-up VAE and VOE), which estimate the fall risk in probabilistic terms. In fact, all these tools provide interval measures, which demonstrated properties of scientific measuring instruments, differently from almost all the fall risk ordinal instruments available in the literature and described in the introduction, which can more properly be called 'instruments of quantification'. These properties are the *presence of a unit of measure* (the conditio sine qua non for measurement), the *unidimensionality* (only one variable can be measured at any one time), the *invariance* (other characteristics extraneous to the measured variable do not have to influence the measurement process and the unit of measure), and the *linearity* (the linear relationship between the measured entity and the constant unit of measurement)[251]. It is essential to state that only the use of scientific measures holds several practical advantages, such as the possibility to perform mathematical calculations, to correctly interpret the variable change, to predict future values of the

variable, and to properly use parametric statistics, which require the linearity of the employed measures as one of the assumptions for their use[184, 252, 253].

To calibrate the VAE, VOE1, and VOE2 scales, we performed a psychometric profile for each of them, including the descriptive statistics of sample and items, the preliminary assessment of dimensionality, through a CFA and a Mokken analysis, and finally, the Rasch analysis. Concerning the descriptive statistics of items, for all the scales, the average inter-item correlations (AIC) were weak (VAE: 0.161 and VOE: 0.246), whereas quite different was the indication provided by the Cronbach's alpha, which expresses unacceptable internal consistency upon the assumption of unidimensionality for the VAE scale (0.381), and a good internal consistency for the VOE scale (0.860). Given these results, for the VAE, a low AIC corresponds to a low level of alpha, which represents a finding frequently described in the literature[254]. In fact, 'internal consistency' refers to the overall degree to which scale items are intercorrelated, whereas 'homogeneity' and 'unidimensionality' indicate whether or not the scale items assess a single underlying factor or construct[255, 256]. Thus, the internal consistency is a necessary but not sufficient condition for measuring homogeneity or unidimensionality[254, 256]. For the VOE scale, the unusually high value of alpha, coexisting with a low AIC, could be due to the evidence that alpha tends to increase with the size (i.e., the number of items) of an instrument[254, 255, 257], whereas the AIC does not[172, 256]. In conclusion, in both cases, given the hypothesis of internal consistency upon the assumption of unidimensionality for these two sets of items, and that AIC is a straightforward indicator of internal consistency compared to alpha, which needs to stay (virtually as all of the individual inter-item correlations) in a 0.40 and 0.50 range for a valid measure of such a construct (i.e., fall risk)[256], these results seem to constitute first signals of the lack of one of the necessary condition for unidimensionality[254-256].

The subsequent specific preliminary assessment of unidimensionality included a Confirmatory Factor Analysis and a Mokken analysis for each scale, which highlighted a multidimensional structure both for VAE and VOE scales, which was somehow expected given the results of the internal consistency analyses. Regarding the CFA, it evidenced an initial failure to support the unidimensionality of the two scales, with the presence of high values of modification indices[162, 167] between items within each scale, indicating local dependence [192, 200, 203]. It is frequently detected in health outcome scales,

although it is often unreported or inadequately addressed[172, 192, 200, 203]. Allowing the correlation of the error terms of the locally dependent pairs of items within each scale guarantees effective management of this problem[162, 167], permitting to achieve a good fit to a one-factor model for both the scales. Instead, the Mokken analysis results led us to choose two different analysis strategies for the scales. For the VAE scale, taking note of the evidence of a two-dimension structure of items and aware of the probable subsequent problem of unidimensionality in the Rasch analysis context, we decided equally to consider the 7-item set as a unique set of items. The reason was linked to the low number of items and, above all, to our interest in obtaining a single anamnestic indicator. Otherwise, for the VOE scale, the strongly multidimensional scale structure (3 scales and 2 items not scalable) convinced us to submit, separately to the Rasch analysis, only the first two scales revealed by the MA, in order to calibrate two different objective tests administered by an healthcare professional (physiotherapist). The chosen scales contained the highest number of items (respectively 13 on scale 1 and 5 on scale 2), which that were also considered the most relevant from the conceptual content point of view.

Regarding the Rasch analysis, for the three scales VAE, VOE1 (VOE items scalable on Mokken scale 1), and VOE2 (VOE items scalable on Mokken scale 2), we performed the analysis of internal construct validity on each item set, under the assumption that each of them could be sufficiently unidimensional to measure a single underlying construct, and hence to constitute a new assessment tool. The analysis highlighted some problems for the three sets, related to the specific requirements of the Rasch model for model fit (invariance and adherence to a probabilistic Guttman pattern), as well as to some general requirements of item response theory models (monotonicity, local independence, and unidimensionality only for VOE1). The first analytical step in all the three solutions was the 'classical' approach of collapsing adjacent score categories of items with disordered thresholds to obtain an ordered structure for them. This action allowed improving the fit to the model for VAE and VOE1 as usually happens[186], but a worsening for VOE2. Besides, for VOE1 also the separation reliability improved, and the multidimensionality reduced. In fact, according to Pallant et al.[186], 'you would expect, for a well-fitting item, that across the whole range of the trait being measured, each response option would systematically take turns showing the highest probability of

endorsement'. When respondents fail to use the response categories consistently with the measured trait level, disordered thresholds happen. In particular, they occur when respondents have consistently discriminating difficulty between response options because there are too many options, or when the labeling of options is potentially confusing or open to misinterpretation (e.g., the use of terms sometimes, often, frequently)[186], or when the category is observed relatively rarely[258]. In our analyses, for the VAE, we found an example of too many options (VAE02: Medications, seven response option with an insufficient representation of the central categories); for the VOE1, an example of labeling confusing or open to misinterpretation (VOE07: SPPB_Balance with confusion in the distinction between the concepts 'semi tandem' and 'tandem', and VOE13: ADL pre-fall and VOE15: IADL pre-fall with confusion in the distinction between 'supervision' and 'minimal assistance'); for the VOE2 examples of rarely observed categories (VOE11: History of falls with no representation of '0 falls' and underrepresentation of '2 falls', and VOE12: Lesions post-fall with the underrepresentation of 'minor injury requiring medical consultation'). The second common analytical step was to address local dependence between items, which has already been pointed out with the Confirmatory Factor Analysis. In this context, it causes model misfit that can be linked either to response dependence or multidimensionality[172]. For each scale, we followed different approaches: for the VAE, the 'alternative 2-testlet approach'[207]; for the VOE1, a mixed application of the 'traditional testlet approach'[207] and the 'factor loading 2-testlet approach'; for the VOE2, only the 'factor loading 2-testlet approach'. The first one, proposed by Maritz et al. in 2019, 'divides conceptually similar items into two distinct testlets of equal size, taking alternative items in each testlet'[207]. In this way, it focuses on the total score of the scale, rather than the single items or groups of items by emphasizing the similarity of the items, as together they should measure the concept of fall risk. The advantages are linked to the creation of testlets of equal size, as recommended by Andrich[206], and the possibility to perform a conditional test of fit, which remains reliable for sample sizes up to 2,000[209]. Differently, the 'traditional testlet approach' creates 'testlets oriented at conceptually associated items and based on their residual correlations'[207]. It highlights the potential differences, e.g., dimensionality between testlets unifies similar items, such as 'fall risk in walking' or 'fall risk consequences'. Different again is the 'the factor loading 2-testlet approach', which creates two testlets with items sharing a similar

Principal Component Analysis loading factor, e.g., testlet with items with positive factor loading vs. testlet with items with negative factor loading. It shares the advantages of the 'alternative 2-testlet approach', although it predominantly represents a statistical and non-clinical/conceptual approach. For two out of three scales (VOE1 and VOE2), this last modification allowed to obtain Rasch model fitting final solutions, which were unidimensional and also satisfied all the other Rasch model requirements (invariance, local independence, monotonicity). Also, they did not evidence DIF for any tested person factor, and all the subjects' responses fitted to the model (except 4.8% of VOE2 subjects). This was not true for the VAE scale, which demonstrated a uniform DIF for neurological diseases for testlet 1, and so required a further analytical step with the testlet splitting into two different items of different difficulty (one for elderly subjects and one for elderly subjects with associated neurological diseases), to achieve the same desired final solution. Once obtained a final fitting solution for all the three scales, we followed analyzing the targeting (i.e., how well the measurement range of the scale matches the distribution of the calibrating sample[182]). We compared the mean location score, obtained for persons, with that of 'zero' set for the items (mean person location values around zero indicate that the scale is well-targeted), and we evaluated the presence of floor and ceiling effects[182]. The three calibrated scales showed different performances. VAE scale demonstrated a good targeting (mean person location 0.424), associated with the absence of significant floor and ceiling effects, whereas for the two VOE, the targeting was worst. In particular, they both highlighted a high negative mean person location (VOE1: -0.909; VOE2: -0.998), which indicates that the sample, as a whole, was located at a lower level of fall risk than the average of the scale[182]. In addition, VOE2 also showed a high floor effect of 31.5%, indicating that this sample percentage could not be differentiated by the scale because, globally, item difficulty was too high compared to person ability, and the scale considers these people 'all equal in terms of minimum fall risk' even if it could not be true[259]. According to Hagquist et al[259], bad targeting also might bring lower reliability. This happened for the discussed scales, where a worse targeting is associated with worse reliability. In detail, the two VOE demonstrated a value of separation reliability, the Person Separation Index, lower than 0.9, which is considered adequate only for group person measurement and not individual as for VAE[168, 243]. Associated with the PSI value, VOE1 and VOE2 could reliably distinguish between 3.2 and 3 statistically distinct levels of ability, called

strata[194]. The possibility gave by Wright in 2001[196] to calculate a PSI value independent from the distribution (DI-PSI) and related Distinct Levels of Performance Ability (DLPA) allowed to obtain at least for VOE a DI-PSI compatible for precise individual measurement (0.900), related to 3 DLPA. In any case, even if for VOE2 the adequate measurement remained the 'group' level also using the Wright method[196], the use of these instruments in screening algorithms aims to differentiate between a group of persons at fall risk to be treated versus a group which is not at risk. In that sense, a precise risk measurement at the individual level is not required. Then, from the conceptual perspective, it was also interesting to appreciate the item hierarchy provided by each scale, based on the persons' responses to the items. Generally, as already stated, it was consistent with the theoretical expectations about the hierarchy levels of the latent variable, and there were some overlapping concepts between the scales, which makes this observation more objective. For example, in VAE and VOE1 the presence of balance problems (items VAE04 and VOE05), and the inability or difficulty to rise from a chair (VAE05, VOE06 and VOE09), even if assessed differently by the cited items, were the earlier aspects which highlight the presence of fall risk. On the contrary, in VOE1 and VOE2, the level of required assistance in activities of daily living prior to the fall (VOE13), and the modification of this assistance after a fall event (VOE16 and VOE14) represented the latest aspects which are linked to the fall risk, and which alter later, when the subject reaches high levels of fall risk. Finally, we concluded the Rasch analysis of each scale with the provision of a table to convert the scale raw scores into interval-level estimates of fall risk. The already discussed availability of linear measures at the end of the Rasch analysis process, and of the exclusive consequences of their use, compared to that of raw scores, have been highlighted and recommended by respected authors, including Grimby, Tennant, and Tesio in their editorial on the Journal of Rehabilitation Medicine in 2011[260]. They perceived the raw-score to linear measure (and surrounding error) conversion tables as a way to make Rasch-derived instruments more user-friendly, allowing the use of the original raw scores in everyday clinical practice by health professionals. Then, the scores can be converted to interval scaling whenever required, simply consulting the conversion table, or by creating a look-up routine in Excel. The availability of such a device (i.e., the conversion table) could facilitate and encourage researchers to use Rasch analysis and Rasch-derived instruments, both in the development and evaluation of instruments, and in the analysis

of data from ordinal scales in outcome research, to definitively overcome ordinal scale limitations and improve the science of outcome measurement[260].

4.1.2 Diagnostic accuracy of the available screening tools

After the calibration of the scales, we performed the study of their diagnostic accuracy, along with that of the two FRAT-up calculated separately on the VAE and VOE variables in the prediction of three outcomes: at least one fall, two falls, and three falls in the following twelve months. The choice of these outcomes followed, first of all, the recommendations of four of the most important international guidelines, cited in the introduction chapter[12, 43, 113, 114]. All four guidelines suggested to ask older people, once per year, information about fall(s) in the last year, and three out of four (except that of the American Geriatrics Society/British Geriatrics Society (ABS/BGS) [43]) to submit to a multifactorial fall risk assessment those elderly who had one or more falls or demonstrated abnormalities of gate and balance in this period. The indications of the ABS/BGS society were slightly different. For older persons reporting only a single fall in the past year and reporting or demonstrating no difficulty or unsteadiness during the evaluation, they did not provide a fall risk assessment. In fact, they stated that these persons might similarly derive benefit from this kind of assessment and successive interventions, but evidence for that lacked in literature[43]. Secondly, the three systematic reviews on the fall risk tools performed in the last twenty years[117-119] evidenced that the predictive validity and the diagnostic accuracy studied about most instruments, considered as the main outcomes ‘one fall’ and ‘recurrent falls’ (with two or more falls). So, we decided to study the performance of the calibrated scales and the two FRAT-up to predict the three cited outcomes, considering these documents and aiming to compare our findings with them.

Given the continuous interval nature of the considered instruments (calibrated scales and two FRAT-up), we realized the graphical representation of their ROC curves, which offers an illustration of the trade-off between a test sensitivity and specificity, and depicts true positive rate against false positive rate for each cutoff score value. It was followed by the calculation of the AUCs and the definition of the optimal cutoff given the construction of the screening algorithms for each tool. Regarding AUCs, the mean value of all the five instruments for all the predictions (0.678) indicated an overall more

than sufficient discriminant power between fallers vs. non-fallers of the instruments. In two cases, regarding the prediction of at least three falls, the VAE scale and FRAT-up VOE showed an AUC higher than 0.7, classifiable as good. These values were superior to the AUCs reported by Gates[119] for eight screening test for one fall and recurrent falls (one fall [0.51-0.61]; recurrent falls [0.57-0.67]), but lower than the summary AUC reported by Park in his meta-analysis[117] about more than seven others tools for the prediction of one, more than one, and more than two fall(s) [0.76-0.97]. Concerning the three available studies on FRAT-up in the literature[120, 121, 123], our AUCs of the calibrated scales and the two FRAT-up calculated on VAE and VOE variables were comparable to those reported there for the prediction of one fall [0.638-0.646][120, 121, 123], and lower for the prediction of at least two falls [0.713][123].

