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[1,2], Joana Cruz[1,2], Ana Oliveira[1,2], Alda Marques[1,2]

1 Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

2 Institute for Research in Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal

Corresponding author

Alda Marques, PT, MSc, PhD, Senior Lecturer, Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Agras do Crasto - Campus Universitário de Santiago, Edifício 30, 3810-193 Aveiro, Portugal. Email: amarques@ua.pt

Abstract

Background: **The** Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and **the** Brief-BESTest are useful to assess balance, however their psychometric properties 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 males; 75.9±7.1years) 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** 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** patients with and without **a** history of falls.

Results: Balance test scores were significantly correlated with each other (spearman's correlation **[*rho*]=**0.73-0.90) and with the ABC Scale (***rho*=**0.53-0.75). **Balance tests** presented high interrater (Intraclass Correlation Coefficients**[ICCs]=**0.85-0.97) and intrarater reliability (**ICCs=**0.52-0.88), and acceptable MDCs (**MDCs=**3.3-6.3 points). **Although all balance tests were able to identify fall status (area under the curve [AUC]=0.74-0.84), the BBS**

(sensitivity=73%, specificity=77%) and the Brief-BESTest (sensitivity=81%, specificity=73%) had the higher ability.

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

Conclusions: The four balance tests are valid, reliable and valuable to **identify** fall status in patients with COPD. The Brief-BESTest presented slightly higher interrater reliability and **ability to differentiate** patients' falls status.

Word count: 3953 words

Introduction

Chronic obstructive pulmonary disease (COPD) is one of the most prevalent chronic diseases among adults aged 60 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 be at **high** risk of falling. **Recent** literature indicates that approximately 30% to 50% of patients with COPD fall at least once during a 6-12 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 **P**arkinson disease.[13,14] The BBS has shown to be highly sensitive and specific in predicting fall risk in community-dwelling older adults.[15] In patients with **P**arkinson disease, the BESTest **has been** reported to be capable of identifying future recurrent fallers.[16] **These** two balance tests were **also** 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, while the psychometric properties (validity, interrater and intrarater reliability, minimal detectable change) 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 decide if 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 have also gained interest to assess balance in patients with **P**arkinson 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 the ability to identify fall status of t**he BBS, BESTest, Mini-BESTest and the Brief-BESTest** in patients with COPD.

Methods

Study design

A cross-sectional study was conducted. Fifty outpatients with COPD were recruited from two primary care centers and one district hospital between November 2013 and November 2014. Approval for this study was obtained from the institutional ethics committees. The reliability sections of this study were

described following the guidelines for reporting reliability and agreement studies (GRRAS).[21]

Participants

Patients were included if they met the following criteria: (i) **diagnosis** of COPD according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) **criteria**;[23] (ii) **age of** 60 years old or older; (iii) **clinical** stability for 1 month prior to the study (no hospital admissions or exacerbations as defined by the GOLD[23]); (iii) **ability** to ambulate with or without a walking aid and iv) living independently in the community. Patients were excluded if they presented co-existing respiratory diseases (e.g., asthma) or had severe neurological (e.g., **P**arkinson disease, **dementia)**, musculoskeletal (e.g., severe osteoarthritis) or psychiatric **(e.g., 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 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] and **asked about** their history of falls using two 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 collected using the modified Medical Research Council **questionnaire** (mMRC).[26] **This questionnaire comprises five 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] **B**alance confidence **was assessed** using the Activities-specific Balance Confidence (ABC) Scale.[27] **The ABC Scale quantifies an individual's perceived ability to maintain his/her balance under different circumstances, using 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, UK) according to **standardized** guidelines.[28] The GOLD spirometric classification was used to determine the severity of the disease: mild COPD, forced expiratory volume in one second (FEV_1) ≥80% predicted; moderate COPD, 50%≤FEV_1<80% predicted; and severe-to-very severe COPD, 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, performed 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 four balance tests between them and also in two 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 from Bonnet,[29] which has established that a minimum of 21 individuals **were** necessary to estimate an Intraclass Correlation Coefficient (ICC) of 0.9 with a 95% confidence interval width of 0.2[30] (α=0.05 and k=2). As interventions with patients with COPD have considerable dropouts (23%[31] and 31%[32]), a 30% attrition rate was estimated, yielding a sample of 28 individuals.

