Manual Therapy, Posturology & Rehabilitation Journal
formerly REVISTA TERAPIA MANUAL

REVIEW

http://dx.doi.org/10.17784/mtprehabjournal.2016.14.432

# Do functional mobility tests predict the risk of falls in community-dwelling elderly?

Ana Carolina Silva de Souza Moreira[1], Giovana Zarpellon Mazo[2], Eduardo Hauser[3], Paula Bertolini de Paiva[3], Enaiane Menezes[1], Fernando Luiz Cardoso[4]

**ABSTRACT**

**Introduction:** The falls are associated with morbidity and mortality in the elderly. Numerous of functional mobility clinical tests have been created to identify older adults with potential for risk of falls. **Objective:** The purpose of this systematic review was to determine the predictive validity of functional mobility tests to predict the risk of falls in community-dwelling elderly. **Method:** Articles in English were searching in MEDLINE, SCOPUS and CINAHL. We found 18,520 documents and, after applying the inclusion and exclusion criteria, 11 articles were part of the final analysis. All articles analyzed included subjects over 60 years old. **Results:** The results showed that the *TUG Test* has good discriminative validity for elderly non-institutionalized, but it does not provide an adequate predictive validity. The *TUG Test* may not be enough as a unique basic screening tool to detect the risk of elderlies´ falling. **Conclusion:** It is suggested that the *TUG Test* should be used in combination with other predictors of falling risk tools or should it be reconfigured for the different levels of elderly physical active functionality.

**Key Words:** fall risk, older adults, validity tests, review, meta-analysis.

## INTRODUCTION

One of the main causes of the loss of autonomy and independence of the elderly community residents are the consequences of physical falls [1, 2], according to the database of the Unified Health System / Brazilian Ministry of Health[3], in Brazil, between 1996 and 2005, about 24,645 elderly people died due to falls, occupying the third place of mortality and the first place among hospitalizations. In 2005, the prevalence of falls in the elderly was 34.8% in seven Brazilian states, and among those who fell, 55% reported a single fall in the previous year. These data are in agreement with the data found in the United States, where about 30% to 40% of the elderly livings in a community suffer at least one fall in their life, and this percentage increases to 60% when the elderly have experienced a fall in previous year[4]. Therefore, falls in the elderly become a public health problem, considering the social and economic burden they generate[5].

Given these considerations, there is a constant concern of health professionals to prevent and mitigate the consequences of falls in the elderly´s health. In both clinical and community health, we must conform to available diagnostic and / or predictive methods acceptable to the individual. The degree of importance of the instrument used for decision-making in relation to therapeutic and / or diagnostic / predictive conduct, including the cost ratio of the instrument´s application, the risks to which the evaluator and the patient will be submitted and patient´s acceptability to such an instruments should also be considered when the indication and use of these instruments[6, 7].

Numerous instrument model methods (scales, tests, questionnaires, among others) have been created among the efforts to minimize and detect the risk of falls in the elderly population so that health professionals can identify individuals with the potential for falls. The most commonly used tests are the *Balance Berg Scale* (BBS), the *Clinical Test of Sensory Organization and Balance Test* (CTSIB), the *Functional Reach Test* (FRT), the *Tinetti Balance Scale* (TBS), *One Leg Stand* (OLS), *Tandem Stand* Test (TS), *Chair Stand Five times Test* (CS-5), and *The Timed Up and Go Test* (TUG). To predict the risk of falls in the elderly, there is still no reference test considered "gold", since the fall phenomenon is multifactorial, one of the factors that explain the postural balance is the functional mobility of the individual.

**Corresponding author:** Ms. Ana Carolina Silva de Souza – Santa Catarina State University /UDESC. Centro da Ciências da Saúde e do Esporte/CEFID; Rua Pascoal Simone, 358; Coqueiros; Florianópolis/SC - Brazil; CEP: 88080-350. e-mail: anakarolfisio@gmail.com

[1] Ph.D. Candidate in Graduate Program in Human Motion Sciences, Santa Catarina State University (UDESC), Santa Catarina (SC), Brazil.

Full list of author information is available at the end of the article.

**Financial support:** None.

**Submission date 21 September 2016; Acceptance date 14 December 2016; Publication online date 27 December 2016**

In the scientific environment, for clinical and epidemiological research, content, criterion and construct validity are essential requirements to obtain a good psychometric test, according to the sample profile. Since clinical trials are considered essential tools to diagnose and / or predict the possibility of risk of some phenomenon. This systematic review aims to analyze the tests in their validation criterion, essentially predictive one. In the construct concept the test must measure exactly what it intends, but also carry out this measurement with accuracy, without errors (validity of criterion). Predictive validity is part of the criterion validity, which should be well understood for the scientific community to which the sample profile of the instrument is appropriate to diagnose and / or predict the phenomenon. Thus, one can distinguish two types of validity criterion: predictive and concurrent[8, 9].

In this sense, there is a literature in the literature about the most appropriate clinical tests to evaluate the risk of falls in the elderly. Since current studies reporting that these tests are having a "roof" or "floor" effect in detecting the risk of falls in the elderly, such as: BBS [10, 11, 12, 13], TUG[14, 15, 16] and other tests[17, 18]. This entails a lack of evidence on which tests would be best suited for use by health care professionals, since they are poorly studied and analyzed in terms of their psychometric qualities.

Insufficient clarity of instrument validity may mislead researchers to make decisions in the clinical setting and in the development of scientific research, and may underestimate or overestimate the results, resulting in potentially erroneous actions in clinical decision-making or in data conclusions from their empirical research. In addition, it is expected that in the literature would be clear about the indications of each test in relation to the different population profile, providing its construct, the dimension corresponding to the construct, test outcomes, collection time, test familiarization, training and manual detailed instruction to the researchers. Thus, when the aforementioned information is easily available to the researchers, it is believed that it would avoid the traditional biases of publications generated by the lack of information about the suitability and reliability of the instrument used in different elderly populations.

To obtain a more accurate indication of the discriminative ability and diagnostic accuracy of the functional balance tests as a clinical screening instrument for the identification of elderly at risk of falling, a comprehensive systematic review and a meta-analysis were performed. In addition, an accurate analysis of the quality of the studies was also performed. Considering the lack of studies that demonstrate the predictive capacity of functional mobility tests in the elderly, this study aimed to analyze theoretically the predictive validity of these available tests, as well as, to their methodological quality in terms of design and procedures.

## METHOD

### Inclusion Criterion

The following inclusion criteria were used: studies which enrolled elderly samples living in a community aged over 60 years (1); Elderly sample without physical limitations (2); Evaluation of physical balance by at least one clinical test (field work) (3); Presentation of at least one test to compare the results (4); Presentation of sensitivity and specificity values of the tests used (5); Publication of results in English, Spanish and or Portuguese (6).

Experimental or observational studies, studies with samples exclusively of elderly people considered dependent, fragile or with pre-established pathologies (neurological, orthopedic, cardiovascular, vestibular system) were excluded in this study.

### Search Strategies

The research was conducted according to the *Preferred Reporting Items for Systematics Reviews and Meta-Analysis* (PRISMA recommendations) guidelines. The MEDLINE (1966-14/ 10/2014), SCOPUS (1960-16/ 10/2104) and CINAHL (1982-10/ 24/2014) databases were selected. The filters used as data search strategy were divided into three blocks: Elderly: Postural Balance and Study Type (APPENDIX I).

### Article Selection

Two independent evaluators selected the studies after reading the titles, excluding those that were not related to the purpose of the review. After the selection of the titles, the evaluators analyzed the abstracts of the articles to identify those that met the inclusion criteria and, subsequently, the eligible studies were analyzed entirety for later inclusion in the systematic review.

