RELIABILITY AND RESPONSIVENESS OF PHYSICAL ACTIVITY MEASURES IN INDIVIDUALS AFTER TOTAL KNEE ARTHROPLASTY

by

Gustavo J. Almeida

BS, Physical Therapy, Centro Universitario do Triangulo, UNITRI, 2001

Specialist, Rehabilitation of the Musculoskeletal System in Sports Medicine, Universidade Federal de Sao Paulo, UNIFESP, 2002

MS, Sciences, Universidade Federal de Sao Paulo, UNIFESP, 2005

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION SCIENCES

This dissertation was presented

by

Gustavo J. Almeida

It was defended on

November 24, 2015

and approved by

James J. Irrgang, PT, PhD, ATC, FAPTA, Professor, Department of Orthopaedic Surgery, School of Medicine and Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh

G. Kelley Fitzgerald, PT, PhD, FAPTA, Professor, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh

John M. Jakicic, PhD, FACSM, FTOS, Professor and Chair, Department of Health and Physical Activity, School of Education, University of Pittsburgh

Dissertation Advisor: Sara R. Piva, PT, PhD, OCS, FAAOMPT, Associate Professor, Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh

RELIABILITY AND RESPONSIVENESS OF PHYSICAL ACTIVITY MEASURES IN INDIVIDUALS AFTER TOTAL KNEE ARTHROPLASTY

Gustavo J. Almeida, PhD

University of Pittsburgh, 2015

Few instruments that measure physical activity (PA) can accurately quantify PA performed at both light and moderate intensities, which is particularly relevant in older adults. Evidence of their reliability and responsiveness to change is limited. The purposes of this study were to: 1) determine test-retest reliability of the Actigraph (ACT), SenseWear Armband (SWA) and the Community Healthy Activities Model Program for Seniors (CHAMPS) questionnaire in assessing free-living PA; and 2) determine the responsiveness to change in PA measured by the three instruments after an exercise program in individuals with knee osteoarthritis who underwent total knee arthroplasty (TKA).

Test-retest reliability was determined by asking subjects to wear the activity monitors for two weeks and complete the CHAMPS at the end of each week. Reliability was measured using intra-class correlation coefficients ($ICC_{2,k}$). For responsiveness, subjects wore the activity monitors for one week at baseline and at 6-months, and completed the CHAMPS at the end of each week. Changes in PA and standardized response mean (SRM) were calculated and compared across instruments. Weighted-Kappa (*K*) was used to determine agreement between the instruments on identifying changes in PA based on measurement error.

Test-retest reliability ranged from moderate to excellent for ACT (ICC=0.75-0.86), and were excellent for SWA (ICC=0.93-0.95) and CHAMPS (ICC=0.86-0.92) in the subjects who reported

similar PA during two-weeks. The 95% confidence interval of the ICCs from SWA was the only one within excellent reliability range (0.85-0.98). Results for responsiveness revealed small and not statistically significant changes in PA ($p>0.05$) measured by each instrument after the intervention. SRMs indicated small degree of responsiveness (SRM=0.01-0.26). When using each instruments' measurement error as a threshold for change in PA, the ACT and SWA agreed on the identification of changes in PA (K=0.36-0.63) and disagreed with the CHAMPS ($K\leq0.22$).

This study provides evidence that the ACT and SWA have better psychometrics than the CHAMPS, and amongst the activity monitors, the SWA showed better psychometrics. Clinicians and researchers can use the results from our studies to make well-informed decisions when selecting instruments to measure free-living PA in individuals with arthritis of the lower extremities.

ACKNOWLEGEMENTS

First, I would like to thank the funding support from the Pepper Center Scholars Pilot Program (grant no. P30-AG024827) and the SHRS Research Development Fund, both from the University of Pittsburgh.

Thank you to my dissertation committee! I greatly appreciate your guidance and support during this journey. James Irrgang, who introduced me to Pitt in 2004 during a scholar visit, other than teaching me the good clinical practice, you showed me the best way to practice research. Thanks a lot for always clarifying my questions in regards to research methods, statistics and psychometrics. Kelley Fitzgerald, who challenged me all the way through the PhD program, making me think through to understand research in depth and defend my point-of-view on my feet going straight to the point, I really appreciate it. John Jakicic, thanks for your guidance in the process of understanding physical activity measures and with data interpretation. A special thanks to Sara Piva for believing in me and for giving me this opportunity to do my PhD at Pitt. You did not measure efforts to guide me and support me every step of the way! You have most, if not all, the credits for the researcher I am today. Thanks for pushing me further and to show me that a 'great job' is not a 'perfect job'.

A big thanks to all my friends, colleagues, collaborators and staff members around the University of Pittsburgh, Carnegie Mellon University, Chatham University and Federal University of Sao Paulo (Brazil). In special, I would like to express my gratitude for everything that Stella Peccin and Ricardo Giusti have done for me. Other than being my mentors in Brazil (until today), Stella and Ricardo are the reason why I accepted Sara's invitation to come to the U.S. At the time I received the invite, I was doing really well in Brazil, practicing as a Physical Therapist at one of the

biggest Orthopedic Clinics in South America and teaching at several universities. So I was hesitating in accepting the invite but they were very supportive and said to me: 'this is a unique opportunity… embrace it', and that's what I did. Here at Pitt I made many good friends. One of the most important ones who became my big partner in the PhD program is Samannaaz Khoja. Her friendship, collaboration and support throughout the program made my life much easier (if that's really a possibility during the PhD training process).

Eternal gratitude for the great support received from my family. It was very hard to leave all of them behind to go after my dreams, but they were always supportive and cheering with me on my achievements throughout the process. During this journey my family and I went through rough situations, which made me think of giving up many times, but they were there for me and with me all the time. A super special thanks to my parents, Tercio and Alzira, for believing in me and supporting me all the time. Even though it was hard for you to understand why I had to leave the country to do this PhD, you were always supportive, and I hope I made you guys proud of me.

My sincere appreciation for the support received from my wife Yana. People say that 'things do not happen by chance' and I think that it is true. During this journey I met the girl who was going to be by my side for the rest of my life and be the mother of my kids. Thank you very much for your patience and understanding when I was had to be away from home to work on my dissertation. Lastly, thank you God for giving me strength and perseverance to keep moving forward towards this important accomplishment in my career.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 CHAPTER 1 - RESEARCH PROPOSAL

1.1 INTRODUCTION

Individuals who undergo total knee arthroplasty (TKA) are generally older adults with end-stage knee osteoarthritis, who have dealt with pain and functional limitations for a long period of time. Because of these factors, they tend to acquire an inactive lifestyle prior to the surgery.[1,2] Although TKA clearly relieves knee pain and improves functioning in the majority of individuals,[3-5] they keep their sedentary behavior after surgery.[1] It is known that 60% of them do not meet physical activity (PA) recommendations and around 1/3 are obese,[2] which puts them at risk for further disability and chronic diseases. Therefore, improving PA is important in this population and accurately measuring changes in PA in research aimed at promoting a more active lifestyle for these individuals is essential. But before PA can be used as the primary outcome in individuals after TKA, it is important for researchers to know how reliable and responsive the available instruments to measure PA are. It is equally important to compare the reliability and responsiveness of commonly used instruments to measure PA in this population, which has not been done before. Investigating and comparing the reliability and responsiveness of commonly used instruments to measure PA will inform the selection of instruments to use and the interpretation of observed changes in measures of PA in research studies.

1.2 PSYCHOMETRICS OF PA MEASURES IN INDIVIDUALS WITH KNEE OSTEOARTHRITIS WHO UNDERWENT TKA

1.2.1 Investigating Reliability of Measures of PA after TKA

Reliability is the consistency of a score or measurement. The Standards for Educational and Psychological Testing define reliability as "the consistency of measurements when the testing procedure is repeated on a population of individuals or groups".[6] Reliability is the extent to which the same results are obtained on repeated administrations of the same test or measure when no change has occurred in the construct being measured.[7,8] Measures of PA should be reliable if they ought to be used in research and clinical practice to evaluate ongoing progress in a treatment situation. Without reliability, it is not viable to determine whether the instruments used to measure PA are accurately and consistent.

Estimates of reliability determine how much of the variability in test scores is due to errors in the measurement and how much is due to variability in true scores.[9] The concept of measurement error is at the heart of reliability. When one measures a construct, the true score is never known and it must be estimated from the observed score, which provides imperfect information. Calculating an index of reliability requires quantifying the measurement error associated with the observed score. The observed score is composed of the true score and the error score. A true score is the replicable feature of the concept being measured. It is the part of the observed score that would recur across different measurement occasions in the absence of error. However, random error is always present in a measurement and it represents the discrepancies between scores obtained on tests and the corresponding true scores. [10] Random error is easily recognized when assessments are used

repeatedly on stable individuals and the observed scores are not the same on each repeated assessment. Random error arises from multiple sources. Some examples of random error in measures of PA are the inherent unpredictable fluctuations in the measurements provided by devices or the variable perception of the amount of activity one performs and reports in self-reported questionnaires.

There are several forms of reliability estimates. To test the reliability of PA measures, it would be appropriate to use the test-retest approach. The test-retest reliability assesses the degree to which test scores are consistent from one test administration to the next at two different points in time during which the construct being measured is believed to be stable. [11] This kind of reliability is used to assess the consistency of a test across time. Test-retest is appropriate to be used in measures of PA because time is believed to be the major source of error in this measure. People tend to vary their PA behavior across time and this inherent variability can affect the estimates of reliability. There are also several indexes to determine test-retest reliability. One such index is the Intraclass Correlation (ICC) coefficient. The ICC is a reliability coefficient designed for use with interval/ratio data, and it is calculated using variance estimates obtained through an analysis of variance. [12] The coefficients range from 0 to 1, with higher values representing better reliability. According to Fitzpatrick et al, a good test-retest reliability coefficient should be 0.70 or higher if the measure is to be used to evaluate the ongoing progress of an individual in a treatment situation. [11]

To determine the error associated with the measurements of PA is also important. It helps to establish thresholds for interpreting true changes in PA between two assessments. Changes greater than measurement error help assure that measured changes are due to an individuals' change and not due to the error inherent to the instrument. The standard error of the measurement (SEM) is a common coefficient of measurement error. The SEM is useful when interpreting change

because it is calculated in the same unit as the measure. In the proposed study, we used the ICC and the index of test-retest reliability and determined measurement error by using the SEM.

Reliability is also a critical component of responsiveness. Measures with poor reliability are likely to be less responsive because the noise introduced by the error might obscure any real change that has occurred. To date, studies have not determined the reliability or measurement error of PA measures during light intensity activities, nor have compared these psychometric properties between portable activity monitors and self-reported questionnaires commonly used in older adults.

1.2.2 Investigating Responsiveness of Measures of PA after TKA

Another important aspect of the psychometrics of instruments is the responsiveness of the measurement. Responsiveness evaluates the ability of a measure to accurately detect changes in the concept being measured when change has occurred.[7,13] Some methods of responsiveness include internal and external methods.[14] Internal responsiveness characterizes the ability of a measure to change during a pre-specified time frame. In internal responsiveness method, the primary interest is in the measure itself and its distribution of change or magnitude of change over time. Internal responsiveness is often examined by administering a measure before and after a treatment of known efficacy within the context of randomized clinical trials or repeated measures designs. Internal responsiveness can be quantified by indices such as effect size, the standardized response mean and the Guyatt Responsiveness Index.[15,16] Since internal responsiveness is based on mean changes, it is used to evaluate changes over time for the overall group of subjects. External responsiveness reflects the extent to which changes in a measure relate to changes in other measures of health status (e.g., global rating of change). The external responsiveness method is used to establish clinically

meaningful levels of change when the external reference identifies change (improvement) that is perceived as important to the individual.[14] Therefore, external responsiveness is useful to evaluate changes over time for individual individuals. In this case, choosing an appropriate external measure is imperative.

To the best of our knowledge, studies have not investigated the internal or external responsiveness of measures of PA performed at light intensity derived from portable monitors and self-reported questionnaires and have not compared the responsiveness indices across measures of PA from portable monitors and questionnaires.

1.3 INSTRUMENTS TO MEASURE PA

The reference standard methods to measure PA are the Doubly Labeled Water (DLW), the metabolic chamber (room calorimeter) and the indirect calorimeter. DLW is used to measure PA during a free-living situation whereas the other methods are used to measure PA in a more confined environment. Methods that measure PA in free-living conditions are preferred when investigating changes in PA because the person is not restricted to a particular environment and can perform daily tasks at their usual pace and in their natural condition. Although DLW is the reference standard to measure PA in free-living conditions, this method is not practical for use in large scale due to its very high cost (≅ $1000.00 per subject per time point) and it needs an experienced person to implement and interpret the results. Due to the practical limitations of DLW, additional instruments have been developed to measure free-living PA in large-scale research. These include self-reported

questionnaires and portable activity monitors. These instruments are more affordable than DLW and do not require high technical expertise. Yet, none of them is free of flaws.

In individuals after TKA, PA has been traditionally measured by self-reported questionnaires.[1] Advantages of self-reported questionnaires include low price, ease of administration, and the ability to address the frequency, intensity, duration and types of PA such as sports, leisure, recreational, and occupational activities.[17] Disadvantage of self-reported questionnaires of PA include inconsistent subject recall, overestimation of PA, and underestimation of sedentary pursuits. These inconsistencies have resulted in inadequate psychometrics.[18]

To overcome the limitation of self-reported measures of PA, portable activity monitors, such as pedometers, heart rate monitors, and accelerometers, have been developed over the past decades to collect real-time PA information. Advantages of portable activity monitors include ability to measure PA in real-time during free-living conditions, and better accuracy than self-reported instruments in estimating energy expenditure. Portable activity monitors are more affordable (monitor ≅$249.00, and software ≅$1000.00) than DLW and the same device can be used for several subjects. Disadvantage of these monitors include the high price compared to questionnaires and the relative complexity of data processing.

Pedometers and accelerometers are the portable PA monitors most used in research in older adults. Pedometers are the least expensive of the portable monitors, are easy to use, and can provide feedback regarding step counts during ambulatory activity such as walking or running. These monitors count the number of steps taken by using a horizontal spring suspended lever arm that moves up and down in response to vertical accelerations of the hip. This motion opens and closes an electrical circuit that accumulates the number of steps taken and provides a digital display. However, the use of pedometers in measuring PA in older adults has limitations such as

underestimation of steps taken at slower gait speeds [19-21] or with irregular and unsteady gait patterns; [19,20,22-24] and inaccuracy in capturing seated activity, upper-extremity activity, or indoor and outdoor household chores such as pushing, lifting, or carrying objects. [25-28] Another limitation includes the nonexistence of internal clocks, which makes them unable to provide information on the duration of activities. Additionally, they do not take into account the intensity of vertical displacement, which would be needed to differentiate activity intensity level, i.e. between walking and running.

Accelerometers overcome many of the limitations of pedometers. Accelerometers are small electronic sensors able to measure and store real-time estimates of the frequency, intensity, and duration of free-living PA.[29,30] They are able to provide such information because they have an imbedded internal clock. These monitors measure movement through body acceleration.[31,32] Acceleration is the change in speed in relation to time. Typically, acceleration is measured in gravitational units (g's) where 1 g is equal to 9.8 m.s-2. Accelerometers take into account the intensity of displacement and are able to differentiate the intensity of the activity (e.g., light versus moderate intensity). Accelerometers integrate a filtered digitized acceleration signal over user specified time interval called epoch. [33] One minute epoch lengths are often used for measuring PA in adults. Data on acceleration are recorded while the individual is wearing the accelerometer and then processed by software on the computer. Accelerometers can be worn on the wrist, upper arm, waist, low back, or ankle, attached by belts, pouches, belt clips, or Velcro bands. Those worn on the waist, hip or low back are well suited for capturing accelerations during normal ambulatory movement. [31,33-35] Accelerometer-based monitors worn on the arm are able to capture body acceleration during ambulatory activities and activities that do not involve ambulation such as weight lifting, mopping and biking. These activity monitors can also monitor movements in different planes. Uniaxial monitors record vertical acceleration in one plane, biaxial monitors record vertical and horizontal

acceleration in two planes, and triaxial monitors' record acceleration in three planes by three different accelerometers positioned internally at 90 degrees from one another. The output from triaxial accelerometers is a composite value of all three accelerometers, which is believed to provide a better indicator of overall body movements. In summary, accelerometer-based monitors are the most used in field-based research and are considered the best choice amongst the portable objective monitoring devices.

1.3.1 Importance of Measuring PA at Light Intensities

To date, most of the self-reported PA questionnaires and PA portable monitors have measured PA at moderate (e.g. brisk walk) or higher intensity levels (e.g. running). The interest in measuring PA at moderate or higher intensities is rooted in findings that health benefits are acquired when more time is spent in activities performed at moderate or higher intensities. [36] According to the PA guidelines issued by the US Department of Health and Human Services in 2008, [32] major health benefits from PA require the performance of at least 150 min (2.5hr) of moderate-intensity aerobic activity each week.

Older adults with TKA are usually sedentary and perform most of daily activities at light intensities. [5] Therefore, measuring PA only at moderate or higher intensities in this population may not be sensitive to capture clinically important changes in PA due to interventions that intend to promote PA. For example, if the baseline level of PA of an individual is very low (e.g., daily activity consists of 30 min of slow walking inside the house to perform basic ADLs), a clinically important change in PA due to intervention could represent doubling the PA (e.g., the same 30 min of slow walking inside the house plus 30 min of slow walking outside). This change, although seems

important, would not be captured by instruments that measure PA at moderate or higher intensities only, because the person walks slow, which an activity of light intensity. Being able to walk more may be the first important progress towards increasing PA to a point that impacts overall health. We believe that measuring PA performed at light intensities after TKA is clinically relevant, and may allow identification of changes in PA that are beneficial to this population. The PA guidelines also report that there is no minimum PA threshold bellow which there is no health benefits.[32]

In this study we proposed to determine the reliability and responsiveness of PA measured at light and moderate intensities, using both self-reported and real-time accelerometer-based monitors. To that end, in the next sessions of this document we discuss the rational for the choice of instruments to measure PA after TKA that were used in the proposed study.

1.3.2 Choice of Self-reported Questionnaire to Measure PA after TKA

Most of the self-reported questionnaires available to measure PA were developed and tested in young adults, and were designed to measure activities at moderate or high intensities (i.e. jogging, running and/or vigorous sports activities).[37,38] As individuals after TKA are often sedentary older adults, who perform most activities at light intensities (e.g., slow walking, light household chores and gardening), questionnaires used to identify patterns of PA in this population have to be able to capture activities that are relevant to older adults. Only a few questionnaires have been developed to capture activities pertinent to older adult and they include the Physical Activity Scale for the Elderly (PASE), the Yale Physical Activity Survey (YPAS), and the Community Healthy Activities Model Program for Seniors (CHAMPS).

The PASE, was developed to measure age-specific PA levels of older adults.[39,40] It queries subjects about the type and amount of PA during the past 7 days. Subjects provide information about the time spent (hr/week) in activities such as light/heavy housework, home repairs, lawn work or yard care, outdoor gardening, caregiving, work for pay or as volunteer. These data are then converted into a total score that ranges from 0-315 units, based on the weight of each activity (time spent * PASE weight). The measurement unit on PASE is aleatory and does not represent energy expenditure. While the PASE has been shown reliable when administered twice in an interval from 3 to 7 weeks (r= 0.84),[39] its validity is questionable when comparing its PA measures to the measures from the reference standard of DLW (r= 0.28). [41] Additional limitations of this instrument is that it does not differentiate light from moderate intensity activities and its score cannot be directly compared with data on energy expenditure provided by portable activity monitors.

The YPAS was developed to measure age-specific PA of older adults in a typical week in the past month.[42] The YPAS evaluates activities across varying types, including household, recreational, and exercise settings. The survey provides three summary indices and five subscales. The three indices include the Total Time Summary Index (TTSI), the Energy Expenditure Summary Index (EESI), and the Activity Dimensions Summary Score (ADSS). The TTSI, which calculates total time spent for each activity on the checklist, is expressed as hr/week. The EESI is expressed as kilocalories per week (kcal/week) and represents the time spent on each activity multiplied by an intensity code and is summed over all activities. The YPAS has shown moderate test-retest reliability, with correlation coefficients of its indices ranging from 0.42 to 0.65. However, its validity is questionable. The correlation of total energy expenditure during PA calculated by the YPAS and the reference standard of DLW was poor (r = 0.07 to 0.11).[41,43] Similar to the PASE, the YPAS does not provide a simple way to distinguish scores from light to moderate intensity activities.

One more instrument developed to be used in older adults is the CHAMPS. [44,45] The CHAMPS was originally developed to be used as an outcome measure of a PA promotion intervention and it measures PA in a typical week in the past month. Subjects provide information on type and frequency of 41 activities of several intensity levels. From the questions, measures of frequency per week and estimated kcal/week in PAs are derived. For both frequency and caloric expenditure, two measures are derived based on: (1) PAs of moderate or greater intensity (metabolic equivalent [MET] value above 3.0), and (2) all specified PAs that include activities of light intensity in addition to moderate and greater. Thus, estimated caloric expenditure measures are calculated by multiplying the estimated duration of each activity by the MET value (i.e., weighting the time spent by the intensity) and summing these across all relevant activities. Then, by simply subtracting the values of moderate or greater intensity from all specified PAs, values of light intensity activities are generated. The CHAMPS has shown small correlation to the reference standard of DLW (r= 0.28), [43] which is similar to PASE and better than the YPAS. The CHAMPS has shown fair to good reliability, with intraclass correlation (ICC) coefficients ranging from 0.27 to 0.84, when measuring PA in people with musculoskeletal disorder and healthy older adults. [19,43-45] A more detailed review of reliability and responsiveness of the CHAMPS is provided in the next section of this document.

By comparing the overall performance of the three PA questionnaires for older adults, it seems that the CHAMPS is an adequate choice. It combines fair to good reliability with limited evidence of criterion-related validity. In addition, its advantage compared to the PASE and YPAS is that it provides an easy way to calculate PA performed at light intensity and estimates PA in units of energy expenditure, allowing for direct comparison with measures from portable monitors. Thus, we proposed CHAMPS as the self-reported instrument to be used in our research study.

1.3.2.1 Reliability and Responsiveness of the CHAMPS questionnaire

To date, six studies have investigated the test-retest reliability of the CHAMPS questionnaire. Measures of PA from CHAMPS have shown to be reliable after a 1-week interval (ICC= 0.78 and 0.84), [46] after 10-days interval (ICC= 0.64 [moderate-intensity only]), [43] after a 2-weeks interval (ICC= 0.62 and 0.76), [19,47] and after a 6-months interval (ICC= 0.66 and 0.70) [44,48] in healthy older adults. Kaleth et al found fair to moderate reliability after a 2-week interval for moderate-intensity PA and light-to-moderate PA measures in adults with fibromyalgia, ICCs= 0.27 and 0.67, respectively. [47] The relatively lower reliability in Kaleth et al may be due to the higher variability in PA patterns in individuals with fibromyalgia in comparison to healthy older adults, as their pain condition may vary from one week to the other, affecting the way they perform daily activities. The reliability of an instrument should be measured in the population that the measure is intended to be used. Results from estimates of reliability derived from healthy older adults or middle-aged adult with fibromyalgia have limited application in older adults after TKA who may experience more severe physical limitations and have distinct patterns of PA. Furthermore, these studies have not compared the reliability of CHAMPS and portable monitor, which we propose to do in this study.

