Validity, Reliability, and Ability to Identify Fall Status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in Patients With COPD

Cristina Jácome, Joana Cruz, Ana Oliveira, Alda Marques

Background. The Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest are useful in the assessment of balance. Their psychometric properties, however, have not been tested in patients with chronic obstructive pulmonary disease (COPD).

Objective. This study aimed to compare the validity, reliability, and ability to identify fall status of the BBS, BESTest, Mini-BESTest, and the Brief-BESTest in patients with COPD.

Design. A cross-sectional study was conducted.

Methods. Forty-six patients (24 men, 22 women; mean age=75.9 years, SD=7.1) were included. Participants were asked to report their falls during the previous 12 months and to fill in the Activity-specific Balance Confidence (ABC) Scale. The BBS and the BESTest were administered. Mini-BESTest and Brief-BESTest scores were computed based on the participants' BESTest performance. Validity was assessed by correlating balance tests with each other and with the ABC Scale. Interrater reliability (2 raters), intrarater reliability (48–72 hours), and minimal detectable changes (MDCs) were established. Receiver operating characteristics assessed the ability of each balance test to differentiate between participants with and without a history of falls.

Results. Balance test scores were significantly correlated with each other (Spearman correlation rho=.73–.90) and with the ABC Scale (rho=.53–.75). Balance tests presented high interrater reliability (intraclass correlation coefficient [ICC]=.85–.97) and intrarater reliability (ICC=.52–.88) and acceptable MDCs (MDC=3.3–6.3 points). Although all balance tests were able to identify fall status (area under the curve=0.74–0.84), the BBS (sensitivity=73%, specificity=77%) and the Brief-BESTest (sensitivity=81%, specificity=73%) had the higher ability to identify fall status.

Limitations. Findings are generalizable mainly to older patients with moderate COPD.

Conclusions. The 4 balance tests are valid, reliable, and valuable in identifying fall status in patients with COPD. The Brief-BESTest presented slightly higher interrater reliability and ability to differentiate participants' fall status.

C. Jácome, PT, PhD, Lab 3R–Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro, Aveiro, Portugal.

J. Cruz, PT, MSc, Lab 3R–Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro, and Institute for Research in Biomedicine, University of Aveiro.

A. Oliveira, PT, MSc, Lab 3R–Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro, and Institute for Research in Biomedicine, University of Aveiro.

A. Marques, PT, PhD, Lab 3R–Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro, Agras do Crasto–Campus Universitário de Santiago, Edifício 30, 3810-193 Aveiro, Portugal, and Institute for Research in Biomedicine, University of Aveiro, Aveiro, Portugal. Address all correspondence to Dr Marques at: amarques@ua.pt.

[Jácome C, Cruz J, Oliveira A, Marques A. Validity, reliability, and ability to identify fall status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD. *Phys Ther.* 2016;96:1807–1815.]

Published Ahead of Print: April 14, 2016
Accepted: March 31, 2016
Submitted: July 8, 2015

Chronic obstructive pulmonary disease (COPD) is one of the most prevalent chronic diseases among adults aged 60 years and older.[1] This respiratory disease is characterized by a progressive deterioration of pulmonary function and by its systemic effects, which contribute greatly to the decline of patients' functional performance.[2] Skeletal muscle weakness, reduced exercise capacity, slow gait, and reduced physical activity levels are well-known systemic effects in COPD.[3–5] As a result, patients with COPD may experience difficulties in performing activities of daily living that require balance control[6] and may be at high risk of falling. Recent literature indicates that approximately 30% to 50% of patients with COPD fall at least once during a 6- to 12-month period.[7–9]

In patients with COPD, it has been shown that balance impairment is independently associated with falls.[10] Thus, valid, reliable, and clinically feasible tests aimed to assess balance are urgently needed to identify patients at risk of falling and to evaluate the impact of rehabilitation programs.

A number of balance tests have been described in the literature. The Berg Balance Scale (BBS) and the Balance Evaluation Systems Test (BESTest) have been the most commonly used tests in patients with chronic diseases such as stroke[11,12] and Parkinson disease.[13,14] The BBS has been shown to be highly sensitive and specific in predicting fall risk in community-dwelling older adults.[15] In patients with Parkinson disease, the BESTest has been reported to be capable of identifying future recurrent fallers.[16] These 2 balance tests also were able to differentiate patients with COPD from healthy age- and sex-matched controls.[7] However, the ability of the BBS and the BESTest to identify fall status in patients with COPD has not yet been explored.

In addition, both the BBS and the BESTest were able to detect changes after a 6-week intervention of balance training within a pulmonary rehabilitation program in patients with COPD.[17] However, although the psychometric properties (validity, interrater and intrarater reliability, and minimal detectable change [MDC]) of these tests have been established in several specific populations,[18–20] they have not yet been investigated in patients with COPD. Determining the psychometric properties of these tests is fundamental to deciding whether they are appropriate to assess balance impairments in patients with COPD.[21]

In the last few years, shortened versions of the BESTest were developed: the Mini-BESTest[22] and the Brief-BESTest.[20] These balance tests also have gained interest for assessment of balance in patients with Parkinson disease,[16] multiple sclerosis,[20] and balance disorders,[22] as they were faster and easier to use in clinical practice in comparison with the BBS and the BESTest. However, neither the Mini-BESTest nor the Brief-BESTest have been applied, or their psychometric properties studied, in patients with COPD. Therefore, the purpose of this study was to compare the validity, reliability, and ability to identify fall status of the BBS, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD.