Evaluation of the methodological quality of the selected studies

For the methodological evaluation, the authors of the present study developed a scale composed of 10 questions. This scale was developed from existing scales in the literature, such as QUADAS and STARD[19] that did not meet the needs of the study objective. The questions elaborated refer to the presence of methodological aspects of the study evaluated, containing the following questions:

1. Does the article present a sample profile? (E.g., age, sample size, faller and non-fallers, secondary diseases)
2. Does the article present inclusion and exclusion criteria for the study?
3. Does the article provide a reference test?
4. Does the article present the value of the reference test (sensitivity and specificity, cutoff point, etc)?

5. Does the article present cut-off points for analyzed tests?
6. Were the test results interpreted without influence among them?
7. Has the application methodology of the tests been adequately described that could allow their later replication?
8. Were the evaluators previously trained to apply the tests?
9. Were the tests applied on all patients? (If not, on very large sample, at random selection?)
10. Have the tests been applied with rest interval so that there wouldn´t be no influence on each other?

All these questions was answered with the answers "yes", "partially" and "no". With each result yes, it was assigned the value of 1 point whereas for the "partially" 0.5 and results "no" the zero value.

The evaluation process of the selected articles was carried out by four independent reviewers, whose analysis of each question of the methodological evaluation was performed in pairs at each step of the process. To resolve disagreements among reviewers, a third reviewer evaluated all items involved. The systematic review was registered in PROSPERO under registration number CRD 42015026961.

The values of true positive (TP), false positive (FP), true false (TN), false negative (FN), sensitivity (SEN), specificity (SPE), accuracy (ACC), positive predictive value (PPV), negative predictive value (VPN), positive likelihood ratio (PLR) and negative likelihood ratio (NLR), in order to check the fallers. SEN, SPE, ACC, VPP, VPN, PLR e NLR were calculated using the following formulas: 1 to 7 and Table 1.

$$SEN=TP/(TP+FN) \quad (1)$$

$$SPE=TN/(FP+TN) \quad (2)$$

$$ACC=(TP+TN)/(TP+FP+TN+FN) \quad (3)$$

$$PPV=TP/(TP+FP) \quad (4)$$

$$NPV=TN/(FN+TN) \quad (5)$$

$$PLR=SEN/(1-SPE) \quad (6)$$

$$NLR=(1-SEN)/SPE \quad (7)$$

**Table 1.** Shows 2 × 2 (two-by-two) table

| | | | |
|---|---|---|---|
| **Test positive** | True positives(TP) a | False positives (FP) b | Total test positives: a+b |
| **Test negative** | False negative (FN) c | True negatives (TN) d | Total test negatives: c+d |
| | Total diseased: a+c | Total normal: b+d | Total population: a+b+c+d |

## Meta-Analysis

We used the Q Chi-square statistic value to estimate the heterogeneity of the individual studies that contribute to the estimate pooled. The homogeneity was obtained by evaluating the differences between the studies, that is, if they were larger than expected just by chance. The $\rho < 0.05$ in this analysis indicates the presence of greater heterogeneity than would be expected by chance alone.

Due to lack of standardization, different thresholds can be used in included studies to define a positive test result. Differential threshold effect may be the reason for the detectable difference in sensitivities and specificities of the precision test studies. The estimated accuracy of each study in the curve (ROC) space (receiver operating characteristic) and the Spearman Correlation Coefficient between log (SEN) and log (1-SPE) were evaluated for threshold effect. A typical "shoulder arm" plot in the ROC curve space and a strong positive correlation suggested a threshold effect.

Statistical analysis is not always necessary in all systematic reviews to check the test precision studies. The necessary condition for gathering the estimates is that the studies and results should be reasonably homogeneous. Estimates can be grouped by the fixed-effect model (FEM) or the random effects model (REM) to incorporate the variation between the studies, and the output can be represented graphically as funnel plots. If heterogeneity due to the threshold effect is present, accuracy data can be gathered by fitting a summary ROC curve (SROC) and by area under the curve (AUC). The SROC curve summarizes and gathers the true and false positive rates of different diagnostic studies. The overall performance of diagnostic studies can be visualized and reflected by an SROC curve without being affected by a change in limit values. The best diagnostic modality would produce a point in the upper left corner or coordinates (0.1) of the SROC space, representing 100% sensitivity (no false negatives) and 100% specificity (false positives) at the individual subject level. Likewise, AUC ranges from 1 to a perfect test that always diagnoses correctly, to 0 for a test that never does this from individual studies or meta-analyzes. The homogeneity test, threshold effect analysis, pooled and specific weighted sensitivity, SROC curve and sensitivity analysis were performed using Meta-Disc version 1.4.

## RESULTS

In the electronic databases used to search for articles, 2.607 were found in MEDLINE, 5.276 in SCOPUS and 11.464 in CINAHL, totaling 18.520 documents. Of these, 1.890 articles

were duplicated for a total of 16.630 articles. In the analysis of the persistent titles the research was selected 311 articles. After analysis of the abstracts, 97 articles were selected for the entire reading. After analyzing the articles in full, 6 of them we could not get fully access, 80 were excluded because they did not meet the inclusion criteria. Thus, 11 were eligible for the present study (Figure 1).

The methodological quality assessment (Table 2), from the five studies (45.45%) were prospective and six (54.54%) retrospective studies referring to the history of falling of the elderly. It was verified that the articles that scored the most were those performed by Dite et al.[20] (9.5 points), Rose et al.[21] (9 points) and Wrisley and Kumar[22] (9 points). The lowest scores were Trueblood et al.[23] (5 points); Greene et al.[24] (5.5 points) and Tsutsuminoto et al.[25] (6 points). Among the questions that were less pointed out and decreased in the methodological quality of the articles scores are the presentation of simple aspects such as "time to rest" (Question 10) and "whether the evaluators were previously trained" (Question 8). However, two questions were addressed by all articles, whether "the tests were applied to all subjects (or at random)" (question 9) and whether the tests were analyzed without any influence on each other (question 6).

The sample profile and characteristics of the studies are summarized in Table 3. It is noted that 81% of the studies excluded elderly with cognitive deficit using the mini-mental examination, 72.72% reported prevalent diseases, 45.45% the type of medication administered, 63.63% reported the number of elderly people who use gaiters and limitations of their daily activities; 36.36% excluded elderly individuals who presented balance deficiency; and 27.27% evaluated the state of the visual, auditory and / or vestibular system. Complementary evaluation of methodological quality using relevance criteria to investigate the risk of falling (phenomenon) indicated that only two studies had representative sample size. Overall, four studies were classified as high quality studies, of which only one included representative sample.