The successive step to study the diagnostic accuracy of every single tool with multiple values, and to use these tests effectively, was to define an appropriate cutoff, which would allow combining scales in screening algorithms, to distinguish which subjects were at medium/high risk and which are not, about the fall risk condition. Between the several already cited existing methods, we decided to use and report the cutoffs determined through the 'Youden Index' and an 'ad hoc' methods. The choice of the Youden Index was linked to its widespread knowledge and to improve communicability[216], whereas that of the 'ad hoc' method to clinical reasoning and the need for the successive construction of screening algorithms. As already described in the methods chapter, we decided to limit the overall false negative rate to specific percentages (≥1 fall: 30-35%; ≥2 falls: 20-25%; ≥3 falls: 10-15%), based on a 'cost' ratio analysis between false negative and false positive rates. This analysis was conceived in terms of money and/or risk of possible complications and/or risk of an incorrect diagnosis between false negative and false positive rates. Regarding the 'fall risk' diagnosis, a false negative was considered to cost more than a false positive, and this ratio increased in favor of false negatives as the severity of the outcome increased (i.e., predicted number of falls). It was therefore decided to limit progressively, based on the predicted number of falls, the number of false negatives (i.e., subjects defined 'not at risk' but that could fall) against the risk of treating 'unnecessarily' subjects defined as false positives, for the non-invasive characteristics of the successive treatments (assessment and treatment of multiple fall risk factors, including balance

exercises, home environmental risk assessments, pharmacological therapy optimization, etc.). The use of the Youden Index method provided cutoffs which minimized the chance to find false positives and false negatives so that all the tools showed high values of sensitivity [68.7%-96.6%] compared to medium/low values of specificity [38.2%-58.1%] for all the predicted outcome, except VOE1 scale, which demonstrated the opposite behavior (sensitivity [51%-62.1%], specificity [67.3%-70.3%]). The use of the 'ad hoc' method, which limited the false negative rate to specific percentages, guaranteed instruments with even higher sensitivity values [75.6%-96.6%] against low specificity values [18.7%-52.2%], and this was valid also for VOE1. In terms of positive and negative predictive values (PPV and NPV), both methods allowed to obtain comparable higher NPV [66.5%-98.4%] than PPV [17%-56.9%] and, consequently, to correctly define the subjects not at risk with higher probability, even if it is crucial to remember their significant dependence on the prevalence of the disease in the population tested[122, 224, 225]. Only for VOE1, the 'ad hoc' method determined lower PPV and positive likelihood ratio (LR+), and higher NPV and negative likelihood ratio (LR-) than the 'Youden Index' ones. Then, both positive and negative likelihood ratios (LR+ and LR-) of all the tests for both methods and all the outcomes were far from the shared values (LR+: >10; LR-: <0.1)[122, 225] considered to provide strong evidence to respectively rule in or rule out diagnoses in most circumstances[122, 225] (LR+ [1.14-2.04], LR- [0.09-0.7]). Despite this, globally, the LR- were closer to them, determining a higher decrease of the disease probability in case of a negative test result, compared to the increasing of that probability in case of a positive test result. For this reason, even these measures confirmed the capacity of these tools to correctly define the subjects not at risk with higher probability. The findings of high values of sensitivity and medium/low of specificity were also reported by Perell[118] and Park[117] in their systematic reviews, whereas the opposite results, similar to the Youden Index VOE1 cutoffs, by Gates[119]. In all cases, however, our diagnostic accuracy analyses confirmed the conclusions by the authors of the cited systematic reviews on screening tools[117-119]: the use of single instruments did not allow to predict elderly fallers with optimal accuracy, and frequently they did not lead to post-test disease probability large enough to cross the threshold for intervention. This insufficient diagnostic accuracy, then, constituted our justification for the construction and validation of the fall risk serial screening algorithms, composed of two or more tools.

4.1.3 Generation of the screening algorithms and their diagnostic accuracy

For each predicted outcome, as assumed in the introduction chapter, we constructed and validated two serial screening algorithms based on the calibrated scales with the 'AND rule', which allowed us to perform three screening steps. The first one, constituted by the 'anamnestic' VAE scale, administered through a phone call by a trained health professional, enabled to rule out the subjects not at risk and to convene as outpatients only those people at risk. The second step was constituted by the 'objective' VOE1 scale, administered by a physiotherapist in an outpatient setting, which ruled out another percentage of people considered not at risk according to this tool, and the final step by the 'objective' VOE2 scale, also administered in this case by a physiotherapist in an outpatient setting, which permitted to rule out the last percentage of elderly not at risk. Those resulted positive at all the three screening steps were considered at medium/high fall risk (for ≥1, ≥2, or ≥3 falls in the following twelve months based on the chosen algorithm), and could be addressed to a targeted multicomponent and multifactorial intervention. As expected, compared to the use of single tools, these serial screening algorithms permitted to obtain higher diagnostic accuracy in the prediction of the fall risk. In particular, the availability of cumulative post-test probabilities derived from the combination of the scales, in which the post-test probability of one test became the pre-test probability for the next test, was interesting because it opened to the possibility of crossing the threshold for intervention/no intervention, which was not achievable with a single screening test. Moreover, the interpretation of the disease in terms of probability, compared to other diagnostic accuracy measures, facilitated clinicians' comprehension and decision-making.

For each predicted outcome, the two screening algorithms were created using both the Youden Index and the 'ad hoc' methods to determine scale cutoffs, and this choice conducted to some differences in terms of diagnostic accuracy. First of all, the use of the 'serial' method with the 'AND' rule for the assembled algorithms led, per se, to higher overall specificity and positive predictive value, and an increase in false negative, compared to the administration of only one of the tests[227-230]. Consequently, on the one hand, the combination of single scales with high values of sensitivity and medium/low values of specificity, and with good NPVs and LR- associated to the

'Youden Index' cutoffs, led to 'Youden Index algorithms' which privileged very high values of specificity (FP rate [13.8%-15.7]), low values of sensitivity (FN rate [52.9%-61.9%]), medium PPV, high NPV, and higher LR+ compared to the 'ad hoc' ones. On the other hand, the combination of single scales with very high values of sensitivity and low values of specificity, and with good NPVs and LR- associated to the 'ad hoc' cutoffs, allowed to define the 'ad hoc' algorithms, which expressed more balanced values of sensitivity and specificity (FP rate [29%-40%] with limitation of false negative to specific rates [20.7%-37.6%]), medium/low PPV, very high NPV, and lower LR- compared to the 'Youden Index' ones. False negative rates of 50-60% obtained for the 'Youden Index algorithms' were considered excessive to accept, considering the specific predicted outcome 'fall risk' and the 'cost' ratio analysis described above, in which a false negative was considered to cost more than a false positive. Besides, lower LR- available for the 'ad hoc algorithms' permitted to reduce significantly, up to minimum percentages, the post-test fall risk probability of those who resulted negative to the tests and, so, to rule out the 'fall risk' exposition more effectively for them. Hence, given these results and considerations, the 'ad hoc' algorithms were preferred to the 'Youden Index' ones.

For 'external validation', we compared the serial screening algorithms, including the calibrated scales (CS), with serial algorithms, including the two calculated FRAT-up (FU), using the latter instrument as a comparator, being already validated and published. As already pointed out in the results chapter, the comparison between the algorithms based on the 'ad hoc' method cutoff definition, which we preferred, was realized using the same limitations of the false negative rates for both of them and, so, the same values of sensitivity. Regarding the other diagnostic accuracy measures (specificity, predictive values), the CS algorithms showed comparable values to the FU algorithms, except for predicting at least one fall, where the CS algorithms gave evidence of superior performances in the listed properties. The superiority was also confirmed for the cumulative post-test probabilities, highlighting a more effective capacity in ruling in and ruling out the disease for all the predicted outcomes. These results can be considered a satisfactory external validation of the serial algorithms based on the 'ad hoc' method, including the calibrated scales.

4.1.4 Construction of additional screening algorithms based on a logistic regression model

The final way to study the diagnostic accuracy of the calibrated scales and their use in screening algorithms, performed through the logistic regression technique, provided further information. The advantages of the use of this analysis in the construction of the screening algorithms were linked to the potential inclusion in the algorithm of other four available fall risk factors that the psychometric analyses excluded from the calibration of the scales (medications, psychological status, cognitive status, visual acuity), and to the possibility to draw a ROC curve with the respective AUC of the model-based algorithms. In particular, the model, including the three calibrated scales for all the predicted outcomes, showed a significantly higher discriminant capacity compared to the single tools, with AUCs well above 0.7 and classifiable as 'good'[122]. On the contrary, the disadvantages were the unique administration of all the scales (parallel use of tests[227, 229, 230]), with the necessity of the convocation of all the subjects to be screening in an outpatient setting, the loss of efficiency in avoiding unnecessary testing, and the more difficult handling and interpretation of the screening results by clinicians, who are generally more familiar with measurement scales than logistic models. For all the three predicted outcomes, it was not possible to include the other four available fall risk factors because their addition did not improve the model-based algorithm's discriminatory capacity.

Moreover, the three final model-based algorithms, which showed a goodness-of-fit to the sample data and included only two of the three calibrated scales, evidenced a non-significant increasing of the discriminatory power of the model if the third scale was added, and non-significant higher AUCs compared to the multivariable models based on the two calculated FRAT-up. The non-significant increase of the discriminatory power of the model if the third scale was added was a different result compared to that obtained with the algorithms based on the serial combination of all the three calibrated scales. In fact, this combination allowed for better probabilities in ruling in and ruling out the disease using three instruments rather than only two, through their positive and negative likelihood ratios. The resulting discrepancy may be due to the 'serial' use with the 'AND rule' of the scales vs. the 'parallel' use with the 'AND rule', where the instruments are administered at the same time, and to which the model-based algorithm

can be assimilated[227, 229, 230]. The serial combination exploits the increasing or decreasing disease probability (according to the positive or negative test results) after the first administered test (post-test probability), which becomes the new pre-test probability of the second one[231]. In the parallel combination case, the pre-test probability remains the same for all three test administrations.

4.1.5 Generated algorithms from the perspective of screening characteristics

The described serial algorithms, including the calibrated scales VAE, VOE1, and VOE2, could constitute the first component of an effective fall prevention program in adults older than 65 years old, as advocated by the World Health Organization[54] and by Holland[55]. In terms of content, these algorithms operate through a preventive operative mode[62], i.e., preventing falls by finding and removing the fall risk factors, and aiming to identify these factors, such as a history of falls, impaired balance, and reduced mobility, which may impact not only on determining a fall but also on elderly's activities and quality of life[55, 99]. This kind of screening could be delivered as a program screening, also known as organized screening or population-based screening, consisting of contacting people over 65 years old who are thought to be at increased fall risk to attend the evaluation[55, 62]. A fundamental role should be played by general practitioners to signal patients potentially at risk, detected through regular surveillance and case finding in primary case of their assisted persons[55]. Regarding the screening criteria defined by Wilson and Jungner in 1968, and the following review proposed over the past 40 years, the implementation of these algorithms respect these criteria in terms of: fall risk as an important health problem[2-4], and a recognize need of the population, also demonstrated by the high rate of elderly's self-reporting for the PRE.C.I.S.A. study; the identification of people at medium-high fall risk (at least one, two, or three fall(s) in the following twelve months) to be treated to reduce the risk in older adults over 65 years as the screening objective in a defined target population; the subjects interested by the screening are correctly informed to accept or not to participate in a confidential way, and are free to decline the invitation; the possibility of evidence-based effective and accepted interventions to reduce falls[2, 44]; the existence of adequate facilities for the diagnosis and the treatment as outpatient settings; equity and access to the screening for

the entire target population can be assured, exploiting the first anamnestic step administered by phone; the existence of recognizable early symptomatic stage, like initial muscular strength impairment, reduced mobility or impaired balance, as identified by the calibrated scales; the algorithm as a suitable test acceptable to the population, as undirected proved during the PRE.C.I.S.A. study; the known natural history of the disease; a program screening with regular surveillance and case finding; the screening algorithm as a part of a program that can integrate education, clinical services, and program management; the screening program can be evaluated in its quality, to minimize potential risks of screening; the overall benefits of the fall risk screening outweigh the potential harms (e.g., fall risk during the screening evaluation, cardiovascular complications to excessive physical exercise, increased mobility which could expose people, particularly those poorly aware of their own instability, at an even higher fall risk).

Regarding the economic evaluation, a fall risk screening based on our validated serial algorithms could be submitted to such an evaluation in terms of cost-effectiveness, cost-benefit, or cost-utility analyses. A recent example was given by Franklin and Hunter in 2019, who 'model the cost-effectiveness of a fall risk assessment between two fall risk screening tests (Quantified Timed Up&Go compared to Timed Up&Go), with referral to one of four fall-prevention interventions compared to no care pathway, when the decision to screen is based on older age in a primary care setting for community-dwelling people'[261]. They gave evidence of the cost-effectiveness of the first test compared to the second one and that the cost-effectiveness of QTUG-based care pathways, compared to no care pathway, was dependent on the age of the cohort screened for fall risk, with the most cost-effective option to screen only people aged 75–89 with referral to a fall-prevention intervention for those identified at fall risk. This happened because only one out of five falls suffered from an injury that required medical care, and for whom a QALY gain and cost-saving could be achieved. These adverse events occur more 3-4 times more often in people aged 75–89, compared to the 65-74 year-group, where the care pathways had a high probability of being cost-effective[261].

Another central point of the screening, discussed in the introduction chapter, is the information. In the case of the proposed fall risk serial algorithms, at the beginning

before the VAE administration, it would be crucial to give evidence-based information about the benefits (e.g., identification of the fall risk factors which can be treated, delivery of a report on modifiable risk factors, together with personalized recommendations for their minimization) and harms (e.g., increased mobility which could expose to a higher fall risk) of this program to all the invited individuals[55]. This would allow achieving informed consent by the older adults through shared decisions based on a balance and understandable picture of the options, and the outcomes should be followed[82]. In particular, concerning our kind of screening, it would be important that the elderly understand and accept that is a periodic screening (one-year recommended) where individuals define at medium/high risk would be addressed to a targeted multicomponent and multifactorial intervention. It is based on constant compliance to health professional's advice and recommendations (e.g., the daily independent performance of strength and balance exercises, the modification of the domestic environment to make it safe, and the willingness to adopt a perspective of personal responsibility in promoting our own healthy and active aging). Only the shared comprehension and acceptance between the health professional and the older adult of the listed elements give sense and efficacy to the fall risk screening, and to the following potential intervention to reduce it[82]. In cancer screening, patient compliance has been studied extensively, together with strategies to improve it. As an example, Subramanian et al., in their review of 2004[262], proposed a theoretical framework, which illustrated the dynamic relationship between physician recommendation and patient adherence. They highlighted factors impacting patient adherence, including physician recommendation, patient demographics, financial enablers, perceived screening risk, and healthcare system interactions. On the other hand, physician recommendations are, in turn, influenced by patient compliance, perceived test effectiveness, physician demographics, guideline awareness, and health system factors. Even evaluated in cancer screening framework, this representation could be valid also for the fall risk screening adherence.