Method

Study Design

A cross-sectional study was conducted. Fifty outpatients with COPD were recruited from 2 primary care centers and one district hospital between November 2013 and November 2014. The reliability sections of this study were described following the Guidelines for Reporting Reliability and Agreement Studies.[21]

Participants

Patients were included if they met the following criteria: (1) diagnosis of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria,[23] (2) age of 60 years or older, (3) clinical stability for 1 month prior to the study (no hospital admissions or exacerbations as defined by the GOLD[23]), (4) ability to ambulate with or without a walking aid, and (5) living independently in the community. Patients were excluded if they presented coexisting respiratory diseases (eg, asthma) or had severe neurological (eg, Parkinson disease, dementia), musculoskeletal (eg, severe osteoarthritis), or psychiatric (eg, psychosis, schizophrenia) impairments that could interfere with the measurements.

Eligible patients were identified and screened by their clinicians and then contacted by the researchers, who explained the purpose of the study and asked about their willingness to participate. When patients agreed to participate, an appointment with the researchers was scheduled at the patients' reference health care center. Written informed consent was obtained prior to data collection.

Data Collection

Sociodemographic, anthropometric (height, weight, body mass index [BMI]) and clinical (comorbidities and number of acute exacerbations of COPD in the preceding year) data were first collected.

Then, patients were provided with a clear definition of falls (an event when you find yourself unintentionally on the ground, floor or lower level[24(p632)]) and asked about their history of falls using 2 standardized questions: (1) "Have you had any falls in the last 12 months?" and, if yes, (2)"How many times did you fall down in the last 12 months?"[25]

Disability resulting from dyspnea was assessed using the modified Medical Research Council (mMRC) questionnaire.[26] This questionnaire comprises 5 grades (0–4), with higher grades indicating greater perceived respiratory limitation. The mMRC is simple to administer and correlates significantly with measures of health status.[23] Balance confidence was assessed using the Activities-specific Balance Confidence (ABC) Scale.[27] The ABC Scale quantifies an individual's perceived ability to maintain his or her balance under different circumstances on a scale of 0% (no confidence) to 100% (total confidence).[27] Participants received explanations about the aim of each questionnaire and were asked to complete them by themselves. For participants who were unable to read, questionnaires were interviewer administered.

Lung function was measured with a portable spirometer (MicroLab 3500, CareFusion, Kent, United Kingdom) according to standardized guidelines.[28] The GOLD spirometric classification was used to determine the severity of the disease: mild COPD, defined as forced expiratory volume in 1 second (FEV_1) ≥80% predicted; moderate COPD, defined as 50%≤FEV_1<80% predicted[29]; and severe-to-very severe COPD, defined as FEV_1<50% predicted.[23]

Lastly, the BBS and the BESTest were performed and participants were encouraged to rest, as needed. Two qualified physical therapists, with at least 4 years of experience in working with patients with COPD, administered the balance assessment. They were experienced using the BBS but had limited experience applying the BESTest in patients with COPD. Therefore, to ensure competency in applying the BESTest, the physical therapists watched the BESTest training video and read the testing procedures. Then they practiced administering the 4 balance tests between them and in 2 patients with COPD prior to the data collection period.

Interrater and intrarater reliability were analyzed in a subsample of the first consecutive 28 participants. This sample size was determined according to the study by Bonett,[30] who established that a minimum of 21 individuals were necessary to estimate an intraclass correlation coefficient (ICC) of .9 with a 95% confidence interval width of .2[31] (α=.05 and required number of raters=2). As interventions with patients with COPD have considerable dropouts (23%[32] and 31%[33]), a 30% attrition rate was estimated, yielding a sample of 28 individuals.

For interrater reliability, the 2 physical therapists rated the participant's performance independently (session 1). For each item of the BBS or BESTest, one rater read the standardized instructions to the participant while the second rater demonstrated how to perform the task. The participant then performed the task with close supervision. Each task was scored immediately after completion by the 2 raters. For intrarater reliability, participants were reassessed by 1 of the 2 physical therapists after a 48- to 72-hour interval (session 2). The order of testing was the same as in the first assessment. An effort was made to keep all factors associated with the testing sessions consistent, specifically the time of the day, location in which the tests were performed, and use of a walking aid (if needed).

Mini-BESTest and Brief-BESTest scores were computed based on the performance of the BESTest tasks. The raters used a custom-designed worksheet to simultaneously record the BESTest and Mini-BESTest item scores. Brief-BESTest scores were extracted from the relevant subset of BESTest items.

Balance Tests

BBS. The BBS is composed of 14 items that assess an individual's performance on specific functional tasks. Each item is scored from 0 to 4, and the maximum test score is 56 points. Higher scores indicate better balance performance. The BBS has high interrater and intrarater reliability in institutionalized older adults[18] and in patients with Parkinson disease[19] and stroke.[11,12] In addition, the BBS has demonstrated ability to identify balance impairments in individuals with vestibular dysfunction, with 75% sensitivity and specificity.[34]

BESTest. The BESTest contains 36 items organized into 6 subsections: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural responses to external perturbations, sensory orientation during stance, and stability in gait.[35] Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment), and the maximum test score is 108 points. The BESTest has high interrater reliability in community-dwelling older adults and in patients with Parkinson disease.[19,35] Moreover, in the study by Duncan et al,[16] the BESTest was able to identify recurrent fallers in patients with Parkinson disease.