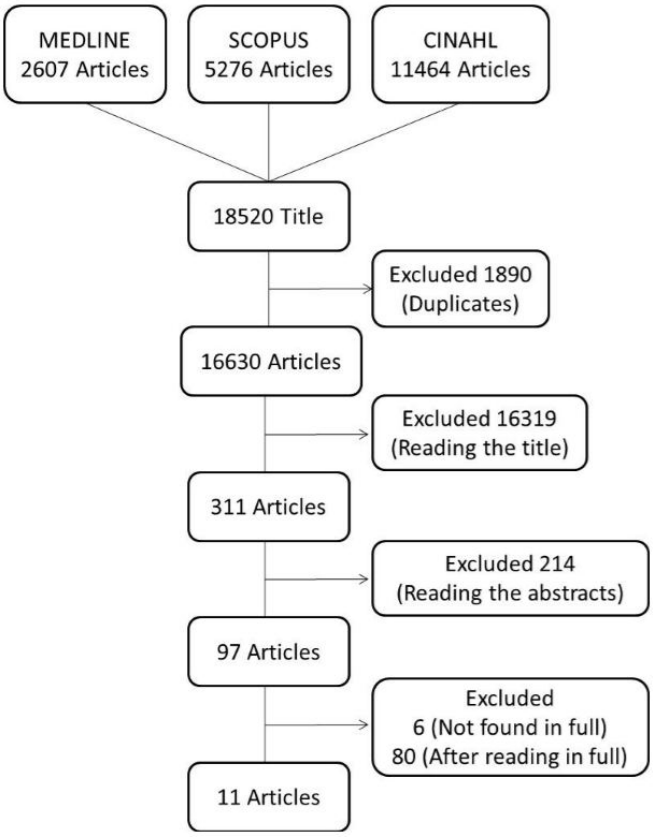


**Figure 1.** Flowchart of the steps of the systematic review studies.

**Table 2.** Methodological quality of systematic review studies.

| Studies | Q1 | Q2 | Q3 | Q4 | Q5 | Q6 | Q7 | Q8 | Q9 | Q10 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Trueblood et al. (2001) [23] | ± | ✓ | ✗ | ✗ | ✓ | ✓ | ± | ✗ | ✓ | ✗ | 5 |
| Shumway-cook et al. (2000) [28] | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | 7 |
| Rose et al. (2002)[21] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | 9 |
| Greene et al. (2010) [24] | ± | ± | ✓ | ✓ | ✗ | ✓ | ± | ✗ | ✓ | ✗ | 5.5 |
| Murphy et al. (2003) [29] | ✓ | ± | ✗ | ✗ | ✓ | ✓ | ± | ✓ | ✓ | ✓ | 7 |
| Wrisley; Kumar (2010) [22] | ± | ✓ | ✓ | ✓ | ✓ | ✓ | ± | ✓ | ✓ | ✓ | 9 |
| Dite et al. (2002) [20] | ✓ | ± | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | 9.5 |
| Chiu et al. (2003) [16] | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | 7 |
| Stel et al. (2003) [26] | ✓ | ± | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | 7.5 |
| Tsutsumimoto et al. (2011)[25] | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | 6 |
| Greene et al. (2012) [27] | ✓ | ± | ✓ | ✓ | ✓ | ✓ | ± | ✗ | ✓ | ✗ | 7 |

Caption: Q1: Question 1; Q2: Question 2; Q3: Question 3; Q4: Question 4; Q5: Question 5; Q6: Question 6; Q7: Question 7; Q8: Question 8; Q9: Question 9; Q10: Question 10; ✓: yes, ±: partial; ✗ not.

**Table 3.** Sample profile of the studies evaluated.

| Studies | Trueblood et al. (2001)[23] | Shumway-cook et al. (2000)[28] | Rose et al.(2002)[21] | Greene et al. (2010)[24] | Murphy et al. (2003)[29] | Wrisley; Kumar (2010)[22] | Dite et al. (2002)[20] | Chiu et al. (2003)[16] | Stel et al. (2003)[26] | Greene et al. (2012)[27] |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age group (years)** | > 60 | > 65 | > 60 | > 60 | > 60 | 60 -90 | > 65 | > 65 | 60 -85 | > 60 |
| **Sample size** | 198 | 30 | 134 | 349 | 45 | 35 | 81 | 34 | 435 | 226 |
| **Secondary or specific diagnoses** | ✗ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ |
| **Medications** | ✗ | ✓ | ✓ | ✗ | ✗ | ✓ | ✗ | ✗ | ✓ | ✓ |
| **Sensorial deficits** | ✗ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✗ | ✓ |
| **Balance deficits** | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ |
| **Cognitive state** | ✓ | ✗ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Limitation of Daily Life Activities** | ✓ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ | ✗ |
| **Gear device for Walking** | ✗ | ✓ | ✓ | ✗ | ✗ | ✗ | ✓ | ✓ | ✓ | ✓ |

Caption: ✓: yes, ±: partial; ✗ not.

Regarding the tests found in the literature, eight studies analyzed the *Timed Up and Go test* (TUG); Four *Balance Berg Scale* (BBS); Three *Functional Reach Test (FRT)*; Three *Chair Stands Five times Test* (CS-5); Two *Tandem Stand Test* (TS); Two *POMA Balance* (POMA-B); One *POMA Gait* (POMA-G); One *POMA Balance and Gait (POMA GT)*; One *Step Test* (ST); One *360º Turns* (360 º T); One *Semi-Tandem Stand Test* (STST); One *Side by Side Stand Test* (SSST); One Clinical Test of Sensory Organization and Balance Test (*CTSIB)*; *One One Leg Stand (OLS)*; One *Four Square Step Test* (FSST); One *Up-to-go* (UG); One *3m Walking Test* (3WT) and; One *Ordered Multi-Stepping Over Hoop* (OMO) (Table 4). Of these, only the TUG and *BERG Tests* can be performed meta-analysis because they have enough studies for the analytical tool.

In the *TUG test*, the combined sample size was composed of 1113 elderly individuals over 60 years old who were able to complete the test. Of these 616 elderly people reported a history of a retrospective or prospective falls. Follow-up periods varied 5 years retrospective and 2 years prospective. Sample sizes of the studies ranged from 30 to 226 participants. Six studies separated their sample between fallers and non-fallers; And two studies divided the sample into multi-fallers (with more than two falls), not multi-fallers (one fall) and non-fallers (TABLE 4). Forest graphs of sensitivities and specificities of 9 subsets of data from all eight studies are shown in Figure 2. Sensitivity values ranged from 10% to 89% with an average of 61.6% (95% CI: 57.2 -65.9%), and the specificity ranged from 62.9 to 100% with a mean of 76.5% (95% CI: 73-79.8%). The heterogeneity test of sensitivity and specificity shows $X^2$ = 59.34 (P <0.0001, $I^2$ = 87.42%) and $X^2$ = 67.2 (P <0.0001, $I^2$ = 85.53%), respectively. The discriminatory TUG values to identify the elderly fallers ranged from 10 to 20 seconds, with a mean of 16.25 seconds. When analyzing the threshold effect (liminal), Spearman's correlation coefficient was equal to -0.075, indicated a very weak association, not being statistically significant (p = 0.85). The post-test probability of an elderly person suffering a fall or not, considered a positive and negative odds ratio was 2.68 (1.73-4.16) and $X^2$ of 34.4 (p <0.001), and 0.41 (95% CI: 0.27-0.64) and $X^2$ of 107.45 (p <0.001), respectively. The SROC curve synthesizes the sensitivity and false positive values (1- specificity), obtaining a value of 0.85 of the area under the curve (AUC) and Cochran Q of 0.78 as the highest common value of the sensitivity and specificity. The overall Odds Ratio of these studies was 8.42 (95% CI: 3.4-20.8) and $X^2$ = 52.25 (p <0.001).

On the four studies found on *BBS*, one did not analyze the predictive validity, the other three studies obtained the combined sample size of 653 elderly individuals over 60 years old who were able to complete the Scale. Of these 329 elderly people reported a history of a retrospective or prospective fall. Follow-up periods varied 5 years retrospective and 2 years prospective studies. Sample sizes of studies ranged from 34 to 329 subjects. Two studies separated their sample between fallers and non-fallers, and one study separate fallers, multi-fallers and non-fallers (TABLE 4). Sensitivity values ranged from 45% to 95.5% with an average of 58.7% (CI = 95%: 53.1-64%, $X^2$ = 29.95, p <0.0001, $I^2$ = 83%), and the specificity ranged from 51% to 95.5% with a mean of 71.3% (95% CI: 66-76.2%, $X^2$ = 30.73, p <0.0001, $I^2$ = 82%). The discriminatory BBS scores to identify the elderly fallers

**Table 4.** Characterization of the studies regarding the objectives, sample, characterization of falls, values of sensitivity and specificity of the tests and conclusion of the studies.