A further point previously discussed has been the ethics linked to the screening. Authors agree that very considerable ethical responsibilities follow the screening application, as it potentially transforms individuals who are supposed to be 'healthy' to a state with some disorder or a potential one. In fact, it is the health service that invites individuals,

who consider themselves healthy, to be tested, to verify the presence of a particular condition at an early stage, either before or very soon after symptoms present, and to treat it with a reversible or at least a containable outcome. Besides, in screening, any abnormality that is identified must be treatable, and the investigation itself must not do any harm[55]. All these conditions can be generally satisfied by our fall risk serial screening algorithms, which identify modifiable risk factors to treat or to contain (e.g., balance exercises to reduce balance impairment, pharmacological optimization to limit the number of medications, footwear advice to improve stability in standing and gait, etc.). It is true that, even in this kind of screening, we could find the presence of non-modifiable risk factors, with whom the individual needs to coexist (e.g., aging, female gender). Then, regarding the disadvantages that screened people can experiment, false positive and negative subjects are present. In our specific case, we have already explained the realized 'cost' ratio analysis in favor of a higher cost of false negatives, which increases according to the outcome severity, and the consequent decision to limit the number of false negatives to specific rates. The justification was based on the acceptable consequences of defining 'at risk' those who are not (false positives) and, hence, treating 'unnecessarily' these subjects for the non-invasive characteristics of the successive treatments. In particular, the anxiety and sometimes morbidity that those with false positive results may experiment, reported by Chamberlain[85] concerning cancer screening, appear difficult to develop for the false positive persons to fall risk screening, except if having been defined as 'at fall risk' does not lead them to reduced mobility and to develop fear of falling, which may, in turn, increase their fall risk. On the other hand, it seems more likely that this definition would lead them to have greater attention and awareness of existing risks that could increase with aging.

Economic consequences related to false positive detection are, instead, present and to be considered in terms of unnecessary resources costs allocated to treat people at low fall risk. However, as generally happens in screening[55], the fall risk screening is perceived as acceptable, safe, and low-risk by older adults. An indirect demonstration was proven, once more, by the high rate of elderly's self-reporting for the PRE.C.I.S.A. study, which aimed to identify and treat people at medium-high fall risk to reduce it.

Concerning screening evaluation, the validated serial algorithms appear to satisfy most of the criteria for evaluation of screening proposed by Cochrane in 1971[67], which are

still valid today. In particular, they are simple to administer and interpret, acceptable to the target population, accurate in giving a true measurement of the studied condition, with good sensitivity and specificity properties already presented in detail. Regarding repeatability of the results and the cost of the test in relation to the benefits of early detection of the disease, evidence should be obtained from new studies, as randomized control trials, which represent the gold standard for the evaluation of screening tools and methods, or high-quality observational studies[55, 62]. Considering biases that can affect screening and highlighted in the introduction chapter, these are mainly linked to the 'survival rate' used as the screening outcome, even if not recommended[62]. However, in the fall risk screening, we consider other kinds of outcomes, like the probability of the event 'fall' in the following twelve months, together with function, activity, and quality of life outcomes (improved balance and muscular strength, increased mobility, reduced fear of falling, improved social participation, etc.). For this reason, potential biases to consider could be the length-biased sampling (when rapidly progressing disease leads a person to consult a health professional immediately, without waiting for the screening call, and this person will have probably a worse prognosis of fall risk: the results will tend to suggest that screening is more effective than is really the case), and the selection bias (those who are most health-conscious of their health condition are likely to participate and to contain better their fall risk in any case, whether they are screened or not).

Also, quality control of the screening addressed to its performance and effectiveness can be developed to implement a fall risk organized screening program, of which our validated algorithms may constitute an important part. The quality, in fact, must exist throughout the screening pathway[97]. Adopting organized screening criteria defined by Hakama[94] and Holland[55], our program can satisfy: the identification of the target population and the individual people to be screened (i.e., older adults over 65 years old); the possible implementation of mechanisms to guarantee high coverage and attendance (i.e., personal letter of information about the importance of fall risk screening, and invitation to answer to the VAE screening step by phone), and to perform screening in pleasant and acceptable conditions for older adults; the maintenance of the best standards in the administration of the test (i.e., personnel training, routine checks of the validity of the tests performed); the presence of facilities (i.e., outpatient setting, trained

and experienced health professionals) to deliver an appropriate intervention of confirmed people at risk; the presence of designed and shared pathway between the participant, the screening, and the intervention centers, which include all the screening program steps; the implementation of regular checks on the feelings of satisfaction of those who have undergone the screening process; the evaluation and monitoring of the total program, in terms of incidence rates and other individual functioning outcomes, including quality of life, and social participation, among those attending.

4.2 Assessment of the effect of neurological diseases on the fall risk screening tests

The assessment of the effect of neurological diseases (elderly and elderly with associated neurological diseases (PD and stroke)) on the fall risk screening tests was realized on the calibrated scales VAE, VOE1, and VOE2, and on the two calculated FRAT-up on VAE and VOE variables. The different adopted analytical techniques, i.e., the analysis of the presence of differential item functioning (DIF) or item bias for the two subgroups of neurological diseases in the Rasch analysis context for the calibrated scales, and an independent sample t-test between the two subgroups of neurological diseases for the FRAT-up gave us different possibilities in the effect management when found.

Hence, regarding the neurological diseases effect highlighted for the VAE scale, the uniform differential item functioning for testlet 1, which demonstrated a large effect on VAE fall risk person estimates, was possible to manage through a splitting analysis, and the consequent preparation of two raw scores into interval-level estimates conversion tables, one for each neurological diseases subgroup. The conversion has been and would be realized before calculating the scale cutoffs and the algorithm results, allowing the use of validated fall risk serial algorithms both for elderly and elderly with associated neurological diseases (PD and stroke), which can be considered neurological diseases effect-free.

Differently, regarding the effect of the neurological disease highlighted for the FRAT-up VOE, the significantly higher fall risk mean probability for the elderly with associated neurological disease group, which determined a small effect, was not possible to handle in the t-test context. For this reason, a small neurological disease

effect on FRAT-up VOE estimated probabilities should be considered in the single use of this tool and/or of algorithms that include it.

4.3 Validation of an ICD&ICF core set for the fall risk in community-dwelling older adults (also with associated neurological diseases)

The new core set for the fall risk in community-dwelling older adults, also with associated neurological diseases, is composed of 103 fall risk factors, which were linked to 74 categories (7 to ICD and 67 to ICF) and 7 ICF 'Personal factors', and it represents the practical implementation of the use of the WHO-FIC as a conceptual framework for the classification of these risk factors. Thanks to the ICF linking rules[135-137], each risk factor was linked to one or more ICD or ICF categories of the different classification domains, which altogether constituted an adapted set of descriptors, i.e., the core set.

The use of the WHO-FIC (ICD and ICF) covers the lack of a common classification framework for the fall risk factors, which can really become a universally accepted and used reference framework even in this field. Besides, the combined and complementary use of ICD and ICF allows providing an etiological framework for health conditions in the ICD and an associated classification for functioning and disability in ICF[124]. With our core set, we demonstrated the substantial independence of the fall risk factors from one or more diseases that predispose to fall. In fact, these diseases (e.g., Parkinson's Disease or stroke) are considered themselves as fall risk factors in literature, but which can be associated with a varied range of other risk factors, different from subject to subject, because different is the subject's functioning in his/her interaction with the environmental and personal factors. As already pointed out in the introduction chapter, it is not the disease itself to expose the subject to a fall (e.g., Parkinson's Disease or stroke), but rather the consequences of the disease, expressed in terms of impairment, activity limitation, and reduced participation, which per se are not pathology-specific. As an example, functioning profiles of two persons with Parkinson's Disease can completely differ because different are the disease manifestations of signs and symptoms classifiable as risk factors, the possible presence of other risk factors due to comorbidities, and the interaction with the environmental and personal factors. The strength of this kind of approach is just the chance to classify the functioning profile of

older adults in terms of fall risk, based on an accurate assessment, and not on pre-established etiological classifications, and to consider in the same unique framework also older adults with associated neurological diseases.

As the other already validated core set available in literature[138], our core set represents a selection of ICF categories from the entire classification for specific health conditions, condition groups, and settings. It aims to facilitate a systematic and comprehensive description of functioning for various purposes and in various settings, including clinical practice and research. Given its length and detail, it can be considered a comprehensive core set for fall risk in older adults also with neurological diseases, which includes a list of ICF categories sufficiently exhaustive to describe in multi- and interdisciplinary assessment the typical spectrum of problems in the functioning of patients with this specific condition. In fact, it encourages members of the team to consider potentially relevant aspects of functioning, even in areas of functioning outside of their respective disciplines. From it, it would be possible to derive a brief version, considerably shorter than the comprehensive one, ideal (although not limited) for use in both clinical studies and single-profession clinical encounters[139, 141].

Compared to the already existing 'ICF core set for falls in acute rehabilitation settings' published by Yen et al. in 2014[143], our core set presents several differences in terms of development method, number and type of included categories, type of setting, and target population. Regarding the development method, we shared the review to identify risk factors and the factor linking to the classification categories, but we did not perform the following Delphi surveys to reach a consensus on the selection of categories. As a replacement, we realized the comparison with the Yen core set, using it as an external comparator. Furthermore, our core set includes a higher number of categories, also derived from the ICD classification. Finally, it addresses to older adults living in the community also with associated neurological disease. In contrast, Yen's core set was focused on an acute rehabilitative setting, which can incorporate other types of disease (e.g., neurological, orthopedic, respiratory, etc.), and also younger people. Common advantages to the core set for fall risk factors have been listed exhaustively by Yen in his paper[143]. In addition to the possibility to use the WHO-ICF as a conceptual framework to systematically organized risk factors, he also affirmed that the highlighted interactions between falls, individual functioning, environmental and personal factors

offer a comprehensive basis for developing multifactorial fall prevention strategies in acute rehabilitation settings[143], and with our core set, also in the community setting. The systematic assessment of fall-related risk factors, provided by the two core sets, is essential to project and implement risk-based fall prevention programs. It is important to note that the ICD&ICF core set only tells us what to measure and not how to measure the respective categories. In this sense, a further possibility given by the linking rules[135-137] is the linking also of the measurement tools used in clinical practice to the ICF categories. The conceptual mapping of all the numerous instruments described in the introduction chapter could help in understanding risk factors (and, hence, categories) that are 'over mapped', and those which, instead, are 'under mapped', to focus on the latter to validate other tools only in case of uncovered conceptual features. The same chance is also provided for the intervention linking, for whom an evaluation of the intervention targets, using the ICF core set, could demonstrate the effectiveness of interventions to reduce the fall risk[143], or the necessity of further interventions to address different risk factors.

4.4 Impact of the three aims on the Public Health domain

Regarding the first aim, with the validation of fall risk serial screening algorithms, and in particular, those based on the clinical 'ad hoc' cutoff, three calibrated scales (VAE, VOE1, and VOE2) would be available, which constitute successive screening steps with good diagnostic accuracy, both in terms of sensitivity and specificity, combined with cost-effectiveness features of the employed resources and the subject's assessment load. Fall risk detection is formulated when the subject results positive to all three tests. This method is particularly cost-effective, has the advantage of avoiding unnecessary tests, but the disadvantage of potentially delaying treatment for diseased patients by lengthening the diagnostic testing period that could remain exposed to a high risk[227-229].

These algorithms could allow implementing a program of 'proactive medicine or healthcare' to identify those at fall risk on a large scale, that could constitute a fall risk screening program. For 'proactive medicine or healthcare' is meant an assistance model for the management of chronic diseases that does not 'wait' for the consequences of the disease to require hospitalization ('reactive medicine or healthcare'), but realizes adequate and differentiated interventions concerning the risk level, also focusing on

prevention and education before diseases arise or worsen. The reference is the 'Chronic care model' (CCM), proposed by E.H. Wagner in 1998[263], which was the first widely disseminated system, and the one that served as the basis for all subsequent models, such as the later Expanded Chronic Care Model (ECCM)[264], used and proposed by the Government of British Columbia of Canada. It emphasizes the importance and relevance of the community context, as well as the importance of prevention and health promotion, aimed at the creation of the Framework for Innovative Care for Chronic Diseases (ICCC) of the World Health Organization (WHO), which is a variant of the original model of Wagner. The ICCC highlights the great importance and urgent need for commitment in the project of health and the social part in an integrated way (key to this project). Besides, it gives evidence to the necessities to sustain, at political and cultural levels, the promotion of health and the prevention of diseases, together with the optimization of the use of existing resources, making the comprehensive management effective and efficient through the formulation of integrated socio-health policies. In Italy, this model was adopted by the Ministero della Salute in its 'Piano Nazionale della cronicità' of 2016[265]. At now, it is widespread in some regions, such as Tuscany and Emilia-Romagna.

The availability of similar fall risk screening algorithms could have many positive implications in the Public Health field. In fact, inside a 'proactive medicine' framework, such a tool, which allows implementing secondary prevention of a widespread and impactful phenomenon in the elderly as the falls, would be available, especially compared to the fragmented, diversified and discontinuous 'usual care', like that of the various Italian healthcare realities. The better identification of those at risk would make possible to deliver targeted and specific interventions, with consequences of a reduction in personal, social, and material costs, but also of increase of older people's social participation, often limited by the fear of falling after an event, and of reduction of morbidity and mortality risks. Regarding the health situation in Emilia-Romagna, it would be conceivable that this screening system would be carried out in dedicated outpatient clinics, like the Case della Salute ('Falls clinic'), and/or in the context of a home physiotherapy assessment for older patients, e.g., over 80. By the way, the activation of a 'falls clinic' and the promotion of a gymnastics program for the older

adults at fall risk at the Case della Salute were two of the declared objectives of the Piano Regionale della Prevenzione of the Regione Emilia-Romagna 2015-2018[266].

Regarding the second aim (a), the study of the effect of the neurological disease on the fall risk screening algorithms including the three calibrated scales showed the presence of a large effect only on one testlet of the first scale (VAE), which was managed in the context of the Rasch analysis. Thanks to this solution, the fall risk serial screening algorithms can be considered valid and extendable to elderly and elderly with associated neurological diseases (PD and stroke). In the Public Health field, the identification of people at fall risk in these populations is of primary importance, because in those people, the risk is even greater for the impairment of the balance dependent on aging, to which that intrinsically due to neurological disease and other risk factors is associated. Therefore, it would be necessary to identify precisely those at a higher fall risk and more exposed to multiple falls, potentially with greater consequences. Our validated algorithms have the advantage to be fall risk screening tools with good diagnostic accuracy, which are independent of the disease that places the subject in a risk situation and, so, neurological diseases effect-free. Consequently, they allow to include in health policies for older age also the detection of people with a higher fall risk due to associated neurological diseases, who are generally not considered.

Regarding the second aim (b), the availability of a comprehensive ICF core set for the fall risk in community-dwelling older adults, including also those with associated neurological diseases, allows the linking of the information contained in the specific fall risk factor to the corresponding ICD/ICF category, according to coded rules. This core set, then, could represent an applicable guide in Public Health clinical practice for a shared and unique description of the fall risk in this population. The main fields in which it could be found its natural use are the geriatric and rehabilitative, which involve community-dwelling individuals over 65 years old, with the possible association of neurological diseases, as PD and stroke.