Upon extensive literature search we identified 10 studies that used the CHAMPS questionnaire to estimate changes in PA in the context of intervention studies. However, only 4 studies described data that enabled for calculation of internal responsiveness indices, [49-52] and other two reported these indices. [53,54] The effect sizes for CHAMPS ranged from small to moderate for activities performed at or above moderate intensity (0.08 and 0.68), and for light-to-moderate PA (0.01 to 0.66).[54,55] The standardized response mean for CHAMPS were moderate for activities performed at or above moderate intensity (0.40 to 0.61), and ranged from small to large for light-to-moderate PA (0.18 to 1.21).[49,51-53] Although most of these studies were done in older adults, which

provides some evidence of the responsiveness of the CHAMPS questionnaire, none have determined external responsiveness of the CHAMPS or compared its responsiveness to accelerometer-based monitors. In addition, none of these studies have assessed reliability or responsiveness of CHAMPS in measuring activities of light intensity.

1.3.3 Choice of Portable Monitors to Measure PA after TKA

In this study, we used accelerometer-based portable monitors to measures PA objectively in subjects after TKA. We have selected the ActiGraph™ (ACT) and the Sensewear™ armband (SWA) for several reasons. First, to our knowledge, the ACT and SWA are the only accelerometer-based monitors that have some evidence of criterion-related validity to measure PA in older adults. [43] Second, due to their small sizes and user friendly wearing characteristics, both have shown to be well tolerated and add minimal burden to the subjects. Third, they both estimate PA at light and moderate intensity levels. Lastly, these monitors are available for me to use free of charge in this study.

The ACT GT3X model, which is worn around the waist, is a small triaxial accelerometer (3.8cm x 3.7cm x 1.8cm in size) that measures body acceleration in the range of 0.05- 2.0 g and converts accelerations into activity counts. The ActiLife 5 software (Actigraph LLC, Pensacola, FL) allows for initializing the monitor and downloading its data, and then, it uses special equations to convert counts into energy expenditure measured across several intensities: sedentary (≤100 counts), lifestyle (101 to 759 counts), light (760 to 1952 counts), moderate (1953 to 5724 counts), and vigorous (5725 to 9498 counts). PA measures from the ACT have been compared to PA energy expenditure from DLW and demonstrated correlation coefficient of 0.49. [43] Studies validating the

ACT in a laboratory setting demonstrated moderate to strong correlations with energy expenditure measured by indirect calorimetry (correlation coefficients ranging from .55 to .85). [55-58] One of the limitations of the ACT, as it is worn on the waist, is its inability to capture activities that involve upper-extremity movement or seated activities. The ACT has shown better accuracy for ambulatory activity (walking and running) while it underestimates energy expenditure during work-related activities (e.g. desk work), certain indoor and outdoor household chores (eg, mopping, gardening), and some recreational tasks. [30,56,57,59-61]

The SWA Pro3 is a multi-sensor wireless monitor (8.6cm x 5.3cm x 2.0cm in size). The monitor is worn over the triceps muscle on the right arm, and combines information from a heat flux, skin temperature, galvanic signal, and biaxial accelerometer, enabling the device to estimate energy expenditure from activities that do not require ambulation. The heat-flux sensor measures the amount of heat being dissipated by the body by measuring the heat loss along a thermally conductive path between the skin and a vent on the side of the armband. Skin temperature and near-armband temperature are also measured by sensitive thermal resistors. The armband also measures galvanic skin response (the conductivity of the wearer's skin) which varies due to physical and emotional stimuli. Lastly, the accelerometers track body acceleration and position. These data are integrated and processed by software using proprietary algorithms utilizing the subject's demographic characteristics (gender, age, height, and weight) to provide minute-by-minute estimates of energy expenditure during different levels of PA (from sedentary to vigorous intensity activities), expressed in kcal/min. PA measures from the SWA have been compared to reference standards such as DLW (correlation coefficients ranging from 0.48 to 0.81) [43,62] and indirect calorimetry (correlation coefficients ranging from 0.11 to 0.92). [55,58,63-67]

1.3.3.1 Reliability and Responsiveness of the ACT and SWA

Only 3 studies examined the reliability of the ACT. One study used a hydraulic shaker table to test the reliability of 43 ACTs in 6 different conditions, in which pairs of combinations between a variety of acceleration and frequency of the shaker were used. [68] This study examined inter and intra-instrument reliability and observed excellent reliability (ICCs ≥0.95). Results from this study cannot be translated to our purposes because it was not done in humans who certainly vary their activities and acceleration patterns throughout the day. This study used tables with the same frequency of acceleration for all monitors, resulting in minimal variability and consequent high reliability.

One study in healthy adults investigated the inter-instrument reliability of the ACT, [69] where the subjects wore two monitors at the same time, one on the right and another one on the left hip for 24-hr, and were instructed to engage in their normal PA habits. ICCs were all above 0.95 indicating excellent association between monitors for PA measures (average activity counts, number of steps, sedentary, light, moderate and vigorous activities). The approach used in this study has decreased the amount of intra-individual variability as data were collected simultaneously using two monitors on the same subject while performing the same activities, resulting in high values of reliability. Calculating reliability simultaneously between two devices on the same individuals do not represent the clinical application intended for the use of portable monitors because usually one device is used to collect data in consecutive days on the same individual. Testing the test-retest reliability of measurements of PA taken during several days of free-living PA in a repeated-measure fashion reflects the clinical application of the measure.

Another study in healthy adults investigated the test-retest reliability in free-living for a 7-day period with 1- to 4-week gap between each period. [70] At least 6 to 10 hours of wearing time per day during awaken hours for 4 days were required. The test-retest reliability was assessed by ICCs

between time 1 and 2 for measures of total activity counts per day, minutes spent in sedentary, light and moderate activities per day. Results of this study showed ICCs ranging from good (0.77 to 0.84) for time spent in sedentary and light intensity activities, to excellent (0.85 to 0.90) for total counts and time spent in moderate activities. Although the results from this study have shown that the ACT is reliable to measure PA over time and that subjects had fairly stable PA levels from week to week, this study was performed in healthy adults. Studies in healthy adults cannot be generalized to older adults after TKA who experience considerable physical limitations and may have patterns of PA very distinct than healthy adults. For example, gait abnormalities in individuals after TKA such as antalgic gait or asymmetric movements due to muscle weakness or joint deformities may alter the acceleration of their bodies and affect accelerometry measures. [71-74]

The reliability of the SWA has been investigated in four studies. Two studies were done in healthy adults, one in obese adults, and one is older adults. One study examined the reliability of the SWA during 40 min resting, with 80 min interval in between two resting occasions. [65] Subjects were in a reclined position and were instructed to rest and remain awake. Results of this study revealed excellent reliability of the device between the two resting occasions (r= 0.93). The other study in healthy adults assessed the test-retest reliability of the SWA while subjects participated in a pre-established daily routine during 13-hr of standardized activities during 2 consecutive days. [75] Activities of daily living were completed in a 3-hr period, where subjects performed 4 types of activities: lying awake, structured sedentary activities, spontaneous and standardized PA. Subjects remained seated for the remaining 10-hr. The reliability for each individual activity and the overall 13-hr period ranged from moderate (ICC = 0.62) to excellent (ICC= 0.98). Two other studies examined the reliability of the SWA during periods of 20 min of resting while awake with 24 hours between tests. One of these studies was done in obese adults and the other in older adults. [66,76]

Results from these studies have shown good to excellent reliability, with coefficients of reliability ranging from 0.88 to 0.98. While these combined studies found moderate to excellent reliability of the SWA between test-retest, three of them were done using resting conditions which do not reflect the utilization of the device during free-living conditions. The intra-individual variability during resting is lower than during free-living activities and would likely result in better reliability. The only study done during free-living conditions asked subjects to perform PA following a pre-defined script, in the same way it is done in a laboratory setting. The pre-defined sequence and time of PA result in lower variability as compared to free-living situation. Although these studies add some evidence to the reliability of the ACT and SWA, none of them investigated the reliability of PA measures at light intensity activities or compared the reliability of ACT, SWA, and self-reported questionnaires.

Our search for studies that used the ACT to assess changes in PA post interventions resulted in 10 studies. Amongst these, only 3 studies provided the means and standard deviation at both the baseline and follow-up that enabled us to calculate indices of responsiveness. A study in children's weight management showed that the ACT has the ability to detect change with standardized response means of 0.28 for time spent in sedentary, 0.25 for time spent in light and 0.50 for time spent in moderate activities after a diet intervention. [77] A study in adults suffering from 3 types of congenital heart disease participating in a progressive walking program found effect sizes ranging from 0.40 to 1.21 for counts per minute, and 0.78 to 1.52 for time spent in moderate activities after 10-week (5 days a week) intervention. [78] A study in adults with coronary disease receiving education and counseling reported effect sizes from 0.40 to 1.66 after 6 weeks of intervention and effect sizes from 1.46 to 2.61 after 12 weeks of intervention. [79] Studies assessing changes in PA post

interventions using the SWA were four in our literature search. [80-83] However, none of them provided summary statistics that could be used to calculated indices of internal responsiveness.

In summary, studies on the responsiveness of the ACT added some evidence to the device's capability of detecting PA changes over-time in sedentary, light and moderate intensities. However, results of those studies have limited application to population of older adults after TKA. Furthermore, the responsiveness of the SWA after an exercise intervention has not been reported and studies have not compared the responsiveness of portable activity monitors and self-reported questionnaires.

1.4 SIGNIFICANCE OF THE PROPOSED STUDY

Literature is scarce on information about reliability, measurement error, and responsiveness of instruments commonly used to measure PA in research settings. Comparing reliability and responsiveness among portable activity monitors (SWA and ACT) and a self-report questionnaire (CHAMPS) during activities of light and moderate intensities performed in free-living conditions will allow for a well-informed choice of instruments for measuring PA in individuals with arthritis of the lower extremities. The results will also help to interpret data in future trials targeting PA promotion to increase activity participation in individuals with arthritis of the lower extremities.

1.5 PRELIMINARY STUDIES

1.5.1 Measurement Timeframe for Individuals with Arthritis to Wear Portable Monitors

As part of my learning experience on measures of PA, I worked in a cross-sectional study to characterize PA and determine the measurement timeframe for wearing portable monitors to obtain consistent estimates of PA in women with rheumatoid arthritis. Results of this study have been published before.[89] In this study, participants wore the SWA for 7 days. Measurements of daily physical activity energy expenditure (PAEE) during activities at or above 1 metabolic equivalent (MET) level (PAEE≥1MET), PAEE during activities at or above 2 METs (PAEE≥2METs), and PAEE during activities at or above 3 METs (PAEE≥3METs) were obtained. Complete data were obtained for 47 participants. Daily usage of the SWA was 98% of the time (23:31 hours/24 hours). The daily average PAEE≥1MET was 1,050 ± 331 kcal/d, for PAEE≥2METs it was 642 ± 309 kcal/d for, and for PAEE≥3METs it was 239 ± 178 kcal/d. Results of intraclass correlation coefficient analyses and multiple linear regressions indicated that 4 days of data collection were needed to reliably estimate PAEE at the three levels.

The above study only included women with RA who are younger then individuals after TKA, which questions the generalizability to the population after TKA. Therefore, we repeated the same protocol and analysis on 20 individuals after TKA (mean age 68 ± 7, BMI 29 ± 4) who were participating in another study. In this study both the ACT and SWA were used. Results indicated that collecting PA data in any combination of 5 consecutive days of the week allowed for consistent PA measures. Similar to individuals with RA, subjects after TKA spent 97.6% of the time performing sedentary to light intensity activities (below 3 METs). Working in these studies allowed

me to gain experience on processing data at several PA intensities from the portable monitors that were used in the proposed research study and provided support that at least full 5 days of data collection is needed when wearing the portable monitors to obtain consistent data. It also seems to support the argument that measuring PA at light intensities is relevant in this population.

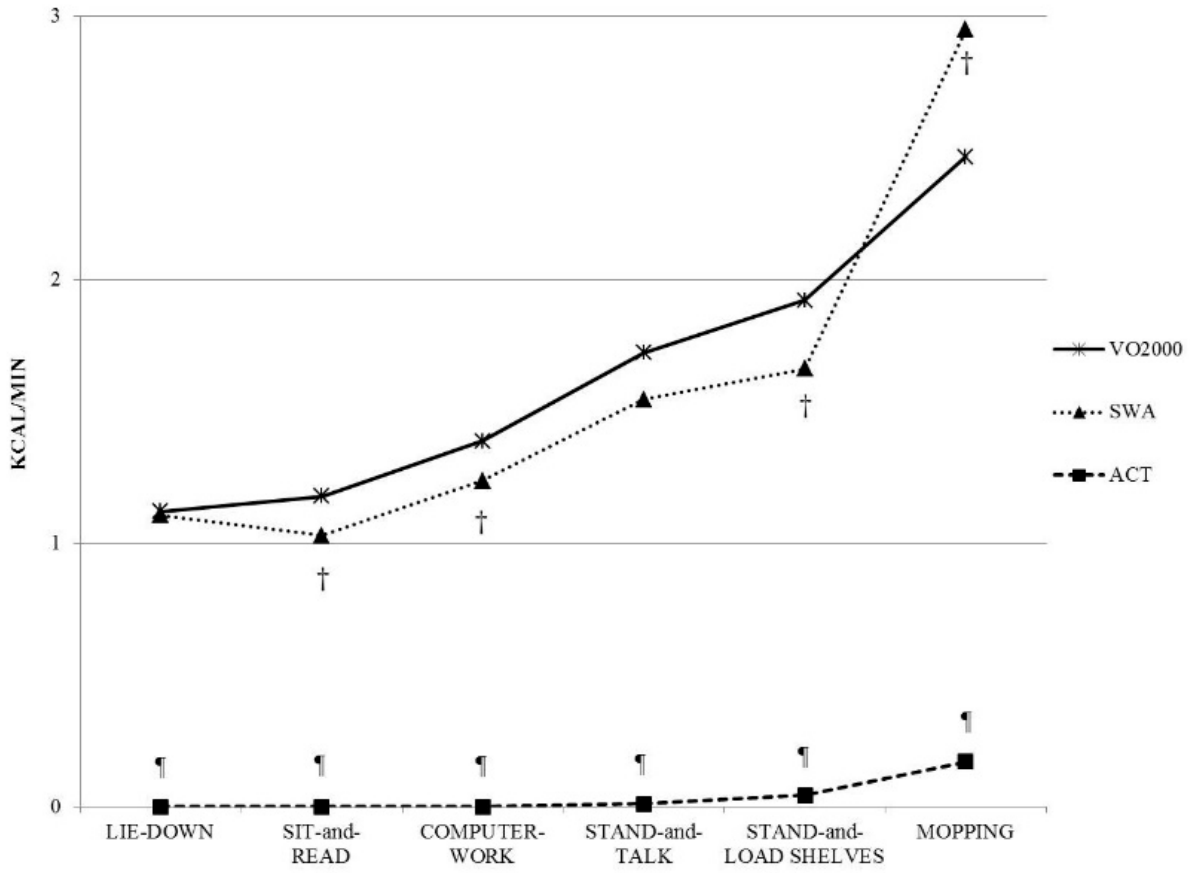
1.5.2 Experience Collecting Data Using the CHAMPS Questionnaire

During the study mentioned above on subjects after TKA we also collected self-reported data on PA using the CHAMPS questionnaire. We administered the CHAMPS during the last day of wearing the ACT and SWA. To match the same period in which they wore the portable activity monitors, when answering the CHAMPS, subjects were asked to recall the last 7 days of PA. We calculated PA in kcal/week, according to scoring instructions, generating values of PAEE≥2METs and PAEE≥3METs. We then calculated the correlations between measures from the CHAMPS and measures from the ACT and SWA. For PAEE≥2METs, the correlation between the CHAMPS and SWA was $r = 0.56$. In this study, we did not have data on PAEE≥2METs from the ACT because we used an older version of the ACT software (ActiLife v.4). For the proposed study, the ActiLife v.5 was used, which allowed determination of PAEE≥2METs. For PAEE≥3METs the correlation between the CHAMPS and ACT was $r=-0.04$, and between CHAMPS and SWA the correlation was $r = 0.51$. This pilot work supports my experience using the CHAMPS questionnaire in subjects after TKA.

1.5.3 Validity of Physical Activity Measures in Individuals after TKA in a Laboratory Setting

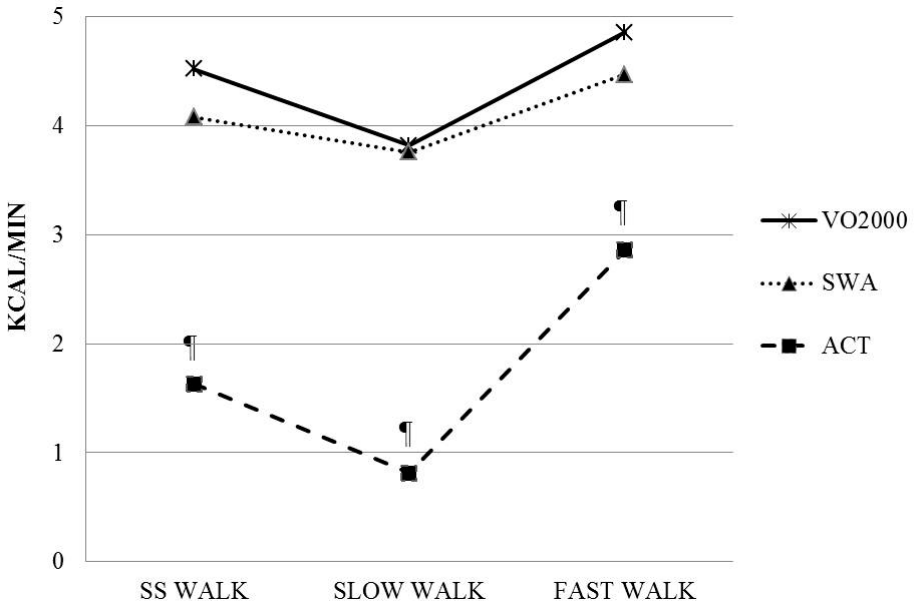
I worked in a study that investigated the validity of ACT and SWA to measure PA in a laboratory setting in 21 subjects after TKA (mean age 68 ± 7, BMI 29 ± 4). Results of this study have been published before.[84] PA was concurrently measured by the two portable monitors and the reference standard of indirect calorimetry (VO2000, Medical Graphics Corporation, St. Paul, MN). Subjects performed several activities during 6 minutes each. The list of activities is depicted in Figure 1. Results indicated that the accuracy of the ACT was lower than the accuracy for the SWA. The SWA showed good accuracy for all activities tested in the study, from lying down to fast walking speed (Figure 1 and 2).

By working in this study I acquired skills at operating the indirect calorimeter and analyzing its measures, and gained experience on the complexities of data processing and management. The study that we are proposing is the next step to investigate the reliability and responsiveness of PA measures during free-living condition after TKA.



(**§**) $p \leq 0.05$ between energy expenditure in kcal/min measured by the ACT and the VO2000; (**‖**) $p \leq 0.05$ between energy expenditure in kcal/min measured by the SWA and the VO2000.

Figure 1. Graph representing energy expenditure measured by the Actigraph (ACT), the SenseWear Armband (SWA), and the reference standard (VO2000) across non-walking activities.



(**§**) $p \leq 0.05$ between energy expenditure in kcal/min measured by the ACT and the VO2000; (**‖**) $p \leq 0.05$ between energy expenditure in kcal/min measured by the SWA and the VO2000.

Figure 2. Graph representing energy expenditure measured by the Actigraph (ACT), the SenseWear Armband (SWA), and the reference standard (VO2000) across walking speeds.

1.6 OBJECTIVES/SPECIFIC AIMS/RESEARCH QUESTIONS

The purpose of this research study was to determine the reliability and responsiveness of measures of PA at light and moderate intensities taken by the ACT, the SWA, and the CHAMPS, and to compare the psychometric properties across these measurement tools in subjects after TKA.

1.6.1 Specific Aim 1

To determine the reliability and measurement error of measures of PA at light and moderate intensities estimated by the ACT, SWA and CHAMPS, in subjects after TKA.

1.6.1.1 Hypothesis 1

Measures of PA taken during two consecutive weeks will show good test-retest reliability for the three instruments, with intra-class correlation coefficients above 0.70 for activities of both light and moderate intensities. Measurement error will be determined by calculating the standard error of measure for the three measurement tools.

For this aim, PA was measured by the ACT, SWA, and the CHAMPS during two subsequent weeks in the subjects who report stable levels of PA. Measures of PA from week one were compared to measures of PA from week two.

1.6.2 Specific Aim 2

To determine the responsiveness of measures of PA taken at light and moderate intensities estimated by the ACT, SWA and CHAMPS in individuals after TKA undergoing an exercise program.

1.6.2.1 Hypothesis 2a

Measures of PA taken by the ACT, SWA and CHAMPS will be able to identify changes in PA performed at light and moderate intensities from pre to post-intervention.

1.6.2.2 Hypothesis 2b

The SWA will be more sensitive to detect changes in PA than the ACT, and the ACT will be more sensitive to detect changes in PA than the CHAMPS.

1.6.3 Specific Aim 3

To explore the associations between a modified global rating of change, which measures an individual's perceived changes in PA, and PA at light and moderate intensities estimated by the ACT, SWA and CHAMPS in individuals after TKA undergoing an exercise program.

1.6.3.1 Hypothesis 3

Changes in PA measured by the modified global rating of change will be associated with changes in PA measured by the ACT, SWA and CHAMPS (correlation coefficients of at least

moderate magnitude), which will support the use of the global rating of change as an external anchor to determine thresholds of clinical important changes in measures of PA.

For aims 2 and 3, PA will be measured by the ACT, SWA, and the CHAMPS before (baseline) and after subjects undergo an exercise program (6 months). The global rating of change scale will be completed only at the end of the exercise program. Changes in PA at light and moderate intensities estimated by the ACT, SWA and CHAMPS were compared with scores on the global rating of change.

1.7 RESEARCH DESIGN

1.7.1 Study Overview

This is an ancillary study from a randomized trial that investigates the effect of rehabilitation approaches to improve quality of life in individuals following TKA. The proposed study used data from the baseline and follow-up assessments of all subjects in the parent trial to determine the reliability and responsiveness of measures of PA assessed by the ACT, the SWA, and the CHAMPS, and compare the psychometric properties across these measurement tools. Eligible subjects performed baseline measures of outcome, and then were randomly assigned to a Comprehensive Behavioral Intervention (CBI) or Standard Physical Exercise program (SPE). Follow-up assessments were performed 6 months after randomization. Since the effectiveness of interventions is not the aim of the proposed study, the data from the two intervention groups were combined for the data analysis.

For this ancillary study, reliability and measurement error were determined at the baseline. During baseline testing, subjects wore the ACT and the SWA for two consecutive weeks, and completed the CHAMPS at the end of each week. Subjects were asked to perform similar physical activities during the two weeks of measurement. Stability of PA was measured by asking the subjects if they performed more, less or about the same activities in week 2 comparing to week 1. Then, data for subjects with stable patterns of PA (about the same activities) were used to determine reliability.