| Author | Objective | Sample | | | | History of fall | | | Fall period | Test | SEN | SPE | CP | Conclusion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean age /standard deviation | N | M | H | SQ | 1Q | 2 or + f | | | | | | |
| Trueblood et al 2001[23] | Determine the SEN and SPE of fall prediction tests and whether the combination of these tests is the best predictor for falls | Total of sample: 77.9 (±7.26) | 189 | 143 | 37 | 99 | 73 | 26 | Six months (prospective) | POMA-B | 24% | 91% | 12/16 | The tests were accurate to identify individuals without risk of falling. |
| | | | | | | | | | | POMA-G | 21% | 95% | 8/12 | |
| | | | | | | | | | | TUG | 10% | 95% | 20 | |
| shumway-cook, Brauer e Woolacot (2000)[28] | Verify the SEN and SPE of the TUG under conditions of one task and two tasks for identification of the elderly who are prone to falling | With falls: 78(±6). No falls 86.2(±6) | 30 | - | - | 15 | _ | 15 | Six months (retrospective) | TUG | 80% | 100% | 13.3 | TUG is a simple screening test. It is a sensitive and specific measure of probability for falls among the elderly. |
| | | | | | | | | | | TUGc | 80% | 93.3% | 15 | |
| | | | | | | | | | | TUGm | 86.7% | 93.3% | 14.5 | |
| Rose, Jones e Luchese (2002)[21] | Determine if the performance of the *8 Up and Go Test* can discriminate adults with falls and without falls and can be specific and sensitive as a predictor of falls as TUG | With falls: 76.6(±6). No falls: 78.4(± 6.5). | 134 | 112 | 22 | 71 | _ | 63 | One year (retrospective) | UG | 78% | 86% | 8.5 | TUG represents a valid and effective method for discriminating possibility of falls occurrence in adults with different ages, levels of activity and independence. |
| | | | | | | | | | | TUG | 71% | 89% | 10 | |
| | | | | | | | | | | BBS | - | - | - | |
| Greene et al 2010[24] | Show whether the use of kinematic sensors in the body can be used to objectively to quantify the *TUG* and provide the amount of analysis time, march and stability time for each *TUG* segment | With falls: 74 (±7,3); No falls: 70,1 (± 6,9) | 349 | 246 | 103 | 142 | 207 | – | Five years (retrospective) | TUG | 58% | 64.8% | - | The assessment of *Body Balance* can be a fundamental strategy to improve quality of life and can guide to fall prevention strategies. |
| | | | | | | | | | | BBS | 57.8% | 64.2% | | |
| Murphy et al (2003)[29] | Identify the performance measures that have the greatest potential to predict falls in the elderly | No falls: 71.2(±7.3) With falls: 79.6±6.5. | 45 | 33 | 12 | 34 | | 11 | Fourteen months (prospective) | Floor Transfer | 64% | 100% | habil/ nãohabil | Several tests were able to discriminate fallers and not fallares, being the best predictors the test of "*Time of Rising of the Ground*" and "*to Walk 15 meters*" |
| | | | | | | | | | | CS-5 | 82% | 82% | 21s | |
| | | | | | | | | | | TST | 55 | 94 | Habil/ não hábil | |
| | | | | | | | | | | POMA-B | 55% | 97% | 12 | |
| | | | | | | | | | | FRT | 73% | 88% | 20,3 | |

Caption: TUG: Timed Up and Go test; TUGc: Timed Up and Go cognitive; TUGc: Timed Up and Go manual; BBS: Balance Berg Scale; UG: Get-Up and Go test; POMA-B: POMA Balance; POMA-G: POMA Gait; CS-5: Chair Stand Five Times Test; ST: 30 s Sit to Stand Test; DGI: Dynamic Gait Index; FGA: Functional Gait Assessment; EMS: Elderly Mobility Scale; FSST: Four Step Square Test; OMO: Ordered Multi-Stepping Over Hoop; Functional Reach Test: FRT; SEN: Sensibility; SPE: Specificity; CP: cutoff points.

**Table 4.** Continued...

| Author | Objective | Sample | | | | History of fall | | | Fall period | Test | SEN | SPE | CP | Conclusion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean age /standard deviation | N | M | H | SQ | 1Q | 2 or + f | | | | | | |
| Wrisley e Kumar (2010))[22] | The aim of the study was to determine the concurrent, discriminatory and predictive validity of FGA in the elderly | No falls: 71,5 (±7,2); With falls: 77,7±7.8 | 35 | 18 | 17 | 29 | – | 6 | six months (retrospective) | TUG | 83% | 96.6% | 12.34 | FGA seems to be able to predict falls and has good concurrency with other balance assessment tools. |
| | | | | | | | | | | BBS | - | - | - | |
| | | | | | | | | | | DGI | 100% | 75.9% | 19 | |
| | | | | | | | | | | FGA | 100% | 83% | 20 | |
| Dite et al. (2002)[20] | To establish the reliability and validity of a new *Dynamic Clinical Balance Test (FSST)*. To assess the sensitivity, specificity to identify subjects who fall and compare with 3 Tests of Balance and Mobility | Multiples falls: 74±5,68; One fall: 73,78±6.0No falls: 74,14±6,07 | 81 | _ | _ | 27 | 27 | 27 | six months (retrospective) | FSST | 89% | 85% | 15 | The FSST was sensitive and specific and had greater sensitivity and specificity in relation to the *3 Tests* with which it was compared. |
| | | | | | | | | | | TUG | 89% | 63% | 13 | |
| | | | | | | | | | | ST | 81% | 67% | 11S | |
| | | | | | | | | | | FRT | 63% | 59% | 25cm | |
| Chiu, Au- Yeng, Lo. (2003))[16] | Compare the *Balance Scale of Berg* and the *Tinetti, Mobility Scale* in the elderly with the *TUG* in the discrimination of elderly fallers and non-fallers | Two or more falls: (82,86±6.63) Only one fall: (82,12±8,19) No falls group: (81,86±7,82); | 78 | 52 | 26 | 39 | 17 | 22 | six months (retrospective) | BBS | 95,5% | 76,5% | 38 | Although the fall is a multifactorial event, evaluations of this type would require a lot of time for application. Thus, BBS is a strong predictor of falls. |
| | | | | | | | | | | POMA B-G | 95,5% | 95,5% | 17 | |
| | | | | | | | | | | EMS | 90,9% | 95,5% | 19 | |
| | | | | | | | | | | TUG | 77,3% | 88,2% | 24s. | |
| Stel et al (2003)[26] | Examine whether easily measurable measures of balance and muscle strength can predict recurrent falls, as well as, sophisticated measurements. Examine which of the modifiable risk factors were most strongly associated with recurrent falls. | No falls or one fall: (78,1±6,1)Two falls or more (78,9±6,6) | 435 | 238 | 197 | 333*; | | 102 | One year (prospective) | TST | 45% | 74% | 10s | Measurements such as the *Tandem Stand* and *Manual Grip Strength* can predict falls as well as sophisticated measurements of body balance. |

Caption: TUG: Timed Up and Go test; TUGc: Timed Up and Go cognitive; TUGc: Timed Up and Go manual; BBS: Balance Berg Scale; UG: Get-Up and Go test; POMA-B: POMA Balance; POMA-G: POMA Gait; CS-5: Chair Stand Five Times Test; ST: 30 s Sit to Stand Test; DGI: Dynamic Gait Index; FGA: Functional Gait Assessment; EMS: Elderly Mobility Scale; FSST: Four Step Square Test; OMO: Ordered Multi-Stepping Over Hoop; Functional Reach Test: FRT; SEN: Sensibility; SPE: Specificity; CP: cutoff points.

**Table 4.** Continued...

| Author | Objective | Sample | | | | History of fall | | | Fall period | Test | SEN | SPE | CP | Conclusion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean age /standard deviation | N | M | H | SQ | 1Q | 2 or + f | | | | | | |
| Tsutsumimoto et al. (2013)[25] | To investigate the association of the *Ordered Multi Steeping Over Hoop (OMO)* with cognitive and physical function and to investigate if the *OMO* can predict incidences of falls | No falls: (84,0±1,1) With falls: (85,5±1,4) | 59 | 49 | 10 | 45 | 14 | | One year (prospective) | OMO | 86% | 60% | 21.9 | The OMO is a good predictor for identifying fallers and not fallers. |
| Greene et al. (2012)[27] | To determine if a method based on sensor data used in the body can predict future falls in the elderly in the community and compare the performance of the prediction of falls in two standardized methods of the same dataset | No falls 71,37 ±6,6 With falls: 71,75 ±6.9 | 226 | 164 | 62 | 83 | 143 | | One year (prospective) | BBS | 44,73% | 73.62% | 45 | The quantification of *TUG* movements using body sensors can be an alternative to evaluate future falls. |
| | | | | | | | | | | TUG | 42,83% | 62.62% | 15.25s | |

Caption: TUG: Timed Up and Go test; TUGc: Timed Up and Go cognitive; TUGc: Timed Up and Go manual; BBS: Balance Berg Scale; UG: Get-Up and Go test; POMA-B: POMA Balance; POMA-G: POMA Gait; CS-5: Chair Stand Five Times Test; ST: 30 s Sit to Stand Test; DGI: Dynamic Gait Index; FGA: Functional Gait Assessment; EMS: Elderly Mobility Scale; FSST: Four Step Square Test; OMO: Ordered Multi-Stepping Over Hoop; Functional Reach Test: FRT; SEN: Sensibility; SPE: Specificity; CP: cutoff points.

