Minimal clinically important difference of four commonly used balance assessment tools in individuals after total knee arthroplasty: A prospective cohort study

Abstract

Background: Although balance is commonly assessed during the recovery of total knee arthroplasty (TKA), the minimal clinically important difference (MCID) of commonly used balance assessment tools has not been previously established in this population.

Objective: To determine the MCID of four balance tests (i.e., the Balance Evaluation Systems Test (BESTest), Mini-BESTest, Brief-BESTest, and the Berg Balance Scale (BBS) in individuals post-TKA).

Design: Prospective cohort

Setting: Outpatient rehabilitation

Population: Inclusion criteria: 1) first primary TKA with diagnosed knee osteoarthritis; 2) aged 50-85 years. Exclusion criteria: 1) TKA due to rheumatoid arthritis of the knee or traumatic injury; 2) known medical conditions that influence balance ability. 146 participants were recruited, and 134 of them with complete data were included in the analysis.

Interventions: Participants received individualized physiotherapy, consisting of electrotherapy for pain and edema control, mobilization and strengthening exercises, gait and balance training, once or twice per week between assessments.

Main Outcome Measures: Participants were assessed on the BESTest, Mini-BESTest, Brief-BESTest, BBS, and Functional Gait Assessment (FGA) 2 and 4 weeks after surgery. The FGA was used as the anchor reference measure to calculate the MCID of the other four balance tests.

A distribution-based approach was also employed to derive the MCID (i.e., standardized effect size of 0.5).

Results: The BESTest (area under curve (AUC)=0.811, 95%CI: 0.739-0.883) had the highest accuracy in detecting clinically important improvements on the FGA (≥4 points), followed by the Mini-BESTest *(*AUC=0.782, 95%CI:0.704-0.860), Brief-BESTest (AUC=0.701, 95%CI:0.618-0.795), and BBS (AUC=0.586, 95%CI:0.490-0.682). The anchor- and distribution-based MCIDs were: 6-8 for the BESTest, 1-2 for the Mini-BESTest, and 2-3 for the Brief-BESTest.

Conclusion: Improvements exceeding MCIDs established above are indicative of significant progress in balance function post-TKA. The BBS is not a recommended tool due to its low AUC value.

Key words: knee osteoarthritis, rehabilitation, treatment effectiveness, balance.

Introduction

Total Knee Arthroplasty (TKA) is a commonly performed orthopedic procedure for severe end-stage knee osteoarthritis. In the United States, over 400,000 TKA procedures are performed annually,[1] and this number is predicted to increase with the aging population to more than 3.4 million cases by 2030.[2] Given that 74 % of TKA surgeries are performed on individuals over the age of 65,[3] and that 17-40 % of individuals post-TKA experience a fall within 6 months to 1 year following surgery,[4-7] research dedicated to reducing balance problems and fall-risk in this population is important. Targeted interventions have been employed to mitigate balance issues,[8,9] however clinically meaningful differences in commonly used balance assessment tools has not been previously established in individuals post-TKA.

Minimal clinically important difference (MCID), and other measures of clinical importance, are recognized as critical considerations for interpreting the efficacy of changes observed following an intervention. They extend beyond statistical significance and provide an interpretation of clinical meaningfulness by supplying information on the smallest difference in a clinical balance score that would lead a clinician to consider a change in treatment.[10] Previous work has described best practices for calculating the MCID, and recommended an anchor-based method, whereby relevant patient-rated, clinician-rated, and disease-specific variables offer meaningful estimates of an instrument's clinical importance.[11] Support of the anchor-based method can also be afforded by clinical indicators through various distribution-based estimates (e.g., the effect size) in order to derive a single value or range of values for the MCID.[11-14] Importantly, the MCID of commonly used clinical balance assessment tools, such as the various versions of the Balance Evaluation Systems Test (BESTest) and the Berg Balance Scale (BBS) has not been previously established among individuals post-TKA. Furthermore, a reduced

sensory orientation score on the BESTtest has been associated with falls in individuals with TKA.[7] Improving best practices in balance assessment and providing clinicians with insight on interpreting changes in performance is critical for optimal recovery post-TKA. Accordingly, the overarching purpose of this study was to determine the MCID of four balance tests (i.e., the three versions of the BESTest, as well as the BBS) in individuals with TKA.