To determine responsiveness, PA measured by the ACT, SWA, and the CHAMPS before (baseline) and after (6-months) subjects underwent the interventions being tested in the parent study was used. The global rating of change in PA was collected at the 6-month timeframe. Changes in PA at light and moderate intensities estimated by the ACT, SWA and CHAMPS were compared with scores on the global rating of change.

All evaluations and interventions were held in the Physical Therapy Clinical and Translational Research Center (PT-CTRC) at the University of Pittsburgh.

1.7.2 Study Sample

Subjects following a unilateral TKA were recruited from the Orthopedic Program at Magee Women's Hospital at UPMC from a single surgeon. Since this study used subjects after TKA who were participating in a randomized trial, the inclusion-exclusion criteria were similar to the parent study.

Inclusion Criteria – Subjects were asked to participate in the study if they: (1) Underwent unilateral TKA at least three months prior to, but no longer than six months prior to study participation. Three months after the surgery is needed to surpass the sub-acute recovery period

when pain, effusion, and knee motion are clinically improved and no longer restrict more intense exercises; (2) The surgical technique was a minimally invasive (quadriceps sparing) TKA performed by the same experienced surgeon. This was to avoid variability in surgical technique and surgeon's practice; (3) Provided a written medical clearance to participate in the study; (4) Spoke fluent English- necessary to reliably complete the study questionnaires and understand study instructions; (5) Were 50 years and older; (6) Were willing to be randomized to either treatment group. Additional inclusion criteria specific to this ancillary study included: (7) Willingness to wear the portable monitors for 2 weeks during baseline period, and answer the CHAMPS twice; (8) Commitment to perform similar activities during the 2 baseline weeks of data collection. All subjects from the main RCT were invited to participate in this ancillary study.

Exclusion Criteria – Participants were excluded from the study if they: (1) Have had bilateral or TKA revision; (2) Have had hip or ankle joint replacement; (3) Were unable to comfortably bear weight on the surgical knee; (4) Have had 2 or more falls within past year; (5) Had an uncontrolled medical condition that would prevent safe participation in the study (uncontrolled blood pressure, dyspnea at rest, cardiovascular disease, absolute contraindications to exercise, [36] and diabetes); (6) Participating in regular exercise during prior 6 months; (7) Use of beta blockers, (8) Have a neurological condition that affects locomotion; (9) Have had a malignancy, life-threatening illness or surgery in the past six months; and (10) Have had a TKA done using a technique other than the minimally invasive.

1.7.3 Outcome Measures

Physical activity was measured by the ACT, the SWA, and CHAMPS.

The **ACT** (GT3X model) is a small triaxial accelerometer (2 x 1.5 x 0.6 inches in size) worn around the waist, on the hip bone aligned with the anterior superior iliac spine. To initialize the monitor and download its data, the ActiLife 5 software (Actigraph LLC, Pensacola, FL) was used. The software generates activity counts by default and was set to collect data at 1 min intervals. Activity counts/min were converted into kilocalories per min (kcal/min) using the ACT software ActiLife 5. The ACT was worn during 2 weeks at baseline and during 1 week at the 6-month follow-up. Its battery lasts for up to 20 days. The ACT has demonstrated adequate validity as a tool to measure PA in free-living conditions in young healthy adults. [85,86,87,43,88]

The **SWA** (Pro3 model) is a monitor worn on the back of the right arm (midpoint between shoulder and elbow). The SWA combines information from multi-sensors such as biaxial accelerometer, heat flux (heat being dissipated by the body), galvanic signal (onset, peak and recovery of maximal sweat rates) and skin temperature. The information is integrated and processed by the SenseWear Professional software v6.1 (BodyMedia Inc., Pittsburgh, PA) using proprietary algorithms utilizing subjects' demographic characteristics (gender, age, height, and weight) to provide PA data measured in kcal/min. SWA has a battery life of 14 days. The SWA was worn during two weeks at baseline and during 1 week at the 6-month follow-up. The SWA has shown adequate validity for assessing PA in free-living conditions in obese and healthy adults. [43,62,89,90]

The software used to download data from the ACT and the SWA allow researchers to calculate energy cost at any metabolic equivalent (MET) level. Therefore, daily averages of PA energy expenditure (PAEE) at 3 different intensity levels were calculated: (a) PAEE during activities at or above 2 METs (PAEE≥2METs), (b) PAEE during activities between 2 and 3 METs (PAEE=2-3METs), and (c) PAEE during activities at or above 3 METs (PAEE≥3METs). Figure 3 exemplifies the measures of PA at different intensities. As at rest the average person has an oxygen

consumption of 3.5 ml/kg/min (or 1 MET). The measure of PAEE≥2METs represents the daily energy spent at or above light activities such as slow walking (±2mph), light cycling (±5 mph), playing a musical instrument, bartending, cooking, scrubbing, and showering. Finally, the measure of PAEE≥3METs represents the daily energy spent at or above moderate activities such as crisp walking (≈3 mph), cycling (≈5.5 mph), bowling, canoeing, janitorial work, cleaning windows, and climbing stairs. An additional variable processed by the ACT and SWA is total energy expenditure (TEE). Total EE represents energy spent at any MET level, from sleeping peacefully to vigorous PA, which includes all values of PAEE.

The **CHAMPS** is a 41-item self-administered questionnaire with a list of specific activities varying from light to vigorous intensity, to which subjects report their frequency and approximate duration of participation in activities in the last week. The questionnaire assesses frequency in times per week, and classifies duration according to time spent in each activity, ranging from <1hr/week to 9 or more hr/week. The information is used to quantify EE (kcal/week). Values are calculated to quantify PA in kcal/week, according to original scoring instructions, generating values of PAEE≥2METs and PAEE≥3METs.[44] The CHAMPS was completed at the end of each of the two baseline weeks of PA data collection, and at the end of the 6-month, after the week the activity monitors were worn. The CHAMPS has shown reasonable validity to measure PA in older adults [19,21,44]

Stability of PA measures: Subjects who agree to participate in the reliability portion of this study were asked to compare the activities that they performed during the 2nd week to the 1st week of wearing the monitors. They had the options to answer if they were more active, less active, or about the same.

Global Rating of Change in PA: Subjects completed a modified global rating of change regarding change in PA status from the time they entered the study to the follow-up assessment. Subjects rated how they perceived the overall change in PA since the beginning of intervention using a 15-point rating scale described by Jaeschke et al.[91] The global rating of change ranges from -7 ("a very great deal less PA") to 0 ("about the same") to +7 ("a very great deal more PA"). Intermittent descriptors of less or more are assigned values from -1 to -7 and from +1 to +7, respectively.

Additional Measures: Demographics, self-reported physical function (measured by the Western Ontario and McMaster Universities Osteoarthritis Index), quality of life (measured by the SF-36), BMI, and blood pressure, were collected from the parent trial and were used to characterize the sample.

1.7.4 Interventions Investigated in Parent Trial

Subjects were masked from the detailed content of the interventions or study purposes (some general information were provided according to IRB requirements). The exercise interventions were delivered by a physical therapist (GJA) and last for 6 months. Both interventions were designed to improve aerobic capacity and muscle strengthening. It was expected that the SPE group would demonstrate small increase whereas the CBI group would demonstrate moderate increase in PA. Therefore, across subjects in both groups, we expected a large variability in changes in measures of PA, providing a good sample to test the responsiveness of these measures. The interventions' plan is represented in Figure 4.

1.7.4.1 Control Group - Standard Physical Exercise Program (SPE)

The SPE represents the typical rehabilitation after TKA surgery. It was expected to provide small improvement in function and PA.[92] Subjects participated in 12 supervised sessions delivered during 12 weeks, followed by a home exercise program performed 2x/week at the same intensity as done under supervision up to 6 months. The SPE consists of: (a) lower extremity range of motion and stretching exercises, (b) lower extremity strengthening exercises, and (c) endurance exercises using treadmill or stationary bicycle. We followed the ACSM guidelines for older adults for dosing and progression of exercises.[93] Strengthening exercises were progressed from light to moderate loading (40% to 60% of 1 repetition maximum). To assess intensity, for each exercise, we started by adding a light but reasonable load and asked the subject to perform 30 repetitions, which corresponds to 40-50% of maximum strength. Additionally, we asked the subject to grade his/her effort according to a 10-point exertion scale, which must have fallen between 3 to 4, representing light-intensity. [93] Progression to a moderate intensity was made by adding more load and decreasing the repetitions up to 15, with an effort perception not more than 6 points, representing moderate intensity. For the endurance training performed on the treadmill or stationary bicycle, the subject was required to maintain his/her heart rate between 50% and 60% of the predicted value (220 - age), representing light to moderate intensity.

1.7.4.2 Experimental Group - Comprehensive Behavioral Intervention (CBI)

The CBI consisted of regular contacts over 6 month period. It included 2 weekly contacts for weeks 1-6, weekly contacts for weeks 7-8, biweekly contact from months 3 to 4, and once a month from months 5 to 6. There was a combination of 12 supervised exercise sessions, 7 self-management and one nutrition-guidance session, for a total of 20 sessions. The CBI was composed of 4 components:

a. The evidence-based exercise program was based on published research after TKA.[94-96] It combined (a) lower extremity strengthening exercises; (b) endurance exercise on the treadmill; (c) functional task-oriented exercises with exercises such as chair rises and stair climbing; (d) balance techniques such side stepping, forward and backward tandem walking, and single-leg-standing balance. During the 12 supervised exercise sessions (delivered during 12 weeks), subjects were instructed to become independent in performing the exercises at home. We followed the ACSM guidelines for older adults for dosing and progression of exercises.[93] Strengthening exercises were progressed from moderate to high loading (60% to 80% of one repetition maximum). To assess intensity, for each exercise, we started by adding a moderate load and asked the subject to perform 17 repetitions at maximum fatigue, which corresponded to 60% of maximum strength. Additionally, we asked the subject to grade his/her effort according to a 10-point exertion scale, which must have fallen between 5 to 6, representing moderate intensity. [93] Progression to a vigorous intensity was made by using stronger elastic bands and decreasing the number of repetitions up to 8 to maximum fatigue. Reported exertion score was not greater than 8 points, representing vigorous intensity. For the endurance training performed on the treadmill or stationary bicycle, the subject was encouraged to exercise at moderate intensity. Recent studies demonstrated that the effects of cardiorespiratory fitness are comparable between moderate and high intensities. [97,98] Subjects were required to maintain their heart rate between 55% and 70% of the predicted value (220 - age), representing moderate intensity. To avoid excessive stresses over the prosthesis, subjects were limited to walk up to 4.5 mph. Exercise behavior were monitored by diaries.

b. Physical activity promotion was based on the program that has been previously developed and tested in overweight and sedentary women. [99,100] Subjects were instructed to engage in moderate or higher intensity exercise for 5 days a week. They began at 20 min or less of walking per day, and

gradually progressed to 60 min per day, as recommended by leading organizations. [101,102] Progression was gradual (5-10 min/d in 4 week intervals) to maximize adherence and minimize musculoskeletal injury. Physical activity instructions were delivered by physical therapist along with the 12 sessions of supervised exercise program. Physical activity was monitored daily by the SWA, and downloaded into the Professional Software. Subjects wore a wrist watch that synchronized with the SWA and provided real-time feedback of minutes of PA, allowing them to alter their PA behaviors to achieve the intervention goals.

c. Healthy nutrition guidance: Subjects received instructions on healthy nutrition based on the Dietary Guidelines for Americans. Only obese subjects were prescribed an energy restricted diet that has demonstrated to be effective in reducing weight. [100,103] For subjects attempting to lose weight, dietary intake was monitored using food diaries. Overweight (not obese) individuals were not offered a restricted diet since in older adults some extra weight may be beneficial. [104] This intervention was delivered by a registered dietitian in small group sessions (one session for all subjects and an additional session for the obese subjects). Body weight was frequently measured at each session and subjects were encouraged to measure it on their own.

d. Self-management integrated the basic self-management skills advocated by the Arthritis Foundation Self-Help Program and behavioral strategies such as self-monitoring, problem solving, relapse prevention, and goal-setting and feedback into approaches a, b and c from above. [105,106] Subjects were provided educational written materials to supplement the individual or group sessions and to develop the behavioral skills necessary for appropriate exercises, sufficient physical activity, and adequate nutrition. Discussion was facilitated by the interventionist. Monitoring was also an important part of self-management. Exercises, PA data, and body weight were regularly reviewed by the research staff and information was used to provide feedback of accomplishments, generate

discussion to revise goals, and discuss problem solving strategies related to the CBI. Seven sessions were devoted to self-management intervention.

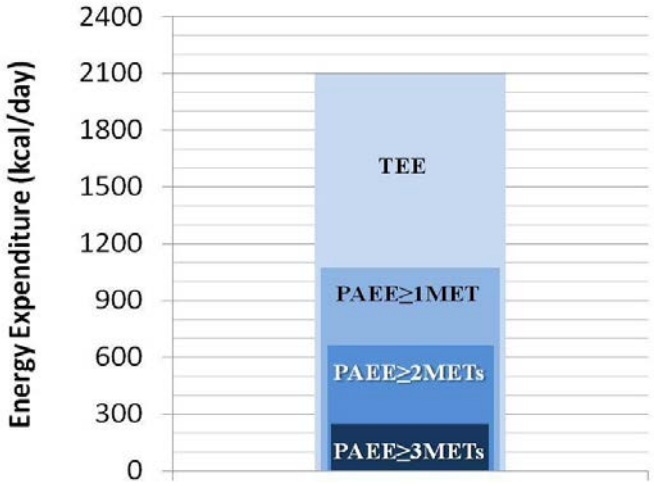


Figure 3. Schematic representation of daily values of energy expenditure.

1.8 ANALYSIS PLAN

Descriptive statistics including means, 95% confidence intervals (CI) and standard deviations, median and 25-75 percentiles, and counts and frequencies were used according to the type and distribution of data. Summaries were provided for subjects' characteristics and measures of PA at baseline (week 1 and week 2) and 6-month follow-up.

1.8.1 Hypothesis 1

Test-retest reliability of PA measures taken during 2 consecutive weeks will show good consistency for the three instruments, with intraclass correlation (ICC) coefficients above 0.70. Data from subjects who agreed to

perform similar activities for two consecutive weeks will be used in the reliability analysis. Test-retest reliability of PA measured by the SWA, ACT and CHAMPS will be determined by ICC. The model $ICC_{2,k}$ will be used, where '2' reflects the test-retest and 'k' the average of all days containing PA data. ICCs will be determined using the absolute agreement definition for variables generated by the ACT and SWA (PAEE≥2METs, PAEE=2-3METs, PAEE≥3METs and number of steps), and CHAMPS questionnaire (PAEE≥2METs and PAEE≥3METs). The 95% confidence intervals around the ICC values will be calculated. We will also use Bland and Altman plots to compare the mean difference between two methods of measurement, and to compute the 95% limits of agreement to allow assessment of systematic difference between the measurements and to identify possible outliers. Measurement error will be calculated as the standard error of measurement (SEM), which is based on the reliability coefficient (ICC) and variance (SD) of measures from week 1 and week 2 from each PA measurement tool using the equation: SEM = SD √1-ICC. We will also use the SEM to derive the minimum detectable change (MDC) for each PA measurement tool. The MDC is calculated as the amount of change needed to be certain, within a defined level of statistical confidence that change is beyond measurement error. [107] The equation 'MDC = χ score level of confidence x √2 x SEM' was used. [108] MDCs were calculated using the standard normal scores of 1.96 (associated with 95% CI - MDC95) and 1.65 (associated with 90% CI - MDC90).

1.8.2 Hypothesis 2a

Measures of PA by the ACT, SWA and CHAMPS will identify significant increase in moderate and light intensity PA from pre to post-intervention (6 months). Significance of changes in PA measured by the ACT, SWA and CHAMPS from pre to post intervention will be calculated using paired t-test. We will perform

the same approach for light and moderate levels of PA. All hypothesis testing will be tested at α=0.05.

1.8.3 Hypothesis 2b

The SWA will be more sensitive to changes in measures of PA than the ACT, and the ACT will be more sensitive to detect changes in measures of PA than the CHAMPS. To test this hypothesis, we will perform two analyses. First, we will use paired t-test to compare the change-scores at 6-month follow-up between each pair of PA measures. (ACT vs. SWA, ACT vs. CHAMPS, and SWA vs. CHAMPS). The same approach will be performed for light and moderate levels of PA. Graphs will be plotted to compare slopes of lines from pre to post intervention for all PA levels measured by each instrument. Second, we will calculate the standardized response mean (SRM) and Guyatt responsiveness index (GRI) to identify which PA measurement tool was more responsive to change after an exercise intervention. The SRM is the ratio of mean change to the standard deviation of the change scores,[108] and the GRI is the ratio of mean change divided by the standard deviation of change in subjects who remained unchanged or worsened.[7]

1.8.4 Hypothesis 3

Changes in PA measured by the modified global rating of change will be associated with changes in PA measured by the ACT, SWA and CHAMPS (correlation coefficients of at least moderate magnitude), which will support the use of the global rating of change as an external anchor to determine thresholds of clinically important changes in measures of PA. Correlations between the modified global rating of change and changes in PA measured by

the ACT, SWA and CHAMPS, will be calculated using Pearson or Spearman correlation coefficient (depending on data distribution). If the association between global rating scores and changes in measures of PA are of at least moderate magnitude, we will attempt to derive cut-offs of clinically important improvement using the method of the area under the receiver operating characteristic (ROC) curve. [14] If associations between global rating scores and changes in measures of PA are below moderate magnitude it will indicate that the modified global rating of change scale may not be appropriate as an external anchor for external responsiveness of PA measures.

1.8.5 Sample Size Estimation

We proposed to recruit 50 subjects to allow for 40 subjects at 6 months (assuming 20% attrition at 6 months). With 40 subjects, $\alpha = 0.05$, we will have 87% power to detect an ICC= 0.75 in the test-retest reliability, with a lower bound of 95%CI as low as 0.40 for all instruments. For the paired t-test, 10 subjects per group were required to detect a small improvement in PA from pre to post intervention. Sample size of 40 would allow above 100% power to identify significant difference in the paired t-test to test the significance of change of each instrument from pre to post, and also to compare the significance of changes among the instruments.

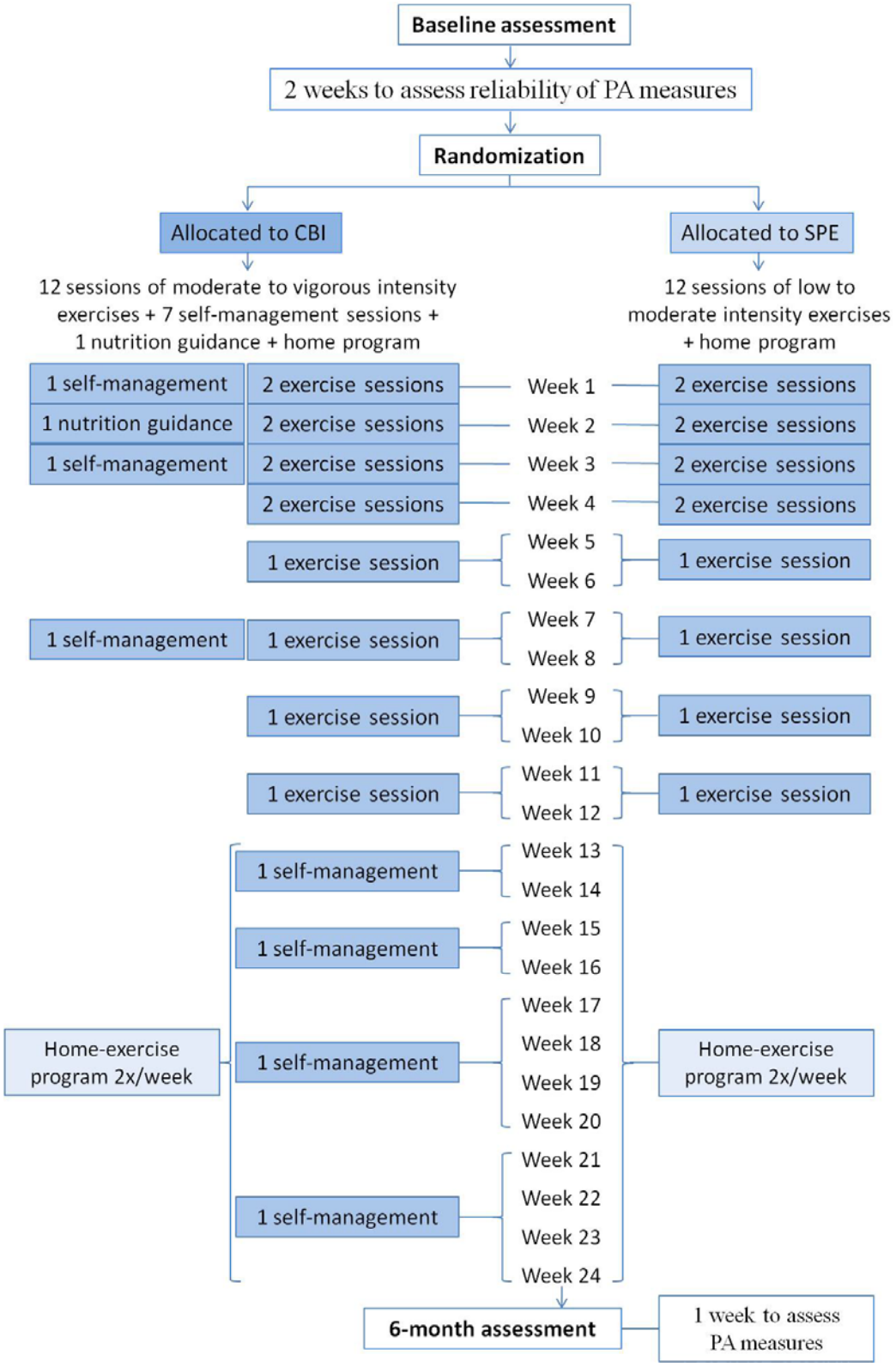


Figure 4. Representation of study flow with the distribution of interventions through 6 months.

2.0 CHAPTER 2 - RELIABILITY OF PHYSICAL ACTIVITY MEASURES DURING FREE-LIVING IN INDIVIDUALS AFTER TOTAL KNEE ARTHROPLASTY

2.1 SUMMARY

Background: Few instruments that measure physical activity (PA) can accurately quantify PA performed at light and moderate intensities, which is particularly relevant in older adults. However, the evidence of their reliability in free-living condition is limited. **Objectives**: To determine test-retest reliability of the Actigraph (ACT), SenseWear Armband (SWA) and the Community Healthy Activities Model Program for Seniors (CHAMPS) questionnaire in assessing free-living PA at light- and moderate-intensities in individuals following total knee arthroplasty (TKA), and compare their reliability in commonly used monitoring timeframes ('24-hours', 'wake-hours', and '10-hours from awakening'). **Design**: One-group, repeated-measures. **Methods**: Subjects wore the activity monitors for two weeks. At the end of each week data were downloaded and the CHAMPS was completed. Measures from week one were compared to those of week two. **Results**: Twenty-eight subjects reported similar PA during the two weeks and were included in the analysis. Mean age was 69(8) years and 75% were female. Test-retest reliability determined by intra-class correlation coefficient (ICC) ranged from moderate to excellent for the ACT (ICC=0.75-0.86), and excellent for the SWA (ICC=0.93-0.95) and CHAMPS (ICC=0.86-0.92). The 95% confidence intervals of the ICCs from SWA were the only ones within the range of excellent reliability (0.85-0.98). The CHAMPS

demonstrated systematic bias with less PA reported in week two. Reliability of PA measures in 'wake-hours' timeframe was comparable to '24-hours' and reflected most PA performed during this period. **Limitations**: Reliability may be lower for time-intervals longer than one week. **Conclusions**: All PA measures had good reliability. Reliability of the ACT was lower than the SWA and CHAMPS. The SWA provided better precision of reliability estimates. Wearing PA monitors during wake-hours provided reliable measures and can reduce subjects' burden.