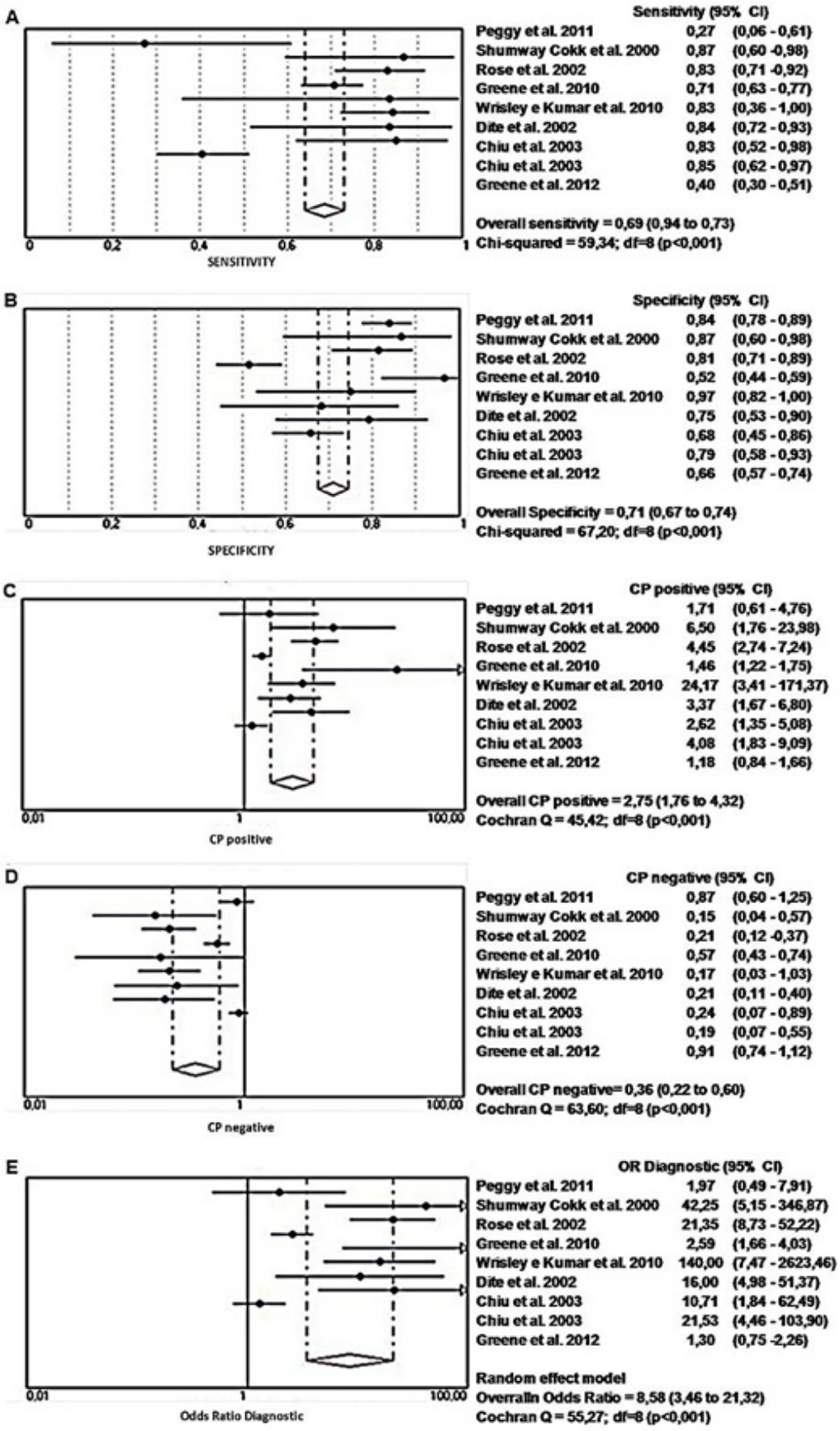


**Figure 2.** Graph of the predictive validity analysis of the *Timed Up and Go Test*.

ranged from 41.62 to 51.48; Multi-fallers 23,18 and; Non-fallers from 48.59 to 54.00, with an average of 41.52 for fallers and 50.69 for non-ones. When analyzing the threshold effect (liminal), Spearman's correlation coefficient was equal to r = -0.800, indicating a strong association and not statistically significant (p = 0.2). The post-test probability of an elderly person suffering a fall or not, considered a positive and negative odds ratio was 2.26 ($X^2$ = 10.52, p = 0.015) and 5.13 ($X^2$ = 17.6; = 0.001), respectively. The SROC curve synthesizes the sensitivity and false positive values (1- specificity), obtaining a value of 0.89 of the area under the curve (AUC) and Cochran Q of 0.82 as the highest common value of sensitivity and specificity. The overall odds ratio of these studies had a value of 6.70 (95% CI: 1.75 - 23.88, $X^2$ = 18.72, p <0.001).

Of the three studies that analyzed FRT, one did not find a prospective difference between fallers and non-fallers, totaling two studies that reported sensitivity of the test. The combined sample size was composed of 126 elderly individuals aged over 60 years were able to complete the scale. Of these 65 elderly people reported a history of a retrospective falls with a time of six months and a prospective falls of 14 months. The sensitivity values were 64.6% ($X^2$ = 20.01, $I^2$ = 46%, p <0.001) and specificity was 75.4% ($X^2$ = 1.88, $I^2$ = 14.9%, p = 0.17). The post-test probability of an elderly person falling or non-falling, considered a positive and negative probability coefficient ratio, was 2.89 and $X^2$ = 6.03 (p = 0.01) and 0.50 and $X^2$ = 1.84 (p = 0.17), respectively. The overall odds ratio of these studies had a value of 6.23 ($X^2$ = 4.49, p = 0.03).

In the *POMA Test*, the combined sample size was composed of 225 elderly individuals over 60 years old who were able to complete the scale. Of these 41 elderly people reported a history of falls. Sensitivity values were 46.4% ($X^2$ = 6.20, p = 0.01) and specificity was 85.8% ($X^2$ = 0.04, p = 0.83). The post-test probability of an elderly patient falling or not, considered a positive and negative probability coefficient ratio, was 3.76 and ($X^2$ = 3.34, p = 0.07), and 0.43 and $X^2$ = 3.94, p = 0.05), respectively. The overall odds ratio of these studies had a value of 8.73 ($X^2$ = 4.13, p = 0.09).

In the *TS*, the combined sample size was composed of 480 elderly subjects over the age of 60 years were able to complete the scale. Of these 113 elderly people reported a history of falls. Sensitivity values were 37% ($X^2$ = 5.12, p = 0.02) and specificity was 82% ($X^2$ = 0.60, p = 0.44). The post-test probability of an elderly person suffering a fall or not, considered a ratio of positive and negative propensity coefficient, was 2.96 ($X^2$ = 4.94, p <0.03) and 0.57 ($X^2$ = 2.96, p = 0.09), respectively. The overall odds ratio of these studies was 5.47 (Q = 4.68, p = 0.03).

Three studies using the *CS-5* did not find a significant difference between the fallers and non-fallers, as did the STST, SSST, CTSIB and OLS.