Methods

These data are part of a larger study described elsewhere.[7,15] We have followed the Strengthening The Reporting of Observational studies in Epidemiology (STROBE) checklist.[16]

Participants. Participants were referred from out-patient physiotherapy rehabilitation at a local hospital between February 2013 and January 2014. A cemented cruciate-substituting prosthetic implant was used for the TKA surgery (Zimmer model: NexGen LPS-Flex, UK). The inclusion criteria consisted of: 1) a first primary TKA with diagnosed knee osteoarthritis; 2) aged 50-85 years; 3) the ability to follow verbal instructions; and 4) the ability to provide informed consent. Exclusion criteria were: 1) a previous operation on the lower limb; 2) TKA due to rheumatoid arthritis of the knee or traumatic injury; and 3) known medical conditions that influence balance ability (e.g., Parkinson's disease). Ethical approval was obtained by the Human Research Ethics Subcommittee of the involved university and hospital. Experimental procedures were in accordance with the Declaration of Helsinki. All participants provided written informed consent.

Sample Size Estimation. An a priori sample size calculation was conducted for individuals with TKA across two assessment time points (MedCalc.ink, Belgium). The null area under curve

(AUC) value was set to 0.700 as it was considered to represent an acceptable discrimination. The effect size was based on a previous study examining the ability of the BESTest, Mini-BESTest, and BBS to discriminate between fallers and non-fallers (AUC=0.840).[17] In order to detect a significant difference in balance scores between two given time points, with a power of 0.90, an alpha of 0.05, and a 10 % attrition rate, a minimum sample of 132 participants was required.

Procedures. The baseline assessment and the first physiotherapy session took place 2 weeks after the TKA surgery in order to provide adequate time for removal of the knee staples. The second assessment took place 2 weeks after the baseline assessment (i.e., 4 weeks after TKA), as most rehabilitation typically takes place during the first month post-surgery, therefore establishing MCID ranges at this time point would be meaningful to clinicians and researchers.[18] The outpatient physiotherapy treatment consisted of individualized one-on-one electrotherapy for pain and edema control, mobilizing and strengthening exercises, as well as gait and balance training, and was provided once or twice per week between the two assessments. Participants were also provided with a brochure post-TKA surgery outlining joint care and home exercises.

Demographic information was collected from medical records and from the patient interview. The balance assessments were performed independently by 1 of the 3 raters who each had more than 10 years of clinical experience. The therapists worked at the same clinic, met at the beginning of the study to ensure consistent scoring, and consulted with each other at weekly meetings to ensure consistency. Participants performed the BESTest, BBS and Functional Gait Assessment (FGA) at both assessments. The test items in the Mini-BESTest and Brief-BESTest were extracted from the items from the full BESTest. Both the rater and sequence of balance assessments were randomized. Importantly, the 3 BESTests, the BBS and the FGA have

exhibited excellent interrater reliability (ICC: 0.96-0.99), intrarater reliability (ICC: 0.92-0.97), concurrent validity (*r*: 0.67-0.93), convergent validity (*r*: 0.34-0.48), and internal consistency (Cronbach Alpha: 0.96-0.98) in individuals with TKA.[19]

The 36-item BESTest evaluates 6 systems that contribute to balance dysfunction including: biomechanical constraints, stability limits, postural responses, anticipatory postural adjustments, sensory orientation, dynamic balance during gait, and cognitive effect.[20] Each item was scored on a 4-level ordinal scale from 0 to 3 (0: severely impaired balance or inability to complete a task; 3 no impairment of balance or ability to perform a task successfully).[21] The 16-item Mini-BESTest and the 8-item Brief-BESTest have been derived from the BESTest for time efficiency purposes.[21] Unlike the BESTest, the Mini-BESTest was scored on a 3-level ordinal scale from 0 (severely impaired balance) to 2 (i.e., no balance impairment).[21] Based on the patient's performance on the BESTest, the therapist provided the rating according to the specific scoring criteria of the Mini-BESTest. The scoring method of the Brief-BESTest was identical to the scoring of the BESTest.[22] The BBS is a 14-item functionally-oriented test, and was scored on a scale of 0 to 4 (0: severe impairment of balance or inability to perform; 4: no impairment in balance).[23]