2.2 INTRODUCTION

Regular physical activity (PA) contributes to health enhancement in many ways. It helps maintain weight, prevents chronic diseases such as diabetes, depression and hypertension, and reduces mortality.[109] Individuals who undergo total knee arthroplasty (TKA) for end-stage knee osteoarthritis are typically older adults who have an inactive lifestyle and perform most of their activities at light intensities due to persistent knee pain and functional limitations. [110-112] Thus, measuring PA using reliable measurement tools able to capture light intensity activities is warranted.

Several instruments are available to measure free-living PA in older adults, which is defined as daily activities performed by the individuals at their own pace under no supervised or controlled conditions.[113,114] But only a few instruments are able to accurately distinguish light (e.g., mopping and walking to do errands) from moderate (e.g., lawn mowing and brisk walking) intensities of PA.[19,48,115-118] Among those instruments, two accelerometer-based activity monitors, the Actigraph (ACT; Actigraph LLC, Pensacola, FL) and the Sensewear Armband (SWA; Bodymedia, Pittsburgh, PA), and a self-reported questionnaire, the Community Health Activities Model Program for Seniors

(CHAMPS), have been validated to measure PA in older adults and are widely used in research that include this population.[43,44,84] The ACT is a waist-mounted accelerometer-based device and the SWA is an arm-mounted multi-sensor device. Their advantage is to measure PA in real time as compared to CHAMPS that relies on individuals' recall. Although the costs of activity monitors are substantial compared to questionnaires, their cost is reasonable for large-scale research. Also, they are far less expensive than doubly labeled water that is the reference standard to measure PA in free-living. A drawback of activity monitors is that they need to be worn for several hours per day and several days a week to capture relevant PA information. The advantages of the CHAMPS include information about the type of PA behaviors (e.g., dance, walk to do errands), very low cost, and it is easy to administer with minimal subjects' burden (15 minutes to complete). Yet, PA questionnaires have been found to overestimate moderate PA and underestimate sedentary behavior.[48,119]

To date, there is limited information on the reliability of these instruments to measure PA in older adults with arthritis of the lower extremities. To our knowledge, no studies have determined the reliability of the SWA in free-living conditions. Studies that used the ACT and CHAMPS to measure free-living PA were mainly in healthy individuals.[19,44,46,48,70,120] Results from healthy individuals may not be applicable to those with knee osteoarthritis who usually have pain and functional limitations, which may lead to gait abnormalities, inactive lifestyle, and obesity.[1,121] More importantly, studies have not compared the performance of the ACT, SWA and CHAMPS. Therefore, it is key to concurrently compare the reliability between measures of PA from activity monitors that use distinct technologies and a self-report questionnaire. This comparison will provide evidence for researchers and clinicians for a well-informed choice of tools to assess PA in individuals with arthritis of the lower extremity.

Specific to accelerometry-based devices, evidence is also limited in terms of the appropriate timeframe to wear the monitors. Most of the studies that investigated the consistency of measures of PA in different monitoring timeframes have done so in terms of number of days, but few on the amount of hours per day.[114,122,123,124] The most commonly used monitoring timeframes in research to assess free-living PA have been 10-hours, wake-hours, and 24-hours.[114,122,123,124] To investigate which monitoring timeframe provides more reliable estimates of PA may help reduce subjects' burden if shorter monitoring periods yield consistent PA measures.

The main purpose of this study was to determine the test-retest reliability of the ACT, SWA, and CHAMPS to assess participation in free-living PA at light- and moderate-intensities in individuals with knee osteoarthritis who underwent TKA. We also aimed to compare the reliability across the three instruments. Additionally, we aimed to determine test-retest reliability of monitoring timeframes that are commonly used to assess PA when using the ACT and SWA.

2.3 METHODS

2.3.1 Design and Subjects

This is an ancillary study that used a one group, repeated-measures design. Subjects participating in a randomized trial that investigated the effect of rehabilitation approaches to improve quality of life in individuals following TKA were invited during baseline assessment to take part in this reliability study. This study took place from October 2011 to March 2013 in the Department of Physical Therapy, University of Pittsburgh. All subjects signed a consent form approved by the University of Pittsburgh Institutional Review Board prior to participation.

In the parent study, individuals were included if they had a unilateral TKA in the past 3-6 months and were at least 50 years old. To be included in this ancillary study, individuals also had to be willing to wear the activity monitors for two consecutive weeks, to complete the CHAMPS twice, and to agree to perform similar activities during the two weeks of data collection. For safety reasons, individuals were excluded from the parent study if they reported more than two falls within the past year, were unable to ambulate a distance of 31 meters without an assistive device, and had medical conditions that precluded safe exercise participation.

2.3.2 Study Protocol

Subjects attended three testing visits. In the first testing visit they completed demographic, pain and physical function self-reported questionnaires, and had their height and weight measured. Knee pain was assessed by a numeric pain scale that ranges from 0 (no pain) to 10 (worse imaginable pain). Physical function was measured using the Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function (WOMAC-PF). The WOMAC-PF is a self-report questionnaire that includes 17 items, each scored from 0 (no limitation) to 4 (extreme limitation), with a total score of up to 68 points. Higher scores indicate worse function. The WOMAC-PF has shown adequate test-retest reliability (associations ranged from 0.68 to 0.92) [125] and validity when compared to measures of patient satisfaction, and health-related quality-of-life such as the SF-36 (associations ranged from 0.48 to 0.64) [125] in people with knee and/or hip osteoarthritis.

At the end of the first visit, subjects were fitted with the ACT and SWA and were instructed to wear the monitors for 7 days (24 hours a day), except during water activities since the activity monitors are not waterproof. Subjects were asked to perform similar activities during the two weeks

of data collection. At the end of week one (test) and week two (retest), subjects returned to download the data from the activity monitors and to complete the CHAMPS (Figure 5). The CHAMPS questionnaire queries PA participation in the past week, which corresponds to the time they wore the monitors. Data from the activity monitors were inspected to assure sufficient data, which was defined as at least 5 days with 10-hour of PA data per day. [122,126,127] Subjects without sufficient data were asked to wear the portable monitors for an additional week and to complete the CHAMPS at the end of that week. To assess stability of PA during the two weeks of data collection subjects were asked to compare their PA participation between the two weeks by answering if they performed more, less, or about the same activities during that period. Data from subjects who reported about the same activities during two weeks, and had at least 5 days with 10-hour data for the ACT and SWA were used in the reliability analysis.

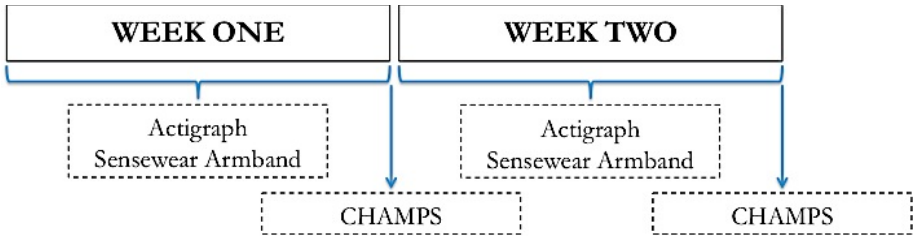


Figure 5. Flow chart of the test-retest reliability protocol.

2.3.3 Measures of Physical Activity

The main outcome of this study was duration and caloric expenditure of daily PA estimated by the ACT, SWA and CHAMPS. We also assessed daily number of steps using the ACT and SWA. The PA measurement tools in use in this study have distinct measurement characteristics to differentiate the intensity levels of PA into light (e.g., recreational activities, slow walk), moderate (e.g., vacuuming, brisk walk), and vigorous (e.g., lawn mowing, doubles tennis). The measures of moderate and vigorous-intensity activities were combined into the moderate category since our sample engaged in negligible amounts of vigorous-intensity activities. The intensity categories of PA compared across the three instruments in this study were light, moderate, and light-to-moderate (combination of light and moderate intensities).

2.3.3.1 Actigraph

The ACT is a small triaxial accelerometer-based monitor (2 x 1.5 x 0.6 inches in size) worn around the waist, at the level of the hip-bone, over the right anterior superior iliac spine. In this study we used the ACT model GT3X and the ActiLife 5 software (Actigraph LLC, Pensacola, FL). The ACT measures body acceleration in activity counts, and was set to collect data at 1-minute intervals. The ACT generates data on activity counts per minute (CPM) and number of steps. To categorize duration of daily activities in minute per day (min/day), the software uses the following CPM cut-points: 760-1951 CPM for light, CPM >1951 for moderate, and ≥760 CPM for light-to-moderate intensities. [128]

The ACT also uses body weight to generate measures of caloric expenditure in kilocalories per day (kcal/day). Freedson equation is used to calculate kcal/day for moderate intensity activities

whereas the Combination equation is used to calculate kcal/day for light-to-moderate intensity PA. [128] To generate values for light intensity activities, we subtracted the values of Freedson equation from Combination equation as recommended by the manufacturer. The ActiLife 5 software identifies non-wear periods using a threshold of 60 consecutive minutes with zero CPM that indicates no movement at all, and allows for up to two minutes of 1-100 CPM in between the 60 minutes. Non-wear periods were also visually analyzed. Measures of PA in free-living from the ACT have demonstrated moderate accuracy (r= 0.49) with doubly labeled water in older adults, [43] and moderate to excellent reliability to assess PA in free-living in healthy adults (ICC= 0.68 to 0.90). [70,120]

2.3.3.2 SenseWear Armband

The SWA is a small multi-sensor device (3.4 x 2.1 x 0.8 inches in size) worn on the back of the right arm, over the triceps muscle, at midpoint between shoulder and elbow. We used the SWA model Pro-3 and the Professional software v6.1 (BodyMedia Inc., Pittsburgh, PA). The SWA combines information from biaxial accelerometer, heat flux (heat being dissipated by the body), galvanic signal (onset, peak and recovery of maximal sweat rates) and skin temperature. The information from the sensors is integrated and processed by the software using proprietary algorithms that account for subjects' demographic characteristics (gender, age, height and weight).

The SWA was set to provide data on number of steps, duration of activities (min/day), and caloric expenditure (kcal/day) in light (2 to 2.9 metabolic equivalents [METs]), moderate (≥3 METs), and light-to-moderate intensities (≥2 METs). The SWA turns off automatically when not in contact with the skin which enables the software to recognize periods of non-wear. Data were also visually screened to identify non-wear periods. The SWA has shown moderate accuracy (r= 0.48) in comparison to doubly labeled water in older adults. [43]

2.3.3.3 Community Health Activities Model Program for Seniors

CHAMPS is a self-reported questionnaire that queries type, frequency and duration of 41 activities usually performed by older adults, ranging from light to vigorous intensities, such as cleaning, gardening and sports activities. Duration in hours per week (hr/week) of each activity is selected from a range of less than one hour per week to 9 or more hours per week, and categorized in two intensity levels according to the CHAMPS activities codebook. [44] Light-to-moderate intensities (≥2 METs) represent all exercise-related activities queried by the questionnaire, such as light to heavy household chores, and recreational and sports activities. Moderate intensity activities (≥3 METs) represent activities such as heavy household chores, calisthenics and sports. To allow direct comparison between CHAMPS and the accelerometer-based monitors, the moderate category was subtracted from light-to-moderate to create a light intensity category for the CHAMPS.

The caloric expenditure from CHAMPS is provided in kcal/week using equations that weigh the intensity of each activity according to the energy requirements for older adults, and also accounts for subjects' body weight. Kcal/week is calculated using the type, frequency, and duration of each activity reported.[44] Scores from the CHAMPS in hr/week or kcal/week were converted into min/day and kcal/day respectively by multiplying the duration score by 60 minutes and then dividing by 7 days. Data on water activities described by the CHAMPS (i.e., items 31 to 33) were excluded from the calculation of PA scores since subjects were required to remove the activity monitors during those activities. CHAMPS has shown reasonable accuracy in comparison to doubly labeled water to measure PA in older adults (r= 0.28), [43] and fair to good reliability (ICC= 0.27 to 0.84) when measuring PA in people with musculoskeletal disorder and healthy older adults. [19,43,44,46-48]

2.3.4 Reliability of Monitoring Timeframes

Test-retest reliability of the monitoring timeframes of the ACT and SWA used three pre-specified timeframes: '24-hours', 'wake-hours' and '10-hour from awakening'. These timeframes are commonly used in research using activity monitors to assess free-living PA.[114] The '24-hours' corresponded a whole day of activities, including sleeping time. The 'wake-hours' was identified from wake-up time to sleep time. The '10-hour from awakening' was determined by counting 10 hours from wake-up time. Wake-up time and sleep time were identified using the SWA. The SWA has shown to be accurate to detect sleep in comparison to the reference standard of polysomnography. [129-131] The same times identified using the SWA software were used to identify wake-up and sleep times on the ACT. Subjects were entered in the timeframe reliability analysis only if they had accelerometry data during '24-hours' and if the monitoring days and times from the ACT matched those identified using the SWA. This approach was used to assure that both monitors were measuring the same activities performed during the same time. The matching approach is illustrated in Figure 6 using data from one subject.

2.3.5 Data Analysis

Descriptive statistics for continuous variables included mean and standard deviation or median and 25-75 percentiles, according to data distribution. Counts and frequencies were used for categorical variables. Demographics and biomedical characteristics between subjects who reported about the same activities during the two weeks and those who reported more/less activities in week two were

compared using Independent Samples T-test or Mann-Whitney U test for continuous data, and Chi-square test for categorical data.

	SenseWear Armband				Actigraph			
	Week One		**Week Two**		**Week One**		**Week Two**	
	Days	**Wear Time**	**Days**	**Wear Time**	**Days**	**Wear Time**	**Days**	**Wear Time**
Subject ID XX	1	10:24 to 23:59	1	**15:04 to 23:59**	1	10:00 to 23:59	1	**15:00 to 23:59**
	2	0:00 to 23:59	2	0:00 to 23:59	2	0:00 to 23:59	2	0:00 to 23:59
	3	0:00 to 23:59	3	0:00 to 23:59	3	0:00 to 23:59	3	1:00 to 23:59
	4	0:00 to 19:50	4	0:00 to 18:24	4	0:00 to 20:59	4	11:00 to 23:59
	5	4:51 to 23:01	5	0:00 to 21:59	5	4:00 to 23:59	5	**NO DATA**
	6	8:25 to 21:16	6	**18:25 to 23:48**	6	0:00 to 23:59	6	4:00 to 23:59
	7	4:47 to 20:26	7	4:59 to 23:59	7	0:00 to 22:59	7	0:00 to 23:59
	8	0:32 to 16:51	8	0:00 to 16:20	8	4:00 to 16:59	8	0:00 to 16:59

The rows highlighted in red indicate days excluded from the analyses. Day 1 was excluded because neither SWA or ACT had sufficient data (<10 hours) during week two. Data from day 5 was not used because the ACT did not collect data in week two. Day 6 was excluded as data from the SWA was not sufficient in week two. The remaining 5 days were used for the timeframe reliability analysis.

Figure 6. Illustration of the approach used to match days of monitoring time between data obtained from the SenseWear Armband (SWA) and Actigraph (ACT) during 24 hours.

Test-retest reliability of PA measures from week one and week two was determined by intra-class correlation coefficient ($ICC_{2,k}$) using absolute agreement. As a general guideline for interpretation, ICC<0.5 means poor reliability, ICC between 0.50 to 0.75 represents moderate reliability, and ICC>0.75 indicates excellent reliability.[108] Measurement error was estimated using the Standard Error of Measurement (SEM). The SEM is calculated based on the reliability coefficient (ICC) and variance (SD) of PA measures between week one and two using the equation: SEM = SD √1-ICC. We also calculated the Minimum Detectable Change (MDC) to provide a threshold within a defined level of statistical confidence that change is beyond measurement error using the following equation: MDC = z score level of confidence x √2 x SEM. [107,108] We used the z scores of 1.96 (associated with 95% confidence - MDC_{95}) and 1.65 (associated with 90% confidence - MDC_{90}). Bland-Altman plots were used to compare the mean difference between the two weeks of measurements, and to assess for systematic bias and outliers.

To compare the differences between the reliability of PA measures across the ACT, SWA and CHAMPS, we examined if the point estimates of the ICC of one of the instruments was contained in the 95% confidence intervals of the ICC from another instrument. If the point estimate is not within the 95% confidence interval, than the reliability of one instrument is statistically significant different from the other. [132]

We also explored if the magnitude of measures of PA differed between the three measurement tools. Paired T-test or Wilcoxon signed-rank test were used according to data distribution (ACT vs. SWA, ACT vs. CHAMPS, and SWA vs. CHAMPS). Only data from week one was used for this analysis. Alpha level was set at 0.05 for all analysis. No attempt was made to correct for multiple comparisons in order to minimize type II error. All the analyses were performed

using the IBM SPSS Statistics 21 (IBM Corporation), and Microsoft Excel 2010 (Microsoft Corporation).

2.4 RESULTS

We invited all 44 individuals from the parent trial to participate in this ancillary study. Two subjects declined participation because they were not willing to wear the activity monitors for two weeks. From the remaining 42 subjects, 7 did not have sufficient data for one of the weeks and were asked to wear the monitors for an additional week. Five of these subjects were excluded due to insufficient data even after wearing the devices for an additional week. Of the remaining 37 subjects, 9 reported more/less activities in week two, leaving 28 subjects with similar PA during both weeks for the reliability analysis (Figure 7). On average, The monitoring wear time for these subjects was 21±3 hours and it was similar during both weeks. The characteristics of subjects who participated in the reliability analysis and the ones excluded from the analysis are compared in Table 1.

Table 2 shows the characterization of PA during the two weeks along with the results from test-retest reliability, SEM, and MDC. The three measurement tools showed good to excellent test-retest reliability (ICC≥ 0.75), with confidence intervals within the range of moderate to excellent reliability (95% CI of ICCs: 0.43 to 0.98). Bland-Altman plots revealed no systematic bias between measures from the two weeks for the ACT and SWA across intensity categories (Figures 8 and 9). However, the plots demonstrated systematic bias for CHAMPS scores in duration of light intensity PA and light-to-moderate PA. The line of equality (zero) was not contained within the 95% CI of the difference between weeks, and indicated that PA values from week one were significantly higher

than those of week two. These differences concur with the significant F-tests from Analysis of Variance (from ICC calculation) that tests the difference between PA measures from week one and week two (Table 2).

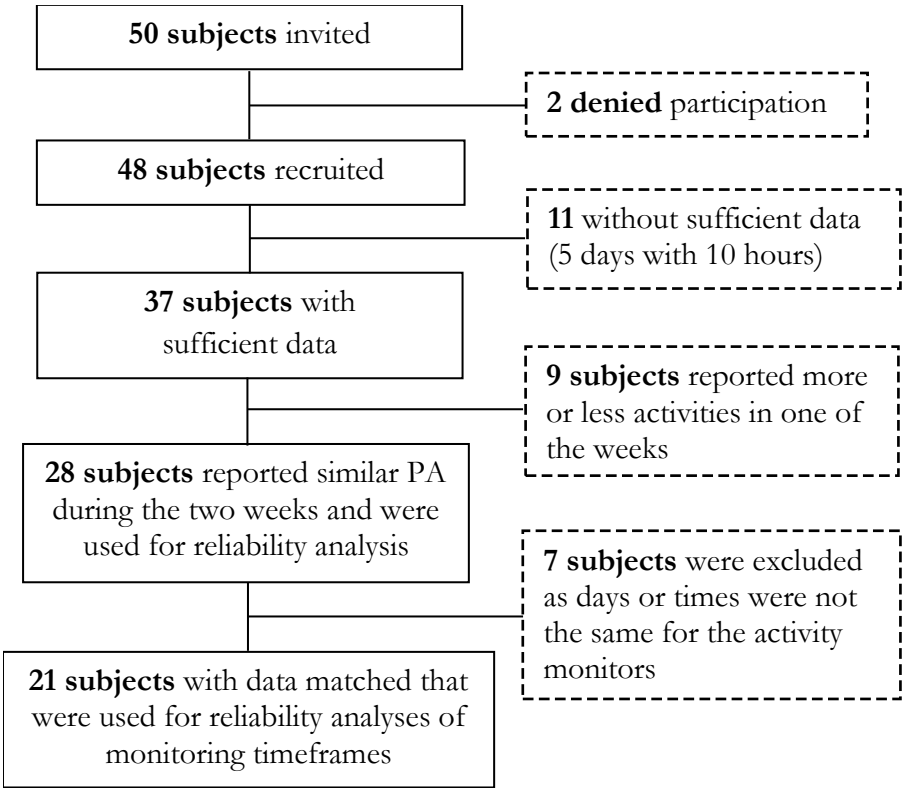


Figure 7. Flow chart of subjects included in the final analysis.

Table 1. Demographic and biomedical characteristics of subjects who participated in the reliability analysis and subjects excluded from this analysis. Numbers represent mean ± SD, unless otherwise indicated.

	Participated in reliability analysis (n= 28)	**Excluded from reliability analysis (n= 22)**	**p-value**
Age in years	68.5 ± 7.4	66.4 ± 5.3	0.383
Sex – female (%)	22 (76)	12 (55)	0.453
BMI in kg/m^2	29.2 ± 3.8	30.0 ± 4.1	0.836
Race – white (%)	27 (93)	18 (82)	0.570
Education – n (%)			0.816
High-school	12 (41)	8 (36)	
College	17 (59)	14 (64)	
1 to 10 years	21 (72)	14 (64)	
More than 10 years	8 (28)	8 (36)	
Time from TKA – n (%)			0.948
3 to 4.9 months	20 (69)	16 (72)	
5 to 6 months	9 (31)	6 (28)	
Knee pain – 0 to 10; median (Q25; Q75)			
Surgical side	2 (1; 3)	3 (0; 3)	0.790
Non-surgical side	3 (0; 6)	2 (0; 3)	0.624
WOMAC-PF	19.5 ± 9.1	18.9 ± 12.9	0.308

P-value of the differences between those who participated and those who were excluded from reliability analysis; BMI: body mass index; TKA: total knee arthroplasty; Q25: quartile 25th; Q75: quartile 75th; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function sub-scale.