## DISCUSSION

This study aimed to analyze the predictive validity of the functional mobility tests to detect the risk of falls in elderly residents in the community and independent., as well as, their methodological qualities in terms of design and procedures. The systematic review evidenced that the only test with scientific discriminative validity for this population is the *TUG*, however it is not plausible to consider it as "gold reference" test, since it does not present an adequate predictive validity. The results of this study may have been influenced by the methodological quality criteria of the studies that were not controlled by this study, although the majority (72.7%) presented good methodological control according the proposed criteria used here (See Table 3). It is noted, however, that there were major biases in the studies that might influence their results.

The review indicated that the mean *TUG* time of the fallers and non-fallers in all studies presented a significant difference, suggesting that the test has the capacity to differentiate the fallers. In investigating the sensitivity and specificity of the test, it was noted that the test has moderate diagnostic capacity. The diagnostic probability ratio (ORD) shows that the *TUG* has a good discriminative capacity to identify the fallers and the non-fallers, discriminative capacity produced with the studies included in this meta-analysis that proves that the *TUG* is a good diagnostic test, with accuracy of 0.85 and the overall OR value of 8.4, besides of the Cronbach Q of 0.78. However, the *TUG* cannot be considered with a predictive test to predict falls in the elderly with precision, that is, cannot be considered a "gold standard test." One of the reasons would be the heterogeneity found in the study, in which the test had a weak and non-significant correlation between sensitivity and specificity, suggesting that there must be other factors resulting in variations in the estimated precision between the individual studies. The variability presented in the test should not be analyzed as an impediment to the analysis of the results, but should be considered in careful evaluations when analyzing the biases of publications.

The variability of the studies may be influencing the results of the systematic review of the *TUG Test*, since of the eight studies five are retrospective, impairing the criterion of predictive validation of the test. In addition, studies conducted under different circumstances, differing in relation to the criteria selection of the sample and the control of intrinsic factors that may influence the performance of the test, indicating that even if we selected studies with the same type of sample, the studies among themselves can present high homogeneity of subjects inside of samples. That is, not all the studies presented intrinsic and extrinsic control of the factors that can have influencing the final result of the test. However, the problem with data homogeneity and inconsistent results, when analyzed systematically, can provide important data on test behavior in clinical practice where the population will be

varied. To test possible interactions of these variables and their influence on accuracy it is necessary to carry out a study with this purpose, with an appropriate design.

At the *BBS*, the four studies found a significant difference in mean score between the fallers and non-fallers, suggesting that the test has the ability to differentiate the fallers. However, of the four studies analyzed only three realized the predictive validity of the instrument. The values of sensitivity (58.7%) and specificity (71.3%) calculated through the meeting of the studies analyzed proved the high accuracy as a diagnostic test. The diagnostic odds ratio indicated had a good discriminative ability to identify the fallers and the non-fallers. However, when analyzing the studies in detail, it can be observed that the results may have been biased in the meta-analysis. Due to the fact that we found two studies[24,27] with the same author in different period distances and design - one retrospective and the other prospective, both presented low test sensitivity and moderate specificity. The other study is by Chiu et al.[16] (2003), with retrospective case-control design, which analyzed between non-fallers and fallers that presented high sensitivity and specificity; And between non-fallers and multi-fallers that obtained excellent sensitivity and specificity (95% in both). In this situation, perhaps the result of the study by Chiu et al.[16], may have influenced the results of the meta-analysis performed here, thus not being able to conclude whether the instrument has good predictive validity.

Several studies have examined the validity of *BBS* in detecting the risk of falls in elderly residents in the community, through factor analysis, ROC curve analysis, and sensitivity and specificity in several languages[30, 31, 32, 33, 34, 35,]. Studies point to a Cronbach's alpha ranging from 0.62 to 0.98; Sensitivity between 42% and 97%; Specificity between 26% and 92%; Cut point between 45 and 50 points[10,13,30,31,32,33,34,35,30,36,37,38,39]. Despite presenting a high alpha of Cronbach indicating a good internal consistency, as well as a good association between item-factor and the other factors, this does not guarantee a good internal consistency in the measurement of the postural control construct. In addition, several studies report that this instrument has a "ceiling effect" [10, 11; 12, 13] and this reason is the justification used in the study by Wrisley and Kumar[22] did not analyze the predictive validity of this instrument, even finding that there was a significant difference between the fallers and non-fallers. In this sense, Souza et al.[39] reports that the *BBS* has no predictive validity in elderly patients with knee osteoarthrosis, exposing some fragility in the test. However, the *BBS* was sensitive to individual differences in the sample, even in the absence of a good index of sensitivity and specificity.

The *POMA B* and *TS* were selected only in two studies each, which did not allow an in-depth analysis in the tests. The tests presented low sensitivity, but it does not clarify the aspect regarding the validity of the tests to detect altered balance and the predisposition to fall for elderly residents in the community.

Three studies used the CS-5, none of which showed discriminative ability to compare the mean between fallers and non-fallers, so the studies did not assess the validity of the criterion. This test is widely used in the elderly as a simple test[40, 41, 42, 43, 44, 45] and in physical performance batteries[46], for indirect measurement of muscle resistance of the knee joint - mainly quadriceps. The *CS-5* has been demonstrated to be a simple and viable physical performance to request a decrease in daily living activity (ADL) and falls in the elderly[47, 48, 49]. Nakazono et al.[50], concluded that *CS-5* and *Chair Test* (30 seconds) are simple and viable physical performance tests for the elderly and rehabilitation patients in the clinical context, to evaluate their muscular endurance.

In addition to the other tests (*360ºT*, *FSST*, *UG*, *Step Test Five Times*, *Step Test 15s*, OMO Test), they presented a single study, indicated a coefficient of precision above of 0.80, suggesting a probable excellent predictive validity, being considered "gold standard" [51]. Therefore, it is clinically important to develop studies that evaluate the predictive validity of these tests to identify the risk of falls in this population.

When analyzing the convergent validity of the tests in the studies, the *TUG* presented greater frequency of convergent validity with the other tests. The *TUG Test* showed weak correlation with *CTSIB* (on firm and unstable surface with closed eyes), moderate correlation with OLS, *FRT* and *FGA*; and high correlation with the *BBS*, *FSST* and *ST*. In relation to the convergent validity *BBS* presented moderate correlation with *FGA* (r = 0.53) and high correlation with *TUG* (r = -0.76). It is understood that the Spearmam and Pearson correlation tests will have a higher correlation index the closer to the dimension of the construct between the tests. Thus, it is expected that the *TUG* presents a high correction between the *FSST*, *ST* and *BBS tests*, since the *TUG* motor tasks are included or are similar to the motor tasks analyzed by the *FSST*, *ST* and *BBS*.

The systematic review has demonstrated the tendency of the studies[22, 29, 22, 26] to publish only diagnostic tests with good predictive validity, which leads to a scientific literature with bias. Studies with negative results should be stimulated for their publications, with the intuition to offer to the scientific environment the behavior of the tests in sample diversity, protecting the researcher to use tests not advisable for its population to be studied. Furthermore, in order to avoid this bias, studies should be designed with sufficient sample size of individuals for results to be credible and adequately disseminated. The accuracy of a diagnostic / predictive test is best determined by comparison with a standard reference test, considered to be the gold standard. Currently, in the literature, the standard reference test to predict falls in the elderly has not yet been clarified. Even so, the review proposed to select studies that presented a reference test

(adjusted test), that is, that verified the predictive validity of the target test and another test. The presence of a reference test reduces sampling rates in studies with non-representative sample size, especially in a control case study that the number of individuals with the phenomenon to be studied is equal to those that do not present what corresponds to a 50% prevalence of phenomenon, so that the researcher can evaluate the usefulness of the test according to its reality in clinical practice, that is, ascertain if the test underestimates or overestimates the ability of the test to detect the phenomenon in the population.