The 10-item FGA was chosen as the anchor reference measure, as walking is not only an important activity of daily living, but it has also been identified as the most common activity during which falls occur in individuals with TKA.[7] Previous work has used the FGA as an anchor measure to calculate the responsiveness among the BESTtest, Mini-BESTest, Brief-BESTtest, and the BBS in individuals with TKA.[15] Moreover, the reference measure should have a nontrivial relationship with the outcome measures,[24] and given that a moderate relationship observed between the FGA and BBS in other patient populations, such as Parkinson's disease,[24]

the FGA was considered to be an appropriate anchor measure. The FGA was therefore used to calculate the MCID for the BESTest, Mini-BESTest, Brief-BESTest, and BBS. The FGA was scored on an ordinal scale from 0 to 3 (0: severe impairment of balance or inability to perform; 3: no impairment in balance).[25]

Balance confidence as measured by the Activities-Specific Balance Confidence Scale-16, pain as measured by the Numeric Pain Rating Scale-11, knee joint proprioception, knee range of motion, and knee muscle strength were also tested 2 weeks and 4 weeks after TKA, and have been described in detail elsewhere.[7]

Data Analysis. In order to ensure that the association between the reference measure and balance tests was nontrivial,[26] a spearman's *r* statistic was computed. Wilcoxon signed-ranks test were used to compare the FGA, BESTest, Mini-BESTest and Brief-BESTest scores between baseline and follow-up as normal distribution assumption were not fulfilled. An anchor-based approach was first used to establish the MCID score of the different balance tests. A MCID of 4 points for the FGA was previously identified in the elderly population,[27] which was then used as an anchor. Using the receiver operator characteristic (ROC) curve analysis, the Youden's index was used to identify the optimal cutoff score of the five balance tests to separate individuals who achieved a FGA change score ≥4 from those with a change score <4. The AUC was used to indicate the discrimination accuracy of the different balance tests (outstanding discrimination: AUC>0.900, excellent discrimination: AUC=0.800-0.900; acceptable discrimination: AUC=0.700-0.800; poor discrimination: AUC=0.700-0.800).[28] A *z* statistic was computed to compare whether significant differences in the AUC existed between the different balance tests (α<0.05; MedCalc.ink, Ostend, Belgium).

The MCID should be interpreted with reference to its variability.[29] Although the distribution-based approach cannot offer an accurate and direct indication of the magnitude of change, it can provide a suggested MCID range, thereby affording further insight into a patient-centered interpretation of meaningful change.[11] Therefore, the distribution-based approach was also used to determine the MCIDs. A standardized effect size of 0.5 was considered to represent a clinically important change, as recommended by previous studies.[9,30,31] This was calculated using the following formula: $0.5 \times SD_{pooled,}$ where the pooled SD represented $\sqrt{[(SD_{pre\text{-}treatment})^2 + (SD_{post\text{-}treatment})^2 / 2]}$.[32]

Next, we deduced a MCID range for each balance measure using both anchor- and distribution-based approaches.[11-13,32] The MCID range was formed from the smaller to larger values of these two approaches. A relative MCID range was calculated by dividing the MCID range by the total score of each balance test.[12]

Results

Initially, 146 participants were recruited, and 12 participants withdrew due to an inability to commit (n=9), a relocation to another city (n=2), and a worsening of health (n=1), leaving 134 participants with complete data. Participant characteristics are reported in Table 1. No walking aids were used during testing. The mean FGA scores and the five balance test scores at the follow-up increased relative to baseline (p<0.001: Table 1; p<0.01: Table 2). According to the ROC curve analysis, the BESTest (AUC=0.811, 95% CI: 0.739-0.883) demonstrated highest discrimination accuracy for minimal clinically important change, followed by Mini-BESTest (AUC=0.782, 95% CI: 0.704-0.860) and Brief-BESTest (AUC=0.706, 95% CI: 0.618-0.795). The discrimination ability of the BBS (AUC=0.586, 95% CI: 0.490-0.682) was the worst.