For interrater reliability, **the** two physical therapists rated the patient'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 **performed** the task. **The** participant then performed the task with close supervision. Each task was scored immediately after completion by the two raters. For intrarater reliability, participants were reassessed by 1 of the 2 physical therapists, after a 48–72h 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. A custom designed worksheet was used by raters to simultaneously record **the** BESTest and Mini-BESTest **item scores.** Brief-BESTest scores were extracted from the relevant subset of BESTest items.

Balance tests

Berg Balance Scale. The BBS is composed of 14 items which 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 **P**arkinson disease[19] and stroke.[11,12] In addition, the BBS **has** demonstrated to be able to identify balance impairments in individuals with vestibular dysfunction, with 75% sensitivity and specificity.[33]

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.[34] Each item is scored from 0 (no balance impairment) to 3 (severe 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 **P**arkinson disease.[19,34] Moreover, the BESTest was **able to identify** recurrent fallers in patients with **P**arkinson disease.[16]

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.[35] Two of the 14 items, namely **stand on one leg** and **compensatory stepping correction–lateral**, are scored bilaterally. Each item is scored from 0 (no balance impairment) to 2 (severe 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.[13,22,36] In patients with **P**arkinson disease, the **Mini-BESTest has showed** high sensitivity (89%) and specificity (81%) in **identifying abnormal** postural responses.[37]

Brief-BESTest. The Brief-BESTest is a 6-item balance test which contains 1 item from each of the 6 subsections of the BESTest.[20] Similarly to the Mini-BESTest, **two items** are scored bilaterally. Each item is scored from 0 (no balance impairment) to 3 (severe balance impairment) and the maximum possible score is 24 points. Higher scores **indicate better** balance performance.[20] This balance test **has showed high interrater** reliability **(ICC=0.99)** in individuals with and without neurological diseases.[20] The Brief-BESTest was found to be **able to identify** recurrent fallers in patients with **P**arkinson disease.[16]

Statistical analysis

All statistical analyzes were performed using IBM SPSS Statistics version 20.0 (IBM Corporation, Armonk, NY, USA) and plots created using GraphPad Prism version 5.01 (GraphPad Software, Inc., La Jolla, CA, USA). The level of significance was set at 0.05.

Descriptive statistics were used to describe the sample. A z-test was applied for normality test using skewness and kurtosis.[38] **Characteristics** were compared between **patients** 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 non-normally distributed (**comorbidities,** BBS, BESTest, Mini-BESTest and Brief-BESTest) and ordinal data (mMRC)**,** and Chi-square tests for categorical data (gender, e**xacerbations of COPD in the preceding year and** GOLD spirometric classification). Patients with **a** history of falls were defined as those

who reported at least one fall during the past year; patients 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 **patients** with and without **a** history of falls were found, effect sizes were computed. Cohen's *d* was used[39] and interpreted as small (***d*≥**0.2), medium (***d*≥**0.5), or large (***d*≥**0.8) effect[40] (G*Power 3.1, University Düsseldorf, Düsseldorf, DE).

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.[41]

Validity

Spearman's 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.[42] Interrater reliability was computed using the scores obtained from the 2 raters in session 1 and intrarater reliability using the scores from 1 rater in sessions 1 and 2. The $ICC_{2,1}$ was used and interpreted as excellent (**ICC**>0.75), moderate to good (**ICC=**0.4-0.75) or poor (**ICC**<0.4).[43]

Minimal Detectable Change

To determine the minimal detectable change (MDC), first the standard error of measurement (SEM) was calculated. The SEM indicates the extent to which a

score varies on repeated measurements[44] and was calculated using the equation 1:

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

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 (equation 2):

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

The MDC was also expressed as a percentage (MDC%), defined as (equation 3):

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

where "*mean*" is the mean of the scores obtained in the two testing sessions. A MDC% below 30% was considered acceptable.[45]

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. Area under the curves (AUC) and the 95% confidence interval were determined.[46] 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 have not a history of falls.[47] AUC** was interpreted as follows: AUC=0.5 no discrimination; 0.7≤AUC<0.8 acceptable discrimination; 0.8≤AUC<0.9 excellent discrimination and AUC≥0.9 outstanding discrimination.[48] The positive and negative likelihood ratios **(LR+ and LR-)** were also computed.[49]

Results

Participants

Fifty patients were contacted and invited to participate in the study. However, 3 were unable to attend the health center and 1 did not complete the assessment. Therefore, 46 participants (24 males) were enrolled in the study. On average, participants were 75.9±7.1 years old, with a mean BMI of 28.4±4.7kg/m^2. 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 summarized in Table 1.