4.5 Study limitations and future developments

4.5.1 Study limitations

The principal study limitations concern the sample used for the calibration of the screening algorithms, the definition of the outcome 'fall' and its detection, the scale cutoff definition, and the type of core set validation. In detail, we used a 'convenience' sample derived from the PRE.C.I.S.A study in which the validation of a fall risk serial screening algorithm was a secondary aim. For this reason, the sample included individuals who could be considered at higher risk compared to the population of community-dwelling older adults because one of the indications given to health professionals to signal an individual for the PRE.C.I.S.A. study was that 'it could be at fall risk?'. This could be led to avoid individuals' signaling without evident fall risk factors and, hence, obtain a sample most affected by the fall risk. Besides, it is important to remember that the sample was constituted by the union of the two groups of randomized subjects of the PRE.C.I.S.A. study with further subjects excluded from that study because 'at low fall risk/not satisfied criteria'. Although this was justified based on the PRE.C.I.S.A. primary endpoint analysis, which did not show a significant difference between the two groups, we cannot exclude a significant improvement at the individual level among those who had been subjected to the experimental intervention. In addition, even if we also included older adults with associated neurological diseases (PD and stroke), not all kinds of these possible diseases most frequently found in old age were considered (e.g., vascular parkinsonism, diabetic neuropathy, medullary canal stenosis, etc.).

A further critique point regards the definition of the outcome 'fall' and its detection. As already pointed out in the introduction chapter, many definitions of 'fall' still exist in the literature, and, at now, there is no universal consensus on that[2, 9, 31]. In this project, we followed the adoption of Lamb's definition[32] by the PRE.C.I.S.A. study of which we used the data. Still, we cannot exclude that participants have also reported other events not included in this definition of 'fall'. Besides, as described in the PRE.C.I.S.A. final report[146], an additional factor that may have significantly affected the outcome of the trial is related to the method of detection of falls. Given the known poor reliability of the elderly's fall retrospective self-reporting, which leads to underestimating the

incidence of falls and overestimating falls in multiple fallers, the measures adopted to prevent it were a partial personalization of the falls diary and the monthly telephone follow-up. Since it was not possible to offer economic incentives to participants for the monthly delivery of completed falls diary, in the same way, it was not possible to verify monthly the actual compilation of them. Hence, it is plausible that, in the context of the monthly telephone follow-up, subjects' data were provided by a part of them not based on the information recorded prospectively in the diary, but rather of 'remember' information[146]. Thus, recall bias and bias due to missing data could have influenced the PRE.C.I.S.A. outcome detection, and so the outcome was used in the current project.

Regarding the definition of the optimal cutoff for predicting the described outcome for all the tools considered in the thesis, we selected the positivity cutoffs after performing the tests, choosing the ones that maximized test performances according to our clinical needs. According to the STARD 2015 guidelines for reporting diagnostic accuracy studies[237, 238], in this way, there is an increased risk that the resulting accuracy estimates are overly optimistic, especially in small studies, and that subsequent studies may fail to replicate the findings.

Finally, as described in the methods and discussion chapters, for the validation of the ICD&ICF core set for the fall risk, we performed a review to identify risk factors and the factor linking to the classification categories as Yen did in his core set[143], but we realized the comparison with the Yen core set, using it as an external comparator, instead of the Delphi surveys, to reach a consensus on the selection of categories. Compared to the standardized and scientifically-based process to develop core sets proposed by the ICF Research Branch[138] and by Selb[139], our validation lacks some phases of the described process in terms of a preparatory empirical multicenter study, of experts' survey, and of testing the proposed core set on a validation sample.

4.5.2 Future developments

The future developments of this project concern, first of all, the necessity to confirm the obtained diagnostic accuracy of the serial screening algorithms on a new validation sample, using the appropriate design study (i.e., prospective), also with the improvement of the representativeness of the population, including community-dwelling older adults without any prior clinical hypothesis on their potential augmented

risk, and also older adults with associated neurological diseases with other pathologies in addition to PD and stroke. Regarding the outcome, it should be improved the detection efficacy, considering other strategies besides those of the PRE.C.I.S.A. study, like the monthly return of falls diaries by post[267], associated with an economic incentive for such redelivery[268], the personalization of these diaries[268], or the use of new technologies (e.g., wearable inertial sensors) which can register the 'fall signal' and, so, to document its real occurence[269].

Once the screening properties are confirmed on a new validation sample, the algorithms should be submitted to economic evaluation, in terms of cost-effectiveness, cost-benefit, or cost-utility analyses, following the example of Franklin and Hunter in 2019[261].

In conclusion, concerning the comprehensive ICD&ICF core set for fall risk, future validation studies based on empirical data collected from community-dwelling older adults, also with associated neurological diseases, are required. After the validation, the brief version of the core set may be extracted.

5 Conclusions

In this thesis, we validated a fall risk serial screening algorithm with a good level of diagnostic accuracy, including three calibrated scales VAE, VOE1, and VOE2, on which the identification of optimal cutoffs, using an 'ad hoc' clinical method, allows the prediction of at least one, two, and three falls in the following twelve months in a sample of community-dwelling older people, also including subjects with Parkinson's Disease and stroke sequelae.

We also demonstrated that the effect of the neurological disease (PD and stroke) on the performance of these tools is minor and manageable in a modern psychometric technique context, like that of Rasch analysis.

These algorithms could help implement an effective Public Health program of 'proactive medicine or healthcare', identifying older adults at fall risk on a large scale (e.g., fall risk screening program). They could also be applied in elderly with associated neurological diseases (PD and stroke) subjects, giving the possibility to use the algorithms in most elderly populations.

Besides, we validated a comprehensive ICD&ICF core set for the fall risk in community-dwelling older adults, also with associated neurological diseases.

Further study projects are desirable to replicate all these findings in the context of larger, multicenter validation studies, improving the sample representativeness, in terms of fall risk level and neurological diseases, of the 'fall' detection modality, and then providing an economic evaluation of the proposed screening algorithm. Also, the comprehensive core set requires to confirm its validity in a new validation study.

6 Appendix A – FRAT-up conversion tables for VAE and VOE variables

A1.FRAT-up conversion table for VAE variables

FRAT-up item and score	Corresponding VAE item and score	Conversion rule
FRAT-up01.Does the subject use a walking aid? Yes/No	n/a	Use of prevalence: 0.18[270]
FRAT-up02.Does the subject use antiepileptics? Yes/No	n/a	Use of prevalence: 0.01[271]
FRAT-up03.Urinary incontinence last year? Yes/No	n/a	Use of prevalence: 0.19[270]
FRAT-up04.Does the subject suffer any pain? (generic pain, e.g., muscular cramps), pain at feet, stomach pain, chest pain, pain in legs, back pain, pain at hips or knees) Yes/No	n/a	Use of prevalence: 0.30[272]
FRAT-up05.Does the subject live alone? Yes/No	n/a	Use of prevalence: 0.32[270]
FRAT-up06.Does the subject use sedatives? Yes/No	n/a	Use of prevalence: 0.14[271]
FRAT-up07.Does the subject suffer from Parkinson? Yes/No	n/a	Use of prevalence: 0.008[273]
FRAT-up08.History of previous falls? (last 12 months) Yes/No	VAE01.Is there a history of any fall in the previous year? Yes/No	If VAE01.Is there a history of any fall in the previous year? = Yes → FRAT-up08.History of previous falls? (last 12 months) = Yes
FRAT-up09.Diabetes blood glucose 126 or suspected diabetes? Yes/No	n/a	Use of prevalence: 0.11[273]
FRAT-up10.Does the subject suffer rheumatic disease? Yes/No	n/a	Use of prevalence: 0.47[274]
FRAT-up11.Dizziness or unsteadiness last year? Yes/No	n/a	Use of prevalence: 0.20[270]
FRAT-up12.Does the subject use antihypertensives? Yes/No	n/a	Use of prevalence: 0.32[275]
FRAT-up13.History of previous strokes? Yes/No	n/a	Use of prevalence: 0.13[273]
Gender recorded in the demographic data		If Gender = female → FRAT-up14.Is the subject female? = Yes

FRAT-up item and score	Corresponding VAE item and score	Conversion rule
FRAT-up15.Fear of falling (activities performed either in the home environment or a community environment)?[276] Yes/No	VAE06.Does the patient feel a fear of falling? Yes/No	If VAE06.Does the patient feel a fear of falling? = Yes → FRAT-up15.Fear of falling = Yes
FRAT-up16.Hearing impairment? (Do you have any trouble hearing (TH)?) 0= No 1=Slight deafness 2=Severe deafness 3=Conversation impossible Hearing impairment cutoff >= 1	n/a	Use of prevalence: 0.36[271]
FRAT-up17.Visual stereognosis (number of tests passed) Score: 0-9 Impairment cutoff ≤3	n/a	Use of prevalence: 0.19[271]
FRAT-up18.Number of Activities of Daily Living (ADL) disabilities (washing face and arms; controlling urination and bowel movements; dressing and undressing; getting in and out of bed; eating, e.g., holding a fork, cutting food, drinking from a glass; using the toilet) Score: 0-6 Physical disability cutoff ≥1	n/a	Use of prevalence: 0.11[277]
FRAT-up19.Mini-Mental State Examination score [278] Score: 0-30 Cognitive impairment cutoff ≤20	n/a	Use of prevalence: 0.19[277]
FRAT-up20.Revised Walking Subscore Score: 0-10 Gait problems cutoff ≤5	n/a	Use of prevalence: 0.42[279]
FRAT-up21.Center for Epidemiologic Studies Depression Scale (CES-D)[280] Score: 0-60 Depression cutoff >20	n/a	Use of prevalence: 0.13[271]
FRAT-up22.Contrast sensitivity? Score: 1-19 Impairment cutoff ≤16	n/a	Use of prevalence: 0.19[271]
FRAT-up23.How does the subject feel (How would you evaluate your current health? How do you feel now?) 1=Very poor 2=Poor 3=Fair (so-so) 4=Good	n/a	Use of prevalence: 0.20[270]

FRAT-up item and score	Corresponding VAE item and score	Conversion rule
5=Very good Poor health-perceived health status cutoff ≤2		
FRAT-up24.Visual acuity (3 meters) (Monoyer's scale) 1=1/10 2=2/10 3=3/10 4=4/10 5=5/10 6=6/10 7=7/10 8=8/10 9=9/10 10=10/10 11=11/10 Impairment cutoff ≤5	n/a	Use of prevalence: 0.19[271]
FRAT-up25.Number of drugs used by the subject Score:0-10	VAE02.Is the patient/client on four or more medications per day? (Number of medications)	If VAE02.Is the patient/client on four or more medications per day? = X → FRAT-up25.Number of drugs used by the subject = X
FRAT-up26.Subject's number of Instrumental Activities of Daily Living (IADL) (using the telephone; using public transportation; cooking a simple meal; doing light housework, e.g., doing dishes, light cleaning; doing heavy housework, e.g., washing windows, floor; taking medications correctly; managing home finances; shopping daily for basic necessities) Score: 0-8 Instrumental disability cutoff >=1	n/a	Use of prevalence: 0.37[273]
FRAT-up27.Physical activity level (last 12 months)[281] 1=Hardly any physical activity 2=Mostly sitting/some walking 3=Light exercise 2-4 hours/week 4=Moderate 1-2 hours or light >4 hours/week 5=Moderate exercise >3 hours/week 6=Intense exercise many times/week 7=Walks 5+ km/day, 5+days/week, 5+ years Physical activity limitation cutoff ≤2	n/a	Use of prevalence: 0.56[281]
FRAT-up28.Age (years) Score: 0-150	Age recorded in the demographic data	If Age = X → FRAT-up28.Age (years) = X

A2.FRAT-up conversion table for VOE variables

FRAT-up item and score	Corresponding VOE item and score	Conversion rule
FRAT-up01.Does the subject use a walking aid? Yes/No	VOE02.Does the patient need a walking aid to perform the 10 meters walking test? Yes/No	If VOE02.Does the patient need a walking aid to perform the 10 meters walking test? = Yes → FRAT01.Does the subject use a walking aid? = Yes
FRAT-up02.Does the subject use antiepileptics? Yes/No	n/a	Use of prevalence: 0.01[271]
FRAT-up03.Urinary incontinence last year? Yes/No	n/a	Use of prevalence: 0.19[270]
FRAT-up04.Does the subject suffer any pain? (generic pain, e.g., muscular cramps), pain at feet, stomach pain, chest pain, pain in legs, back pain, pain at hips or knees) Yes/No	n/a	Use of prevalence: 0.30[272]
FRAT-up05.Does the subject live alone? Yes/No	n/a	Use of prevalence: 0.32[270]
FRAT-up06.Does the subject use sedatives? Yes/No	n/a	Use of prevalence: 0.14[271]
FRAT-up07.Does the subject suffer from Parkinson? Yes/No	n/a	Use of prevalence: 0.008[273]
FRAT-up08.History of previous falls? (last 12 months) Yes/No	VOE11.Number of falls in the past 12 months? 0=No falls 1=1 fall 2=2 falls 3=3 or more falls	If VOE11.Number of falls in the past 12 months? ≥1 → FRAT-up08.History of previous falls? (last 12 months) = Yes
FRAT-up09.Diabetes blood glucose 126 or suspected diabetes? Yes/No	n/a	Use of prevalence: 0.11[273]
FRAT-up10.Does the subject suffer rheumatic disease? Yes/No	n/a	Use of prevalence: 0.47[274]
FRAT-up11.Dizziness or unsteadiness last year? Yes/No	VOE17. When walking and turning, does the person appear unsteady or at risk of losing their balance? 0=No unsteadiness observed 1=Yes, minimally unsteady on walking or turning 2=Yes, moderately unsteady on walking or turning (needs supervision) 3=Yes, consistently and severely unsteady on walking or turning (needs constant hands on assistance)	If VOE17.When walking and turning, does the person appear unsteady or at risk of losing their balance? ≥ 1 → FRAT-up11.Dizziness or unsteadiness last year? = Yes
FRAT-up12.Does the subject use antihypertensives? Yes/No	n/a	Use of prevalence: 0.32[275]

FRAT-up13.History of previous strokes? Yes/No	n/a	Use of prevalence: 0.13[273]
FRAT-up14.Is the subject female? Yes/No	Gender recorded in the demographic data	If Gender = female → FRAT-up14.Is the subject female? = Yes
FRAT-up15.Fear of falling (activities performed either in the home environment or a community environment)?[276] Yes/No	n/a	Use of prevalence: 0.33[275]
FRAT-up16.Hearing impairment? (Do you have any trouble hearing (TH)?) 0= No 1=Slight deafness 2=Severe deafness 3=Conversation impossible Hearing impairment cutoff >= 1	n/a	Use of prevalence: 0.36[271]
FRAT-up17.Visual stereognosis (number of tests passed) Score: 0-9 Impairment cutoff ≤3	n/a	Use of prevalence: 0.19[271]
FRAT-up18.Number of Activities of Daily Living (ADL) disabilities (washing face and arms; controlling urination and bowel movements; dressing and undressing; getting in and out of bed; eating, e.g., holding a fork, cutting food, drinking from a glass; using the toilet) Score: 0-6 Physical disability cutoff ≥1	VOE13.Prior to this fall, how much assistance was the individual requiring for personal care activities of daily living (e.g., dressing, grooming, toileting)? *(NOTE: If no fall in last 12 months, rate current function)* 0=None (completely independent) 1=Supervision 2=Some assistance required 3=Completely dependent	If VOE13.Prior to this fall, how much assistance was the individual requiring for personal care activities of daily living (e.g., dressing, grooming, toileting)? >0 → FRAT-up18.Number of Activities of Daily Living (ADL) disabilities ≥1
FRAT-up19.Mini-Mental State Examination score [278] Score: 0-30 Cognitive impairment cutoff ≤20	VOE21.Cognitive status (according to the Abbreviated Mental Test Score by Hodkinson, 1972[150]) 1=AMTS 9-10 or intact 2=AMTS 7-8 mildly impaired 3=AMTS 5-6 moderately impaired 4=AMTS 4 or less severely impaired	If VOE21.Cognitive status = 3 o 4 → FRAT-up19.Mini-Mental State Examination score ≥20
FRAT-up20.Revised Walking Subscore Score: 0-10 Gait problems cutoff ≤5	VOE01.10 Metres Walking test (seconds)	If VOE01.10 Metres Walking test > 12 seconds → FRAT-up20. Revised Walking Subscore ≤5
FRAT-up21.Center for Epidemiologic Studies Depression Scale (CES-D)[280] Score: 0-60 Depression cutoff >20	VOE20.Psychological status (anxiety, depression, loss of cooperation, of insight or judgment) 1=Does not appear to have any of these 2=Appears mildly affected by one or more 3=Appears moderately affected by one or more 4=Appears severely affected by one or more	If VOE20.Psychological status > 1 → CES-D > 20
FRAT-up22.Contrast sensitivity?	n/a	Use of prevalence: 0.19[271]