Further, the AUC of the BBS was significantly smaller than the BESTest (95% CI: 0.123-0.328, $p<0.01$), Mini-BESTest (95% CI: 0.089-0.304, $p<0.01$), and Brief-BESTest (95% CI: 0.011-0.230, $p=0.03$). The AUC of BESTest was also significantly higher than the Brief-BESTest (95% CI: 0.037-0.172, $p<0.01$; Table 3 and Figure 1). Based on the Youden's Index, the optimal cutoff scores to accurately detect the individuals who attained a FGA change score ≥4 from those who did not (MCID) were determined for the four balance tests (i.e., anchor-based approach; Table 4). Fifty-four percent of participants ($n=72$) exhibited an FGA change score ≥4 from week 2 to week 4. The MCID values derived from the distribution-based approach (i.e., standardized effect size of 0.5) were also calculated. The MCID values obtained from the above two approaches were then used to infer MCID ranges and are also shown in Table 4. The MCID range of the BBS was not presented because its MCID value derived from the anchor-based approach was not useful due to the low AUC value (0.586).

Discussion

This is the first study to establish MCID values for the BESTest, Mini-BESTest, and Brief-BESTest in individuals with TKA. Based on the anchor-based approach, the BESTest demonstrated excellent discrimination (AUC=0.811), whereas the Mini-BESTest and Brief-BESTest showed acceptable discrimination (AUC=0.782 and 0.706, respectively) to accurately identify those who attained a clinically important change in the FGA score (≥4 points). The inclusion of a variety of balance systems and more challenging tasks in the three BESTests (e.g., hip and trunk lateral strength, and postural responses to external perturbations) may have accounted for the higher AUC values. In particular, the "stability in gait" domain is assessed in

all three BESTests, which explains its high correlation with FGA, which also evaluates the ability to maintain equilibrium in performing dynamic gait activities.

On the other hand, the BBS exhibited the poorest discrimination (AUC=0.586) compared to the other three BESTests. As a consequence, the MCID value of BBS derived from the anchor-based approach may not be useful due to its lower similarity to the FGA. Perhaps the BBS may not accurately predict balance performance improvements because the majority of BBS items involve relatively simple balance tasks. This may explain the ceiling effect in performance and attenuated variability. More specifically, the BBS measures static balance and transfers, and does not include dynamic balance. The absence of these tasks may render the BBS inadequate at predicting performance in dynamic activities, including those assessed in the FGA.

Among the four balance tests, the BBS also demonstrated the lowest relative MCID value (3.2 %) when the distribution-based approach was employed. It is possible that a ceiling effect in the BBS emerged as the baseline scores may have already been high, as 99.3 % of participants scored in the top 20 % on the BBS at baseline (i.e., a score equal to, or greater than 45/56); thus, most of the participants had only small changes in BBS score during the follow-up period, resulting in a small MCID value derived from the distribution-based approach. Ceiling effects in the BBS have been observed in various populations,[33,34] as well as less internal and external responsiveness in the TKA population relative to each version of the BESTest.[15] According to the distribution-based values, the BESTest and the Mini-BESTest exhibited smaller relative MCID values compared to the Brief-BESTest. A similar pattern of relative MCID values among these tests has also been shown in older adults living in the community.[35] The wide distribution and large variability in the change score of the Brief-BESTest may indicate that it has better ability in discriminating individuals with different degrees of balance recovery relative to the

Mini-BESTest and BESTest. Overall, although the BBS is commonly administered, based on our findings it is not a recommended balance assessment tool in individuals post-TKA, especially at later stages of rehabilitation, as it may not accurately detect improvements in balance recovery.