All balance tests were able to significantly differentiate between participants with and without **a** history of falls ($p<0.01$) (Table 1). The largest effect sizes were found for the BBS (d=1.02) and for the Brief-BESTest (d=1.01). The effect sizes for the BESTest and for the Mini-BESTest were also large (d=0.87 and d=0.81). 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).

(table 1)

Validity

All balance tests were strongly correlated **with each other**, with *rho* ranging from 0.73 to 0.90 (p<0.001). **The ABC Scale** was significantly correlated with the BBS (*rho*=0.75), BESTest (*rho*=0.61), Mini-BESTest (*rho*=0.55) and **the** Brief-BESTest (*rho*=0.53) (p<0.001) (Figure 1).

(Figure 1)

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 **the** Brief-BESTest. Excellent interrater relative reliability was observed **for** all balance tests (**ICCs>0.85**). Good interrater agreement was verified for all four balance tests, with mean differences **close to zero** (Table 2).

(table 2)

The BBS had moderate to good relative intrarater reliability (ICC=0.52), while the other balance tests had excellent reliability (**ICCs=0.82-0.87)** (Table 2). Bland-Altman plots revealed no systematic bias**, with mean differences ranging from -0.7 to 0.7** (Figure 2).

(Figure 2)

Minimal detectable change

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 the 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/good **ability of all four balance tests to**

identify fall status. The higher AUCs were found for the BBS (AUC=0.84; 95%CI**=0.72-0.96)** and for the Brief-BESTest (AUC=0.78; 95%CI**=0.64-0.92)** (Table 3). The **sensitivity of the Brief-BESTest** (81%) was 8%, 13% and 17% higher **than 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 positive (**LR+=**3 and **LR+=**3.20) and **the** lower negative (**LR-=**0.25 and **LR-=**0.35) likelihood ratios (Table 3).

(table 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 of 52.5 (**sensitivity=73%; specificity=77%)** for the BBS **were identified** (Figure 3).

(Figure 3)

Discussion

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

Findings showed that among the **four** 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.**[11,37] Thus, **caution should be taken when selecting BBS to assess balance in patients with COPD who have mild balance dysfunction (e.g., score on balance clinical measures worse than 1 standard deviation from the mean score published for healthy older people)[47], as it may not be able to detect meaningful changes. In these**

specific cases, the use of the Brief-BESTest, the BESTest or of the Mini-BESTest may be recommended.

The four balance tests were significantly associated with each other and with the ABC Scale, **demonstrating** good concurrent and convergent validity. These findings are in agreement with studies conducted in other specific populations.[22,36,50]

Balance tests presented high interrater relative reliability (**ICCs>0.8**), however, slightly lower ICCs were found for intrarater relative reliability (**ICCs>0.5**). It is common **to find lower intrarater than interrater reliability**.[13,22] However, while the high interrater reliability values were in accordance with previous findings in other populations, the values found for intrarater reliability were not (**ICCs=0.88-0.96[13,22]**). This may be related with 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.[51] As most daily life tasks involve dynamic balance, dyspnea may have played a role in patients' performance **during the** two 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 relibility of the Timed Up and Go in patients with advanced COPD.[52] In terms of absolute reliability, no systematic bias was found for interrater nor intrarater reliability and, thus**, it seems that** clinicians can be confident in using these four 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[12,22,53,54]), BESTest (range 6.2-6.9[26,50,55]) and Mini-BESTest (range 2.4-3.7[22,36,50,55]). **For the Brief-BESTest, the MDC found was slightly higher compared to the MDCs established for older cancer survivors (MDC=2.6 points)[50] and patients submitted to total knee arthroplasty (MDC=3.2 points).[55]** These differences may be population-specific, but may also be related with **the samples used**. In the present study, participants' mean age was 76 years old and 52% of them were male. **In the reported studies,** the mean ages were between 68[50] and 69[55] **years old and most participants were female** (71%[50] and 74%[55]). The MDCs determined are acceptable,[45] 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 four 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-.** These cutoff **points** were similar to those reported in other populations for the **BBS (52 points[37])** and for the Brief-BESTest (11 points[16]). **However,** when adding the information of the celling effect and of the reliability, the Brief-