Mini-BESTest. The Mini-BESTest includes 14 items from sections of the BESTest related to anticipatory postural adjustments, reactive postural responses, sensory orientation, and stability in gait.[36] Two of the 14 items (stand on one leg and compensatory stepping correction–lateral) are scored bilaterally. Each item is scored from 0 (severe balance impairment) to 2 (no balance impairment), and the maximum possible score is 28 points. Higher scores indicate better balance performance. High interrater and intrarater reliability have been found for the Mini-BESTest in patients with balance disorders, chronic stroke, and Parkinson disease.[13,22,37] In patients with Parkinson disease, the Mini-BESTest has showed high sensitivity (89%) and specificity (81%) in identifying abnormal postural responses.[38]

Brief-BESTest. The Brief-BESTest is a 6-item balance test that contains 1 item from each of the 6 subsections of the BESTest.[20] Similarly to the Mini-BESTest, 2 items are scored bilaterally. Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment), and the maximum possible score is 24 points. Higher scores indicate better balance performance.[20] This balance test has shown high interrater reliability (ICC=.99) in individuals with and without neurological diseases.[20] The Brief-BESTest was found to be able to identify recurrent fallers in patients with Parkinson disease.[16]

Data Analysis

All statistical analyses were performed using IBM SPSS Statistics version 20.0 (IBM Corp, Armonk, New York) and plots created using GraphPad Prism version 5.01 (GraphPad Software Inc, La Jolla, California). The level of significance was set at .05.

Descriptive statistics were used to describe the sample. A z test was applied for normality test using skewness and kurtosis.[39] Characteristics were compared between participants with and without a history of falls and between those included in the reliability analysis and the remaining sample using independent t tests for normally distributed data (age, BMI, ABC Scale, and FEV_1), Mann-Whitney U tests for nonnormally distributed data (comorbidities, BBS, BESTest, Mini-BESTest, and Brief-BESTest) and ordinal data (mMRC), and chi-square tests for categorical data (sex, exacerbations of COPD in the preceding year, and GOLD spirometric classification). Participants with a history of falls were defined as those who reported at least one fall during the past year; participants without a history of falls were defined as those who reported no falls during the past year. When significant differences on the performance of balance tests between participants with and without a history of falls were found, effect sizes were computed. Cohen's *d* was used[40] and interpreted as small ($d \geq 0.2$), medium ($d \geq 0.5$), or large ($d \geq 0.8$) effect[41] (G*Power 3.1, University of Düsseldorf, Düsseldorf, Germany).

The skewness of the distribution of scores was assessed for each balance test to verify the occurrence of ceiling and floor effects. A positive skewness value greater than 1 indicates a substantial floor effect, and a negative value lower than −1 indicates a substantial ceiling effect.[42]

Validity

The Spearman correlation (rho) was used to examine the relationship among balance tests (concurrent validity) and between each balance test and the ABC Scale (convergent validity).

Reliability

As recommended for reliability studies, both the relative and absolute reliability were determined with the ICC and the Bland and Altman method, respectively.[43] Interrater reliability was computed using the scores obtained from the 2 raters in session 1, and intrarater reliability was computed using the scores from 1 rater in sessions 1 and 2. The ICC (2,1) was used and interpreted as excellent (ICC>.75), moderate-to-good (ICC=.4–.75), or poor (ICC<.4).[44]

MDC

To determine the MDC, first the standard error of measurement (SEM) was calculated. The SEM indicates the extent to which a score varies on repeated measurements[45] and was calculated using the equation:

$$(1) \qquad SEM = SD\ \sqrt{(1 - ICC)}$$

where SD is the standard deviation of the scores obtained from all individuals, and ICC is the intrarater reliability coefficient.

The MDC at the 95% level of confidence (MDC_{95}) was calculated as follows:

$$(2) \qquad MDC_{95} = SEM \times 1.96 \times \sqrt{2}$$

The MDC also was expressed as a percentage (MDC%), calculated as:

$$(3) \qquad MDC\% = (MDC_{95}/mean) \times 100,$$

where "mean" is the mean of the scores obtained in the 2 testing sessions. An MDC% below 30% was considered acceptable.[46]

Ability to Identify Fall Status

Receiver operating characteristic (ROC) analysis was used to assess the ability of each balance test to differentiate between patients with and without a history of falls. The cutoff for each balance test was chosen as the point where the sensitivity and specificity were simultaneously maximized. The area under the curve (AUC) and the 95% confidence interval were determined.[47] The AUC is the probability of correctly identifying a patient with COPD who has a history of falls in randomly selected pairs of patients who have and do not have a

Table 1.
Participants' Characteristics[a]

Characteristic	Total (N=46)	Participants Without a History of Falls (n=23)	Participants With a History of Falls (n=23)	*P*
Age (y)	75.9 (7.1)	74.6 (5.9)	77.2 (8)	.21
Sex				
Male	24 (52.2%)	14 (60.9%)	10 (43.5%)	.38
Female	22 (47.8%)	9 (39.1%)	13 (56.5%)	
BMI (kg/m^2)	28.4 (4.7)	28.4 (4.8)	28.3 (4.8)	.91
mMRC, M (IQR)	2 (1–2)	2 (1–2)	2 (1–3)	.28
Exacerbations in the previous year				
0	28 (60.9%)	16 (69.6%)	12 (52.2%)	.18
>1	18 (39.1%)	7 (30.4%)	11 (47.8%)	
Comorbidities, M (IQR)	2 (1–3)	2 (0–3)	2 (1–3.75)	.40
FEV_1 (% predicted[29])	69.4 (19.9)	68.8 (21)	70.1 (19.2)	.83
GOLD spirometric classification				
Mild	13 (28.3%)	8 (34.8%)	5 (21.7%)	.55
Moderate	21 (45.7%)	9 (39.1%)	12 (52.2%)	
Severe-to-very severe	12 (26.1%)	6 (26.1%)	6 (26.1%)	
ABC Scale	64.1 (25.7)	84.8 (11.7)	43.3 (17.8)	<.001
BBS	50.1 (5.5)	53.3 (4.3)	48.3 (5.4)	≤.001
BESTest	77.8 (12.5)	82.8 (11.4)	72.7 (11.7)	.01
Mini-BESTest	20.8 (4.9)	22.6 (4.4)	18.9 (4.7)	.01
Brief-BESTest	15.7 (4.9)	18.0 (4.2)	13.5 (4.7)	.01

[a] Values shown as $\overline{X}$ (SD) or n (%), unless otherwise indicated. ABC=Activities-specific Balance Confidence, BBS=Berg Balance Scale, BESTest=Balance Evaluation Systems Test, BMI=body mass index, FEV_1=forced expiratory volume in 1 second, GOLD=Global Initiative for Chronic Obstructive Lung Disease, IQR=interquartile range, M=median, mMRC=modified Medical Research Council dyspnea scale.