Similar anchor- and distribution-based strategies have been used to calculate the MCID for the various versions of the BESTest and BBS in a variety of populations.[36-38] For example, previous research has reported an anchor-based MCID range of 10.2 to 17.4 points for the BESTest in individuals with chronic obstructive pulmonary disease,[36] while other work has reported a MCID of 4 points for the Mini-BESTest in individuals with neurological disorders.[38] Additionally, a MCID of 3 points has been produced for the BBS in individuals with multiple sclerosis,[37] 7 points in individuals with neurological disorders,[38] and 3.5 to 7.1 points in individuals with chronic obstructive pulmonary disease.[36] Other work has established the minimal detectable change (MDC) for the various versions of the BESTest and BBS and found higher MDC values for the Mini-BESTest and Brief-BESTest than the MCID estimates calculated in the current study.[39,40] Theoretically, the MCID should be greater than the MDC, as the MDC is a threshold for measurement error,[41] while the MCID is the smallest clinically meaningful change in score and related to a beneficial change in health status perceived by the patient.[11,42] One reason for the discrepancy in MDC and MCID estimates may stem from changes in other constructs as a function of rehabilitation, such as sensory or cognitive, that may differentially contribute to modifications in disability level across patient groups.[27] Another reason could be younger and/or less disabled populations may have higher expectations related to their outcomes.[27] Because values for MDC and MCID can vary from patient group to patient group,[43,44] clinically meaningful changes should be calculated for each population as they are not

likely to be generalizable across populations.[11] Our MCID ranges are only estimates; thus further research is needed to establish MCID ranges across patient groups.

The current study incorporates a performance-based balance improvement (i.e., FGA) and a distribution-based measure (i.e., a standardized effect size of 0.5) to identify a MCID range, thereby offering critical information to clinicians assessing individuals post-TKA. Notably, improvements above 6 to 8 points in the BESTest, 1 to 2 points in the Mini-BESTest, and 2 to 3 points in the Brief-BESTest are suggestive of functional gains post-TKA. Researchers and clinicians should consider improvements in power, range of motion, pain, and balance confidence when interpreting meaningful improvements in balance (Table 1). Because the BBS showed an unacceptable discrimination level as revealed by the AUC value, it is not a recommended balance assessment tool for assessing recovery in balance function among individuals post-TKA during the later stages of rehabilitation (2 weeks post-TKA and onward).

The various versions of the BESTest have a wealth of online resources, such as training videos, a list of equipment required, and multiple translated versions (www.bestest.us). Other systems important for balance control may be missing from these tests and the full BESTest is time consuming to administer, which may be potential barriers.[45] Nevertheless, the various versions of the BESTest are easy to administer and has are valid and reliable tools in individuals with TKA.[19]

Limitations. This study has several limitations. The MCID range is dependent on the anchor applied, the diagnosis, and baseline patient status. It is possible that patients were not rated by the same examiner at their second visit and the examiners were not blinded to the visit number; however, the balance tests have shown excellent test-retest reliability, and the clinicians regularly

consulted with each other, which may decrease erroneous data from subjective scoring. Extracting the Mini-BESTest and Brief-BESTest scores from the BESTest may have led to different scores than if they were all performed and rated separately. Participants received 1-2 physical therapy sessions per week, and this variability could have contributed to differences in follow-up scores. Our findings are only generalizable to individuals with knee osteoarthritis between week 2 and 4 after TKA surgery with no medical conditions that may affect balance (e.g., Parkinson's disease). In order to develop a more comprehensive MCID range, future work should consider including a self-perceived change scale (e.g., Global Rating Scale of Change), as well as longer follow-up periods. Lastly, we did not include a comparison group who did not receive treatment.

Conclusion

This study has established the MCID values for the BESTest, Mini-BESTest, and Brief-BESTest for individuals with TKA. The BBS is not a recommended balance assessment tool as it exhibited low sensitivity and specificity to detect changes in balance function. Altogether, these findings may be used to form a basis for clinicians to interpret whether changes in commonly used balance assessment tools have reached a clinically important threshold in response to rehabilitation.

Conflict of Interest Statement

There is no conflict of interest among authors.