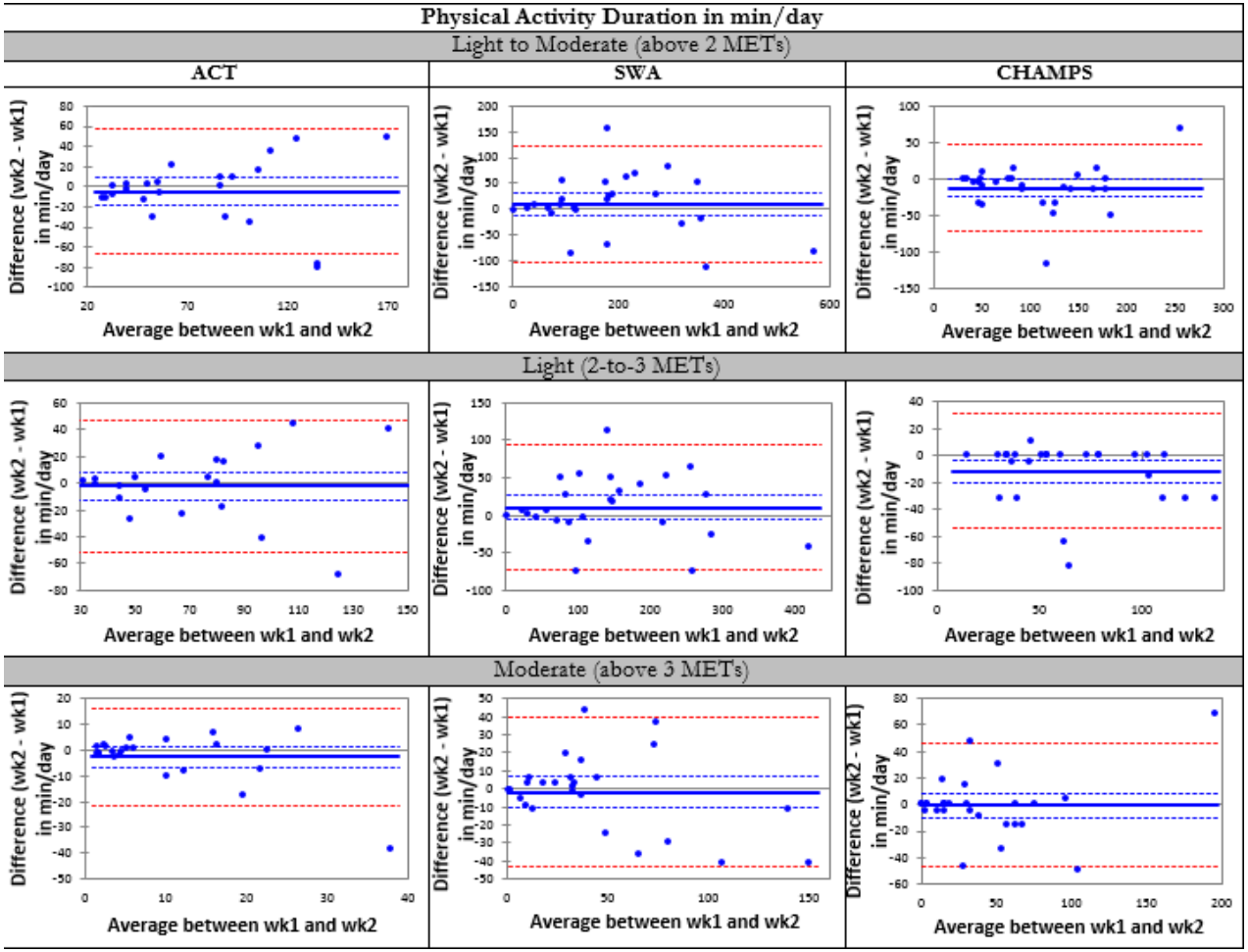
Table 2. Average duration of physical activity (min/day) and caloric expenditure (kcal/day) measured by the Actigraph (ACT), Sensewear Armband (SWA) and CHAMPS questionnaire during two weeks. Intra-class correlation coefficient (ICC) with its corresponding 95% confidence interval (95% CI), standard error of measurement (SEM), and minimal detectable change (MDC).

		PA intensity	Week 1	Week 2	Diff (95% CI)	Significance F-test	ICC (95% CI)	SEM	MDC 90%	MDC 95%
ACT	Min/day	Light	67.3±33.0	65.0±36.3	2.2 (-8.5; 13.0)	0.670	0.86* (0.66; 0.94)	13.0	30.3	36.0
		Moderate	11.7±13.0	9.0±8.2	2.6 (-1.5; 6.8)	0.200	0.75* (0.43; 0.89)	5.3	12.4	14.7
		Light to moderate	78.9±42.9	74.0±43.2	4.9 (-8.8; 18.6)	.467	0.85* (0.64; 0.94)	16.9	39.4	46.8
	Kcal/day	Light	249.7±95.0	246.0±103.4	3.7 (-24.8; 32.2)	0.790	0.88 (0.72; 0.95)	34.4	80.2	95.3
		Moderate	73.4±85.0	56.5±60.5	16.9 (-10.5; 44.3)	0.213	0.77 (0.47; 0.90)	34.9	81.4	96.7
		Light to moderate	323.1±162.0	302.4±153.6	20.6 (-27.4; 68.6)	0.382	0.86 (0.67; 0.94)	59.0	137.8	163.7
	Number of Steps		4676.8±2031.6	4413.9±1693.9	262.9 (-368.2; 894.1)	0.394	0.85* (0.63; 0.94)	709.2	1654.8	1965.7
SWA	Min/day	Light	138.9 ± 102.8	149.7 ± 98.7	-10.8 (-27.9; 6.3)	0.205	0.95§ (0.90; 0.98)	22.5	52.6	62.5
		Moderate	45.3 ± 44.7	43.5 ± 36.6	1.8 (-6.7; 10.3)	0.669	0.93 (0.85; 0.97)	10.8	25.1	29.8
		Light to moderate	184.3 ± 138.8	193.2 ± 127.4	-9.0 (-32.1; 14.2)	0.433	0.95 (0.89; 0.98)	29.8	69.5	82.5
	Kcal/day	Light	400.9 ± 266.4	445.7 ± 280.5	-44.9 (-102.1; 12.3)	0.119	0.93§ (0.83; 0.97)	77.3	180.5	214.4
		Moderate	213.7 ± 199.7	205.7 ± 180.4	8.0 (-40.5; 56.4)	0.738	0.89 (0.76; 0.95)	63.0	147.1	174.7
		Light to moderate	614.8 ± 427.0	651.5 ± 421.0	-36.7 (-125.2; 51.8)	0.401	0.93 (0.84; 0.97)	112.2	261.8	311.0
	Number of Steps		6261.8±3647.5	6371.5±3432.3	-109.7 (-763.2; 543.8)	0.166	0.95 (0.88; 0.98)	936.6	2185.5	2596.1

Table 2 (continued)

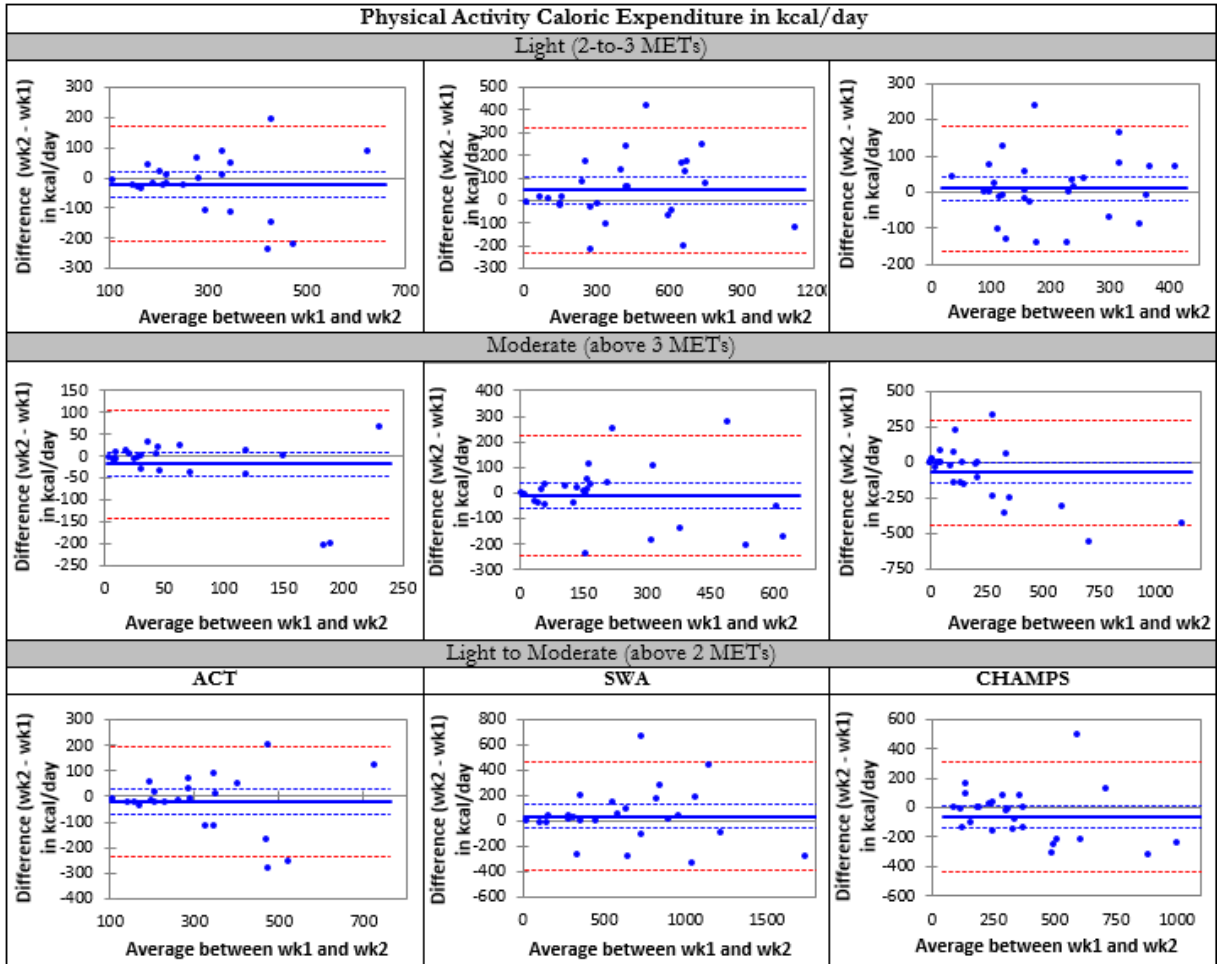
CHAMPS	**Min/day**	**Light**	67.8±37.4	58.5±31.6	11.4 (3.0; 19.8)	**0.010**	0.86 (0.66; 0.94)	12.9	30.1	35.8
		Moderate	41.1±39.8	40.0±46.1	0.61 (-8.6; 9.8)	0.892	0.92¶ (0.83; 0.96)	12.2	28.4	33.7
		Light to moderate	108.8±57.6	98.5±61.4	12.0 (0.18; 23.8)	**0.047**	0.92 (0.82; 0.97)	16.8	39.3	46.7
	Kcal/day	**Light**	196.5±106.4	206.4±113.8	-9.9 (-43.9; 24.1)	0.555	0.82 (0.60; 0.92)	46.7	109.0	129.5
		Moderate	237.8±322.8	165.4±199.4	72.4 (-0.8; 145.6)	0.052	0.85 (0.66; 0.93)	101.1	236.0	280.3
		Light to moderate	434.4±359.3	371.9±269.1	62.5 (-10.8; 135.8)	0.091	0.90 (0.78; 0.95)	99.4	231.9	275.4

* The point estimates of the ICC from ACT were not contained in the 95% CIs of the ICC from SWA; § The point estimates of the ICC from SWA were not contained in the 95% CIs of the ICC from CHAMPS; ¶ The point estimates of the ICC from CHAMPS were not contained in the 95% CIs of the ICC from ACT; Diff: raw difference between week one and two; Significance of F-test from the Analysis of Variance used to calculate ICCs;



Blue solid line indicates the mean of the difference between measures from week 1 to week 2. Blue dashed line indicates 95% confidence interval around the difference between weeks. Red dashed line indicates the 95% limit of agreement between pooled mean of physial activity measures in both weeks.

Figure 8. Bland and Altman plots from measures of physical activity duration of subjects included in the reliability analysis.

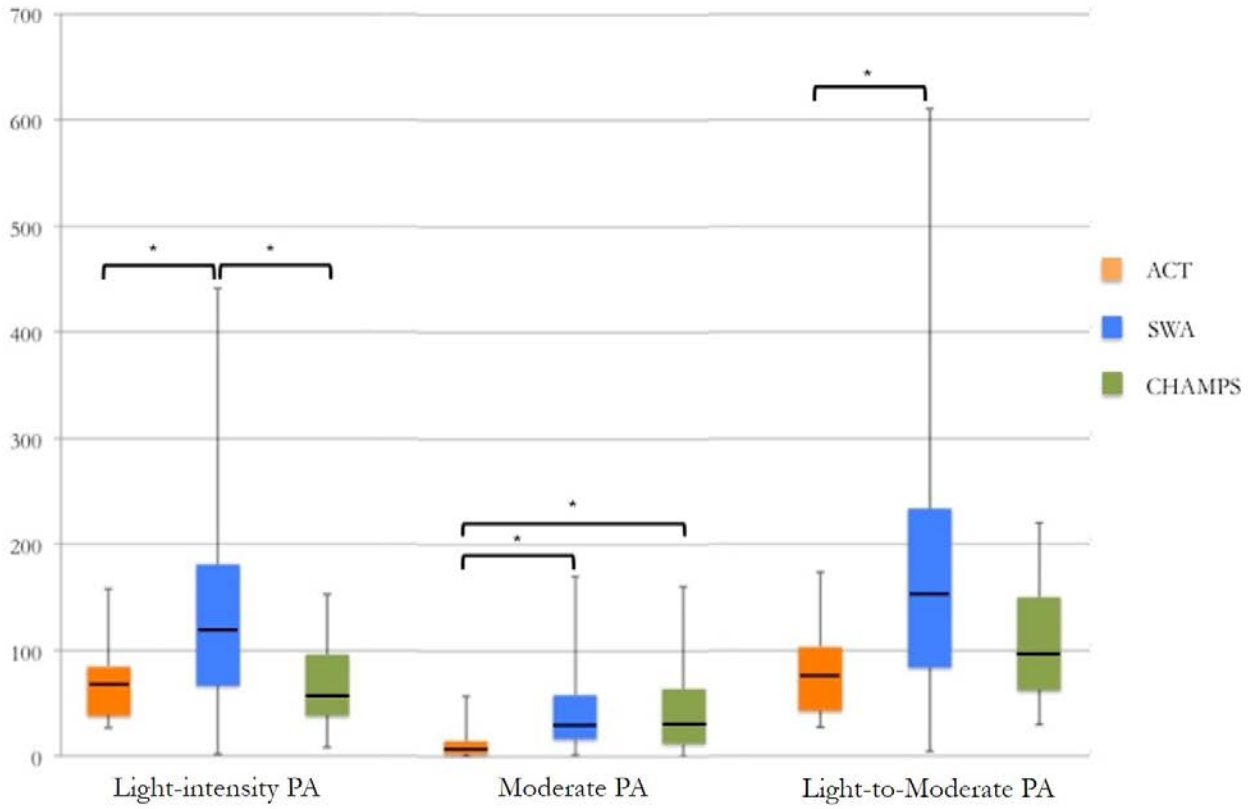


Blue solid line indicates the mean of the difference between measures from week 1 to week 2. Blue dashed line indicates 95% confidence interval around the difference between weeks. Red dashed line indicates the 95% limit of agreement between pooled mean of physial activity measures in both weeks.

Figure 9. Bland and Altman plots from measures of physical activity caloric expenditure of subjects included in the reliability analysis.

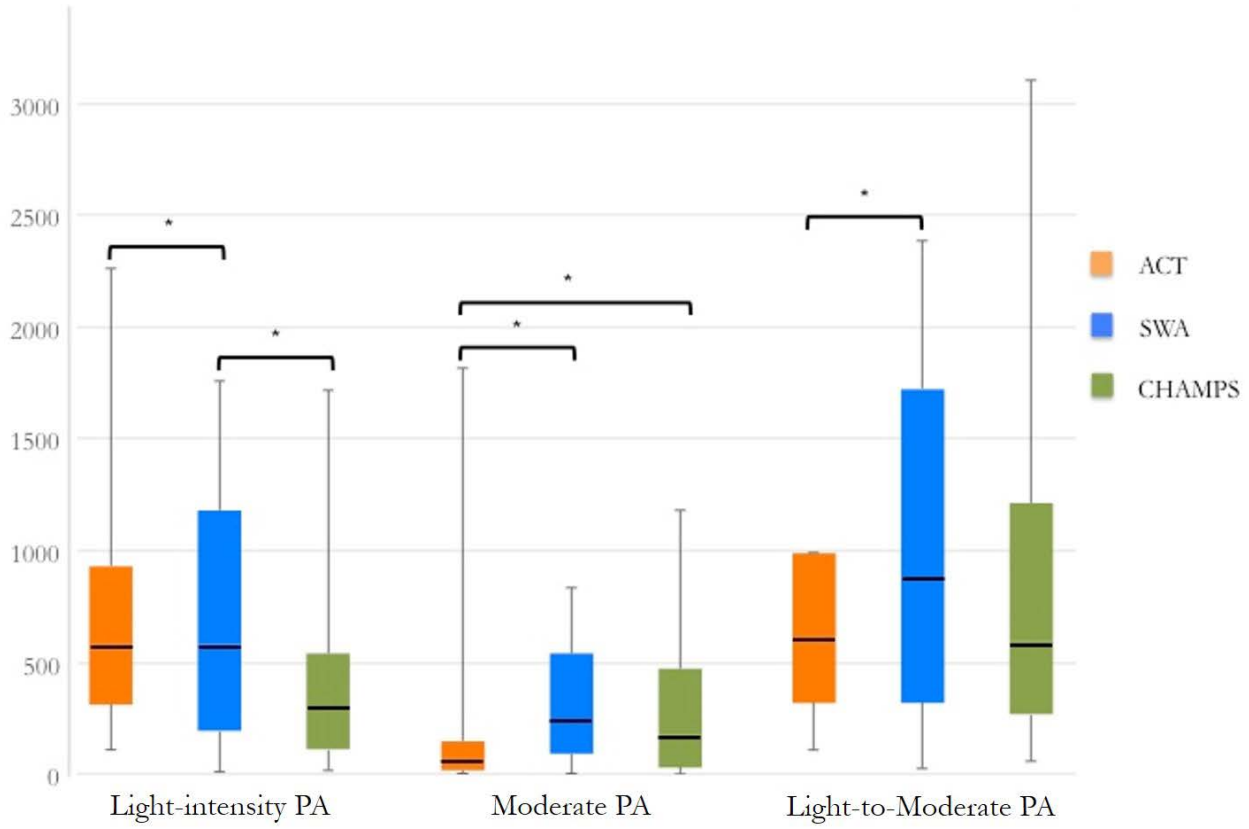
The comparison of ICC values across measurement tools demonstrated significant lower reliability of the ACT as compared to the SWA (Table 2). The ICC estimates of PA duration for the ACT in light (ICC= 0.86), moderate (ICC= 0.75), low-to-moderate (ICC= 0.85) intensities, and number of steps (ICC= 0.85) were not contained within the 95% CIs of the ICC values for the SWA in those intensity categories ([0.89; 0.98], [0.85; 0.97] and [0.90; 0.98], respectively), and number of steps (0.88; 0.98). The reliability of the ACT (ICC= 0.75) was also significantly lower compared to CHAMPS (95% CI [0.83; 0.96]) in duration of moderate PA, but not for the other variables. The reliability of the CHAMPS was significantly lower than that of the SWA for duration (ICC= 0.86 vs. 95% CI [0.90; 0.98]) and caloric expenditure (ICC= 0.82 vs. 95% CI [0.83; 0.97]) of light intensity PA.

Figures 10 and 11 show duration and caloric expenditure, respectively, measured by the ACT, SWA and CHAMPS during week one. Duration of PA and caloric expenditure measured by the ACT were significantly lower than that of the SWA across all intensity categories ($p<0.050$). PA measures of duration and caloric expenditure from the ACT were significantly lower than CHAMPS in moderate PA ($p=0.001$ and $p=0.004$, respectively), duration in light-to-moderate PA ($p=0.020$), and caloric expenditure in light PA ($p=0.001$). The CHAMPS scores in duration and caloric expenditure were significantly lower than measures from SWA in light PA ($p=0.004$ and $p=0.001$, respectively), and light-to-moderate PA ($p=0.018$ and $p=0.020$, respectively). The same analysis was run using data from week two and generated very similar results.



* Duration score significantly different between measurement tools ($p<0.017$);

Figure 10. Comparison between the magnitude of the duration of physical activity across instruments during week one. Actigraph (ACT), Sensewear Armband (SWA) and CHAMPS



* Caloric expenditure score significantly different between measurement tools ($p<0.017$);

Figure 11. Comparison between the magnitude of the caloric expenditure of physical activity across instruments during week one. Actigraph (ACT), Sensewear Armband (SWA) and CHAMPS.

2.4.1 Reliability of Monitoring Timeframes

Seven of the 28 subjects from the reliability analysis either did not have data during '24-hours,' or their PA data from ACT and SWA did not match the same monitoring days and times. Amongst the 21 subjects included in this analysis, the minimum wear-time was 22:47 hr/day. The reliability of

monitoring timeframes is described in Tables 3, 4 and 5 ('24-hours,' '10-hour from awakening,' and 'wake-hours,' respectively). Reliability of the ACT and SWA in each of the monitoring timeframes ranged from moderate to excellent (ICC= 0.75 to 0.96). To simplify interpretation, these tables report PA data in min/day only, since PA data in min/day and kcal/day were highly correlated across all timeframes (Spearman's rho from 0.85 to 0.96).

The reliability of the ACT and SWA in monitoring timeframes ranged from good to excellent (ICC= 0.75 to 0.96). The ICCs for each instrument were similar across all timeframes and indicated no statistical differences. However, there was a trend towards higher ICCs with narrower confidence intervals in the 'wake-hours' and '10-hours from awakening' in comparison to '24-hours' timeframe. Bland-Altman plots revealed no systematic bias between PA measures from the two weeks for the ACT and SWA across intensity categories and timeframes (Figures 12-14).

2.5 DISCUSSION

This is the first study to concurrently determine the test-retest reliability of the ACT, SWA, and CHAMPS during free-living PA performed in light and moderate intensities after TKA. Results of the reliability analysis revealed good test-retest reliability for the three instruments. Measures of PA from the SWA have shown to be more reliable than those of the ACT and CHAMPS. PA duration measured by the SWA showed better reliability than measures from the ACT across all PA categories, and was also better than measures of light-intensity PA from the CHAMPS. The better reliability of PA measures from the SWA is further supported by the lower bounds of the 95% CIs that were consistently above the threshold of excellent reliability, i.e., ICC>0.75.

Table 3. Daily physical activity duration (min/day), intra-class correlation coefficient (ICC) and measurement error from 21 subjects with data matched between Actigraph (ACT) and Sensewear Armband (SWA) in the 24-hour timeframe during the two weeks of data collection.