Score: 1-19 Impairment cutoff ≤16		
FRAT-up23.How does the subject feel (How would you evaluate your current health? How do you feel now?) 1=Very poor 2=Poor 3=Fair (so-so) 4=Good 5=Very good Poor health-perceived health status cutoff ≤2	n/a	Use of prevalence: 0.20[270]
FRAT-up24.Visual acuity (3 meters) (Monoyer's scale) 1=1/10 2=2/10 3=3/10 4=4/10 5=5/10 6=6/10 7=7/10 8=8/10 9=9/10 10=10/10 11=11/10 Impairment cutoff ≤5	n/a	Use of prevalence: 0.19[271]
FRAT-up25.Number of drugs used by the subject Score:0-10	VOE19.Medications (sedatives, anti-depressants, anti-Parkinson's, diuretics, anti-hypertensives, hypnotics) 1=Not taking any of these 2=Taking one 3=Taking two 4=Taking more than two	If VOE19.Medications = 1 → FRAT-up25.Number of drugs used by the subject = 0 If VOE19.Medications = 2 → FRAT-up25.Number of drugs used by the subject = 1 If VOE19.Medications = 3 → FRAT-up25.Number of drugs used by the subject = 2 If VOE19.Medications = 4 → FRAT-up25.Number of drugs used by the subject >2
FRAT-up26.Subject's number of Instrumental Activities of Daily Living (IADL) (using the telephone; using public transportation; cooking a simple meal; doing light housework, e.g., doing dishes, light cleaning; doing heavy housework, e.g., washing windows, floor; taking medications correctly; managing home finances; shopping daily for basic	VOE15.Prior to this fall, how much assistance was the individual requiring for instrumental activities of daily living (e.g., shopping, housework, laundry)? *(NOTE: If no fall in last 12 months, rate current function)* 0=None (completely independent) 1=Supervision 2=Some assistance required	If VOE15.Prior to this fall, how much assistance was the individual requiring for instrumental activities of daily living (e.g., shopping, housework, laundry)? >0 → FRAT-up26.Subject's number of Instrumental Activities of Daily Living (IADL) ≥1

necessities) Score: 0-8 Instrumental disability cutoff >=1	3=Completely dependent	
FRAT-up27.Physical activity level (last 12 months)[281] 1=Hardly any physical activity 2=Mostly sitting/some walking 3=Light exercise 2-4 hours/week 4=Moderate 1-2 hours or light >4 hours/week 5=Moderate exercise >3 hours/week 6=Intense exercise many times/week 7=Walks 5+ km/day, 5+days/week, 5+ years Physical activity limitation cutoff ≤2	n/a	Use of prevalence: 0.56[281]
FRAT-up28.Age (years) Score: 0-150	Age recorded in the demographic data	If Age = X → FRAT-up28.Age (years) = X

7 Appendix B – ICD&ICF core set for the fall risk in community-dwelling older adults (also with associated neurological diseases)

Category[135-137]	**Descriptor**	**Risk factor**	**Effect measure (OR, HR, RR with CI 95%) for all fallers**	**Effect measure (OR, HR, RR with CI 95%) for recurrent fallers**
Diseases (ICD-11 version 04/2019)[249]				
8A00.0	Parkinson disease	Parkinson disease	2.71 (1.08-6.84)[41]	2.84 (1.77-4.58)[41]
8A00.2	Secondary parkinsonism			
8C00-8C03	*Polyneuropathies and other disorders of the peripheral nervous system*	*Other comorbidities (peripheral neuropathy)*[20]	-	-
8B00-8B26	Cerebrovascular diseases	History of stroke	1.61 (1.31-1.98)[41]	1.79 (1.51-2.13)[41]
FA00-FA38	Arthropathies	Rheumatic disease	1.47 (1.28-1.70)[41]	1.57 (1.42-1.73)[41]
FA00-FA05	*Arthrosis*	*Arthrosis*[20]	-	-
FB83.1	*Osteoporosis*	*Osteoporosis*[20]	-	-
Body functions (ICF 2017 - English)[131]				
b110	Consciousness functions*	*Confusion/disorientation*[143]	-	-
b114	Orientation functions*			
b140-b189	Specific mental functions	Cognitive impairment[20]	1.36 (1.12-1.65)[41] 3.38 (1.31-8.67)[23, 282] 0.92[+] (0.86-0.99) [23, 283] 0.80 (0.63-1.02)[23, 284] 1.75 (1.02-2.99)[30]	1.56 (1.26-1.94)[41] 3.44 (1.05-11.5)[19, 23]
b140	Attention functions*	Attention[143]	3.32[+] (1.40-7.90)[23, 285]	-
b1402	*Dividing attention*	*Dual tasking*[20]	-	-
b1441	Long-term memory	Semantic fluency	0.90° (0.84-0.98)[23, 24]	-
b152	Emotional functions	Depression[20]	1.63 (1.36-1.94)[41] 1.05 (1.02-1.08)[23, 286] 1.11 (1.01-1.21)[23, 250] 2.11 (1.18-3.75)[30]	1.86 (1.45-2.38)[41]
		Fear of falling	1.55 (1.14-2.09)[41] 4.24 (1.29-14)[23, 25] 1.04 (1.01-1.08)[23, 284]	2.51 (1.78-3.54)[41]
		Anxiety[20]	-	-
b1565	Visuospatial perception*	*Hemineglect*[143]	-	-
b1600	Pace of thought	Central processing speed (cognitive reaction time)	1.22[+] (1.05-1.43)[23, 285]	-
b164	Higher-level cognitive functions	Executive functions	6.29 (2.75-14.4)[23, 282]	-

Category[135-137]	Descriptor	Risk factor	Effect measure (OR, HR, RR with CI 95%) for all fallers	Effect measure (OR, HR, RR with CI 95%) for recurrent fallers
			1.63+ (1.01-2.64)[23, 283]	
b210	Seeing functions*	Visual impairment[20, 143]	1.35 (1.18-1.54)[41]	1.60 (1.28-2.00)[41]
b21022	Contrast sensitivity	Visual contrast	0.66 (0.45-0.98)[23, 282]	-
b2152	*Functions of external muscles of the eye*	*Oculomotor impairment*[20]	-	-
b230	Hearing functions	Hearing impairment	1.21 (1.05-1.39)[41]	1.53 (1.33-1.76)[41]
b235	Vestibular functions*	*Balance and posture stability*[143]	-	-
b240	Sensations associated with hearing and vestibular function*	Dizziness and vertigo[20, 143]	1.80 (1.39-2.33)[41]	2.28 (1.90-2.75)[41]
b2402	Sensation of falling*	*Fear of falling*[143]	-	-
b260	Proprioceptive function*	Proprioception[143]	1.06+ (1.01-1.13)[23, 283]	-
b280	Sensation of pain	Pain	1.39 (1.19-1.62)[41]	1.60 (1.44-1.78)[41]
b4101	*Heart rhythm*	*Cardiac arrhythmia*[20]	-	-
b420	Blood pressure functions*	*Postural hypotension*[143]	-	-
b4201	*Decreased blood pressure**	*Postural/orthostatic hypotension*[20, 143]	-	-
b4552	Fatiguability	Perceived effort to complete tasks	1.04 (1.01-1.07)[23, 25]	-
b620	Urination functions	Urinary incontinence[20]	1.40 (1.26-1.57)[41]	1.67 (1.45-1.92)[41]
b730	Muscle power functions*	*Motor status related to stroke*[143]	-	-
b7303	Power of muscles in lower half of the body	Knee extensor strengths of the weaker leg	0.19 (0.09-0.41)[23, 282] 0.98+ (0.97-1.00)[23, 283]	-
		Knee extensor strength of the stronger leg	0.99+ (0.98-1.00)[23, 283]	-
		Knee flexor strength of the weaker leg	0.63 (0.42-0.94)[23, 282]	-
		Knee extensor strength	0.97 (0.94-0.99)[23, 284] 0.99+ (0.98-1.00)[23, 283]	-
		Ankle dorsiflexor strength of the weaker leg	0.68 (0.46-0.99)[23, 282]	-
		Ankle dorsiflexor strength of the stronger leg	0.67 (0.45-0.99)[23, 282]	-
b7306	*Power of all muscles of the body*	*Weakness due to inactivity*[20]	-	-
b735	Muscle tone functions*	*Motor status related to stroke*[143]	-	-
b7355	*Tone of muscles of trunk*	*Axial rigidity*[20]	-	-
b740	Muscle endurance functions*	*Motor status related to stroke*[143]	-	-
b755	Involuntary movement reaction functions*	Postural instability/altered balance[20]	3.16 (1.91-5.22)[23, 282] 1.63+ (1.34-1.98)[23, 283] 3.87 (2.39-6.26)[30]	-
		Postural abnormalities[20]	-	-
		Balance and posture stability[143]	-	-
b760	Control of voluntary movement functions*	*Balance and posture stability*[143]	-	-
b765	Involuntary movement functions*			

Category[135-137]	Descriptor	Risk factor	Effect measure (OR, HR, RR with CI 95%) for all fallers	Effect measure (OR, HR, RR with CI 95%) for recurrent fallers
b7650	*Involuntary contractions of muscles*	Dyskinesia[20]	1.31+ (1.05-1.65)[23, 283]	-
b770	Gait pattern functions*	Freezing of gait[20]	5.84 (2.55-13.41)[23, 282] 1.16 (1.08-1.25)[23, 284]	4.64 (2.0-10.7)[19, 23]
			1.05+ (1.03-1.07)[23, 283]	
		Shuffling and small-scaled gait[20]	-	-
		Gait stability[143]	-	-
Body structures (ICF 2017 - English)[131]				
s750	Structure of lower extremity*	*Amputation*[143]	-	-
s770	Additional musculoskeletal structures related to movement*			
Activities and participation (ICF 2017 - English)[131]				
d360	Using communication devices and techniques	Instrumental disability	1.46 (1.20-1.77)[41]	2.04 (1.41-2.95)[41]
d410	Changing basic body position*	*Balance and posture stability*[143]	-	-
d415	Maintaining a body position*			
d420	Transferring oneself*	Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Transfers[20]	-	-
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Level of functional independence[143]	-	-
d450	Walking*	Gait problems	2.06 (1.82-2.33)[41]	2.16 (1.47-3.19)[41]
		Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Impaired mobility[20]	4.36 (2.68-7.10)[30]	-
		Level of functional independence[143]	-	-
d455	Moving around*	Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Impaired mobility[20]	4.36 (2.68-7.10)[30]	-
		Level of functional independence[143]	-	-
d4551	Climbing	Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Impaired mobility[20]	4.36 (2.68-7.10)[30]	-
d460	Moving around in different locations*	Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Impaired mobility[20]	4.36 (2.68-7.10)[30]	-
		Level of functional independence[143]	-	-
d465	Moving around using equipment*	Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
		Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Impaired mobility[20]	4.36 (2.68-7.10)[30]	-

Category[135-137]	Descriptor	Risk factor	Effect measure (OR, HR, RR with CI 95%) for all fallers	Effect measure (OR, HR, RR with CI 95%) for recurrent fallers
		Level of functional independence[143]	-	-
d470-d489	Moving around using transportation	Instrumental disability	1.46 (1.20-1.77)[41]	2.04 (1.41-2.95)[41]
		Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
d510	Washing oneself	Difficulty or assistance with ADL	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
		Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
d520	Caring for body parts	Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
		Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
d530	Toileting*	Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
		Urinary incontinence[143]	-	-
		Physical disability	1.56 (1.22-1.99)[41]	2.42 (1.80–3.26)[41]
d540	Dressing	Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
		Environmental hazards[20]	-	-
d550	Eating	Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
d560	Drinking	Difficulty or assistance with ADLs	-	1.13 (1.04–1.22)[23, 287]
		Disability in self-care	2.30 (1.51-3.49)[30]	-
d570	Looking after one's health	Disability in self-care	2.30 (1.51-3.49)[30]	-
d620	Acquisition of goods and services	Instrumental disability	1.46 (1.20-1.77)[41]	2.04 (1.41-2.95)[41]
d630	Preparing meals		1.46 (1.20-1.77)[41]	2.04 (1.41-2.95)[41]
d640	Doing housework			
d860-d879	Economic life			
d920	Recreation and leisure	Physical activity levels	0.54 (0.36–0.80)[23, 250]	-
Environmental factors (ICF 2017 - English)[131]				
		No. medications (for 1-drug increase)	1.06 (1.04-1.08)[41]	1.06 (1.04-1.08)[41]
		Polypharmacy (use of>3 drugs other than anti-PD)[20]	-	-
e1101	Drugs*	*Medications (antidepressant, diuretics, sedatives, antihypertensive, anti-Parkinsonism, antiepileptics, neuroleptics, polypharmacy (>4 medications))*[143]	-	-
		Use of sedatives[20]	1.38 (1.15-1.66)[41]	1.53 (1.34-1.75)[41]
		Use of sedative and psychotropic medications	3.19 (1.36-7.48)[30]	2.23 (1.18-4.23)[30]
		Use of antihypertensives	1.25 (1.06-1.48)[41]	1.23 (1.05-1.44)[41]

Category[135-137]	**Descriptor**	**Risk factor**	**Effect measure (OR, HR, RR with CI 95%) for all fallers**	**Effect measure (OR, HR, RR with CI 95%) for recurrent fallers**
		Use of antiepileptics	1.88 (1.02-3.49)[41]	2.68 (1.83-3.92)[41]
		Higher total doses of levodopa[20]	-	-
		Use of dopamine agonists, anticholinergic[20]	-	-
e1108	*Products or substances for personal consumption, other specified*	*Daily use of alcohol*[20]	-	-
e120	Products and technology for personal indoor and outdoor mobility and transportation*	*Activities when falling*[143]	-	-
e1201	Assistive products and technology for personal indoor and outdoor mobility and transportation	Walking aid use[20]	2.18 (1.79-2.65)[41] 5.17 (2.27–11.75)[23, 250]	3.09 (2.10-4.53)[41]
e150	*Design, construction and building products and technology of buildings for public use**	*Environmental hazards*[20]	-	-
		Light[143]	-	-
e155	*Design, construction and building products and technology of buildings for private use*	*Environmental hazards*[20]	-	-
e240	*Light**	*Environmental hazards*[20]	-	-
		Light[143]	-	-
e310	Immediate family	Living situation (alone vs. not alone)	1.33 (1.21-1.45)[41]	1.25 (1.10-1.43)[41]
Personal factors (ICF 2017 - English: not classified)[131]				
Pf*		Age (5-year increase)[20, 143]	1.12 (1.07-1.17)[41]	1.12 (1.07-1.18)[41]
Pf		Sex (women vs. men)	1.30 (1.18-1.42)[41]	1.34 (1.12-1.60)[41]
		Female gender[20]	-	-
Pf*		Fall history[20]/previous fall[143]	2.77 (2.37-3.25)[41] 1.67 (1.03-2.72)[30]	3.46 (2.85-4.22)[41] 11.8 (2.9-48.2)[19, 23] 3.02 (1.23–7.44)[23, 287] 4.19 (2.50-7.01)[30]
Pf		Self perceived health status (poor vs. good)	1.50 (1.15-1.96)[41]	1.82 (1.26-2.61)[41]
Pf		Comorbidity (increment of 1 condition)	1.23 (1.16-1.30)[41]	1.48 (1.25-1.74)[41]
Pf		*Disease severity (PD)*[20]	-	-
Pf		*Functional neurosurgery (particularly STN DBS)*[20]	-	-

NOTES: in *italics* fall risk factors not associated with any measure of effect. These measures, when not specified, are expressed in OR (odds ratio fallers vs. non-fallers).