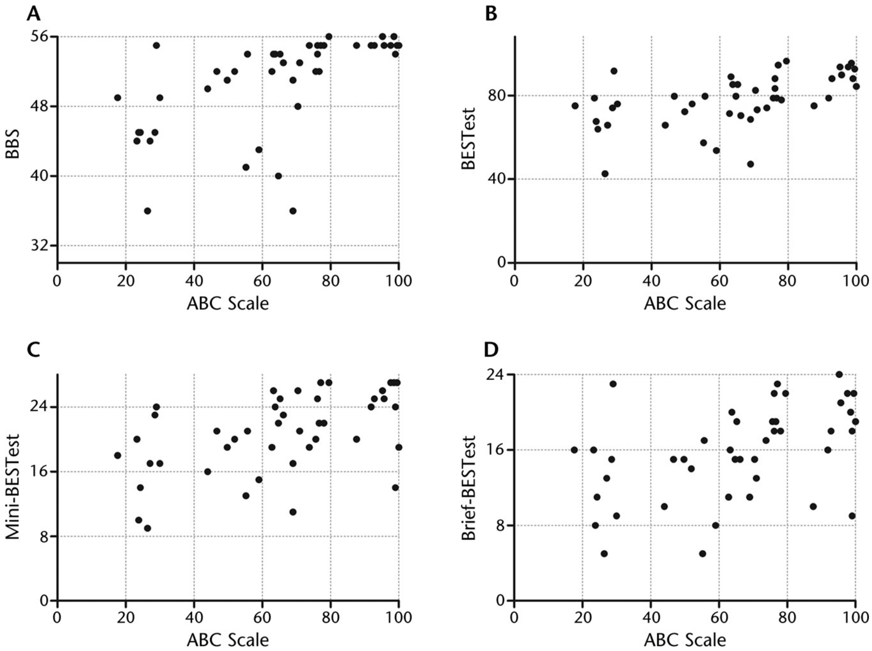

Figure 1.
Scatterplots showing the relationship between the Activities-specific Balance Confidence (ABC) Scale and (A) the Berg Balance Scale (BBS), (B) the Balance Evaluation Systems Test (BESTest), (C) the Mini-BESTest, and (D) the Brief-BESTest (N=46).

history of falls.[48] The AUC was interpreted as follows: AUC=0.5 indicates no discrimination, 0.7≤AUC<0.8 indicates acceptable discrimination, 0.8≤AUC <0.9 indicates excellent discrimination, and AUC≥0.9 indicates outstanding discrimination.[49] The positive and negative likelihood ratios (LR+ and LR−) also were computed.[50]

Role of the Funding Source

This work was partially supported by O Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE and Fundação para a Ciência e a Tecnologia (FCT) under project UID/BIM/04501/2013 and by National Funds through Fundação para a Ciência e a Tecnologia, in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.

Results

Participants

Fifty patients were contacted and invited to participate in the study. However, 3 patients were unable to attend the health center, and 1 patient did not complete the assessment. Therefore, 46 participants (24 men, 22 women) were enrolled in the study. On average, participants were 75.9 years of age (SD=7.1), with a mean BMI of 28.4 kg/m^2 (SD=4.7). The median mMRC grade was 2 ("I walk slower than people of the same age on the level because of the breathlessness" or "I have to stop for breath when walking on my own pace on the level"). According to the GOLD spirometric classification, 28.3% (n=13) of the participants had mild COPD, 45.7% (n=21) had moderate COPD, and 26.1% (n=12) had severe-to-very severe COPD (n=12). No significant differences regarding any of the sociodemographic, anthropometric, and clinical characteristics were found between participants with and without a history of falls. Participants' characteristics are shown in Table 1.

All balance tests were able to significantly differentiate between participants with and without a history of falls (P<.01) (Tab. 1). The largest effect sizes were found for the BBS (d=1.02) and the Brief-BESTest (d=1.01). The effect sizes for the BESTest and the Mini-BESTest also were large (d=0.87 and d=0.81, respectively). The BBS had the highest ceiling effect (skewness=−1.31). The Brief-BESTest was less skewed (skewness=−0.44) than the BESTest (skewness=−0.77) and the Mini-BESTest (skewness=−0.79).

Validity

All balance tests were strongly correlated with each other, with rho ranging from .73 to .90 (P<.001). The ABC Scale was significantly correlated with the BBS (rho=.75), BESTest (rho=.61), Mini-BESTest (rho=.55) and the Brief-BESTest (rho=.53) (P<.001) (Fig. 1).