BESTest had the best performance. These results are important for clinical practice since 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 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) and** 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,[56] exercise capacity,[1] gait speed[4] and physical activity levels.[38] These impairments may have also contributed to the balance deficits and risk of falling found in 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 patients' performance in 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 of 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] has used a sample size of 46 to estimate a LR- of 0.13, with 90% specificity and 80% specificity.[57]** Moreover, as there are false positives and false negatives **in** all four 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 non-fallers. Finally, as this was a cross-sectional study, the ability of the balance tests to identify fallers among patients with COPD was only possible to be analyzed retrospectively. Longitudinal studies should be conducted in order to assess the **prospective** ability of these tests in identifying recurrent fallers.

The BBS, BESTest, Mini-BESTest, and **the** 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.

Acknowledgements

The authors would like to 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.

The authors have no conflicts of interest to declare.

Funding support: None.

References

1. Fleg JL, Morrell CH, Bos AG, Brant LJ, et al. Accelerated Longitudinal Decline of Aerobic Capacity in Healthy Older Adults. *Circulation.* 2005;112(5):674-682.

2. Barnes PJ, Celli BR. Systemic manifestations and comorbidities of COPD. *Eur Respir J.* 2009;33(5):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. *Deutsches Ärzteblatt International.* 2010;107(17):306-316.

5. Ilgin D, Ozalevli S, Kilinc O, Sevinc C, et al. Gait speed as a functional capacity indicator in patients with chronic obstructive pulmonary disease. *Annals of Thoracic Medicine.* 2011;6(3):141-146.

6. Cruz J, Marques A, Jácome C, Gabriel R, et al. Global Functioning of COPD Patients With and Without Functional Balance Impairment: An Exploratory Analysis Based on the ICF Framework. *COPD.* 2015;12(2):207-216.

7. Beauchamp MK, Sibley KM, Lakhani B, Romano J, et al. Impairments in systems underlying control of balance in COPD. *Chest.* 2012;141(6):1496-1503.

8. Roig M, Eng JJ, MacIntyre DL, Road JD, et al. Falls in people with chronic obstructive pulmonary disease: an observational cohort study. *Respir Med.* 2011;105(3):461-469.

9. Marques A, Cruz J, Quina S, Regêncio M, et al. Reliability and Minimal Detectable Change of the Timed Up & Go and the 10-Meter Walk tests in Patients with COPD. *COPD.* 2015;Dec 7 [Epub ahead of print]:1-9.

10. Beauchamp MK, Hill K, Goldstein RS, Janaudis-Ferreira T, et al. Impairments in balance discriminate fallers from non-fallers in COPD. *Respir Med.* 2009;103(12):1885-1891.

11. Mao HF, Hsueh IP, Tang PF, Sheu CF, et al. Analysis and comparison of the psychometric properties of three balance measures for stroke patients. *Stroke.* 2002;33(4):1022-1027.

12. Flansbjer UB, Blom J, Brogardh 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(3):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(2):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(6):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(9):896-903.

16. Duncan RP, Leddy AL, Cavanaugh JT, Dibble LE, 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(4):542-550.

17. Beauchamp MK, Janaudis-Ferreira T, Parreira V, Romano JM, et al. A randomized controlled trial of balance training during pulmonary rehabilitation for individuals with COPD. *Chest.* 2013;144(6):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(2):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(1):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(9):1197-1207.

21. Kottner J, Audigé L, Brorson S, Donner A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106.

22. Godi M, Franchignoni F, Caligari M, Giordano A, 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(2):158-167.

23. Vestbo J, Hurd SS, Agustí AG, Jones PW, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med.* 2013;187(4):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(8):631-636.

25. Zhang F, Ferrucci L, Culham E, Metter EJ, 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(3):478-492.

26. Doherty DE, Belfer MH, Brunton SA, Fromer L, et al. Chronic Obstructive Pulmonary Disease: Consensus Recommendations for Early Diagnosis and Treatment. *J Fam Pract.* 2006;Supplement:S1-S8.

27. Branco PS. Validation of the Portuguese Version of the “Activities-specific Balance Confidence Scale”. *RSPMFR.* 2010;19(2):20-25.

28. Miller MR, Hankinson J, Brusasco V, Burgos F, et al. Standardisation of spirometry. *Eur Respir J.* 2005;26(2):319-338.

29. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21(9):1331-1335.

30. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17(1):101-110.

31. Fischer MJ, Scharloo M, Abbink JJ, van 't Hul AJ, et al. Drop-out and attendance in pulmonary rehabilitation: the role of clinical and psychosocial variables. *Respir Med.* 2009;103(10):1564-1571.

32. Garrod R, Marshall J, Barley E, Jones PW. Predictors of success and failure in pulmonary rehabilitation. *Eur Respir J.* 2006;27(4):788-794.

33. Cohen HS, Kimball KT. Usefulness of some current balance tests for identifying individuals with disequilibrium due to vestibular impairments. *J Vestib Res.* 2008;18(5-6):295-303.

34. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest) to Differentiate Balance Deficits. *Phys Ther.* 2009;89(5):484-498.

35. Franchignoni F, Horak F, Godi M, Nardone A, et al. Using psychometric techniques to improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med.* 2010;42(4):323-331.

36. 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(8):1102-1115.

37. King LA, Priest KC, Salarian A, Pierce D, et al. Comparing the Mini-BESTest with the Berg Balance Scale to Evaluate Balance Disorders in Parkinson's Disease. *Parkinsons Dis.* 2012;2012:375419.

38. Troiano RP, Berrigan D, Dodd KW, Masse LC, et al. Physical activity in the United States measured by accelerometer. *Med Sci Sports Exerc.* 2008;40(1):181-188.

39. 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(2):175-191.

40. Cohen J. *Statistical power analysis for the behavioral sciences.* 2 ed. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

41. Bulmer MG. *Principles of statistics.* New York: Dover Publications; 1979.

42. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil.* 1998;12(3):187-199.

43. Fleiss J. Reliability of measurements. In: Fleiss J, ed. *The design and analysis of clinical experiments*. 1 ed. New York: John Wiley & Sons; 1986.

44. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-240.

45. Huang S-LH, Ching-Lin; Wu, Ruey-Meei; Tai, Chun-Hwei; Lin, Chin-Hsien; Lu, Wen-Shian. Minimal Detectable Change of the Timed "Up & Go" Test and the Dynamic Gait Index in People With Parkinson Disease. *Phys Ther.* 2011;91(1):114-121.

46. 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(3):837-845.

47. Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med.* 2000;45(1-2):23-41.

48. Hosmer DW, Lemeshow S. *Applied logistic regression.* 2nd ed. New York: Wiley; 2000.

49. Deeks JJ, Altman DG. Diagnostic tests 4: likelihood ratios. *BMJ.* 2004;329(7458):168-169.

50. Huang MH, Miller K, Smith K, Fredrickson 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.* 2015;Feb 18 [Epub ahead of print].

51. Kessler R, Partridge MR, Miravitlles M, Cazzola M, et al. Symptom variability in patients with severe COPD: a pan-European cross-sectional study. *Eur Respir J.* 2011;37(2):264-272.

52. Mesquita R, Janssen DJA, Wouters EFM, Schols JMGA, 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(11):2131-2138.

53. 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(7):1201-1208.

54. 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(5):343-346.

55. Chan AC, Pang MY. Assessing Balance Function in Patients With Total Knee Arthroplasty. *Phys Ther.* 2015;95(10):1397-1407.

56. 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(7):1121-1129.

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

58. Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, 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;Suppl 16(6):5-40.

1

Table 1 - Participants' characteristics (n=46).

Characteristics	**Total (n=46)**	**Without a history of falls (n=23)**	**With a history of falls (n=23)**	**p-value**
Age (years)	75.9 (7.1)	74.6 (5.9)	77.2 (8)	0.21
Gender				
Male	24 (52.2%)	14 (60.9%)	10 (43.5%)	0.38
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)	0.91
mMRC, M [IQR]	2 [1, 2]	2 [1, 2]	2 [1, 3]	0.28
Exacerbations in the previous year				
0	**28 (60.9%)**	**16 (69.6%)**	**12 (52.2%)**	**0.18**
> 1	**18 (39.1%)**	**7 (30.4%)**	**11 (47.8%)**	
Comorbidities, M [IQR]	**2 [1, 3]**	**2 [0, 3]**	**2 [1, 3.75]**	**0.40**
FEV_1 (% predicted[58])	69.4 (19.9)	68.8 (21)	70.1 (19.2)	0.83
GOLD spirometric classification				
Mild	13 (28.3%)	8 (34.8%)	5 (21.7%)	0.55
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)	<0.001