°HR (Hazard ratio fallers vs. non-fallers); [+]RR (relative risk fallers vs. non-fallers); Pf (Personal factors).

*Category present only or also in the 'ICF core set for falls in acute rehabilitation settings[143]'.

8 References

1. Peel NM, Kassulke DJ, McClure RJ. Population based study of hospitalised fall related injuries in older people. Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention 2002;8:280-283.
2. Gillespie LD, Robertson MC, Gillespie WJ, et al. Interventions for preventing falls in older people living in the community. Cochrane Database Syst Rev 2012;9:CD007146.
3. Campbell AJ, Borrie MJ, Spears GF, Jackson SL, Brown JS, Fitzgerald JL. Circumstances and consequences of falls experienced by a community population 70 years and over during a prospective study. Age and ageing 1990;19:136-141.
4. Rubenstein LZ, Josephson KR. The epidemiology of falls and syncope. Clinics in geriatric medicine 2002;18:141-158.
5. Tinetti ME, Kumar C. The patient who falls: "It's always a trade-off". JAMA 2010;303:258-266.
6. Allen NE, Schwarzel AK, Canning CG. Recurrent falls in Parkinson's disease: a systematic review. Parkinson's disease 2013;2013:906274.
7. https://www.epicentro.iss.it/passi-argento/dati/cadute#dati [online].
8. Rubenstein LZ. Falls in older people: epidemiology, risk factors and strategies for prevention. Age and ageing 2006;35 Suppl 2:ii37-ii41.
9. Organization WH. WHO Global Report on Falls Prevention in Older Age. 2007.
10. ISTAT. Report "Gli incidenti domestici"2014.
11. http://dati.istat.it/Index.aspx?DataSetCode=DCIS_CMORTE1_EV [online].
12. Brigoni P, De Masi S, Di Franco M, et al. Prevenzione delle cadute da incidente domestico negli anziani. Programma nazionale per le linee guida (PNLG) 2007;13.
13. Salute Md. Appropriatezza diagnostica e terapeutica nella prevenzione delle fratture da fragilità da osteoporosi. Quaderni del Ministero della Salute 2010.
14. Rossini M, Piscitelli P, Fitto F, et al. [Incidence and socioeconomic burden of hip fractures in Italy]. Reumatismo 2005;57:97-102.
15. Yardley L, Smith H. A prospective study of the relationship between feared consequences of falling and avoidance of activity in community-living older people. Gerontologist 2002;42:17-23.
16. Stolze H, Klebe S, Zechlin C, Baecker C, Friege L, Deuschl G. Falls in frequent neurological diseases--prevalence, risk factors and aetiology. Journal of neurology 2004;251:79-84.
17. Homann B, Plaschg A, Grundner M, et al. The impact of neurological disorders on the risk for falls in the community dwelling elderly: a case-controlled study. BMJ open 2013;3:e003367.
18. de Rijk MC, Launer LJ, Berger K, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 2000;54:S21-23.
19. Camicioli R, Majumdar SR. Relationship between mild cognitive impairment and falls in older people with and without Parkinson's disease: 1-Year Prospective Cohort Study. Gait & posture 2010;32:87-91.
20. Fasano A, Canning CG, Hausdorff JM, Lord S, Rochester L. Falls in Parkinson's disease: A complex and evolving picture. Movement disorders : official journal of the Movement Disorder Society 2017;32:1524-1536.
21. Kerr GK, Worringham CJ, Cole MH, Lacherez PF, Wood JM, Silburn PA. Predictors of future falls in Parkinson disease. Neurology 2010;75:116-124.

22. Voss TS, Elm JJ, Wielinski CL, et al. Fall frequency and risk assessment in early Parkinson's disease. Parkinsonism & related disorders 2012;18:837-841.
23. Canning CG, Paul SS, Nieuwboer A. Prevention of falls in Parkinson's disease: a review of fall risk factors and the role of physical interventions. Neurodegenerative disease management 2014;4:203-221.
24. Parashos SA, Wielinski CL, Giladi N, Gurevich T, National Parkinson Foundation Quality Improvement Initiative I. Falls in Parkinson disease: analysis of a large cross-sectional cohort. Journal of Parkinson's disease 2013;3:515-522.
25. Ashburn A, Stack E, Pickering RM, Ward CD. Predicting fallers in a community-based sample of people with Parkinson's disease. Gerontology 2001;47:277-281.
26. Vossius C, Nilsen OB, Larsen JP. Parkinson's disease and hospital admissions: frequencies, diagnoses and costs. Acta neurologica Scandinavica 2010;121:38-43.
27. Dorsey ER, Constantinescu R, Thompson JP, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. Neurology 2007;68:384-386.
28. Weerdesteyn V, de Niet M, van Duijnhoven HJ, Geurts AC. Falls in individuals with stroke. J Rehabil Res Dev 2008;45:1195-1213.
29. Verheyden GS, Weerdesteyn V, Pickering RM, et al. Interventions for preventing falls in people after stroke. Cochrane Database Syst Rev 2013;5:CD008728.
30. Xu T, Clemson L, O'Loughlin K, Lannin NA, Dean C, Koh G. Risk Factors for Falls in Community Stroke Survivors: A Systematic Review and Meta-Analysis. Arch Phys Med Rehabil 2018;99:563-573 e565.
31. Yoshida S. A global report on falls prevention, epidemiology of falls. . Ageing and Life Course, Family and Community Health, WHO 2007.
32. Lamb SE, Jorstad-Stein EC, Hauer K, Becker C. Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus. J Am Geriatr Soc 2005;53:1618-1622.
33. de Negreiros Cabral K, Perracini MR, Soares AT, et al. Effectiveness of a multifactorial falls prevention program in community-dwelling older people when compared to usual care: study protocol for a randomised controlled trial (Prevquedas Brazil). BMC Geriatr 2013;13:27.
34. Berg WP, Alessio HM, Mills EM, Tong C. Circumstances and consequences of falls in independent community-dwelling older adults. Age and ageing 1997;26:261-268.
35. Tinetti ME, Speechley M, Ginter SF. Risk factors for falls among elderly persons living in the community. The New England journal of medicine 1988;319:1701-1707.
36. Ashburn A, Stack E, Ballinger C, Fazakarley L, Fitton C. The circumstances of falls among people with Parkinson's disease and the use of Falls Diaries to facilitate reporting. Disability and rehabilitation 2008;30:1205-1212.
37. Gazibara T, Pekmezovic T, Tepavcevic DK, et al. Circumstances of falls and fall-related injuries among patients with Parkinson's disease in an outpatient setting. Geriatric nursing 2014;35:364-369.
38. Mak MK, Pang MY. Parkinsonian single fallers versus recurrent fallers: different fall characteristics and clinical features. Journal of neurology 2010;257:1543-1551.
39. Hyndman D, Ashburn A, Stack E. Fall events among people with stroke living in the community: circumstances of falls and characteristics of fallers. Arch Phys Med Rehabil 2002;83:165-170.
40. NHS Centre for Reviews and Dissemination UoY. Effective Helth Care, Bulletin for decision makers on the effectiveness of health service interventions 1996.
41. Deandrea S, Lucenteforte E, Bravi F, Foschi R, La Vecchia C, Negri E. Risk factors for falls in community-dwelling older people: a systematic review and meta-analysis. Epidemiology 2010;21:658-668.

42. Da Silva Gama Z, Gómez-Conesa A. Risk factors for falls in the elderly: systematic review. Rev Saúde Pública 2008;42:946-956.
43. Panel on Prevention of Falls in Older Persons AGS, British Geriatrics S. 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;59:148-157.
44. Guirguis-Blake JM, Michael YL, Perdue LA, Coppola EL, Beil TL. Interventions to Prevent Falls in Older Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA 2018;319:1705-1716.
45. Allen NE, Sherrington C, Paul SS, Canning CG. Balance and falls in Parkinson's disease: a meta-analysis of the effect of exercise and motor training. Movement disorders : official journal of the Movement Disorder Society 2011;26:1605-1615.
46. Tomlinson CL, Patel S, Meek C, et al. Physiotherapy versus placebo or no intervention in Parkinson's disease. Cochrane Database Syst Rev 2013;7:CD002817.
47. Shen X, Wong-Yu IS, Mak MK. Effects of Exercise on Falls, Balance, and Gait Ability in Parkinson's Disease: A Meta-analysis. Neurorehabilitation and neural repair 2016;30:512-527.
48. Hirsch MA, Iyer SS, Sanjak M. Exercise-induced neuroplasticity in human Parkinson's disease: What is the evidence telling us? Parkinsonism & related disorders 2016;22 Suppl 1:S78-81.
49. Canning CG, Sherrington C, Lord SR, et al. Exercise therapy for prevention of falls in people with Parkinson's disease: a protocol for a randomised controlled trial and economic evaluation. BMC Neurol 2009;9:4.
50. Morris ME, Menz HB, McGinley JL, et al. Falls and mobility in Parkinson's disease: protocol for a randomised controlled clinical trial. BMC Neurol 2011;11:93.
51. Morris ME, Martin C, McGinley JL, et al. Protocol for a home-based integrated physical therapy program to reduce falls and improve mobility in people with Parkinson's disease. BMC Neurol 2012;12:54.
52. Denissen S, Staring W, Kunkel D, et al. Interventions for preventing falls in people after stroke. Cochrane Database Syst Rev 2019;10:CD008728.
53. Dean CM, Rissel C, Sharkey M, et al. Exercise intervention to prevent falls and enhance mobility in community dwellers after stroke: a protocol for a randomised controlled trial. BMC Neurol 2009;9:38.
54. https://www.who.int/en/news-room/fact-sheets/detail/falls [online].
55. Holland WW, Stewart S. Screening in Disease Prevention: What Works? Oxford2005.
56. Organization WH. https://www.who.int/cancer/prevention/diagnosis-screening/screening/en/.
57. Illness CoC. Chronic Illness in the USA. Volume I. Prevention of chronic illness. Cambridge, MA.1957.
58. McKeown T. Screening in Medical Care: reviewing the evidence. Oxford1968.
59. Service UKNH. https://www.gov.uk/guidance/nhs-population-screening-explained [online].
60. Salute Md. http://www.salute.gov.it/portale/salute/p1_5.jsp?lingua=italiano&id=25&area=Screening [online].
61. Hulka BS. Cancer screening. Degrees of proof and practical application. Cancer 1988;62:1776-1780.
62. Bretthauer M, Kalager M. Principles, effectiveness and caveats in screening for cancer. The British journal of surgery 2013;100:55-65.
63. https://www.epicentro.iss.it/politiche_sanitarie/40-anni-ssn-screening [online].
64. Wilson JMG, Jungner G. Principles and practice of screening for disease. . Geneva: WHO 1968.

65. Andermann A, Blancquaert I, Beauchamp S, Dery V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. Bulletin of the World Health Organization 2008;86:317-319.
66. https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme [online].
67. Cochrane AL, Holland WW. Validation of screening procedures. British medical bulletin 1971;27:3-8.
68. Sackett DL, Holland WW. Controversy in the detection of disease. Lancet 1975;2:357-359.
69. Andermann A, Blancquaert I, Beauchamp S, Costea I. Guiding policy decisions for genetic screening: developing a systematic and transparent approach. Public health genomics 2011;14:9-16.
70. Commission E. Cancer Screening in the European Union. Report on the implementation of the Council Recommendation on cancer screening2017.
71. Getz L, Sigurdsson JA, Hetlevik I. Is opportunistic disease prevention in the consultation ethically justifiable? Bmj 2003;327:498-500.
72. Sassi F. The outcomes of medical diagnosis on economic perspective. London: London School of Economics and London University., 2000.
73. Sassi F, Le Grand J, Archard L. Equity versus efficiency: a dilemma for the NHS. If the NHS is serious about equity it must offer guidance when principles conflict. Bmj 2001;323:762-763.
74. Mushlin AI, Ruchlin HS, Callahan MA. Costeffectiveness of diagnostic tests. Lancet 2001;358:1353-1355.
75. The trouble with screening. Lancet 2009;373:1223.
76. Skrabanek P. The physician's responsibility to the patient. Lancet 1988;1:1155-1157.
77. Law M. Screening without evidence of efficacy (editorial). Bmj 2004;328:301-302.
78. Kosters JP, Gotzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database Syst Rev 2003:CD003373.
79. Shepperd S, Charnock D, Gann B. Helping patients access high quality health information. Bmj 1999;319:764-766.
80. Stacey D, Legare F, Lewis K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017;4:CD001431.
81. Barratt A, Trevena L, Davey HM, McCaffery K. Use of decision aids to support informed choices about screening. Bmj 2004;329:507-510.
82. Berger ZD, Brito JP, Ospina NS, et al. Patient centred diagnosis: sharing diagnostic decisions with patients in clinical practice. Bmj 2017;359:j4218.
83. Rose G. Strategy of prevention: lessons from cardiovascular disease. Br Med J (Clin Res Ed) 1981;282:1847-1851.
84. Marshall KG. Prevention. How much harm? How much benefit? 4. The ethics of informed consent for preventive screening programs. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 1996;155:377-383.
85. Chamberlain JM. Which prescriptive screening programmes are worth while? Journal of epidemiology and community health 1984;38:270-277.
86. Force UPST. Screening for Prostate Cancer: US Preventive Services Task Force Recommendation Statement. JAMA 2018;319:1901-1913.
87. Shaw C, Abrams K, Marteau TM. Psychological impact of predicting individuals' risks of illness: a systematic review. Social science & medicine 1999;49:1571-1598.
88. Marteau TM, Kidd J, Michie S, Cook R, Johnston M, Shaw RW. Anxiety, knowledge and satisfaction in women receiving false positive results on routine prenatal screening: a