	Activity Intensity	Week 1 Mean ± SD	Week 2 Mean ± SD	Diff (95% CI)	Significance F-test	ICC (95% CI)	SEM	MDC 90%	MDC 95%
ACT	**Light**	65.5±34.2	66.3±37.9	0.2 (-11.4; 11.8)	0.970	0.86 (0.64; 0.94)	13.5	31.4	37.3
	Moderate	10.4±12.8	8.7±8.0	1.7 (-2.7; 6.1)	0.429	0.75 (0.46; 0.90)	5.2	12.1	14.4
	Light to moderate	76.9±44.0	75.0±44.8	1.9 (-12.8; 16.6)	0.790	0.84 (0.61; 0.94)	17.7	41.3	49.0
	Steps	4570.1±2041.1	4374.4±1683.6	195.6 (-404.2; 795.5)	0.501	0.89 (0.70; 0.96)	615.7	1436.7	1706.6
SWA	**Light**	135.8±89.8	148.9±92.6	-13.1 (-34.6; 8.4)	0.219	0.93 (0.82; 0.97)	23.3	54.4	64.6
	Moderate	45.3±39.8	43.8±34.1	1.5 (-8.1; 11.1)	0.750	0.91 (0.79; 0.97)	11.0	25.7	30.5
	Light to moderate	181.1±118.1	192.6±116.1	-11.6 (-39.7; 16.5)	0.400	0.93 (0.82; 0.97)	30.1	70.2	83.4
	Steps	6104.1±3505.8	6095.8±3133.6	8.3 (-759.1; 775.6)	0.982	0.92 (0.81; 0.97)	923.4	2154.7	2559.5

Significance of F-test from the Analysis of Variance used to calculate ICCs; SEM: Standard Error of the Measurement; MDC: Minimal Detectable Change.

Table 4. Daily physical activity duration (min/day), intra-class correlation coefficient (ICC) and measurement error from 21 subjects with data matched between Actigraph (ACT) and Sensewear Armband (SWA) in the 10-hours from awakening timeframe during the two weeks of data collection.

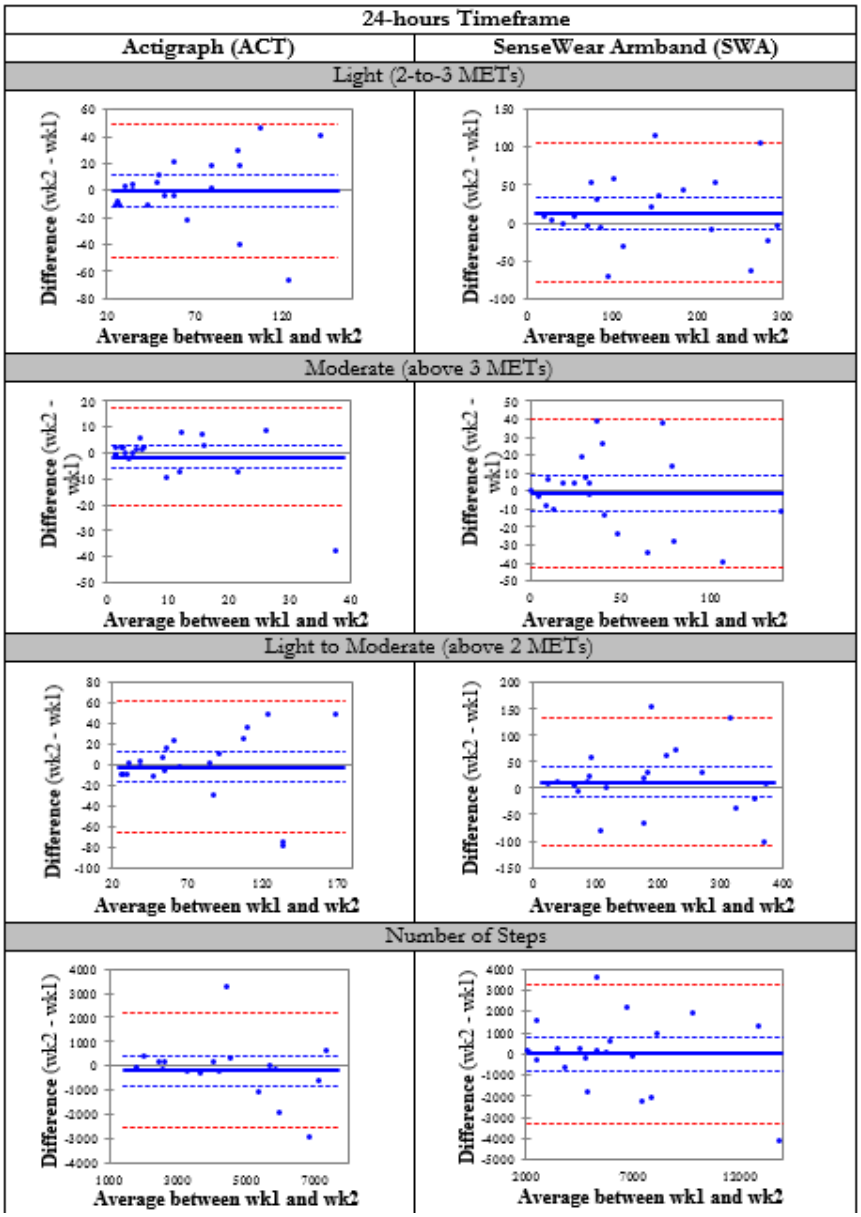
	Activity Intensity	Week 1 Mean ± SD	Week 2 Mean ± SD	Diff (95% CI)	Significance F-test	ICC (95% CI)	SEM	MDC 90%	MDC 95%
ACT	**Light**	50.3±29.7	47.3±28.9	2.9 (-6.7; 12.5)	0.531	0.88 (0.71; 0.95)	9.8	22.8	27.1
	Moderate	7.7±8.5	6.7±6.7	1.0 (-1.7; 3.6)	0.449	0.85 (0.63; 0.94)	3.0	6.9	8.2
	Light to moderate	57.9±35.8	54.0±34.7	3.9 (-7.5; 15.4)	0.484	0.89 (0.72; 0.95)	11.3	26.3	31.2
	Steps	3325.2±1483.5	3237.6±1316.4	87.6 (-286.6; 461.7)	0.628	0.92 (0.80; 0.97)	396.7	925.8	1099.7
SWA	**Light**	102.1±67.2	109.6±62.9	-7.5 (-22.4; 7.5)	0.309	0.94 (0.86; 0.98)	16.3	38.0	45.2
	Moderate	36.5±35.5	34.9±27.7	1.5 (-7.1; 10.2)	0.713	0.92 (0.80; 0.97)	9.2	21.4	25.4
	Light to moderate	138.6±94.7	144.7±83.1	-6.1 (-27.4; 15.1)	0.556	0.94 (0.85; 0.98)	22.3	52.1	61.9
	Steps	4501.2±2894.5	4515.3±2422.0	-14.1 (-666.2; 638.0)	0.965	0.95 (0.87; 0.98)	620.1	1447.0	1718.9

Significance of F-test from the Analysis of Variance used to calculate ICCs; SEM: Standard Error of the Measurement; MDC: Minimal Detectable Change.

Table 5. Daily physical activity duration (min/day), intra-class correlation coefficient (ICC) and measurement error from 21 subjects with data matched between Actigraph (ACT) and Sensewear Armband (SWA) in the wake-hours timeframe during the two weeks of data collection.

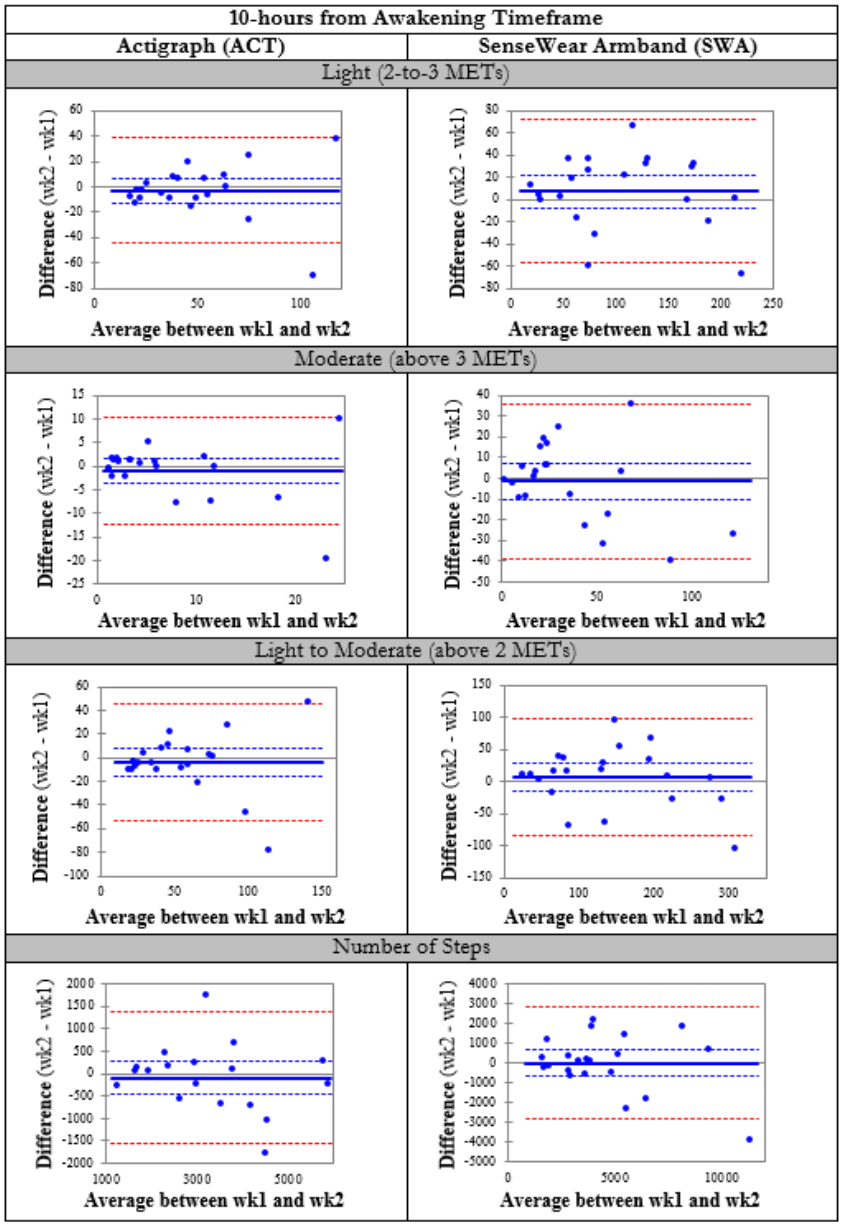
	Activity Intensity	Week 1 Mean ± SD	Week 2 Mean ± SD	Diff (95% CI)	Significance F-test	ICC (95% CI)	SEM	MDC 90%	MDC 95%
ACT	**Light**	64.0±33.3	63.6±37.4	0.3 (-11.5; 12.2)	0.956	0.86 (0.66; 0.94)	13.2	30.7	36.5
	Moderate	9.9±12.8	8.3±7.7	1.6 (-2.8; 5.9)	0.456	0.82 (0.56; 0.93)	4.9	11.4	13.6
	Light to moderate	73.8±42.9	71.9±43.9	1.9 (-13.0; 16.8)	0.794	0.86 (0.65; 0.94)	16.3	37.9	45.1
	Steps	4506.2±2045.4	4352.1±1693.3	154.1 (-446.1; 754.3)	0.595	0.89 (0.70; 0.96)	617.9	1441.2	1712.8
SWA	**Light**	134.0±88.9	147.7±91.0	-12.7 (-34.0; 8.6)	0.226	0.94 (0.85; 0.98)	22.3	52.1	61.9
	Moderate	45.1±39.6	43.8±34.3	1.3 (-8.3; 10.9)	0.780	0.93 (0.83; 0.97)	9.9	23.2	27.5
	Light to moderate	180.1±117.6	191.5±114.8	-11.5 (-39.3; 16.4)	0.402	0.94 (0.85; 0.98)	28.9	67.5	80.2
	Steps	5991.5±3482.7	6001.5±3127.4	-10.0 (-773.6; 753.7)	0.979	0.96 (0.89; 0.98)	686.1	1601.0	1901.8

Significance of F-test from the Analysis of Variance used to calculate ICCs; SEM: Standard Error of the Measurement; MDC: Minimal Detectable Change.



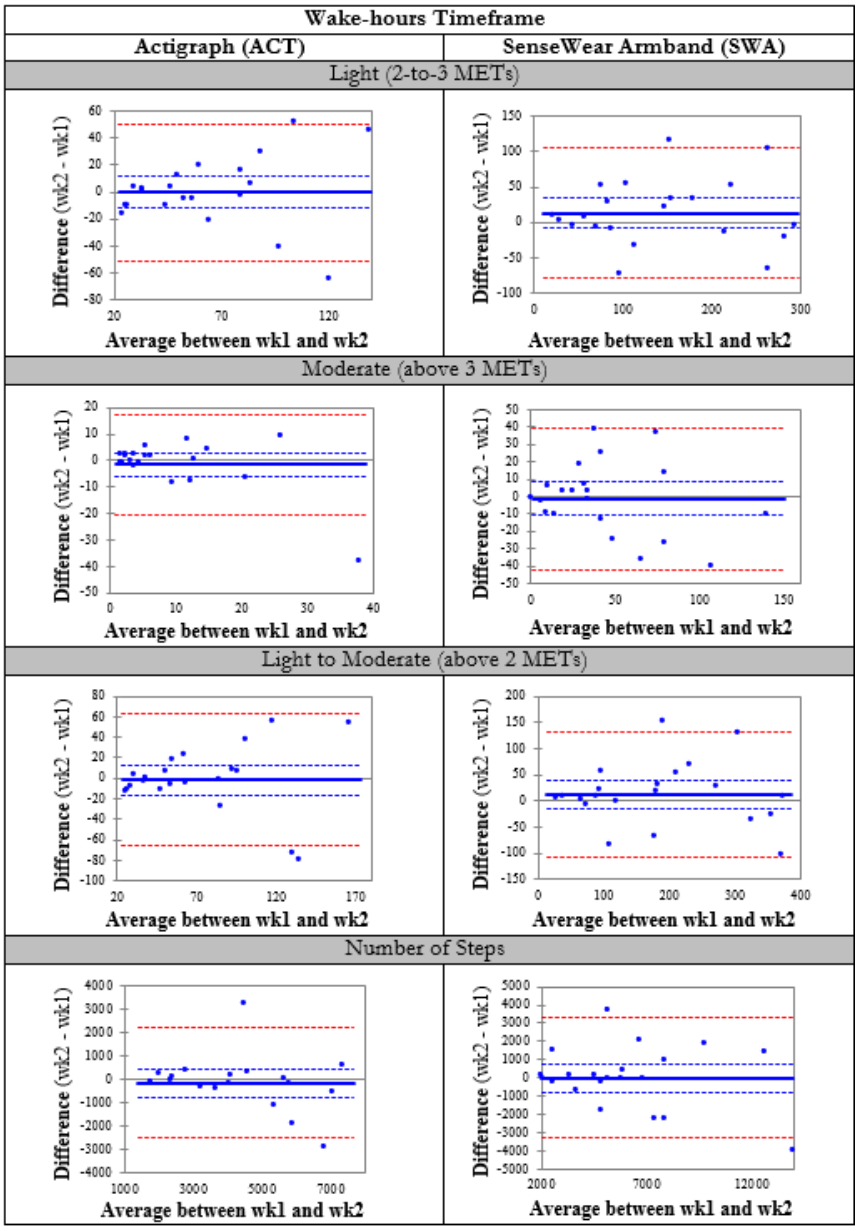
Blue solid line indicates the mean of the difference between measures from week 1 to week 2. Blue dashed line indicates 95% confidence interval around the difference between weeks. Red dashed line indicates the 95% limit of agreement between pooled mean of physial activity measures in both weeks.

Figure 12. Bland and Altman plots from measures of physical activity during 24-hours timeframe.



Blue solid line indicates the mean of the difference between measures from week 1 to week 2. Blue dashed line indicates 95% confidence interval around the difference between weeks. Red dashed line indicates the 95% limit of agreement between pooled mean of physial activity measures in both weeks.

Figure 13. Bland and Altman plots from measures of physical activity during 10-hours from Awakening timeframe.



Blue solid line indicates the mean of the difference between measures from week 1 to week 2. Blue dashed line indicates 95% confidence interval around the difference between weeks. Red dashed line indicates the 95% limit of agreement between pooled mean of physial activity measures in both weeks.

Figure 14. Bland and Altman plots from measures of physical activity during Wake-hours timeframe.

Better reliability of the SWA as compared to the ACT is supported by consistent higher ICC values for the SWA both in duration of PA (ICCs for SWA ranged from 0.93 to 0.95 whereas ICCs for ACT ranged from 0.75 to 0.86), and caloric expenditure (ICCs for SWA ranged from 0.89 to 0.93 whereas ICCs for the ACT ranged from 0.77 to 0.86). Despite the consistent differences in ICC values between activity monitors, these differences were only significant for duration of PA and not for caloric expenditure. The reason for this inconsistency is probably because the differences in ICCs between ACT and SWA for caloric expenditure were slightly smaller than the differences in ICC for duration of PA. Differences in the ICC values from caloric expenditure (SWA minus ACT) were 0.05 for light (0.93 minus 0.88), 0.12 for moderate (0.89 minus 0.77), and 0.07 for light-to-moderate intensity PA (0.93 minus 0.86) whereas for PA duration, these differences were 0.09 for light (0.95 minus 0.86), 0.18 for moderate (0.93 minus 0.75), and 0.10 for light-to-moderate (0.95 minus 0.85). The SWA also demonstrated superior reliability in comparison to CHAMPS in light intensity PA. The better reliability of the SWA is further supported as it was the only instrument with lower bounds of 95% CIs consistently above the threshold of good reliability, i.e., ICC>0.75.

Unlike the ACT and SWA, the CHAMPS had systematic bias with higher scores in week one as compared to week two in light and light-to-moderate PA. Despite high reliability for these categories (ICC> 0.85), this finding questions the ability of this questionnaire to provide consistent scores. High ICC values along with systematic bias can be explained by the relative higher between-subject variability as compared to within-subject variability. In our study, the range of differences in PA duration between weeks (week one minus week two) was much higher for between-subjects than within-subjects. For example, the difference between weeks for light intensity PA between-subjects ranged from -77 to 81 min/day, and within-subject differences ranged from -11 to 81 min/day. We only provide example for light intensity PA because systematic bias in light-to-moderate was

attributed to the contribution of light intensity measures since the moderate intensity category did not demonstrate systematic bias. We believe the differences in self-report of PA between both weeks were not due to true changes in PA behavior since we did not observed systematic bias in the objective measures from the activity monitors. The differences in CHAMPS may have occurred due to the subject's difficulty in recalling light intensity PA. [48,93] Since most of the activities at this intensity are those of daily-living, individuals may not keep track of how many times per week or for how long tasks such as sweeping or walking to grocery store are performed. Whereas, reporting moderate PA such as exercise walking or strength training tend to be easier, since these activities are purposefully done for exercise and may be performed fewer times per week.

Although our results for the SWA cannot be put in perspective with literature because no prior studies determined the reliability of PA measures from the SWA in free-living, they can be compared to studies on the ACT and CHAMPS. Yet, studies on the ACT and CHAMPS investigated the reliability of PA measures from each instrument separately rather than concurrently in the same study. The studies on the ACT were done in healthy middle-aged and older adults, and reported ICCs ranging from 0.74 to 0.90.[70,120] Although the ICCs from our study fall within the range of ICCs from those studies, their findings cannot be generalized to our population, since older adults with knee osteoarthritis who undergo TKA tend to be less active than their healthy and younger counterparts.[1] Several studies in healthy older adults investigated the reliability of CHAMPS and reported lower reliability compared to our study (ICC= 0.62 to 0.81).[19,44,46,48] A study in adults with fibromyalgia also reported lower reliability than ours (ICC= 0.27 to 0.76).[47] Amongst these studies, only one used one-week interval to determine the reliability of PA measured by the CHAMPS and reported estimates (ICC= 0.79 to 0.81)[46] comparable to ours (ICC= 0.82 to 0.92). The lower ICCs reported in the other studies could be attributed to the longer time interval between

test-retest that ranged from 2-weeks to 6-months. Future larger studies should investigate the comparative reliability of activity monitors and questionnaires using different test-retest intervals.

Our study also provided the SEM and MDC values for the PA measures and is unique in this regard. We are not aware of other studies that reported these indices of measurement error of PA measured by the ACT, SWA or CHAMPS. The SEM and MDC use the same units as the measurement, and are clinically useful because they can be used to interpret changes in PA that are beyond measurement error. While the SEM represents the value of measurement error itself, the MDC uses the SEM to compute a threshold within a defined level of statistical confidence that true change has occurred beyond measurement error. We reported the MDC with two degrees of confidence, the MDC_{90} and MDC_{95}, so one can choose how strict to be when making interpretation of changes in PA overtime. For instance, using the MDC_{90} for duration in light-intensity PA measured by the SWA (53 min/day), if a person increases time in light activities by greater than 53 min/day, one can be 90% confident that there was a real change in PA that is beyond measurement error. Although 53 min/day seems to be a large change, it must be noted that activities falling in this intensity category (2 to 2.9 METs), are those including mopping, gardening and slow walking, which are spread throughout the day and are doable for individuals with arthritis of the lower extremities.

When the magnitude of PA values was compared across instruments, we observed that PA values from the ACT were consistently lower than those from the SWA, and lower than those from CHAMPS in moderate PA. To date, a couple of studies in young adults compared measures of PA concurrently using the ACT and SWA in free-living conditions, and reported lower PA values measured by the ACT in comparison to SWA. [55,57,115] To our knowledge, one study compared the magnitude of values between the ACT, SWA and CHAMPS in healthy older adults. [43] The study used the ACT, SWA and CHAMPS to measure PA energy expenditure, which is a measure of all

activities performed excluding resting. Their results indicated that measures of PA from the ACT were higher than the SWA, and that the SWA measures were similar to CHAMPS. The divergence between results might be due to the different methodologies they used to calculate caloric expenditure from the PA instruments. While we used the equations recommended by the ACT, SWA and CHAMPS to calculate PA at different intensities, the authors of that study used questionable approaches. For example, when using the ACT, Freedson equation was used to calculate caloric expenditure for all activities regardless of intensity level. Yet, Freedson equation was developed to calculate energy expenditure of moderate intensity activities. Using the incorrect equation would overestimate energy expenditure. The same authors calculated energy expenditure for the CHAMPS by subtracting resting energy expenditure from the total score. This approach is not appropriate since CHAMPS total score does not include resting. However, we were able to compare the results of our study to theirs on the number of steps a day measured by the ACT and SWA. Their findings of lower number of steps from the ACT (5917 steps a day) as compared to SWA (7022 steps a day) agreed with our results (4677 vs. 6262). This finding adds to the volume of literature on the limitations of the ACT in measuring number of steps in slow walkers such as older adults with functional limitations. [133-135] Furthermore, a recent study in individuals who underwent TKA has shown that measures from the ACT consistently underestimate the SWA in a variety of activities. [84] Combined, these findings support that the SWA seems to be more appropriate to measure PA in individuals with osteoarthritis of the lower extremities.