Table 2.
Interrater and Intrarater Reliability of the BBS, BESTest, Mini-BESTest, and Brief-BESTest (n=28)[a]

Balance Test	Interrater Reliability			Intrarater Reliability		
	ICC (2,1) (95% CI)	Mean Difference (SD)	95% LA	ICC (2,1) (95% CI)	Mean Difference (SD)	95% LA
BBS	.94 (.88, .97)	0.5 (1.6)	−2.6 to 3.6	.52 (.19, .74)	−0.7 (2.9)	−6.3 to 4.9
BESTest	.85 (.70, .92)	−1.2 (3.8)	−8.6 to 6.2	.87 (.73, .94)	−0.5 (3.7)	−7.7 to 6.8
Mini-BESTest	.85 (.71, .93)	−0.7 (2.1)	−4.7 to 3.3	.88 (.75, .94)	0 (1.7)	−3.4 to 3.4
Brief-BESTest	.97 (.94, .99)	−0.1 (1.0)	−2.1 to 2.0	.82 (.66, .92)	−0.7 (2.5)	−5.6 to 4.2

[a] BBS=Berg Balance Scale, BESTest=Balance Evaluation Systems Test, ICC=intraclass correlation coefficient, 95% CI=95% confidence interval, 95% LA=95% limits of agreement.

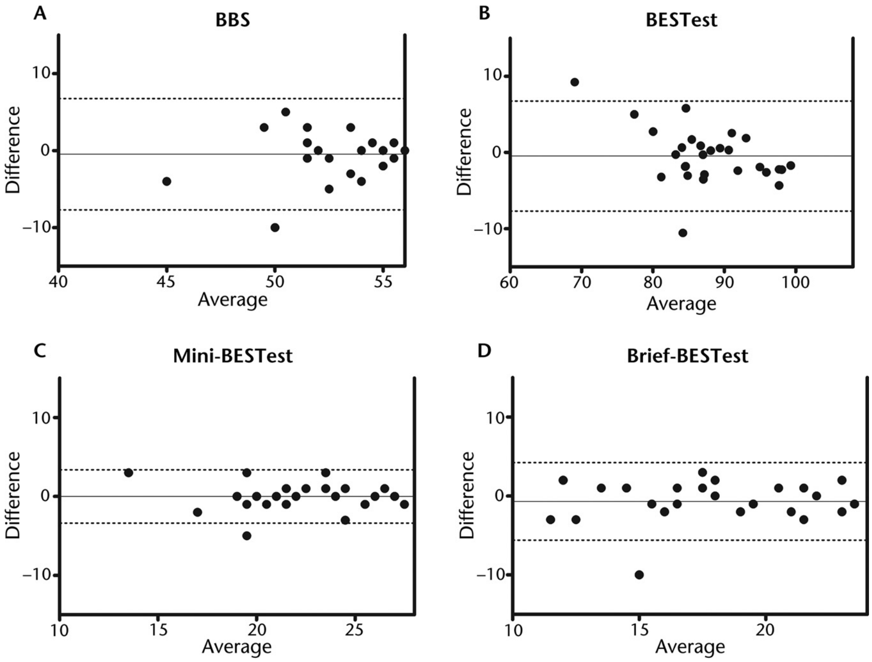

Figure 2.
Bland-Altman plots of (A) Berg Balance Scale (BBS), (B) Balance Evaluation Systems Test (BESTest), (C) Mini-BESTest, and (D) Brief-BESTest between 2 sessions (n=28). The solid line represents the mean difference between sessions 1 and 2, and the dashed lines represent the 95% limits of agreement.

Interrater and Intrarater Reliability

There were no significant differences between participants included in the reliability analysis and the remaining participants. Table 2 presents the relative and absolute interrater and intrarater reliability results of the BBS, BESTest, Mini-BESTest, and Brief-BESTest. Excellent interrater relative reliability was observed for all balance tests (ICC≥.85). Good interrater agreement was verified for all 4 balance tests, with mean differences close to zero (Tab. 2).

The BBS had moderate-to-good relative intrarater reliability (ICC=.52), and the other balance tests had excellent reliability (ICC=.82−.88) (Tab. 2). Bland-Altman plots revealed no systematic bias, with mean differences ranging from −0.7 to 0 (Fig. 2).

MDC

The MDC_{95} was 5.9 (SEM=2.1, $MDC_{\%}$=11.1%), 6.3 (SEM=2.3, $MDC_{\%}$=7.2%), 3.3 (SEM=1.2, $MDC_{\%}$=14.9%), and 4.9 (SEM=1.8, $MDC_{\%}$=26.9%) for the BBS, BESTest, Mini-BESTest, and Brief-BESTest, respectively.

Ability to Identify Fall Status

Table 3 presents the results from the ROC analysis. The AUCs ranged from 0.74 to 0.84, indicating an acceptable-to-good ability of all 4 balance tests to identify fall status. The higher AUCs were found for the BBS (AUC=0.84; 95% CI=0.72, 0.96) and the Brief-BESTest (AUC=0.78; 95% CI=0.64, 0.92) (Tab. 3). The sensitivity of the Brief-BESTest (81%) was 8%, 13%, and 17% higher than that of the BBS (73%), the Mini-BESTest (68%), and the BESTest (64%), respectively. Specificity was similar across balance tests (65%–77%). The Brief-BESTest and the BBS presented the higher LR+ values (LR+=3 and 3.20) and the lower LR− values (LR−=0.25 and 0.35) (Tab. 3).

To differentiate between participants with and without a history of falls, cutoff points of 16.5 (sensitivity=81%, specificity=73%) for the Brief-BESTest and 52.5 (sensitivity=73%, specificity=77%) for the BBS were identified (Fig. 3).