BBS	50.1 (5.5)	53.3 (4.3)	48.3 (5.4)	≤0.001
BESTest	77.8 (12.5)	82.8 (11.4)	72.7 (11.7)	0.01
Mini-BESTest	20.8 (4.9)	22.6 (4.4)	18.9 (4.7)	0.01
Brief-BESTest	15.7 (4.9)	18.0 (4.2)	13.5 (4.7)	0.01

Note: values shown as mean (SD) or n(%), unless otherwise indicated. Abbreviations: ABC, Activities-Specific Balance Confidence; BBS, Berg Balance Scale, BESTest, Balance Evaluation Systems Test; BMI, body mass index; FEV_1, forced expiratory volume in one second; GOLD, Global Initiative for Chronic Obstructive Lung Disease, IQR, interquartile range; M, median; mMRC, modified Medical Research Council dyspnea scale.

Table 2 – Interrater and intrarater reliability of the Berg Balance Scale (BBS), Balance Evaluation Systems Test (BESTest), Mini-BESTest and **the** Brief-BESTest (n=28).

	Interrater reliability			**Intrarater reliability**			
Balance test	**$ICC_{2,1}$ (95% CI)**		**Mean difference (SD)**	**95% LA**	**$ICC_{2,1}$ (95% CI)**		**Mean difference (SD)**	**95% LA**
BBS	0.94	(0.88-0.97)	0.5 (1.6)	-2.6-3.6	0.52	(0.19-0.74)	-0.7 (2.9)	-6.3-4.9
BESTest	0.85	(0.70-0.92)	-1.2 (3.8)	-8.6-6.2	0.87	(0.73-0.94)	-0.5 (3.7)	-7.7-6.8
Mini-BESTest	0.85	(0.71-0.93)	-0.7 (2.1)	-4.7-3.3	0.88	(0.75-0.94)	0 (1.7)	-3.4-3.4
Brief-BESTest	0.97	(0.94-0.99)	-0.1 (1.0)	-2.1-2.0	0.82	(0.66-0.92)	-0.7 (2.5)	-5.6-4.2

Abbreviations: 95% CI, 95% confidence intervals; ICC, Intraclass Correlation Coefficient; 95% LA, 95% limits of agreement; SD, standard deviation.

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).

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

Abbreviations: AUC, area under the curve; 95%CI, 95% confidence intervals; SEM, standard error.

Figure legends

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).