References

1. NIH Consensus. Consensus statement on total knee replacement (2003). NIH Consensus and State-of-the-Science Statements 2008;8-10.
2. Kurtz S, Ong K, Lau E et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. *J Bone Joint Surg Am* 2007; 89(4): 780-785.
3. Praemer A, Furner S and Rice DP. Musculoskeletal conditions in the United States. American Academy of Orthopaedic Surgeons 1992.
4. Swinkels A, Newman JH and Allain TJ. A prospective observational study of falling before and after knee replacement surgery. *Age Ageing* 2009; 38(2): 175-181.
5. Levinger P, Menz HB, Morrow AD, et al. Lower limb proprioception deficits persist following knee replacement surgery despite improvements in knee extension strength. Knee Surgery, Sports Traumatology, *Arthroscopy* 2012; 20(6): 1097-1103.
6. Matsumoto H, Okuno M, Nakamura T, et al. Fall incidence and risk factors in patients after total knee arthroplasty. *Arch Orthop Traum Su* 2012; 132(4): 555-563.
7. Chan ACM, Jehu DAM and Pang MYC. Falls after total knee arthroplasty: frequency, circumstances, and associated factors. A prospective cohort study. *Phys Ther* 2018; 98 (9): 767-778.
8. Minns Lowe CJ, Barker KL, Dewey M, et al. Effectiveness of physiotherapy exercise after knee arthroplasty for osteoarthritis: systematic review and meta-analysis of randomised controlled trials. *BMJ* 2007; 335(7624): 812.

9. Piva SR, Gil AB, Almeida GJ, et al. A balance exercise program appears to improve function for patients with total knee arthroplasty: a randomized clinical trial. *Phys Ther* 2010; 90(6): 880.
10. Guyatt GH, Osoba D, Wu AW, et al. Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc* 2002; 77(4): 371-383.
11. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61(2): 102-109.
12. Yost KJ and Eton DT. Combining distribution-and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof* 2005; 28(2): 172-191.
13. Cella D, Eton DT, Lai JS, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive minimal clinically important differences on the Functional Assessment of Cancer Therapy (FACT) anemia and fatigue scales. *J Pain Symptom Manage* 2002; 24(6): 547-561.
14. Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui KK. Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Eval Health Prof* 2005; 28(2): 160-171.
15. Chan AC, Ouyang XH, Jehu DAM, et al. Recovery of balance function among individuals with total knee arthroplasty: Comparison of responsiveness among four balance tests. *Gait Posture* 2018; 59: 267-271.

16. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg* 2014; 12(12): 1495-1499.
17. Yingyongyudha A, Saengsirisuwan V, Panichaporn W, et al. The Mini-Balance Evaluation Systems Test (Mini-BESTest) demonstrates higher accuracy in identifying older adult participants with history of falls than do the BESTest, Berg Balance Scale, or Timed Up and Go Test. *J Geriatr Phys Ther* 2016; 39(2): 64-70.
18. Zech A, Hendrich S and Pfeifer K. Association between exercise therapy dose and functional improvements in the early postoperative phase after hip and knee arthroplasty: An observational study. *PM R* 2015; 7(10): 1064-1072.
19. Chan AC and Pang MY. Assessing balance function in patients with total knee arthroplasty. *Phys Ther* 2015; 95(10): 1397-1407.
20. Horak FB. Postural orientation and equilibrium: what do we need to know about neural control of balance to prevent falls? *Age Ageing* 2006; 35(2): 7-11.
21. 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(4): 323-331.
22. Padgett PK, JV, Jacobs SL and Kasser. 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.
23. Berg KO, Wooddauphinee SL and Williams JI. Measuring balance in the elderly - validation of an instrument. *Can J Public Health* 1992; 83: 7-11.