Table 3.
Ability to Identify Fall Status of the Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and the Brief-BESTest (n=46)[a]

Balance test	AUC (SEM)	95% CI	Cutoff Point	% Sensitivity/ % Specificity	Positive/Negative Likelihood Ratios
BBS	0.84 (0.06)	0.72, 0.96	52.5	73/77	3.20/0.35
BESTest	0.75 (0.07)	0.61, 0.90	76.9	64/77	2.8/0.47
Mini-BESTest	0.74 (0.07)	0.60, 0.89	21.5	68/65	1.96/0.49
Brief-BESTest	0.78 (0.07)	0.64, 0.92	16.5	81/73	3/0.25

[a] BBS=Berg Balance Scale, BESTest=Balance Evaluation Systems Test, AUC=area under the curve, 95% CI=95% confidence interval, SEM=standard error of measurement.

Discussion

This was the first study, to our knowledge, to investigate the validity, reliability, and ability to identify fall status of the BBS, BESTest, Mini-BESTest, and Brief-BESTest in patients with COPD.

Findings showed that among the 4 balance tests, the Brief-BESTest had the lowest ceiling effect (as indicated by the degree of skewness), followed by the BESTest and the Mini-BESTest. Conversely, similarly to previous studies, the BBS showed a high ceiling effect.[11,38] Thus, caution should be taken when selecting the BBS to assess balance in patients with COPD who have mild balance dysfunction (eg, score on balance clinical measures worse than 1 standard deviation from the mean score published for healthy older people),[48] as it may not be able to detect meaningful changes. In these specific cases, the use of the Brief-BESTest, the BESTest, or the Mini-BESTest may be recommended.

The 4 balance tests were significantly associated with each other and with

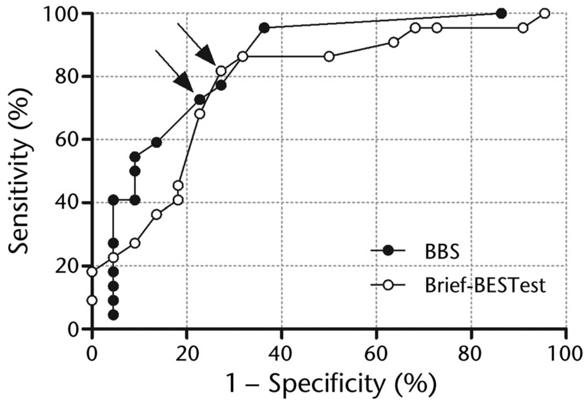

Figure 3.
Receiver operating characteristics of the Berg Balance Scale (BBS) and the Brief-Balance Evaluation Systems Test (Brief-BESTest) to differentiate between participants with and without a history of falls. The points corresponding to cutoff points are indicated by arrows.

the ABC Scale, demonstrating good concurrent and convergent validity. These findings are in agreement with studies conducted in other specific populations.[22,37,51]

Balance tests presented high interrater relative reliability (ICC>.8); however, slightly lower ICCs were found for intrarater relative reliability (ICC>.5). It is common to find lower intrarater reliability than interrater reliability.[13,22] However, although the high interrater reliability values were in accordance with previous findings in other populations, the values found for intrarater reliability were not (ICC=.88-.96[13,22]). This finding may be related to the between-days symptom variation of patients with COPD. It is well known that, in patients with COPD, the perception of symptoms, mainly dyspnea, vary over the week and have a negative impact on patients' activities of daily living, such as washing, dressing, drying after bathing, and getting out of bed.[52] As most daily life tasks involve dynamic balance, dyspnea may have played a role in participants' performance during the 2 sessions. Future studies should investigate intrarater reliability of the analyzed balance tests within the same day to reduce the variability of patients' health status. This has been done to explore intrarater reliability of the Timed "Up & Go" Test in patients with advanced COPD.[53] In terms of absolute reliability, no systematic bias was found for interrater or intrarater reliability. Thus, it seems that clinicians can be confident in using these 4 balance tests to assess balance impairments in patients with COPD.

The established MDCs were within the range described in other populations: BBS (range=3.3-6.3[12,22,54,55]), BESTest (range=6.2-6.9[26,51,56]), and Mini-BESTest (range=2.4-3.7[22,37,51,56]). For the Brief-BESTest, the MDC found was slightly higher compared with the MDCs established for older survivors of cancer (MDC=2.6 points)[51] and patients with total knee arthroplasty (MDC=3.2 points).[56] These differences may be population-specific, but they also may be related to the samples used. In the present study, participants' mean age was 76 years, and 52% of them were male. In the reported studies, the mean ages were between 68[51] and 69[56] years, and most participants were female (71%[51] and 74%[56]). The MDCs determined are acceptable[46] and can be used by clinicians to identify a true change in balance over time or in response to interventions in patients with COPD. Moreover, the MDCs found can strengthen the results obtained in previous studies.[17]

Determining the ability of balance tests to identify fall status in patients with COPD is crucial to allow clinicians to detect risk of falling before a fall occurs and implement effective interventions. The results showed that all balance tests were able to significantly differentiate between patients with and without a history of falls, although the largest effect sizes were found for the BBS and the Brief-BESTest. When analyzing the ROC curves, it was verified that all 4 balance tests had an acceptable ability to differentiate between patients with and without a history of falls. Yet, the cutoff points of the BBS and of the Brief-BESTest demonstrated higher sensitivity and specificity and, simultaneously, higher LR+ and LR− values. These cutoff points were similar to those reported in other populations for the BBS (52 points[38]) and for the Brief-BESTest (11 points[16]). However, when adding the information of the ceiling effect and of the reliability, the Brief-BESTest had the best performance. These results are important for clinical practice because they suggest that, if equipment or time to perform a balance test is limited, clinicians may confidently rely on the Brief-BESTest. It is not known, however, whether the differences in the ability to identify fall status among balance tests are clinically meaningful, and this issue needs to be explored in future studies.