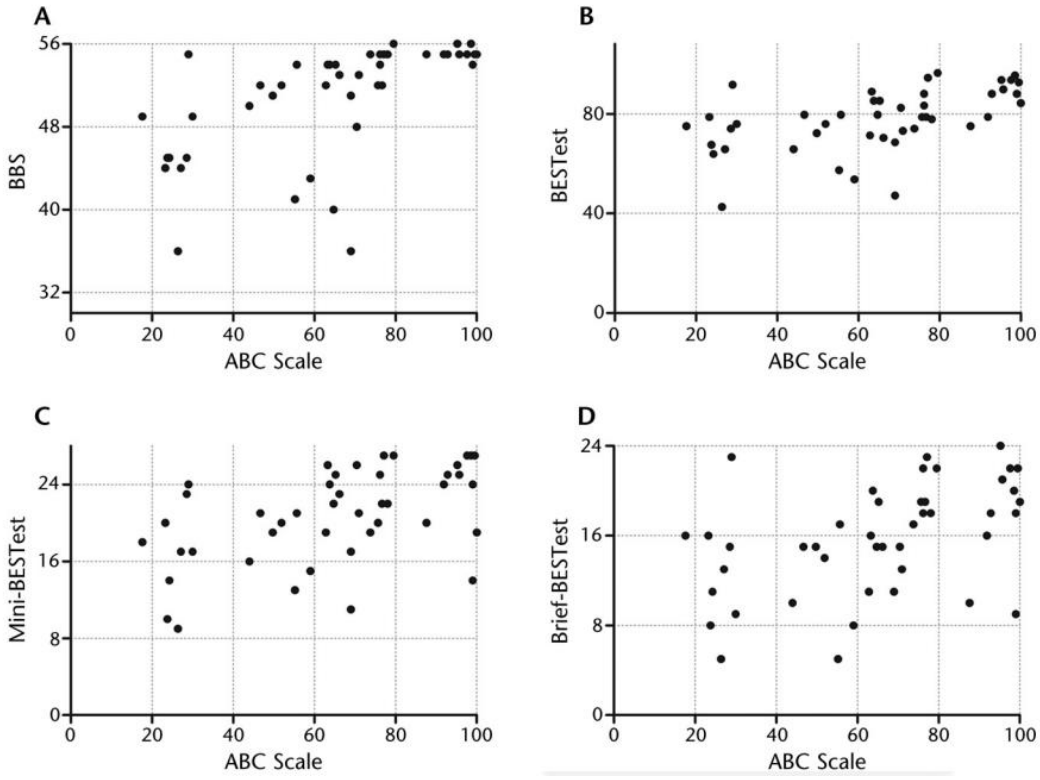

Figure 2 - Bland and Altman plots of the (A) Berg Balance Scale (BBS), (B) Balance Evaluation Systems Test (BESTest), (C) Mini-BESTest and (D) Brief-BESTest between two sessions (n=28). The bold line represents the mean difference between sessions 1 and 2 and the dotted lines the 95% limits of agreement.

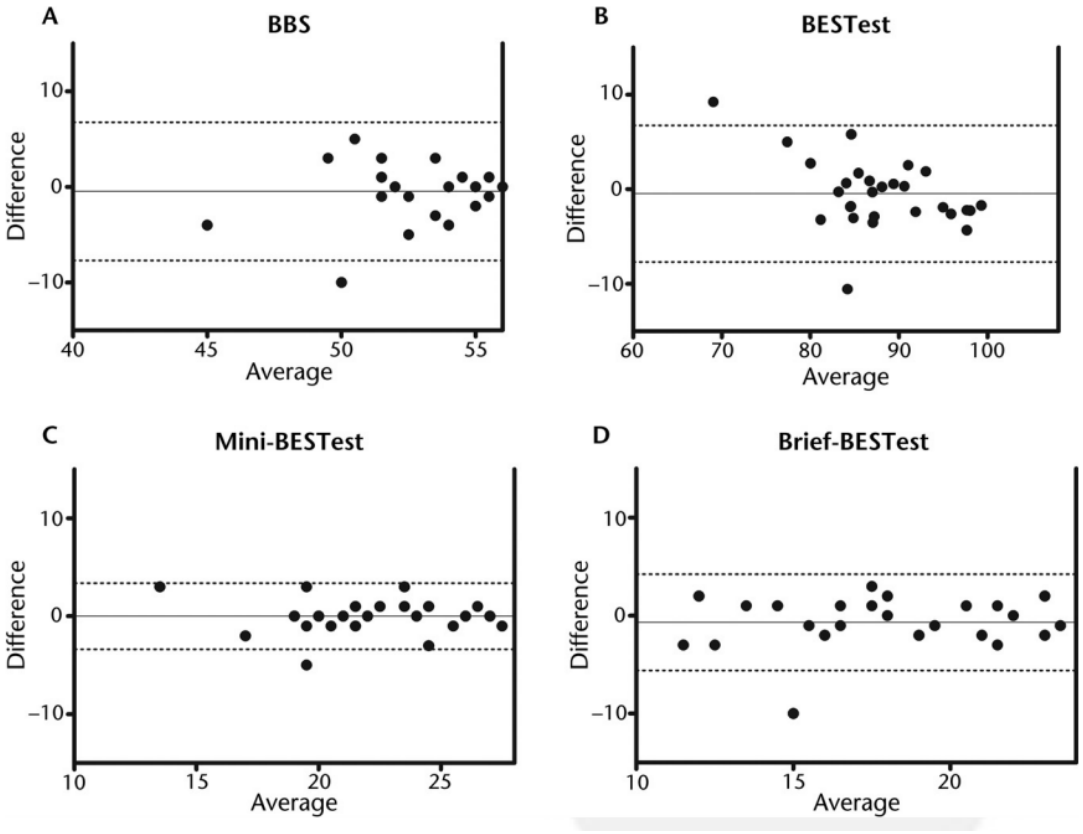

Figure 3 – Receiver operator characteristics (ROC) of the Berg Balance Scale (BBS) and the Brief-Balance Evaluation Systems Test (Brief-BESTest) to **differentiate** between participants with and without **a** history of falls. The points corresponding to cutoff points are indicated by arrows.