24. Leddy AL, Crowner BE and 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.
25. Wrisley DM, Marchetti GF, Kuharsky DK, et al. Reliability, internal consistency, and validity of data obtained with the Functional Gait Assessment. *Phys Ther* 2004; 84(10): 906-918.
26. Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and reporting results based on patient-reported outcomes. *Value Health* 2007; 10: 116-124.
27. Beninato M, Fernandes A and Plummer LS. Minimal clinically important difference of the functional gait assessment in older adults. *Phys Ther* 2014; 94(11): 1594-603.
28. Hosmer DW and Lemeshow S. Applied logistic regression. New York, US: Wiley 2000.
29. Portney LG and Watkins MP. Foundations of clinical research: Applications to practice (Third Edition). US: Prentice Hall;2014
30. Norman GR, Sloan JA and Wyrwich KW. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care* 2003; 41(5): 582-592.
31. Osoba D, Bezjak A, Brundage M, et al. Analysis and interpretation of health-related quality-of-life data from clinical trials: basic approach of the National Cancer Institute of Canada Clinical Trials Group. *Eur J Cancer* 2005; 41(2): 280-287.
32. Lin K., Hsieh YW, Wu CY, et al. Minimal detectable change and clinically important difference of the Wolf Motor Function test in stroke patients. *Neurorehabil Neural Repair* 2009; 23(5): 429-434.

33. Tanji H, Gruber-Baldini AL, Anderson KE, et al. A comparative study of physical performance measures in Parkinson's disease. *J Mov Disord 2*008; 23(13): 1897-1905.

34. Tsang CS, Liao LR, Chung RC, et al. Psychometric properties of the Mini-Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling individuals with chronic stroke. *Phys Ther* 2013; 93(8): 1102.

35. Marques A, Almeida S, Carvalho J, et al. Reliability, validity, and ability to identify fall status of the balance evaluation systems test, mini-balance evaluation systems test, and brief-balance evaluation systems test in older people living in the community. *Arch Phys Med Rehabil* 2016; 97(12): 2166-2173.

36. Beauchamp MK, Harrison SL, Goldstein RS, et al. Interpretability of change scores in measures of balance in people with COPD. *Chest* 2016; 149(3): 696-703.

37. Gervasoni E, Jonsdottir J, Montesano A, et al. Minimal clinically important difference of Berg Balance Scale in people with multiple sclerosis. *Arch Phys Med Rehabil* 2017; 98(2): 337-340.

38. 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(2): 158-167.

39. Viveiro LAP, Gomes GCV, Bacha JMR, et al. Reliability, validity, and ability to identity fall status of the Berg Balance Scale, Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest in older adults who live in nursing home. *J Geriatr Phys* Ther 2018; Nov 6. Epub ahead of print.

40. Jácome C, Cruz J, Oliveira A, et al. 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(11): 1807-1815.
41. Beaton DE, Tarasuk V, Katz JN, et al. "Are you better?" A qualitative study of the meaning of recovery. *Arthritis Rheum* 2001; 45: 270–279.
42. Hays RD and Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research (How meaningful is it?). *Pharmacoeconomics* 2000; 18: 419-423.
43. Beninato M, Portney LG. Applying concepts of responsiveness to patient management in neurologic physical therapy. *J Neurol Phys Ther* 2011; 35: 75–81.
44. Liang MH, Lew RA, Stucki G, et al. Measuring clinically important changes with patient-oriented questionnaires. *Med Care* 2002; 40: II45–II51.
45. Horak FB, Wrisley DM and Frank J. The Balance Evaluation Systems Test (BESTest) to differentiate balance deficits. *Phys Ther* 2009; 89(5): 484-498.

Table 1. Participant characteristics 2 weeks (Baseline) and 4 weeks (Follow-Up) post Total Knee Arthroplasty.