Our study is also the first to provide evidence on the comparative reliability of the activity monitors across different monitoring timeframes. While evidence on the number of days to wear activity monitors is robust, little is known about the consistency of PA measures across different daily wear-periods. Our results indicated that the reliability of the ACT and SWA seems comparable

across timeframes; however, the information given by '10-hours from awakening' may not provide a good representation of daily activities. In comparing the amount of PA data contained in the '10-hours from awakening' to the '24-hours', the 10-hours had 19% ([36.5 min/day from 10-hours]/[45.3 min/day from 24-hours in moderate PA from SWA]) to 29% ([47.3 min/day from 10-hours]/[66.3 min/day from 24-hours in light intensity PA from ACT]) less moderate and light intensity PA data respectively. Whereas, the 'wake-hours' had only 0% to 5% less PA data compared to '24-hours'. These differences between timeframes are in line with a study in healthy adults that measured PA using the ACT and reported that PA measured during a 10-hours period resulted in up to 42% less PA data as compared to a 14-hours period (wake-hours). [114] Using data from periods shorter than wake-hours result in underrepresentation of daily PA since many people tend to spread their activities throughout the day. For example, a person wakes up at 6:30 am, does some household chores such as cleaning, goes to work at 8 am, sits at the office desk until 5 pm, and goes to the gym at 6 pm. If PA is measured during '10-hours from awakening', all the activities carried out after 4:30 pm would not be measured, including going to the gym, which represented a relevant PA in this example. Therefore, the 'wake-hours' seems to be the most appropriate monitoring timeframe to use since it will capture most of the activities performed in 24-hour period and provides reliable PA measures. In addition, measuring PA during wake-hours will reduce burden on subjects, as most of them feel uncomfortable wearing the monitors while sleeping. Also, the data processing for 'wake-hours' is less cumbersome than for '10-hours from awakening'.

We acknowledge that our study had a small sample size. Although our primary analysis had only 28 subjects, it did not seem to have negatively affected the ICC results as the values had generally narrow confidence intervals. Another limitation was that indices of measurement error were calculated based on one-week interval. We also acknowledge that estimates of reliability and

measurement error may be different for longer time intervals. In addition, although our sample was a good representation of individuals with knee osteoarthritis who undergo TKA, our results may not be generalized to healthy older adults without dysfunctions of the lower extremities.

2.6 CONCLUSION

Our results indicated that duration of PA measured by the SWA is more reliable than measures from the ACT across intensity categories, and more reliable than measures of light-intensity PA from CHAMPS. Measuring daily PA using the ACT or CHAMPS may not be ideal because the ACT measures significantly lower PA than the SWA and CHAMPS, and measures of light-intensity PA from the CHAMPS were significantly different between weeks. Furthermore, it seems that monitoring PA during wake-hours provides reliable data that resemble the 24-hours timeframe. Using the wake-hours monitoring timeframe may help reduce burden on subjects without compromising reliability when wearing activity monitors.

3.0 CHAPTER 3 - RESPONSIVENESS OF PHYSICAL ACTIVITY MEASURES FOLLOWING EXERCISE PROGRAMS AFTER TOTAL KNEE ARTHROPLASTY

3.1 SUMMARY

Background: Few instruments that measure physical activity (PA) can accurately quantify PA performed at light and moderate intensities, which is particularly relevant to older adults. However, evidence of their responsiveness after an intervention is limited. **Objectives**: To determine and compare the responsiveness of two activity monitors (Actigraph [ACT]) and Sensewear Armband [SWA]) and one questionnaire (Community Healthy Activities Model Program for Seniors [CHAMPS]) in assessing PA in light and moderate intensities after an exercise intervention in individuals following total knee arthroplasty (TKA). **Design**: One-group pretest-posttest. **Methods**: Changes in duration of daily PA from baseline to 6 months and the standardized response mean (SRM) were calculated to assess internal responsiveness, and were compared across instruments. Correlations between measured changes by the three instruments and perceived changes in PA assessed by a global rating of change at 6 months were used to test external responsiveness. Agreement between instruments in identifying changes in PA based on measurement error was determined using weighted-Kappa. **Results**: Thirty subjects, mean age 67(6) and 73% female, were included in the analysis. Changes in PA measured by each instrument were small and not significant ($p>0.05$). SRMs indicated small degree of responsiveness (SRM<0.30). Scores in the global rating of

change did not correlate with changes measured by the instruments (rho=0.05-0.28, p>0.05). ACT and SWA agreed on identifying changes in moderate-intensity PA (K=0.60) and number of steps (K=0.63). There was no agreement between CHAMPS and the activity monitors (K≤0.22). **Conclusion**: The measurement tools had similar small degree of internal responsiveness. Using measurement error as a threshold allowed identification of individual changes in PA and might be an alternative to investigate changes in PA in future studies.

3.2 INTRODUCTION

The benefits of regular physical activity (PA) to improve general health are well known. Being physically active can lower the risk for certain comorbid conditions such as heart disease, stroke, type 2 diabetes, depression, some cancers, and obesity.[109] Individuals who undergo total knee arthroplasty (TKA) for end-stage knee osteoarthritis are typically older adults who are less active than their healthier counterparts and are at greater risk of developing co-morbidities.[110-113] Interventions to increase PA and prevent co-morbidities in individuals after TKA have received increased attention among the research community lately.[136,137] To test the effectiveness of interventions to increase PA, researchers need measurement tools that are sensitive to changes in PA behavior. However, information on the responsiveness of instruments that measure PA behavior is limited.

Responsiveness evaluates the ability of a measurement tool to accurately detect changes in the construct being measured when change has occurred and it can be determined by internal and external responsiveness methods.[7,13,14] The internal responsiveness method is based on the

distribution of the data and represents the ability of a group measure to change over time in response to an intervention. This method can be quantified by indices such as effect sizes that are calculated using the magnitude of the change and its variability.[14] Since this method is based on group change, it is suitable to determine changes that occur at the group level. For the external responsiveness method, an external anchor that assesses perceived changes in PA by the individual is used to compare to individuals changes in PA detected by the measurement tools being tested. This method is used to determine the amount of change in the measurement that corresponds to change that is perceived as important to the individual and can be used to calculate clinically meaningful thresholds for changes at the individual level.

There are numerous tools available to measure PA, including activity monitors and self-report questionnaires, but evidence of their ability to capture change in PA over time is limited. Two activity monitors and one self-report questionnaire have been commonly used in research investigating older population such as those after TKA. The Actigraph (ACT; Actigraph LLC, Pensacola, FL), worn at the waist level, and the Sensewear Armband (SWA; Bodymedia, Pittsburgh, PA), worn on the arm, are activity monitors that have been validated to measure PA in older adults.[43,84] The Community Health Activities Model Program for Seniors (CHAMPS) was specifically developed to measure self-reported PA in older adults.[44] These three instruments have the advantage to measure PA at light intensities such as household chores and slow walking, which represent most of daily activities performed by older adults who undergo TKA.[19,48,115-118] While the activity monitors are costly and need to be worn for several days to provide representative data of daily activities, the questionnaire has low cost and takes 15-20 minutes to complete and recalls activities from past week. On the other hand, the activity monitors measure PA in real time, which eliminates the problem of recall bias that is commonly faced with self-report PA measures.[48,119]

To our knowledge, no studies have determined the responsiveness of the SWA and the few studies on the responsiveness of the ACT and CHAMPS did not include older adults with mobility problems.[44,138-140] Results from these studies may not be applicable to older adults who underwent TKA, since pain and functional limitations associated with knee osteoarthritis make them less active than younger people and their healthy older counterparts.[1] Additionally, studies have not concurrently compared the responsiveness of these three commonly used instruments. Concurrently comparing the responsiveness of PA measures estimated by the ACT, SWA and CHAMPS will provide information for a well-educated choice of tools to assess changes in PA behavior over time in individuals with osteoarthritis of the lower extremities. Hence, the purpose of this study was to determine and compare the responsiveness of the ACT, SWA, and CHAMPS in assessing PA from light, moderate and vigorous intensities after rehabilitation that involved exercise and physical activity promotion programs in individuals who underwent TKA.

3.3 METHODS

3.3.1 Design and Subjects

This was an ancillary study that used a one-group pretest-posttest design. Baseline and 6-month data on PA from subjects participating in a randomized trial that investigated the effect of rehabilitation approaches to improve physical function in individuals following TKA were analyzed.[137] This study took place from October 2011 to August 2013 in the Department of Physical Therapy, University of Pittsburgh. All subjects signed a consent form approved by the University of Pittsburgh Institutional Review Board prior to participation.

The inclusion and exclusion criteria followed that of the parent study.[137] Individuals with unilateral TKA in the past 3-6 months and at least 50 years of age were included. Those who reported more than two falls within the past year, were unable to ambulate a distance of 31 meters without an assistive device, and had medical conditions that precluded safe exercise participation were excluded. Specific to this ancillary study, individuals were included only if they were willing to wear the activity monitors for 7 days during baseline and 6-month follow-up.

3.3.2 Study Protocol

Subjects attended two testing visits, one prior to the rehabilitation program (baseline), and one at the end of the program 6 months after baseline. During the baseline assessment, in order to characterize the sample, the subjects completed questionnaires of demographics, pain and physical function. Pain was assessed by an 11-point numeric pain scale and physical function by the 17-item Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function subscale. Subjects' height and weight were also measured. After finishing the testing procedures at baseline, subjects were fitted with the ACT and SWA and were instructed to wear the monitors for 7 days, except during water activities since the activity monitors are not waterproof. At the end of the 7-day period, subjects returned to the research facility to download the data from the activity monitors and to complete the CHAMPS. The CHAMPS questionnaire queries PA participation in the past week, which corresponds to the time they wore the monitors. Data from the activity monitors were inspected to assure sufficient data, which was defined as at least 5 days with 10-hours of PA data per day.[122,126] Subjects without sufficient data were asked to wear the portable monitors for an additional week and to complete the CHAMPS at the end of that week.

After completing the baseline measures of PA, subjects were randomized into two exercise groups, which were combined for the purpose of this study. Both exercise programs included endurance and strengthening for the lower extremities, with the experimental group performing more intense training than the control. In addition, the experimental group was exposed to balance and functional exercises as well as a behavioral PA promotion program. The exercise programs consisted of 12 supervised sessions delivered within a 3-month period followed by a home exercise program for another 3 months.[137]

The PA measures from baseline were repeated at 6-month follow-up. In addition, subjects were asked to rate how they perceived the overall change in PA since the beginning of the exercise program using a modified global rating of change scale. The scale consists of 15 points that range from -7 ("a very great deal less active") to 0 ("about the same") to +7 ("a very great deal more active"). Descriptors of improvement range from +1 to +7, and of worsening, range from –1 to –7. Subjects with a rating of +3 ("somewhat more active") or higher were classified as 'more active'. Subjects with a rating between +2 ("a little bit more active") to –2 ("a little bit less active) were classified as 'not changed'. Subjects with a rating of –3 ("somewhat less active") or lower were classified as 'less active'.

3.3.3 Measures of Physical Activity

The ACT is a small triaxial accelerometer-based monitor (2 x 1.5 x 0.6 inches in size) worn around the waist at the level of the hip-bone and over the right anterior superior iliac spine. In this study we used the ACT model GT3X and the ActiLife 5 software (Actigraph LLC, Pensacola, FL). The ACT measures body acceleration in activity counts, and was set to collect data at 1-minute intervals. The

ACT generates data on activity counts per minute (CPM) and number of steps. To categorize duration of daily activities in min/day, the software uses the following CPM cut-points: 760-1951 CPM for light, CPM 1952-5724 for moderate and >5724 CPM for vigorous intensities. [128] The ACT uses body weight to generate measures of caloric expenditure in kcal/day. The equations recommended by the manufacturer were used to convert PA duration into caloric expenditure in kcal/min. The non-wear periods are calculated by the software following the manufacturer's recommendation and were also visually analyzed. Measures of PA from the ACT have demonstrated moderate accuracy as compared to doubly labeled water in older adults and good reliability in older adults after TKA. [43,141]

The SWA is a small multi-sensor device (3.4 x 2.1 x 0.8 inches in size) worn on the back of the right arm over the triceps muscle and at midpoint between shoulder and elbow. We used the SWA model Pro-3 and the Professional software v6.1 (BodyMedia Inc., Pittsburgh, PA). This device has shown to be accurate in comparison to doubly labeled water in older adults and good reliability in older adults after TKA. [43,141] The SWA combines information from biaxial accelerometer, heat flux (heat being dissipated by the body), galvanic signal (onset, peak and recovery of maximal sweat rates) and skin temperature. The information from the sensors is integrated and processed by the software using proprietary algorithms that account for subjects' demographic characteristics (gender, age, height and weight). The SWA was set to provide data on number of steps, duration of activities (min/day), and caloric expenditure (kcal/day) in light (2 to 2.9 metabolic equivalents [METs]), moderate (3 to 5.9 METs) and vigorous intensities (≥6 METs), and number of steps. The SWA turns off automatically when not in contact with the skin, which enables the software to recognize periods of non-wear. Data were also visually screened to identify non-wear periods.

The CHAMPS is a self-reported questionnaire that queries type, frequency and duration of 41 activities usually performed by older adults, ranging from light to vigorous intensities. Duration in hours per week (hr/week) of each activity is selected from a range of less than one hour per week to 9 or more hours per week and categorized in two intensity levels according to the CHAMPS activities codebook. [44] Light-to-moderate PA (≥2 METs) represent all exercise-related activities, such as light to heavy household chores, and recreational and sports activities. Moderate intensity activities (≥3 METs) represent activities such as heavy household chores, calisthenics and sports. To allow direct comparison between the CHAMPS and the activity monitors, the score from moderate activities was subtracted from that of light-to-moderate to create a light-intensity category for the CHAMPS. The CHAMPS also calculates caloric expenditure using equations that take into account the intensity of each activity and accounts for subjects' body weight. [44] Scores from the CHAMPS in hr/week and kcal/week were converted into min/day and kcal/day respectively by multiplying the duration score by 60 minutes and then dividing by 7 days. CHAMPS has shown significant association when compared to doubly labeled water in older adults, and moderate reliability in people with musculoskeletal disorder and healthy older adults. [43,47,141]

The main outcome measure of this study was PA in minutes per day (min/day) estimated by the ACT, SWA and CHAMPS during waking hours. For each instrument PA was estimated at light (e.g., recreational activities, slow walk), moderate (e.g., vacuuming, brisk walk), and vigorous (e.g., lawn mowing, doubles tennis) intensity. The measures of moderate and vigorous-intensity activities were combined into the moderate category since our sample engaged in negligible amounts of vigorous-intensity activities (< 1 min/day). We also assessed daily number of steps using the ACT and SWA. Therefore, the PA intensity categories compared across the three instruments in this

study were light, moderate, and light-to-moderate (combination of light and moderate intensities); and number of steps was compared between the ACT and SWA only.

3.3.4 Data Analysis

Descriptive statistics for continuous variables included mean and standard deviation or median and 25-75 percentiles, according to data distribution. Counts and frequencies were used for categorical variables. To visually assess changes in PA we calculated magnitude of changes in PA from pre to post intervention (6-months minus baseline) and plotted histograms with those data. Internal responsiveness was determined using paired t-tests or Wilcoxon signed-rank tests comparing PA values from baseline to 6-month follow-up. Alpha level was set at 0.05. Additionally, the standardized response mean (SRM) was calculated as an index of responsiveness. The SRM is calculated as the ratio of mean change to the standard deviation of the change scores. [108] Values obtained from the SRM are interpreted as trivial degree of responsiveness (<0.20), small degree of responsiveness (0.20 to 0.49), moderate degree of responsiveness (0.50 to 0.79) and high degree of responsiveness (≥0.80). [108]

To determine external responsiveness, we first explored if the modified global rating of change was suitable as an external anchor to capture perceived changes in PA, meaning that its scores would correlate at least moderately (≥0.30) with changes measured by the ACT, SWA and CHAMPS. Correlations were calculated using Pearson or Spearman rho correlation coefficients according to data distribution. If the scores on the global rating of change correlated with changes in PA measured by the three instruments, cut-offs of clinically important change were calculated based on receiver operating characteristic (ROC) curves. [14]

To compare responsiveness across the PA measurement tools one-way repeated measures Analysis of Variance (ANOVA) was used for each PA intensity category to determine if the magnitude of change obtained from the three instruments were statistically different from each other. Since the three instruments were used to assess changes in PA in the same subjects, changes in PA at each intensity category measured by each instrument was used as the repeated factor. If the ANOVAs revealed significant differences, paired samples t-tests were used for the post hoc comparisons between PA measures applying Bonferroni adjustments. The alpha level was set at 0.016 to account for the three comparisons (ACT vs. SWA, SWA vs. CHAMPS, ACT vs. CHAMPS). Responsiveness was also compared by examining the 95% confidence intervals of the SRM. If the 95% confidence intervals did not overlap, the responsiveness of one instrument was deemed significantly different from the other. [132] In addition, weighted Kappa was used to investigate the agreement between instruments on identifying subjects who were had less, more, or about the same PA after the intervention, based on the standard error of the measurement previously published. [141] Values obtained from weighted Kappa are interpreted as poor (<0.20), fair (0.21-0.40), moderate (0.41-0.60), good (0.61-0.80) and very good (0.81-1.00). [142] Analyses were performed using the IBM SPSS Statistics 21 (IBM Corporation) and Microsoft Excel 2010 (Microsoft Corporation).

3.4 RESULTS

Forty-two subjects completed the randomized clinical trial, of which 30 had baseline and follow-up data for the three PA measurement tools and were included in the responsiveness analyses. The

demographic and biomedical characteristics between subjects included and those excluded from the responsiveness analyses were similar (Table 6). Subjects included in the responsiveness analyses were on average 67±6 years old, predominantly females (73%) and obese (body mass index = 30±4 kg/m^2). The daily monitoring wear time was on average 15±2 hours per day and was similar at baseline and follow-up.

Table 6. Baseline characteristics of study sample.

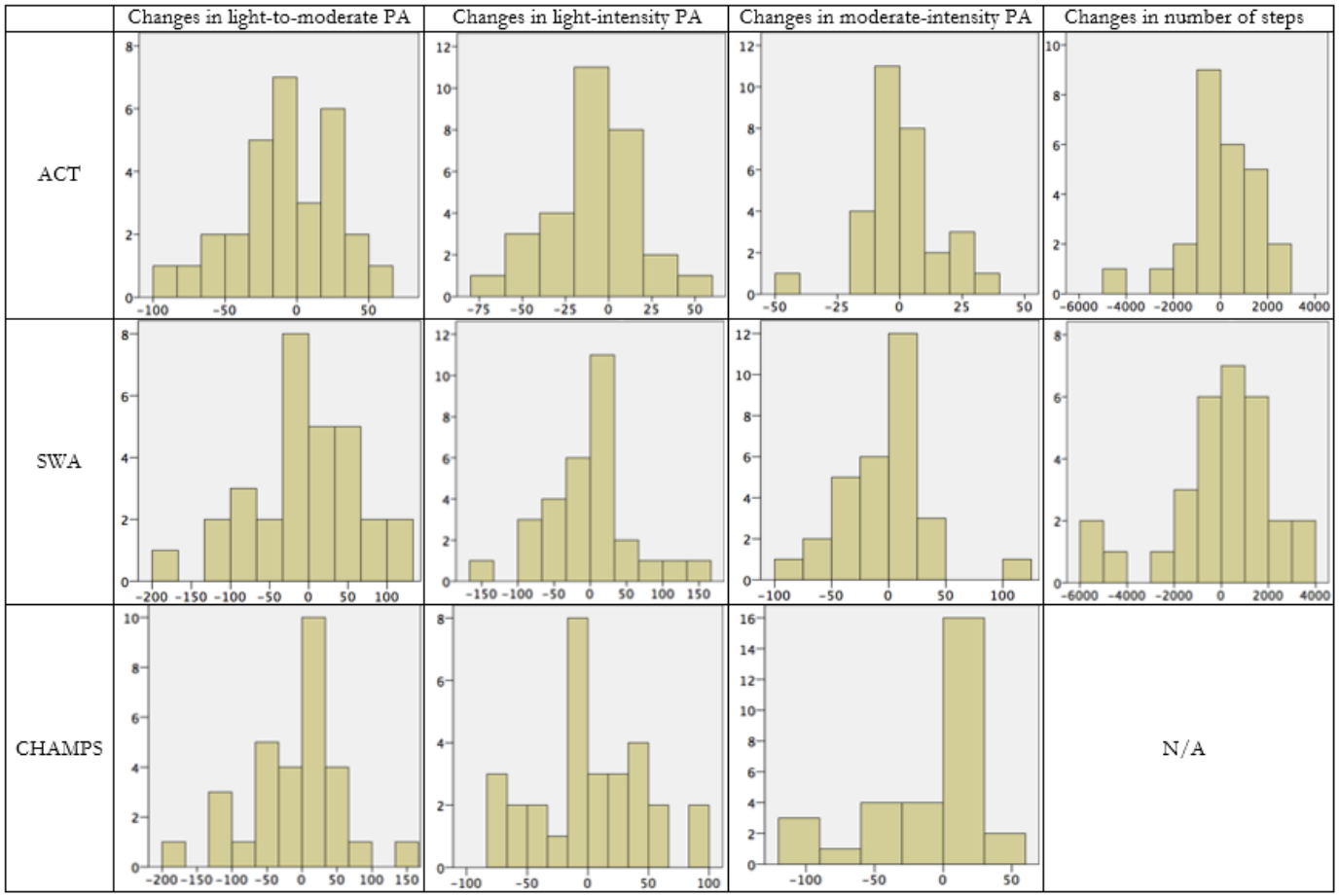
Variables	Included in Responsiveness Analysis n=30	Not Included in Responsiveness Analysis n=12
Age in years	67.3 (6.2)	69.8 (7.2)
Sex – female (%)	22 (73.3)	8 (66.7)
BMI in kg/m^2	29.9 (4.1)	31.0 (3.6)
Race – white (%)	27 (90.0)	10 (83.3)
Education – n (%)		
High-school	12 (40.0)	4 (33.3)
College	18 (60.0)	8 (66.7)
Time from TKA – n (%)		
3 to 4 months	7 (23.3)	4 (33.3)
4 to 5 months	13 (43.3)	4 (33.3)
5 to 6 months	10 (33.4)	4 (33.4)
Knee pain – 0 to 10; median (Q25; Q75)		
Surgical side	2 (1; 3)	3 (1; 3)
Non-surgical side	3 (0; 5)	3 (1; 6)
WOMAC-PF – 0 to 68;	19.1 (9.5)	18.6 (10.5)

SD: standard deviation; BMI: body mass index; TKA: total knee arthroplasty; Q25: quartile 25th; Q75: quartile 75th; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function sub-scale;

Figure 15 depicts the distribution of changes in PA from baseline to follow-up. The graphs indicate that the number of subjects who became less active, more active, and those who did not change PA was similar. For example, based on zero as a threshold for no change, measures from the ACT in light-to-moderate PA revealed that 18 subjects became less active and 12 more active. As per the SWA, 14 subjects became less active and 16 more active. Scores from the CHAMPS indicated that 16 subjects became less active and 14 more active. We also visually observed the number of subjects who changed beyond a magnitude that most clinicians would agree to be an important change (e.g., 20 min/day in light-to-moderate PA). Based on that, measures from the ACT revealed that 10 subjects became more active, 9 less active, and 11 did not change; based on the SWA, 11 subjects became more active, 13 less active, and 6 did not change; and based on the CHAMPS, 7 subjects became more active, 11 less active, and 12 did not change. Figure 16 shows the slopes of lines from baseline to follow-up for each individual subject.