It is not possible with a single clinical trial to predict the risk of falling in the elderly because the fall is a multifactorial phenomenon. However, it is essential to have the scientific knowledge which discriminatory clinical tests are valid for the population profile to be studied. The systematic review pointed out that the only test with scientific discriminative validity for elderly residents in the community is the *TUG*, but it is not possible to consider it as a gold reference test because it does not present an adequate predictive validity. *TUG* may not be sufficient as a simplified screening tool to detect the risk of falls, we suggest the use of *TUG* associated with other predictors of fall risk to provide additional information on the identification of the elderly with the potential to suffer future falls. The meta-analysis of fall risk factors points to fear of falls, dementia, vertigo, gait deficiency, antiepileptic drugs, muscle weakness, balance deficiency and environmental risk are predictors of falls in community-dwelling elderly people[52]. It is suggested that longitudinal studies be conducted in to determine the causes and risk factors for falls in the elderly, verify the validity of the contents of functional mobility tests to detect the risk of falls. Besides that it is suggested studies to determine if the proposed tests and risk factors have predictive validity over time to detect the risk of falls in the elderly and studies to monitor risk factors and functional mobility tests remain effective over time.

## CONCLUSION

Although the systematic review evidenced that the studies present good methodological quality, they present great biases that compromised the result of the systematic review, such as: Types of studies found - many studies presented a retrospective design compromising the predictive validity of the test; Non-representative sample size; Selection criteria; Control of intrinsic and extrinsic factors; Concurrent validity of the tests to ascertain the predictive validity of the target test and of another test; Besides other biases.

Although the studies presented important biases, it can be evidenced that the only test with discriminative scientific validity for elderly residents in the community is the *TUG test*, but it cannot be considered as "gold reference test", since its predictive validity does not allow for this purpose. Considering that the phenomenon of fall in the elderly is multifactorial, it is suggested that the *TUG test* can be used in association with other predictors of fall risk to provide additional information on the identification of the elderly with the potential to suffer future falls.

**AUTHOR'S CONTRIBUTION**

ACSSM: Systematic review and writing of the article; PBP: Systematic review and writing of the article; EH: Systematic review; ECM: Systematic review; GZM: Correction of the manuscript; FLC: Correction and translation of the manuscript.

**CONFLICTS OF INTEREST**

The authors declare that they have no conflicts of interest in the research.

**AUTHOR DETAILS**

2. Professor in Graduate Program in Human Motion Sciences, Santa Catarina State University (UDESC), Santa Catarina (SC), Brazil. 3. Master in Human Motion Sciences, Santa Catarina State University (UDESC), Santa Catarina (SC), Brazil. 4. Professor in Graduate Program in Human Motion Sciences and Graduate Program on Education, Santa Catarina State University (UDESC), Santa Catarina (SC), Brazil.

## REFERENCES

1. Silvestre JA, Costa Neto MM. Abordagem do idoso em programas de saúde da família. Cad Saúde Pública. 2003;19(3):839-47.
2. Perracini MR, Ramos LR. Fatores associados a quedas em uma coorte de idosos residentes na comunidade. Rev Saúde Pública. 2002; 36(6): 709-16.
3. Datasus. Banco de dados do Sistema Único de Saúde. 2010
4. Aschkenasy MT, Rothenhaus TC. Trauma and falls in the elderly. Emerg Med Clin North Am, 2006; 24: 413-32.
5. Englander F. Economic dimensions of slip and fall injuries. J Forensic Sci. 1996; 41(5): 733-46.
6. Zuckerman D. Hype in health reporting: "checkbook science" buys distortion of medical news. Int. J. Health Serv. 2003; 33(2): 383-89.
7. Georg AE, Duncan BB, Toscano CM, Schmidt MI, Mengue S, Duarte C, Polanczyk CA. Análise econômica de programa para rastreamento do diabetes mellitus no Brasil. *R.* Saúde Públ. 2005; 39(3): 452-60.
8. Pasquali L. Instrumentação psicológica: Fundamentos e práticas. Porto Alegre, Brasil: Artmed, 2010.
9. Pasquali L. Psicometria. Rev. Esc. Enferm. 2009; 43: 992-99.
10. Newton R. Balance screening of an inner city older adult population. Arch Phys Med Rehabil. 1997; 78: 587-91.
11. Muir SW; Berg K; Chesworth B; SPEECHLEY M. Use of the Berg Balance Scale for predicting multiple falls in community-dwelling elderly people: a prospective study. Phys Ther. 2008; 88(4): 449-59.
12. Santos GM, Souza ACS, Virtuoso JF, Tavares GMS, Mazo GZ. Valores preditivos para o risco de quedas em idosos praticantes e não praticantes de atividade física por meio do uso da Escala de Equilíbrio de Berg. Braz J Phys Ther. 2011, 15(2): 95-101.
13. Alzayer L, Beninato M, Portney LG. The accuracy of individual Berg Balance Scale items compared with the total Berg score for classifying people with chronic stroke according to fall history. J Neurol Phys Ther. 2009, 33(3): 136-43.
14. Morris R, Harwood RH, Baker R, Sahota O. Armstrong S Masud T. A comparison of different balance tests in the prediction of falls in older women with vertebral fractures: a cohort study. Age Ageing. 2007, 36(1): 78-83.
15. Thrane G, Joakimsen RM, Thornquist E. The association between timed up and go test and history of falls: the Tromsø study. BMC Geriatr 2007, 12(7): 1.