The results from this study should be interpreted in light of the following limitations. The sample included older patients (age >60 years) primarily with moderate COPD, which limits the generalizability of the results to the overall COPD population. It is known that older adults frequently present reduced skeletal muscle strength,[57] exercise capacity,[1] gait speed,[4] and physical activity levels.[39] These impairments also may have contributed to the balance deficits and risk of falling found in the patients with COPD. Moreover, it is unclear whether factors related to COPD, such as severity of dyspnea, number of comorbidities, and acute exacerbations, have contributed to risk of falling, as differences between patients with and without a history of falls were not statistically significant. Future studies should include a more balanced sample of COPD grades and compare the balance impairment and risk of falling between patients with COPD and healthy controls in order to clarify these issues.

Another potential limitation is that the order of testing was not randomized, so fatigue may have affected participants' performance on some of the tests. However, participants were given frequent resting breaks. In addition, the Mini-BESTest and Brief-BESTest scores were derived from the BESTest performance. Considering the length of the BESTest, it is possible that inter-item influences may have occurred. Future studies should assess the psychometric properties of the Mini-BESTest and the Brief-BESTest when performed separately from the BESTest. The small number of participants used to perform ROC analysis may be seen as another limitation of the present study. Nevertheless, previous research applying the BESTest in patients with and without neurological conditions (LR−=0.27)[20] used a sample size of 46 to estimate an LR− of 0.13, with 90% specificity and 80% specificity.[58] Moreover, as there are false positives and false negatives in all 4 balance tests, cutoff points should be considered indicators of risk of falling to assist clinical decision making, instead of definitive points to classify fallers and nonfallers. Finally, as this was a cross-sectional study, only retrospective analysis of the ability of the balance tests to identify fallers among patients with COPD was possible. Longitudinal studies should be conducted to assess the prospective ability of these tests in identifying recurrent fallers.

The BBS, BESTest, Mini-BESTest, and Brief-BESTest are valid, reliable, and valuable tests to differentiate fall status in patients with COPD. If equipment or time is limited, clinicians may confidently rely on the Brief-BESTest. The MDC established for these balance tests can be used by clinicians to identify a true change in balance in patients with COPD.

Mrs Jácome, Mrs Cruz, and Dr Marques provided concept/idea/research design. All authors provided writing, data collection, and consultation (including review of manuscript before submission). Mrs Jácome and Dr Marques provided data analysis and project management. Dr Marques provided facilities/equipment. The authors acknowledge the patients who participated in this study. A special thanks to Mariana Couto, Mariana Lopes, Sílvia Ignácio, and Ana Teixeira for their valuable contribution in data collection.

Approval for this study was obtained from the Ethic Committee of the Research Unit of Health Sciences at the School of Nursing in Coimbra.

This work was partially supported by O Fundo Europeu de Desenvolvimento Regional (FEDER) through COMPETE and Fundação para a Ciência e a Tecnologia (FCT) under project UID/BIM/04501/2013 and by National Funds through Fundação para a Ciência e a Tecnologia, in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.

DOI: 10.2522/ptj.20150391

References

1 Fleg JL, Morrell CH, Bos AG, et al. Accelerated longitudinal decline of aerobic capacity in healthy older adults. *Circulation.* 2005;112:674–682.

2 Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *Eur Respir J.* 2009;33:1165–1185.

3 Decramer M, De Benedetto F, Del Ponte A, Marinari S. Systemic effects of COPD. *Respir Med.* 2005;99(suppl 2):S3–S10.

4 Jahn K, Zwergal A, Schniepp R. Gait disturbances in old age: classification, diagnosis, and treatment from a neurological perspective. *Dtsch Arztebl Int.* 2010;107:306–316.

5 Ilgin D, Ozalevli S, Kilinc O, et al. Gait speed as a functional capacity indicator in patients with chronic obstructive pulmonary disease. *Ann Thorac Med.* 2011;6:141–146.

6 Cruz J, Marques A, Jácome C, et al. Global functioning of COPD patients with and without functional balance impairment: an exploratory analysis based on the ICF framework. *COPD.* 2015;12:207–216.

7 Beauchamp MK, Sibley KM, Lakhani B, et al. Impairments in systems underlying control of balance in COPD. *Chest.* 2012;141:1496–1503.

8 Roig M, Eng JJ, MacIntyre DL, et al. Falls in people with chronic obstructive pulmonary disease: an observational cohort study. *Respir Med.* 2011;105:461–469.

9 Marques A, Cruz J, Quina S, et al. Reliability and minimal detectable change of the Timed Up & Go and the 10-Meter Walk tests in patients with COPD. *COPD.* 2016;13:279–287.

10 Beauchamp MK, Hill K, Goldstein RS, et al. Impairments in balance discriminate fallers from non-fallers in COPD. *Respir Med.* 2009;103:1885–1891.

11 Mao HF, Hsueh IP, Tang PF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke.* 2002;33:1022–1027.

12 Flansbjer UB, Blom J, Brogårdh C. The reproducibility of Berg Balance Scale and the Single-leg Stance in chronic stroke and the relationship between the two tests. *PM&R.* 2012;4:165–170.

13 Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and BESTest sections for balance assessments in individuals with Parkinson disease. *J Neurol Phys Ther.* 2011;35:90–97.

14 Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-Item Short-Form Health Survey, and the Unified Parkinson Disease Rating Scale in people with parkinsonism. *Phys Ther.* 2008;88:733–746.

15 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.

16 Duncan RP, Leddy AL, Cavanaugh JT, et al. Comparative utility of the BESTest, mini-BESTest, and brief-BESTest for predicting falls in individuals with Parkinson disease: a cohort study. *Phys Ther.* 2013;93:542–550.