Measures	Baseline	Follow-Up	*p*-value
Age (years)	66.3 ± 6.6	NA	NA
Sex % (*n*)			
Male	29.1 (39)	NA	NA
Female	70.9 (95)	NA	NA
Side of Operation % (*n*)			
Left	41.8 (56)	NA	NA
Right	58.2 (78)	NA	NA
Body Mass Index (kg/m^2)	26.7 ± 3.8	NA	NA
FGA (reference measure)	16.0 ± 5.0	20.3 ± 5.7	p<0.001*
BBS	51.8 ± 4.2	53.8 ± 2.7	p<0.001*
BESTest	67.7 ± 10.9	76.0 ± 9.9	p<0.001*
Mini-BESTest	16.4 ± 4.2	19.5 ± 3.7	p<0.001*
Brief BESTest	11.8 ± 4.4	14.7 ± 4.6	p<0.001*
ABC score, 0%–100%	53.9 ± 20.1	65.9 ± 17.1	p<0.001*
Pain intensity measured with NPRS, 0–10	2.3 ± 1.3	1.8 ± 1.1	p<0.001*
Operated knee proprioception, º	1.9 ± 1.3	1.6 ± 1.0	p=0.01*
Operated knee flexion ROM, º	101.0 ± 11.3	108.4 ± 9.8	p<0.001*
Nonoperated knee flexion ROM, º	117.2 ± 12.4	117.6 ± 12.2	p=0.27
Operated knee extension ROM, º	-4.1 ± 10.8	-3.5 ± 5.0	p=0.49
Nonoperated knee extension ROM, º	-1.9 ± 4.9	-1.9 ± 5.1	p=1.00
Operated knee flexion strength, N/kg		1.8 ± 0.5	

Nonoperated knee flexion strength, N/kg	2.2 ± 0.5
Operated knee extension strength, N/kg	2.1 ± 0.7
Nonoperated knee extension strength, N/kg	2.9 ± 0.7

** significant difference between baseline and follow-up*

FGA: Functional Gait Assessment; BBS: Berg Balance Scale; BESTest: Balance Evaluation Systems Test; ABC: Activities-specific Balance Confidence Scale; ROM: Range of Motion; N/kg: Newtons per kilogram

Table 2. Correlations between the FGA and four balance tests.

	Correlation Coefficient	
Balance tests	**Baseline score**	**Change score**
BESTest	0.820*	0.551*
Mini-BESTest	0.816*	0.516*
Brief BESTest	0.755*	0.402*
BBS	0.730*	0.153

** Significantly correlated to the FGA ($p<0.01$).*

Table 3. Area under the curve (AUC) of the four balance tests.

Test Result Variable(s)	AUC	95% Confidence Interval	
		Lower Bound	Upper Bound
BESTest	0.811*†	0.739	0.883
Mini-BESTest	0.782*	0.704	0.860
Brief BESTest	0.706*	0.618	0.795
BBS	0.586	0.490	0.682

** represents significantly greater AUC compared to the BBS ($p<0.05$).*

† represents a significantly greater AUC compared to the Brief BESTest ($p<0.05$)

Table 4. Anchor- and distribution-based MCIDs of the four balance tests.

Balance test (maximum possible score)	**MCID (Anchor-based)**		**MCID (Distribution-based)**		**MCID range (points)**	**MCID range (%)**
	Absolute value (points)	**Relative value (%)**	**Absolute value (points)**	**Relative value (%)**		
BESTest (108)	8	7.5	6	5.2	6 to 8	5.2 to 7.5
Mini-BESTest (28)	2	5.4	1	4.3	1 to 2	4.3 to 5.4
Brief-BESTest (24)	3	10.4	2	9.2	2 to 3	9.2 to 10.4
BBS (56)	5	8.0	2	3.2	NA*	NA*

**the MCID range of the BBS is not presented because its MCID value derived from the anchor-based approach is not useful due to the low area under the curve value (0.586).*

Figure Legend

Figure 1. Receiver operating characteristic curve of the Balance Evaluation Systems Test (BESTest), Mini-BESTest, Brief-BESTest, and Berg Balance Scale (BBS). The area under the curve (AUC) of the BBS was smaller than the BESTest, Mini-BESTest, and Brief-BESTest. The AUC of BESTest was also higher than the Brief-BESTest.

Supplementary Figure 1. Individual change scores on the Functional Gait Assessment (FGA) between week 2 and week 4.

Supplementary Figure 2. Individual change scores on the Balance Evaluation Systems Test (BESTest) between week 2 and week 4.

Supplementary Figure 3. Individual change scores on the Mini Balance Evaluation Systems Test (Mini-BESTest) between week 2 and week 4.

Supplementary Figure 4. Individual change scores on the Brief Balance Evaluation Systems Test (Brief-BESTest) between week 2 and week 4.

Supplementary Figure 5. Individual change scores on the Berg Balance Scale (BBS) between week 2 and week 4.