16. Chiu AY, Au-Yeung SS, Lo SK. A comparison of four functional tests in discriminating fallers from non-fallers in older people. Disabil Rehabil. 2003, 25(1): 45-50.
17. Pardasaney PK, Latham NK, Jette AM, Wagenaar RC, Ni P, Slavin MD, Bean JF. Sensitivity to change and responsiveness of four balance measures for community-dwelling older adults. Phys Ther. 2012, 92(3): 388-97.
18. Farrell MK, Rutt RA, Lusardi MM, Williams AK. Are scores on the physical performance test useful in determination of risk of future falls in individuals with dementia? J Geriatr Phys Ther. 2011, 34(2): p57-63.
19. Oliveira MRF, Gomes AC, Toscano C M. QUADAS e STARD: Avaliação da qualidade de estudos de acurácia de testes diagnósticos Rev Saude Publica, 2011, 45(2): 416-22.
20. Dite W, Temple VA. A clinical test of stepping and change of direction to identify multiple falling older adults Arch Phys Med Rehabil. 2002, 83(11): 1566-71.
21. Rose DJ, Lucchese N, Wiersma LD. Development of a multidimensional balance scale for use with functionally independent older adults Arch Phys Med Rehabil. 2006, 87(11): 1478-85.
22. Wrisley DM, Kumar NA. Functional gait assessment: concurrent, discriminative, and predictive validity in community-dwelling older adults. Phys Ther. 2010, 90(5): 761-73.
23. Trueblood PR, Hodson-Chennault N, McCubbin A, Youngclarke D. Performance and impairment-based assessments among community dwelling elderly: sensitivity and specificity. Aging (Albany NY) 2001, 24(1):2–6
24. Greene BR, O'Donovan A, Romero-Ortuno R. Quantitative falls risk assessment using the timed up and go test. Biomed Tech (Berl) 2010, 57(12): 2918-26.
25. Tsutsumimoto K, Doi T, Misu S, Ono R, Hirata S. Can the Ordered Multi-Stepping Over Hoop test be useful for predicting fallers among older people? A preliminary 1 year cohort study Aging Clin Exp Res. 2013, 25(4): 427-32.
26. Stel VS, Smit JH, Plujim SMF. Balance and mobility performance as treatable risk factors for recurrent falling in older persons. J Clin Epidemiol. 2003, 56(7): 659-68.
27. Greene BR, Doheny EP, Walsh C. Evaluation of falls risk in community-dwelling older adults using body-worn sensors. Gerontology 2012, 58(5): 472-80.
28. Shumway-Cook A, Brauer S, et al. Predicting the probability for falls in community-dwelling older adults using the Timed Up & Go Test. Phys Ther. 2000, 80(9): 896-903.
29. Murphy SL, Dubin JA, Gill TM. The development of fear of falling among community-living older women: predisposing factors and subsequent fall events. J Gerontol A Biol Sci Med Sci. 2003, 58(10): M943-M947.
30. Matsushima M, Yabe I, Uwatoko H, Shirai S, Hirotani M, Sasaki H. Reliability of the Japanese version of the Berg balance scale. Intern Med. 2014, 53(15): 1621-4.
31. Salavati M, Negahban H, Mazaheri M, Soleimanifar M, Hadadi M, Sefiddashti L, Zahraee MH, Davatgaran K, Feizi A. The Persian version of the Berg Balance Scale: inter and intra-rater reliability and construct validity in elderly adults Disabil Rehabil. 2012, 34(20): 1695-98.
32. De Figueiredo KM, De Lima KC, Cavalcanti MAC, Guerra RO. Interobserver reproducibility of the Berg Balance Scale by novice and experienced physiotherapists Physiother Theory Pract. 2009, 25(1):30-6
33. Sahin F, Yilmaz F, Ozmaden A, Kotevolu N, Sahin T, Kuran B. Reliability and validity of the Turkish version of the Berg Balance Scale J Geriatr Phys Ther. 2008, 31(1): 32-37.
34. Wang CY, Hsieh CL, Olson SL, Wang CH, Sheu CF, Liang CC. Psychometric properties of the Berg Balance Scale in a community-dwelling elderly resident population in Taiwan. J Formos Med Assoc. 2006, 105(12): 992-1000.
35. Miyamoto ST, Lombardi Junior I, Berg KO, Ramos LR, Natour J. Brazilian version of the Berg balance scale Braz J Med Biol Res. 2004, 37(9): 1411-21.
36. Muir SW, Berg K, Chesworth B. Balance impairment as a risk factor forfalls in community-dwelling older adults Who are high functionin: a prospective study Phys Ther. 2010, 90(3): 338-47.
37. Holbein-Jenny MA, Billek-Sawhney B, Beckman E, Smith T. Balance in personal care home residents: a comparison of the Berg Balance Scale, the Multi-Directional Reach Test, and the Activities-Specific Balance Confidence Scale J Geriatr Phys Ther. 2005, 28(2): 48-53.
38. Riddle DL, Stratford PW. Interpreting validity indexes for diagnostic tests: an illustration using the Berg balance test. Phys Ther. 1999, 79(10); 939-48.
39. Souza ACS, Santos GM. Sensibilidade da Escala de Equilíbrio de Berg em indivíduos com osteoartrite Motriz: Rev Educ Fis. 2012, 18(2): 307-18.
40. Reider NC, Naylor PJ, Gaul C Sensitivity and specificity of the minimal chair height standing ability test: a simple and affordable fall-risk screening instrument. J Geriatr Phys Ther. 2015, 38(2):90-5.
41. Hirase T, Inokuchi S, Matsusaka N, Nakahara K, Okita M. A modified fall risk assessment tool that is specific to physical function predicts falls in community-dwelling elderly people. J Geriatr Phys Ther. 2014, 37(4): 159-65.
42. Buatois S, Perret GC, Gueguen R, Miget P, Vançon G, Perrin P, Benetos A. A simple clinical scale to stratify risk of recurrent falls in community-dwelling adults aged 65 years and older. Phys Ther. 2010, 90(4); 550-60.
43. Tiedemann A, Shimada H, Sherrington C, Murray S, Lord S. The comparative ability of eight functional mobility tests for predicting falls in community -dwelling older people Age Ageing. 2008, 37(4): 430-35.
44. Buatois S, Miljkovic D, Manckoundia P, Gueguen R, Miget P, Vançon G, Perrin P, Benetos A. Five times sit to stand test is a predictor of recurrent falls in healthy community-living subjects aged 65 and older. J Am Geriatr Soc. 2008, 56(8):1575-77.
45. Schaubert KL, Bohannon RW. Reliability and validity of three strength measures obtained from community-dwelling elderly persons J Strength Cond Res. 2005, 19(3): 717-20.
46. Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol. 1994, 49(2): M85-94.
47. Ward RE, Leveille SG, Beauchamp MK, Travison T, Alexander N, Jette AM, Bean JF. Functional performance as a predictor of injurious falls in older adults. J Am Geriatr Soc. 2015, 63(2): 315-20.
48. Cheng YY, Wei SH, Chen PY, Tsai MW, Cheng IC, Liu DH, Kao CL. Can sit-to-stand lower limb muscle power predict fall status? Gait Posture 2014, 40(3): 403-7.
49. Zhang M, Kim JC, Li Y, Shapiro BB, Porszasz J, Bross R, Feroze U, Upreti R, Martin D, Kalantar-Zadeh K, Kopple JD. Relation between anxiety, depression, and physical activity and performance in maintenance hemodialysis patients J Ren Nutr. 2014, 24(4): 252-60.
50. Nakazono T, Kamide N, Ando M. The Reference Values for the Chair Stand Test in Healthy Japanese Older People: Determination by Meta-analysis. J Phys Ther Sci. 2014, 26(11):1729-31.
51. Noronha APP, Alchieri JC. Reflexões sobre os instrumentos de avaliação psicológica Em R Primi (Org) Temas em avaliação psicológica. Campinas: Impressão Digital do Brasil 2002.
52. 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(5): 658-68.

**APPENDIX I.** Search terms MEDLINE, SCOPUS and CINAHL.

| MEDLINE Terms | SCOPUS Terms | CINAHL Terms |
|---|---|---|
| **BLOCK ELDERLY:**<br>1 Aged/<br>2 agetw<br>3 elderlytw<br>4 oldertw<br>5 (old* adj2 people)tw<br>6 (old* adj2 adult*)tw<br>7 aging/<br>8 or/1-7 | **BLOCK ELDERLY**<br>Aged<br>Age<br>Elderly<br>Older<br>old* PRE/2people<br>old* PRE/2 adult<br>aging | **BLOCK ELDERLY:**<br>S1 (MH "Aged")<br>S2 TX Age<br>S3 TX Elderly<br>S4 TX Older<br>S5 TX (Old* n2 People)<br>S6 TX (Old* n2 adult*)<br>S7 (MH "Aging")<br>S8 S1 or S2 or S3 or S4 or S5 or S6 or S7 |
| **BLOCK STUDY TYPE**<br>9 exp "Sensitivity and Specificity"/<br>10 sensitivitytw<br>11 specificitytw<br>12 ((pre-test or pretest) adj probability)tw<br>13 post-test probabilitytw<br>14 predictive value$tw<br>15 likelihood ratio$tw<br>16 or/9-15 | **BLOCK STUDY TYPE**<br>"Sensitivity and Specificity"<br>Sensitivity<br>specificity<br>((pre-test OR pretest) PRE/2 probability)<br>predictive value$<br>likelihood ratio$<br>post-test probability | **BLOCK STUDY TYPE**<br>S9 (MH "Sensitivity and Specificity")<br>S10 TX Sensitivity<br>S11 TX Specificity<br>S12 TX Post-test probability<br>S13 (MH "Predictive Value of Tests")<br>S14 TX (likelihood ratio*)<br>S15 S9 or S10 or S11 or S12 or S13 or S14 |
| **BLOCK BALANCE**<br>17 (control adj2 postur*)tw<br>18 equilibriumtw<br>19 (postur* adj2 equilibr*)tw<br>20 balancetw<br>21 (postur* adj2 balance)tw<br>22 Postural Balance/<br>23 or/17-22 | **BLOCK BALANCE**<br>postural balance<br>postur* PRE/2 balance<br>balance<br>postur* PRE/2 equilibr*<br>equilibrium<br>control PRE/2 postur* | **BLOCK BALANCE**<br>S16 TX (Control n2 postur*)<br>S17TX (Equilibrium)<br>S18 TX (Postur* n2 Equilibr*)<br>S19 TX Balance<br>S20 TX Balance, Postural<br>S21 S16 or S17 or S18 or S19 or S20 |