randomized controlled trial. Journal of psychosomatic obstetrics and gynaecology 1993;14:185-196.
89. Austoker J. Gaining informed consent for screening. Is difficult--but many misconceptions need to be undone. Bmj 1999;319:722-723.
90. Hoffmann TC, Del Mar C. Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. JAMA internal medicine 2015;175:274-286.
91. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. Bmj 2001;323:157-162.
92. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. Psychological Science in the Public Interest 2007;8:53-96.
93. Lauby-Secretan B, Scoccianti C, Loomis D, et al. Breast-cancer screening—viewpoint of the IARC Working Group. New England journal of medicine 2015;372:2353-2358.
94. Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. Evaluation of screening programmes for gynaecological cancer. British journal of cancer 1985;52:669-673.
95. Madlensky L, Goel V, Polzer J, Ashbury FD. Assessing the evidence for organised cancer screening programmes. European journal of cancer 2003;39:1648-1653.
96. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. Report on a Workshop of the UICC Project on Evaluation of Screening for Cancer. International journal of cancer 1990;46:761-769.
97. Miller AB. Quality assurance in screening strategies. Virus research 2002;89:295-299.
98. DESA U. United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects 2019: Highlights. 2019.
99. Sazlina SG. Health screening for older people-what are the current recommendations? Malaysian family physician : the official journal of the Academy of Family Physicians of Malaysia 2015;10:2-10.
100. Moyer V, LeFevre M, Siu A. High priority evidence gaps for clinical preventive services. Third annual report to Congress on High-Priority Evidence Gaps for Clinical Preventive Services US Preventive Services Task Force Retrieved from http://www uspreventiveservicestaskforce org/Page/Name/third-annual-report-to-congress-on-high-priority-evidence-gaps-for-clinical-preventive-services; 2013.
101. The Royal Australian College of General Practitioners. Guidelines for preventive activities in general practice. 9th edn, updated. East Melbourne, Vic: RACGP, 2018.
102. Thilagaratnam S, Ding YY, Au Eong KG, et al. Health Promotion Board-Ministry of Health clinical practice guidelines: functional screening for older adults in the community. Singapore medical journal 2010;51:518-521; quiz 522.
103. https://canadiantaskforce.ca/guidelines/published-guidelines/ [online].
104. https://www.nhs.uk/conditions/nhs-screening/ [online].
105. Pepersack T, College of G, the Belgian Society for G, Geriatrics. Minimum geriatric screening tools to detect common geriatric problems. The journal of nutrition, health & aging 2008;12:348-352.
106. Epstein AM, Hall JA, Fretwell M, et al. Consultative geriatric assessment for ambulatory patients. A randomized trial in a health maintenance organization. JAMA 1990;263:538-544.
107. Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. Lancet 1993;342:1032-1036.
108. Spalding MC, Sebesta SC. Geriatric screening and preventive care. American family physician 2008;78:206-215.
109. Bulpitt CJ, Benos AS, Nicholl CG, Fletcher AE. Should medical screening of the elderly population be promoted? Gerontology 1990;36:230-245.

110. https://www.epicentro.iss.it/anziani/StudioArgento [online].
111. Maggi S, Zucchetto M, Grigoletto F, et al. The Italian Longitudinal Study on Aging (ILSA): design and methods. Aging 1994;6:464-473.
112. https://www.epicentro.iss.it/anziani/anziani-bisogni [online].
113. Health NIf, Excellence C. Falls in older people: assessing risk and prevention: NICE, 2013.
114. https://www.safetyandquality.gov.au/sites/default/files/migrated/Guidelines-COMM.pdf.
115. Health NIf, Excellence C. Surveillance of falls in older people: assessing risk and prevention (NICE guideline CG161)2019.
116. Council) NNHaMR. A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines, Australian Government, Canberra. 1999.
117. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. Aging clinical and experimental research 2018;30:1-16.
118. Perell KL, Nelson A, Goldman RL, Luther SL, Prieto-Lewis N, Rubenstein LZ. Fall risk assessment measures: an analytic review. The journals of gerontology Series A, Biological sciences and medical sciences 2001;56:M761-766.
119. Gates S, Smith LA, Fisher JD, Lamb SE. Systematic review of accuracy of screening instruments for predicting fall risk among independently living older adults. J Rehabil Res Dev 2008;45:1105-1116.
120. Palumbo P, Klenk J, Cattelani L, et al. Predictive Performance of a Fall Risk Assessment Tool for Community-Dwelling Older People (FRAT-up) in 4 European Cohorts. Journal of the American Medical Directors Association 2016;17:1106-1113.
121. Cattelani L, Palumbo P, Palmerini L, et al. FRAT-up, a Web-based fall-risk assessment tool for elderly people living in the community(ffrat.farseeingresearch.eu/). Journal of medical Internet research 2015;17:e41.
122. Simundic AM. Measures of Diagnostic Accuracy: Basic Definitions. Ejifcc 2009;19:203-211.
123. Palumbo P, Palmerini L, Bandinelli S, Chiari L. Fall Risk Assessment Tools for Elderly Living in the Community: Can We Do Better? PloS one 2015;10:e0146247.
124. Jakob R, Ustun B, Madden R, Sykes C. The WHO Family of International Classifications. Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz 2007;50:924-931.
125. https://www.who.int/classifications/en/ [online].
126. https://www.who.int/classifications/icd/en/ [online].
127. https://www.reteclassificazioni.it/portal_main.php?portal_view=public_custom_page&id=12 [online].
128. https://www.who.int/classifications/icd/revision/icd11faq/en/ [online].
129. http://www.salute.gov.it/portale/p5_1_1.jsp?lingua=italiano&id=126 [online].
130. https://www.who.int/classifications/icf/en/ [online].
131. Organization WH. International classification of functioning, disability and health. Retrieved from https://apps.who.int/classifications/icfbrowser/. 2017.
132. ICF e-learning tool, ICF Research Branch, available at https://www.icf-elearning.com/wp-content/uploads/articulate_uploads/ICF%20e-Learning%20Tool_2018%20-%20Storyline%20output/story_html5.html [online].
133. Lord SR, C. S, H.B. M. Falls in Older People: Risk Factors and Strategies for Prevention. MA: Cambridge University Press2007.
134. Stucki G, Cieza A, Ewert T, Kostanjsek N, Chatterji S, Ustun TB. Application of the International Classification of Functioning, Disability and Health (ICF) in clinical practice. Disability and rehabilitation 2002;24:281-282.

135. Cieza A, Brockow T, Ewert T, et al. Linking health-status measurements to the international classification of functioning, disability and health. J Rehabil Med 2002;34:205-210.
136. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. J Rehabil Med 2005;37:212-218.
137. Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF Linking Rules to strengthen their potential for establishing comparability of health information. Disability and rehabilitation 2019;41:574-583.
138. https://www.icf-research-branch.org/icf-core-sets-projects2 [online].
139. Selb M, Escorpizo R, Kostanjsek N, Stucki G, Ustun B, Cieza A. A guide on how to develop an International Classification of Functioning, Disability and Health Core Set. Eur J Phys Rehabil Med 2015;51:105-117.
140. Cieza A, Ewert T, Ustun TB, Chatterji S, Kostanjsek N, Stucki G. Development of ICF Core Sets for patients with chronic conditions. J Rehabil Med 2004:9-11.
141. Geyh S, Cieza A, Schouten J, et al. ICF Core Sets for stroke. J Rehabil Med 2004:135-141.
142. https://www.icf-core-sets.org/ [online].
143. Yen TH, Lin LF, Wei TS, Chang KH, Wang YH, Liou TH. Delphi-based assessment of fall-related risk factors in acute rehabilitation settings according to the International Classification of Functioning, Disability and Health. Arch Phys Med Rehabil 2014;95:50-57.
144. Papelard A, Daste C, Alami S, et al. Construction of an ICF core set and ICF-based questionnaire assessing activities and participation in patients with systemic sclerosis. Rheumatology 2019;58:2260-2272.
145. Vitacca M, Giardini A, Corica G, et al. Implementation of a real-world based ICF set for the rehabilitation of respiratory diseases: a pilot study. Minerva medica 2019.
146. La Porta F. Programma di Ricerca Regione-Università: report sui risultati raggiunti alla conclusione del progetto PRUA2-2013-000020562018.
147. Nandy S, Parsons S, Cryer C, et al. Development and preliminary examination of the predictive validity of the Falls Risk Assessment Tool (FRAT) for use in primary care. Journal of public health 2004;26:138-143.
148. Russell MA, Hill KD, Day LM, Blackberry I, Gurrin LC, Dharmage SC. Development of the Falls Risk for Older People in the Community (FROP-Com) screening tool. Age and ageing 2009;38:40-46.
149. Stapleton C, Hough P, Oldmeadow L, Bull K, Hill K, Greenwood K. Four-item fall risk screening tool for subacute and residential aged care: The first step in fall prevention. Australasian journal on ageing 2009;28:139-143.
150. Hodkinson HM. Evaluation of a mental test score for assessment of mental impairment in the elderly. Age and ageing 1972;1:233-238.
151. Collen FM, Wade DT, Bradshaw CM. Mobility after stroke: reliability of measures of impairment and disability. International disability studies 1990;12:6-9.
152. Podsiadlo D, Richardson S. The timed "Up & Go": a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991;39:142-148.
153. Rossiter-Fornoff JE, Wolf SL, Wolfson LI, Buchner DM. A cross-sectional validation study of the FICSIT common data base static balance measures. Frailty and Injuries: Cooperative Studies of Intervention Techniques. The journals of gerontology Series A, Biological sciences and medical sciences 1995;50:M291-297.
154. Rikli RE, Jones CJ. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. Journal of Aging and Physical Activity 1999;7:129-161.

155. Jones CJ, Rikli RE, Beam WC. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. Research quarterly for exercise and sport 1999;70:113-119.
156. 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. Journal of gerontology 1994;49:M85-94.
157. Duncan PW, Weiner DK, Chandler J, Studenski S. Functional reach: a new clinical measure of balance. Journal of gerontology 1990;45:M192-197.
158. Bowden MG, Balasubramanian CK, Behrman AL, Kautz SA. Validation of a speed-based classification system using quantitative measures of walking performance poststroke. Neurorehabilitation and neural repair 2008;22:672-675.
159. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther 2000;80:896-903.
160. STEADI. 30-second chair stand cut-off available at https://www.cdc.gov/steadi/pdf/STEADI-Assessment-30Sec-508.pdf [online].
161. Thomas JI, Lane JV. A pilot study to explore the predictive validity of 4 measures of falls risk in frail elderly patients. Arch Phys Med Rehabil 2005;86:1636-1640.
162. Muthén LK, Muthén BO. MPlus User's Guide 6th edition [online]. Available at: http://www.statmodel.com/download/usersguide/.
163. Long JS. Confirmatory factor analysis: a preface to LISREL. In: Newbury L, New Delhi: Sage Pubblications, ed. Lewis-Beck MS, editor Quantitative applications in the social sciences1983.
164. Hoyle RH. Confirmatory factor analysis. . In: Press seSDA, ed. Tinsley H, Brown S, editors Handbook of applied multivariate statistics and mathematical modeling2000.
165. La Porta F, Franceschini M, Caselli S, Cavallini P, Susassi S, Tennant A. Unified Balance Scale: an activity-based, bed to community, and aetiology-independent measure of balance calibrated with Rasch analysis. J Rehabil Med 2011;43:435-444.
166. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6:1-55.
167. Byrne BM. Structural equation modeling with AMOS. Basic Concepts, Applications, and Programming - 2nd edition. New York: Taylor & Francis, 2010.
168. Revicki DA, Chen W, Tucker CA. Developing item banks for patient-reported health outcomes. In: Reise PS, Revicki DA, eds. Handbook of Item Response Theory Modeling: applications to typical performance assessments. New York2014.
169. Lundgren-Nilsson A, Jonsdottir IH, Ahlborg G, Jr., Tennant A. Construct validity of the Psychological General Well Being Index (PGWBI) in a sample of patients undergoing treatment for stress-related exhaustion: a Rasch analysis. Health and quality of life outcomes 2013;11:2.
170. Watt T, Groenvold M, Deng N, et al. Confirmatory factor analysis of the thyroid-related quality of life questionnaire ThyPRO. Health and quality of life outcomes 2014;12:126.
171. Pellicciari L, Piscitelli D, Caselli S, La Porta F. A Rasch analysis of the Conley Scale in patients admitted to a general hospital. Disability and rehabilitation 2018:1-10.
172. Basagni B, Piscitelli D, De Tanti A, et al. The unidimensionality of the five Brain Injury Rehabilitation Trust Personality Questionnaires (BIRT-PQs) may be improved: preliminary evidence from classical psychometrics. Brain injury 2020;34:673-684.
173. Mokken RJ. A Theory and Procedure of Scale Analysis. 1971.
174. van Abswoude AAH, van der Ark LA, Sijtsma K. A Comparative Study of Test Data Ddimensionality Assessment Procedures Under Nonparametric IRT Models. Applied Psychological Measurement 2004;28:3-24.

175. Hardouin J-B. MSP: Stata module to perform the Mokken Scale Procedure. STATA journal 2004:1-21.
176. Molenaar IW, Sijtsma K. User's Manual MSP5 for Windows. The Netherlands2000.
177. Sijtsma K, Van der Ark LA. A tutorial on how to do a Mokken scale analysis on your test and questionnaire data. British Journal of Mathematical & Statistical Psychology 2017;70:137-158.
178. Andries van der Ark L. Mokken Scale Analysis in R. Journal of Statistical Software 2007;20.
179. Andrich D. Rasch models for measurement. London: Sage Publications., 1988.
180. La Porta F, Caselli F, Ianes A, et al. Can we scientifically and reliably measure the level of consciousness in Vegetative and Minimally Conscious States? Rasch analysis of the Coma Recovery Scale Revised. Arch Phys Med Rehabil 2013;94 (3):527-535.
181. Tesio L. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J Rehabil Med 2003;35:105-115.
182. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis and rheumatism 2007;57:1358-1362.
183. Kucukdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in Physical and Rehabilitation Medicine: An educational review. J Rehabil Med 2011;43:661-672.
184. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health technology assessment 2009;13:iii, ix-x, 1-177.
185. Tennant A, Penta M, Tesio L, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. Med Care 2004;42:I37-48.
186. Pallant JF, Tennant A. An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). The British journal of clinical psychology 2007;46:1-18.
187. Ottonello M, Pellicciari L, Giordano A, Foti C. Rasch analysis of the Fatigue Severity Scale in Italian subjects with multiple sclerosis. J Rehabil Med 2016;48:597-603.
188. Wright BD. Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? Rasch Measure Trans 1998;12:641-642.
189. La Porta F, Giordano A, Caselli S, Foti C, Franchignoni F. Is the Berg Balance Scale an effective tool for the measurement of early postural control impairments in patients with Parkinson's disease? Evidence from Rasch analysis. Eur J Phys Rehabil Med 2015;December 51:705-716.
190. Kreiner S, Christensen KB. Person parameter estimation and measurement in Rasch Models. Christensen KB, Kreiner S, Mesbah M, editors Rasch models in health Applied mathematics series London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc., 2013.
191. Mesbah M, Kreiner S. Rasch models for ordered polytomous categories. Christensen KB, Kreiner S, Mesbah M, editors Rasch models in health Applied mathematics series London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc.2013.
192. Marais I. Local dependence. Christensen KB, Kreiner S, Mesbah M, editors Rasch models in health Applied mathematics series: London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc., 2013.
193. Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of applied measurement 2002;3:205-231.
194. Wright B, Masters GN. Rating scale analysis. Chicago: MESA Press1982.