17 Beauchamp MK, Janaudis-Ferreira T, Parreira V, et al. A randomized controlled trial of balance training during pulmonary rehabilitation for individuals with COPD. *Chest.* 2013;144:1803–1810.

18 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;28:48–53.

19 Leddy AL, Crowner BE, Earhart GM. Functional Gait Assessment and Balance Evaluation System Test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. *Phys Ther.* 2011;91:102–113.

20 Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. *Phys Ther.* 2012;92:1197–1207.

21 Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64:96–106.

22 Godi M, Franchignoni F, Caligari M, et al. Comparison of reliability, validity, and responsiveness of the mini-BESTest and Berg Balance Scale in patients with balance disorders. *Phys Ther.* 2013;93:158–167.

23 Vestbo J, Hurd SS, Agustí AG, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2013;187:347–365.

24 Biderman A, Cwikel J, Fried A, Galinsky D. Depression and falls among community dwelling elderly people: a search for common risk factors. *J Epidemiol Community Health.* 2002;56:631–636.

25 Zhang F, Ferrucci L, Culham E, et al. Performance on five times sit-to-stand task as a predictor of subsequent falls and disability in older persons. *J Aging Health.* 2013;25:478–492.

26 Doherty DE, Belfer MH, Brunton SA, et al. Chronic obstructive pulmonary disease: consensus recommendations for early diagnosis and treatment. *J Fam Pract.* 2006;55:S1–S8.

27 Branco PS. Validation of the Portuguese version of the "Activities-specific Balance Confidence Scale." *Revista da Sociedade Portuguesa de Medicina Física e de Reabilitação.* 2010;19:20–25.

28 Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J.* 2005;26:319–338.

29 Quanjer PH, Tammeling GJ, Cotes JE, et al. Lung volumes and forced ventilatory flows. Report Working Party Standardization of Lung Function Tests, European Community for Steel and Coal. Official Statement of the European Respiratory Society. *Eur Respir J Suppl.* 1993;16:5–40.

30 Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21:1331–1335.

31 Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17:101–110.

32 Fischer MJ, Scharloo M, Abbink JJ, et al. Drop-out and attendance in pulmonary rehabilitation: the role of clinical and psychosocial variables. *Respir Med.* 2009;103:1564–1571.

33 Garrod R, Marshall J, Barley E, Jones PW. Predictors of success and failure in pulmonary rehabilitation. *Eur Respir J.* 2006;27:788–794.

34 Cohen HS, Kimball KT. Usefulness of some current balance tests for identifying individuals with disequilibrium due to vestibular impairments. *J Vestib Res.* 2008;18:295–303.

35 Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther.* 2009;89:484–498.

36 Franchignoni F, Horak F, Godi M, et al. Using psychometric techniques to improve the Balance Evaluation Systems Test: the Mini-BESTest. *J Rehabil Med.* 2010;42:323–331.

37 Tsang CS, Liao LR, Chung RC, Pang MY. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Phys Ther.* 2013;93:1102–1115.

38 King LA, Priest KC, Salarian A, et al. Comparing the Mini-BESTest with the Berg Balance Scale to evaluate balance disorders in Parkinson's disease. *Parkinsons Dis.* 2012;2012:375419.

39 Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40:181–188.

40 Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* 2007;39:175–191.

41 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

42 Bulmer MG. *Principles of Statistics.* New York, NY: Dover Publications; 1979.

43 Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil.* 1998;12:187–199.

44 Fleiss J. Reliability of measurements. In: Fleiss J, ed. *The Design and Analysis of Clinical Experiments*. New York, NY: John Wiley & Sons Inc; 1986.

45 Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19:231–240.

46 Huang SL, Hsieh CH, Wu RM, et al. Minimal detectable change of the Timed "Up & Go" Test and the Dynamic Gait Index in people with Parkinson disease. *Phys Ther.* 2011;91:114–121.

47 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.

48 Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med.* 2000;45:23–41.

49 Hosmer DW, Lemeshow S. *Applied Logistic Regression.* 2nd ed. New York, NY: John Wiley & Sons Inc; 2000.

50 Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ.* 2004;329:168–169.

51 Huang MH, Miller K, Smith K, et al. Reliability, validity, and minimal detectable change of Balance Evaluation Systems Test and its short versions in older cancer survivors: a pilot study. *J Geriatr Phys Ther.* 2016;39:58–63.

52 Kessler R, Partridge MR, Miravitlles M, et al. Symptom variability in patients with severe COPD: a pan-European cross-sectional study. *Eur Respir J.* 2011;37:264–272.

53 Mesquita R, Janssen DJA, Wouters EFM, et al. Within-day test-retest reliability of the Timed Up & Go Test in patients with advanced chronic organ failure. *Arch Phys Med Rehabil.* 2013;94:2131–2138.

54 Hiengkaew V, Jitaree K, Chaiyawat P. Minimal detectable changes of the Berg Balance Scale, Fugl-Meyer Assessment Scale, Timed "Up & Go" Test, gait speeds, and 2-minute walk test in individuals with chronic stroke with different degrees of ankle plantarflexor tone. *Arch Phys Med Rehabil.* 2012;93:1201–1208.

55 Donoghue D, Stokes EK. How much change is true change? The minimum detectable change of the Berg Balance Scale in elderly people. *J Rehabil Med.* 2009;41:343–346.

56 Chan AC, Pang MY. Assessing balance function in patients with total knee arthroplasty. *Phys Ther.* 2015;95:1397–1407.

57 Moreland JD, Richardson JA, Goldsmith CH, Clase CM. Muscle weakness and falls in older adults: a systematic review and meta-analysis. *J Am Geriatr Soc.* 2004;52:1121–1129.

58 Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform.* 2014;48:193–204.