195. Kreiner S. The Rasch model for dichotomous items. . Christensen, KB, Kreiner, S, Mesbah, M: Rasch models in health Applied mathematics series: London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc, 2013.
196. Wright BD. Separation, Reliability and Skewed Distributions: Statistically Different Levels of Performance. Rasch Measure Trans 2001;14(4).
197. Fisher WPj. Rating Scale Instrument Quality Criteria. Rasch Measurement Transactions 2007;21:1:1095.
198. Linacre JM. Optimizing rating scale category effectiveness. Journal of applied measurement 2002;3:85-106.
199. Wainer H, Kiely G. Item clusters and computer adaptive testing: a case for testlets. J Educ Meas 1987;24:185-202.
200. Lundgren Nilsson A, Tennant A. Past and present issues in Rasch analysis: the functional independence measure (FIM) revisited. J Rehabil Med 2011;43:884-891.
201. Geri T, Piscitelli D, Meroni R, et al. Rasch analysis of the Neck Bournemouth Questionnaire to measure disability related to chronic neck pain. J Rehabil Med 2015;47:836-843.
202. Giacino JT, Kalmar K. Diagnostic and prognostic guidelines for the vegetative and minimally conscious states. Neuropsychol Rehabil 2005;15:166-174.
203. La Porta F, Caselli S, Susassi S, Cavallini P, Tennant A, Franceschini M. Is the Berg Balance Scale an Internally Valid and Reliable Measure of Balance Across Different Etiologies in Neurorehabilitation? A Revisited Rasch Analysis Study. Arch Phys Med Rehabil 2012;93 (7):1209-1216.
204. Panella L, La Porta F, Caselli S, Marchisio S, Tennant A. Predicting the need for institutional care shortly after admission to rehabilitation: Rasch analysis and predictive validity of the BRASS Index. Eur J Phys Rehabil Med 2012;48:443-454.
205. Andrich D, Sheridan BS, Luo G. RUMM2030: Rasch Unidimensional Models for Measurement. RUMM Laboratory: Perth, Western Australia. www.rummlab.com.2015.
206. Andrich D. Components of Variance of Scales With a Bifactor Subscale Structure From Two Calculations of α. Educational Measurement: Issues and Practice 2016;35:25-30.
207. Maritz R, Tennant A, Fellinghauer C, Stucki G, Prodinger B. The Functional Independence Measure 18-item version can be reported as a unidimensional interval-scaled metric: Internal construct validity revisited. J Rehabil Med 2019;51:193-200.
208. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological methods 2016;21:137-150.
209. Mueller MK, S. Item Fit Statistics in Common Software for Rasch Analysis. Research Report 15-06, Department of Biostatistics, University of Copenhagen.2015.
210. Tennant A, Pallant JF. DIF matters: A practical approach to test if Differential Item Functioning makes a difference. Rasch Measure Trans 2007;20:1082-1084.
211. Chiari P, Mosci D, Naldi E. Evidence-Based Clinical Practice. La pratica clinico-assistenziale basata su prove di efficacia. Milano: McGraw-Hill, 2011.
212. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochemia medica 2016;26:297-307.
213. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.
214. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839-843.
215. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982;143:29-36.

216. Rota M, Antolini L. Finding the optimal cut-point for Gaussian and Gamma distributed biomarkers. Computational Statistics & Data Analysis 2014;69:1-14.
217. López Ratón M, Rodríguez Álvarez MX, Cadarso Suárez CM, Gude Sampedro F. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. 2014: American Statistical Association.
218. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-35.
219. Perkins NJ, Schisterman EF. The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. American journal of epidemiology 2006;163:670-675.
220. Liu X. Classification accuracy and cut point selection. Statistics in medicine 2012;31:2676-2686.
221. Miller R, Siegmund D. Maximally selected chi square statistics. Biometrics 1982:1011-1016.
222. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 4 edition ed: Lippincott Williams & Wilkins, 2019.
223. van Belle G, Fisher LD, P.J. H, Lumley T. Biostatistics: A Methodology For the Health Sciences, 2nd edition ed: Wiley-Interscience, 2004.
224. Sedighi I. Interpretation of Diagnostic Tests: Likelihood Ratio vs. Predictive Value. Iranian journal of pediatrics 2013;23:717.
225. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. Bmj 2004;329:168-169.
226. Safari S, Baratloo A, Elfil M, Negida A. Evidence Based Emergency Medicine; Part 4: Pre-test and Post-test Probabilities and Fagan's nomogram. Emergency 2016;4:48-51.
227. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. AJR American journal of roentgenology 2005;184:14-19.
228. Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: Wiley & Sons, 2002.
229. Hershey JC, Cebul RD, Williams SV. Clinical guidelines for using two dichotomous tests. Medical decision making : an international journal of the Society for Medical Decision Making 1986;6:68-78.
230. Cebul RD, Hershey JC, Williams SV. Using multiple tests: series and parallel approaches. Clinics in laboratory medicine 1982;2:871-890.
231. Lusardi MM, Fritz S, Middleton A, et al. Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability. Journal of geriatric physical therapy 2017;40:1-36.
232. Mayer D. Bayes' theorem, predictive values, post-test probabilities, and interval likelihood ratios. Essential Evidence-Based Medicine. New York, NY: Cambridge University Press, 2009.
233. Aldrich JH, Nelson FD. Linear Probability, Logit, and Probit Models. Newbury Park: CA: Sage, 1984.
234. Long JS, Freese J. Regression Models for Categorical Dependent Variables Using Stata: CollegeStation, TX: Stata Press., 2014.
235. Hosmer DW, Jr., , Lemeshow SA, Sturdivant RX. Applied Logistic Regression: Hoboken, NJ: Wiley, 2013.
236. Hosmer DW, Jr.,, Lemeshow S. Goodness-of-fit tests for the multiple logistic regression model. Communications in Statistics: Theory and Methods, Part A 1980;9:1043–1069.
237. Bossuyt PM, Cohen JF, Gatsonis CA, Korevaar DA, group S. STARD 2015: updated reporting guidelines for all diagnostic accuracy studies. Annals of translational medicine 2016;4:85.
238. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ open 2016;6:e012799.

239. Tabachnick BG, Fidell LS. Using multivariate statistics 5th edition. Boston: Pearson, 2013.
240. Watson R, Egberink IJL, Kirke L, Tendeiro JN, Doyle F. What are the minimal sample size requirements for Mokken scaling? An empirical example with the Warwick- Edinburgh Mental Well-Being Scale. Health Psychology and Behavioral Medicine 2018;6:203-213.
241. Linacre JM. Sample size and item calibration [or Person Measure] stability. Rasch Measurement Trasactions 1994;7:328 - accessed on 325/302/2013 from http://www.rasch.org/rmt/rmt2074m.htm.
242. Bland J, Altman D. Multiple significance tests: the Bonferroni method. British Medical Journal 1995;310:170.
243. Brodersen J, Doward LC, Thorsen H, Mckenna SP. Writing health-related items for Rasch models - Patient-Reported Outcome Scales for Health Sciences: from medical paternalism to patient autonomy. In: Christensen KB, Kreiner S, Mesbah M, eds. in Rasch Models in Health. London UK, Hoboken NJ: ISTE Ltd and John Wiley & Sons, Inc, 2013.
244. Goksuluk D, Korkmaz S, Zararsiz G, AE K. easyROC: An Interactive Web-tool for ROC Curve Analysis Using R Language Environment. The R Journal 2016;8:213-230.
245. La Porta F, Franceschini M, Caselli S, Susassi S, Cavallini P, Tennant A. Unified Balance Scale: classic psychometric and clinical properties. J Rehabil Med 2011;43:445-453.
246. Pallant J. SPSS survival manual: McGraw-Hill Education (UK), 2013.
247. Cohen JW. Statistical power analysis for the behavioral sciences (2nd edn). NJ: Lawrence Erlbaum Associates1988.
248. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care 2007;45:S22-31.
249. Organization WH. International statistical classification of diseases and related health problems (11th Revision). Retrieved from https://icd.who.int/browse11/l-m/en. 2018.
250. Matinolli M, Korpelainen JT, Korpelainen R, Sotaniemi KA, Virranniemi M, Myllyla VV. Postural sway and falls in Parkinson's disease: a regression approach. Movement disorders : official journal of the Movement Disorder Society 2007;22:1927-1935.
251. Emerson W. On quantity calculus and units of measurement. Metrologia 2008;45:134.
252. Forrest M, Andersen BR. Ordinal scale and statistics in medical research. Br Med J (Clin Res Ed) 1986;292:537-538.
253. Piscitelli D, Pellicciari L. Responsiveness: is it time to move beyond ordinal scores and approach interval measurements? Clinical rehabilitation 2018;32:1426-1427.
254. Tavakol M, Dennick R. Making sense of Cronbach's alpha. International journal of medical education 2011;2:53-55.
255. Cortina JM. What is coefficient alpha? An examination of theory and applications. Journal of applied psychology 1993;78:98.
256. Clark LA, Watson D. Constructing validity: New developments in creating objective measuring instruments. Psychol Assess 2019;31:1412-1427.
257. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Research in Science Education 2018;48:1273-1296.
258. Linacre JM. Disordered Rasch-Andrich thresholds [online]. Available at: https://www.winsteps.com/facetman/disordered.htm.
259. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. Int J Nurs Stud 2009;46:380–393.
260. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end malpractice? J Rehabil Med 2012;44:97-98.

261. Franklin M, Hunter RM. A modelling-based economic evaluation of primary-care-based fall-risk screening followed by fall-prevention intervention: a cohort-based Markov model stratified by older age groups. Age and ageing 2019;49:57-66.
262. Subramanian S, Klosterman M, Amonkar MM, Hunt TL. Adherence with colorectal cancer screening guidelines: a review. Prev Med 2004;38:536-550.
263. Wagner EH. Chronic disease management: what will it take to improve care for chronic illness? Effective clinical practice : ECP 1998;1:2-4.
264. Barr VJ, Robinson S, Marin-Link B, et al. The expanded Chronic Care Model: an integration of concepts and strategies from population health promotion and the Chronic Care Model. Hospital quarterly 2003;7:73-82.
265. Ministero della Salute I. Piano nazionale della cronicità2016.
266. https://assr.regione.emilia-romagna.it/pubblicazioni/rapporti-documenti/prevenzione-cadute-anziani-2016 [online].
267. Cameron ID, Murray GR, Gillespie LD, et al. Interventions for preventing falls in older people in nursing care facilities and hospitals. Cochrane Database Syst Rev 2010:CD005465.
268. Stark SL, Silianoff TJ, Kim HL, Conte JW, Morris JC. Tailored Calendar Journals to Ascertain Falls Among Older Adults. OTJR : occupation, participation and health 2015;35:53-59.
269. Ahn S, Kim J, Koo B, Kim Y. Evaluation of inertial sensor-based pre-impact fall detection algorithms using public dataset. Sensors 2019;19:774.
270. Gassmann KG, Rupprecht R, Freiberger E, Group IZGS. Predictors for occasional and recurrent falls in community-dwelling older people. Zeitschrift fur Gerontologie und Geriatrie 2009;42:3-10.
271. Tromp AM, Pluijm SM, Smit JH, Deeg DJ, Bouter LM, Lips P. Fall-risk screening test: a prospective study on predictors for falls in community-dwelling elderly. Journal of clinical epidemiology 2001;54:837-844.
272. Pluijm SM, Smit JH, Tromp EA, 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 : a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA 2006;17:417-425.
273. Stone KL, Ancoli-Israel S, Blackwell T, et al. Actigraphy-measured sleep characteristics and risk of falls in older women. Archives of internal medicine 2008;168:1768-1775.
274. Clough-Gorr KM, Erpen T, Gillmann G, et al. Multidimensional geriatric assessment: back to the future preclinical disability as a risk factor for falls in community-dwelling older adults. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences 2008;63:314-320.
275. Chu LW, Chi I, Chiu AY. Incidence and predictors of falls in the chinese elderly. Annals of the Academy of Medicine, Singapore 2005;34:60-72.
276. Deshpande N, Metter EJ, Lauretani F, Bandinelli S, Ferrucci L. Interpreting fear of falling in the elderly: what do we need to consider? Journal of geriatric physical therapy 2009;32:91-96.
277. Capon A, Di Lallo D, Mastromattei A, Pavoni N, Simeoni S. [Incidence and risk factors for accidental falls among general practice elderly patients in Latina, Central Italy]. Epidemiologia e prevenzione 2007;31:204-211.
278. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. Journal of psychiatric research 1975;12:189-198.
279. Covinsky KE, Kahana E, Kahana B, Kercher K, Schumacher JG, Justice AC. History and mobility exam index to identify community-dwelling elderly persons at risk of falling. The journals of gerontology Series A, Biological sciences and medical sciences 2001;56:M253-259.

280. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. Applied psychological measurement 1977;1:385-401.
281. Townsend N, Wickramasinghe K, Williams J, Bhatnagar P, Rayner M. Physical activity statistics 20152015.
282. Latt MD, Lord SR, Morris JG, Fung VS. Clinical and physiological assessments for elucidating falls risk in Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society 2009;24:1280-1289.
283. Paul SS, Sherrington C, Canning CG, Fung VS, Close JC, Lord SR. The relative contribution of physical and cognitive fall risk factors in people with Parkinson's disease: a large prospective cohort study. Neurorehabilitation and neural repair 2014;28:282-290.
284. Paul SS, Canning CG, Sherrington C, Lord SR, Close JC, Fung VS. Three simple clinical tests to accurately predict falls in people with Parkinson's disease. Movement disorders : official journal of the Movement Disorder Society 2013;28:655-662.
285. Allcock LM, Rowan EN, Steen IN, Wesnes K, Kenny RA, Burn DJ. Impaired attention predicts falling in Parkinson's disease. Parkinsonism & related disorders 2009;15:110-115.
286. Balash Y, Peretz C, Leibovich G, Herman T, Hausdorff JM, Giladi N. Falls in outpatients with Parkinson's disease: frequency, impact and identifying factors. Journal of neurology 2005;252:1310-1315.
287. Matinolli M, Korpelainen JT, Sotaniemi KA, Myllyla VV, Korpelainen R. Recurrent falls and mortality in Parkinson's disease: a prospective two-year follow-up study. Acta neurologica Scandinavica 2011;123:193-200.