Internal responsiveness analyses demonstrated that none of the PA measurement tools detected significant group changes in PA from pre to post intervention ($p \geq 0.163$) for both duration of PA (Table 7) or caloric expenditure (Table 8). Additionally, the values of SRM ranged from 0.01 to 0.26, indicating that the measurement tools had a trivial to small degree of responsiveness across PA categories.

For external responsiveness, while the graphical representation of change in PA (Figure 15) indicated that similar number of subjects became less or more active, none of the subjects reported being less active based on the global rating of change in PA (Table 9). According to the global rating of change, 5 subjects (17%) reported having the same activity level at 6 months and the remaining 25 subjects (83%) reported being at least a little bit more active. Consequently, the scores from the global rating of change in PA were not significantly correlated (rho from 0.05 to 0.28) with the



ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire.

Figure 15. Changes in physical activity (PA) duration measured by the ACT, SWA and CHAMPS questionnaire pre to post intervention across the three PA intensity categories, and number of steps (ACT and SWA only). Numbers on the Y-axis represent frequency and numbers on the X-axis represent PA duration in min/day.

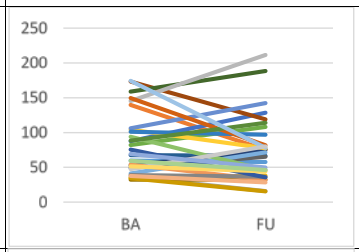
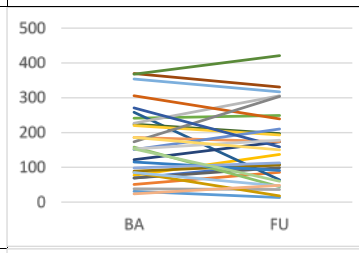
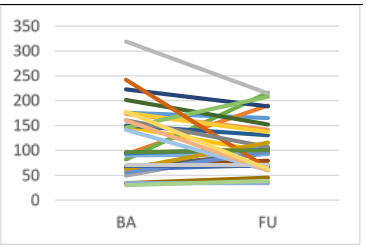
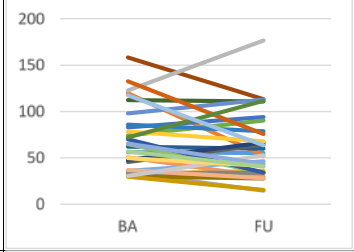
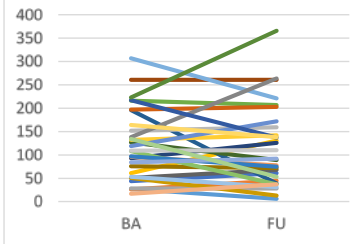
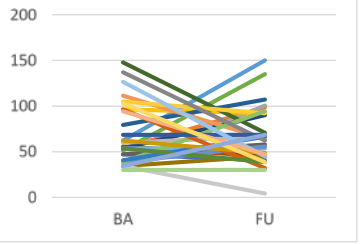
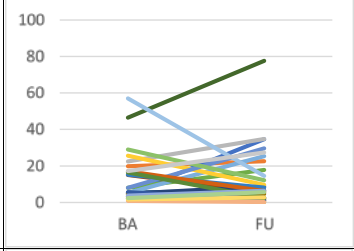
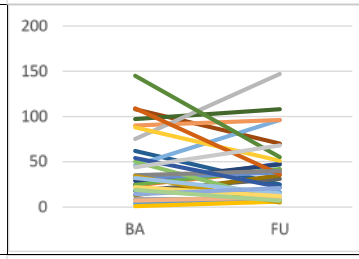
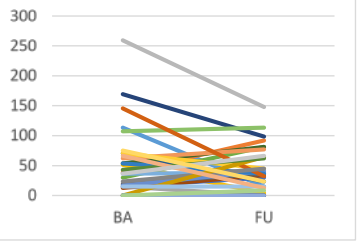
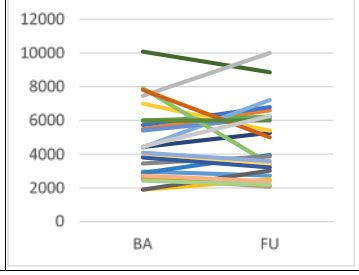
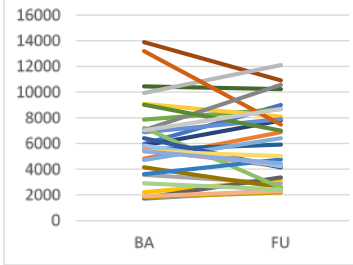
PA Categories	ACTIGRAPH	SENSEWEAR ARMBAND	CHAMPS
Light-to-moderate PA (min/day)			
Light-intensity PA (min/day)			
Moderate-intensity PA (min/day)			
Number of steps/day			Not applicable

Figure 16. Slopes of lines from baseline to follow-up across all physical activity categories measured by each instrument for each individual subject.

changes in PA duration measured by any of the three instruments (Table 9). The correlations were similar and also not significant for measures of PA in caloric expenditure (Data not shown). Due to these poor correlations, ROC curves were not pursued.

Since results on PA duration were similar to those on PA caloric expenditure, for the comparison of responsiveness across the PA measurement tools we present data on PA duration only. Table 10 shows the comparison of magnitude of changes in PA duration between instruments. Results indicated no statistically significant differences across PA categories of the number of steps ($p \geq 0.124$). Moreover, it was observed that the 95% confidence interval of the SRMs across instruments largely overlapped (Tables 7 and 8). When the measurement error was used as a threshold for change in PA (Table 11), weighted Kappa (*K*) indicated moderate agreement between the ACT and SWA on identifying changes in PA that were beyond the standard error of the measurement in activities of moderate-intensity ($K=0.60$), good agreement in number of steps ($K=0.63$), and fair agreement in light-to-moderate PA ($K=0.36$), all of which were statistically significant. Agreement was fair between ACT and SWA in measures of light-intensity PA ($K=0.25$) as well as between ACT and CHAMPS in measures of light-to-moderate PA ($K=0.22$), but were not statistically significant. There was poor agreement between CHAMPS and the activity monitors in measures of light-to-moderate PA ($K \leq 0.07$) and light-intensity PA ($K \leq 0.12$). Agreement was also poor between CHAMPS and SWA in identifying changes in moderate-intensity PA ($K=0.12$). Table 12 depicts the number of times the measurement tools agreed with each other on classifying individuals who had more, less or about the same level of PA using the measurement error as a threshold for change in PA.

Table 7. Duration of physical activity in min/day measured by the ACT, SWA and CHAMPS questionnaire, the magnitude of changes scores between baseline and follow-up, and the standardized response mean. Data represent mean±SD, unless otherwise indicated.

Instrument	PA categories	Baseline	Follow-up	Change in PA	P-value ‡	SRM (95% CI)
ACT	Light-to-moderate (min/day)	81.5±44.4	75.3±47.3	-6.2±36.6 95% CI: -19.9; 7.4	0.358	-0.17 (-0.50; 0.20)
	Light-intensity (min/day)	69.6±35.1	62.4±36.2	-7.2 ±27.5 95% CI: -17.5; 3.1)	0.163	-0.26 (-0.57; 0.11)
	Moderate-intensity (min/day)	11.9±13.4	12.6±16.2	0.6±14.3 95% CI: -4.7; 6.0	0.814	0.04 (-0.33; 0.40)
	Number of Steps (steps/day)	4676±2151	4667±2109	9±1526 95% CI: -607; 625	0.976	-0.01 (-0.35; 0.37)
SWA	Light-to-moderate (min/day)	163.6±104.7	158.6±108.3	-5.0±70.5 95% CI: -31.4; 21.3	0.698	-0.07 (-0.48; 0.23)
	Light-intensity (min/day)	119.3±77.4	117.1±88.8	-2.3±60.2 95% CI: -24.7; 20.3	0.844	-0.04 (-0.40; 0.33)
	Moderate-intensity (min/day)	44.2±37.8	41.4±36.7	-2.8±37.0 95% CI: -16.6; 11.1	0.686	-0.08 (-0.43; 0.29)
	Number of Steps (steps/day)	6003±3311	5960±2995	-42.8±2266.3 95% CI: -889.1; 803.5	0.918	-0.02 (-0.34; 0.37)
CHAMPS	Light-to-moderate (min/day)	121.7±70.4	110.1±53.4	-11.6±64.6 95% CI: -35.7; 12.6	0.335	-0.18 (-0.51; 0.19)
	Light-intensity (min/day)	68.4±57.9	67.0±37.5	-1.4±46.0 95% CI: -18.5; 15.8	0.870	-0.03 (-0.39; 0.33)
	Moderate-intensity (min/day)	53.4±33.3	44.2±31.8	-9.1±46.5 95% CI: -26.5; 8.2	0.290	-0.20 (-0.52; 0.17)

ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire; PA: physical activity; 95% CI: 95% confidence interval; ‡ p-values from analysis of the change scores between baseline and follow-up from Paired t-test; SRM: standardized response mean.

Table 8. Physical activity caloric expenditure in kcal/day measured by the ACT, SWA and CHAMPS questionnaire, the magnitude of changes scores between baseline and follow-up, and the standardized response mean. Data represent mean±SD, unless otherwise indicated.

Instrument	PA categories	Baseline	Follow-up	Change in PA	P-value ‡	SRM (95% CI)
ACT	Light-to-moderate	323.6±157.7	309.4± 190.9	-14.3±114.6 95% CI: -57.1; 28.6	0.501	-0.13 (-0.47; 0.24)
	Light	247.6±99.4	233.9±105.2	-13.7±66.2 95% CI: -38.5; 11.0	0.266	-0.21 (-0.53; 0.16)
	Moderate	76.0±86.0	75.5±112.5	-0.5±72.7 95% CI: -27.7; 26.6	0.969	-0.01 (-0.37; 0.35)
SWA	Light-to-moderate	573.4±351.1	557.8±356.5	-15.6±218.9 95% CI: -97.3; 66.1	0.699	-0.07 (-0.42; 0.30)
	Light	348.1±200.8	343.7± 234.8	-4.4±154.0 95% CI: -61.9; 53.1	0.876	-0.03 (-0.39; 0.33)
	Moderate	217.8±186.0	214.1± 198.4	-3.8±152.5 95% CI: -60.7; 53.2	0.894	-0.03 (-0.39; 0.33)
CHAMPS	Light-to-moderate	492.4±376.1	480.3±280.9	-12.2±289.4 95% CI: -120.2; 95.9	0.820	-0.04 (-0.40; 0.33)
	Light	217.0±110.7	227.9±119.9	10.9±115.7 95% CI: -32.3; 54.2	0.609	-0.09 (-0.43; 0.28)
	Moderate	275.4±328.4	252.4±235.7	-23.1±251.7 95% CI: -117.1; 70.9	0.619	-0.09 (-0.43; 0.28)

95% CI: 95% confidence interval; ‡ p-values from analysis of the change scores between baseline and follow-up from Paired t-test.

Table 9. Change in duration of physical activity measured by the ACT, SWA and CHAMPS questionnaire across descriptors from the global rating of change, and the correlations between changes in PA measured by the instruments and scores from global rating of change.

	ACT	**SWA**	**CHAMPS**
Frequency of descriptors from the GRC-PA and measured changes in light-to-moderate PA in min/day. Median (Q25; Q75)			
About the same active (n=5)	-15.6 (-72.0, 29.9)	- 37.8 (-100.2, 65.7)	0.0 (-93.2, 27.9)
A little bit more active (n=2)	4.3 (-22.0, 30.6)	-18.7 (-29.2, -8.2)	-15.0 (-32.1, 2.1)
Somewhat more active (n=6)	12.0 (-18.1, 27.2)	-17.2 (-29.7, 57.2)	-15.0 (-65.9, 9.1)
Moderately more active (n=5)	-0.6 (-16.1, 0.8)	-11.8 (-110.9, 29.1)	-19.3 (-69.6, 0.0)
Quite a bit more active (n=6)	0.6 (-30.6, 34.1)	59.2 (22.5, 90.6)	20.4 (-51.4, 51.4)
A great deal more active (n=2)	-44.6 (-66.4, -18.8)	-69.1 (-694, -68.8)	-70.7 (-177.9, 36.4)
A very great deal more active (n=4)	21.4 (-26.8, 39.6)	9.7 (-82.3, 29.7)	18.2 (2.1, 108.2)
Spearman's rho correlations between GRC-PA and changes in PA measured by the instruments			
Light-to-moderate PA (min/day)	0.15	0.28	0.14
Light-intensity PA (min/day)	0.13	0.20	0.28
Moderate-intensity PA (min/day)	0.05	0.07	0.21
Number of Steps (min/day)	0.13	0.24	-----

ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire; GRC-PA: global rating of change in physical activity; Q25: quartile 25th; Q75: quartile 75th;

Table 10. Comparison of physical activity change scores across the three measurement tools at each intensity category.

PA category	Δ from ACT Mean ± SD	Δ from SWA Mean ± SD	Δ from CHAMPS Mean ± SD	*df*	*F*	P-value¶
Light-to-moderate (min/day)	0.6 ± 14.3	-2.8 ± 37.0	-9.1 ± 46.5	1, 29	1.940	0.174¶
Light-intensity (min/day)	-6.2 ± 36.6	-5.0 ± 70.5	-11.6 ± 64.6	1, 29	0.622	0.437¶
Moderate-intensity (min/day)	-7.2 ± 27.5	-2.3 ± 60.2	-1.4 ± 46.0	1, 29	1.648	0.209¶
Number of Steps	9 ± 1526	-43 ± 2266	-----	-----	-----	0.124§

ACT: Actigraph, SWA: Sensewear Armband; CHAMPS: Community Health Activities Model Program for Seniors; Δ: Changes in physical activity; *df*: degrees of freedom; *F*: F ratios from analysis of variance ¶: P-value: results from Analysis of Variance (ANOVA) between mean changes measured by the instruments; § P-value: results from paired samples t-test between changes in number of steps measured by the ACT and SWA;

Table 11. Number of subjects who were less active, more active, or did not change based on the standard error of the measurement, and weighted Kappa between measures of PA from the ACT, SWA and the CHAMPS questionnaire.

PA Category		ACT n (%)	SWA n (%)	CHAMPS n (%)	Weighted Kappa (95% CI)		
					ACTxSWA	SWAxCHAMPS	CHAMPSxACT
Light-to-Moderate PA	< SEM	11 (37%)	10 (33%)	12 (40%)	0.36 (0.09; 0.62)*	-0.07 (-0.33; 0.19)	-0.04 (-0.32; 0.24)
	No Δ	10 (33%)	11 (37%)	11 (37%)			
	> SEM	9 (30%)	9 (30%)	7 (23%)			
Light-intensity PA	< SEM	12 (40%)	9 (30%)	10 (33%)	0.25 (-0.03; 0.52)	-0.12 (-0.37; 0.14)	-0.02 (-0.29; 0.26)
	No Δ	11 (37%)	16 (53%)	9 (30%)			
	> SEM	7 (23%)	5 (17%)	11 (37%)			
Moderate-intensity PA	< SEM	7 (23%)	11 (37%)	11 (37%)	0.60 (0.38; 0.81)*	0.12 (-0.16; 0.40)	0.22 (-0.05; 0.49)
	No Δ	16 (53%)	10 (33%)	8 (26%)			
	> SEM	7 (23%)	9 (30%)	11 (37%)			
Number of Steps	< SEM	4 (13%)	9 (30%)	-----	0.63 (0.39; 0.86)*	-----	-----
	No Δ	13 (43%)	10 (33%)	-----			
	> SEM	9 (30%)	11 (37%)	-----			

ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire; PA: physical activity; SEM: standard error of the measurement; n (%): number of subjects (percentage); * statistically significant with $p<0.05$; <SEM: less active based on SEM; >SEM: more active based on SEM; No Δ: No changes in PA based on SEM.

SEM for the ACT is: LPA (13.0 min/day), MPA (5.3 min/day), LMPA (16.9 min/day) and number of steps (709 steps/day).

SEM for the SWA is: LPA (22.5 min/day), MPA (10.8 min/day), LMPA (29.8 min/day) and number of steps (937 steps/day).

SEM for the CHAMPS is: LPA (12.9 min/day), MPA (12.2 min/day) and LMPA (16.8 min/day).

Table 12. Agreement between measurement tools based on results from weighted Kappa. The diagonal line in bold font represents the number of times the instruments agreed on classifying individuals who were more, less or about the same active.

PA Categories	**ACT** (yellow) x **SWA** (orange)					**ACT** (yellow) x **CHAMPS** (orange)					**CHAMPS** (yellow) x **SWA** (orange)				
Light-to-moderate PA		worse	same	better	Total		worse	same	better	Total		worse	same	better	Total
	worse	**5**	5	1	11	worse	**4**	4	3	11	worse	**3**	5	4	12
	same	3	**4**	3	10	same	5	**4**	1	10	same	4	**3**	4	11
	better	1	2	**6**	9	better	3	4	**2**	9	better	2	3	**2**	7
	Total	9	11	10	30	Total	12	12	6	30	Total	9	11	10	30
Light-intensity PA		worse	same	better	Total		worse	same	better	Total		worse	same	better	Total
	worse	**5**	5	1	11	worse	**3**	3	5	11	worse	**2**	5	3	10
	same	5	**5**	2	12	same	5	**4**	3	12	same	4	**4**	1	9
	better	0	4	**3**	7	better	2	2	**3**	7	better	4	5	**2**	11
	Total	10	14	6	30	Total	10	9	11	30	Total	10	14	6	30
Moderate-intensity PA		worse	same	better	Total		worse	same	better	Total		worse	same	better	Total
	worse	**5**	2	0	7	worse	**3**	2	2	7	worse	**3**	6	2	11
	same	5	**8**	3	16	same	7	**5**	4	16	same	4	**2**	2	8
	better	0	0	**7**	7	better	1	1	**5**	7	better	3	2	**6**	11
	Total	10	10	10	30	Total	11	8	11	30	Total	10	10	10	30
Number of steps		worse	same	better	Total										
	worse	**3**	2	0	5										
	same	4	**10**	2	16										
	better	0	1	**8**	9										
	Total	7	13	10	30										

3.5 DISCUSSION

This is the first study to determine and compare the responsiveness of the ACT, SWA, and CHAMPS in assessing PA from light to moderate intensities after rehabilitation following TKA. Results of this study indicate that although a good number of subjects increased, decreased or did not change PA participation from baseline to follow-up, when changes in PA were measured at the group level they were small and not statistically significant. Thus, the ACT, SWA and the CHAMPS demonstrated similar trivial to small degree of internal responsiveness. Furthermore, subjects' perception of changes in PA did not correlate with changes in PA measured by the three instruments and precluded the determination of thresholds of clinically important changes. Study results also demonstrate that when using the standard error of the measurement as a threshold for changes in PA at the individual-level, the two activity monitors (ACT and SWA) demonstrated good agreement in identifying changes in PA that were beyond measurement error, whereas the measures of change from the portable devices did not agree with information from the self-reported questionnaire (CHAMPS).

Results from our study agree with prior studies that assessed the responsiveness of the ACT and CHAMPS and reported trivial to small degree of responsiveness in those PA measurement tools. [44,138-140] Lee et al. evaluated the responsiveness of the ACT during a telephone delivered weight loss intervention that promoted PA and dietary change in individuals with type-2 diabetes (20 to 77 years old) and found a SRM of 0.24 for measures of moderate-intensity PA.[138] Swartz et al. determined the responsiveness of the ACT in sedentary healthy adults (21 to 67 years old) during an

intervention to reduce sitting time and found a SRM of 0.18 for activities performed at 1 to 1.9 METs, 0.19 for activities at 2 to 2.9 METs, and 0.30 for activities above 3 METs. [140] Gardiner et al. determined the responsiveness of the ACT in detecting sedentary time in nonworking older adults (average age 73 years old), and found an effect size of 0.36 after an intervention that targeted a reduction of sedentary time. [139] One study investigated the responsiveness of the CHAMPS questionnaire in healthy older adults (65 to 90 years old) during interventions to increase PA participation and found effect sizes of 0.33 and 0.37 for measures of moderate-intensity PA and light-to-moderate PA respectively. [44]

Investigating responsiveness of measures of PA in older adults with arthritis of the lower extremities is challenging due to the difficulty in promoting changes in lifestyle and the high variability in PA measures, both of which lead to small effect sizes. To assess responsiveness of measures of PA, regardless of the population, seems problematic because of the inherent difficulty to promote changes in PA behavior. A recent meta-analysis assessed the overall effectiveness of interventions designed to increase PA in 3,643 healthy older adults included in 33 studies.[39] These studies used a pretest-posttest design and assessed PA utilizing a variety of methods, including activity monitors and questionnaires such as the ACT, SWA and CHAMPS. The authors reported a small pooled effect size of 0.23 (95% CI: 0.15; 0.31),[39] indicating only minimal changes in activity behavior. Thus, the development of interventions able to promote substantial changes in activity behavior is warranted. Recently, funding agencies such as the National Institute of Health have issued announcements seeking innovative interventions to increase PA participation. In terms of high variability in the PA measures, the coefficient of variation for change scores (i.e., ratio of standard deviation to the mean) was very high in our study and across the studies that reported responsiveness indices of the ACT and CHAMPS (cited above).[44,138-140] The coefficients of variation

ranged from 3 to 33 times the mean for change in duration of PA and from 53 to 170 times the mean for change in number of steps. The higher coefficient of variation for changes in number of steps is because the mean change for the group was very small in comparison to the variability of the change. Since variability serves as the numerator of the equation, it resulted in enormous coefficients of variation. The large variability in change scores can be attributed to the fact that about 1/3 of the subjects became more active, 1/3 became less active, and 1/3 did not change their PA participation at follow-up. Consequently, the instruments were not able to detect changes at the group-level, which makes the utilization of methods that investigate changes in PA at the individual-level more appealing.

In this study we attempted to assess responsiveness at the individual-level using a modified global rating of change as an external anchor. However, changes perceived by the subjects had only poor to small correlations with changes in PA measured by the three instruments. We believe that the lack of correlation between perceived and measured changes in PA is partially due to social desirability bias.[143] When completing the global rating of change in PA after the intervention, subjects tended to respond to the question in a socially acceptable direction. This tendency may have led to an overstatement of changes in PA participation since none of them reported less PA. Although our research team diligently attempted to decrease social desirability bias by instructing the subjects to provide honest and objective answer on changes in PA after the intervention, it seems that it was not enough to prevent that. In addition, the small correlations may also be a result of the difficulty that subjects have in adjudicating changes in the amount of activities during a long period of time (6 months) used in this study. The perception of change in PA participation over 6 months may be particularly difficult for older adults with knee osteoarthritis because they usually engage mainly in light-intensity PA and this type of PA has been shown to be the most difficult to recall.[144]

We are not aware of any studies that attempted to use external responsiveness methods to investigate sensitivity of change of the ACT, SWA or CHAMPS to compare our results for external responsiveness.

An alternative to assess changes in PA at the individual-level is the use of the standard error of the measurement as a threshold. Using this approach was a practical step taken to identify changes in PA that were above measurement error, which might be appropriate in situations where patient-perceived changes are unavailable.[145] While measurement error has been used as a threshold in many studies that investigated changes in other outcomes [146,147] this is the first study to discuss changes in PA behavior based on measurement error. By using this approach, we were able to test the agreement between PA measurement tools in classifying individuals who did or did not change their PA participation. While the ACT and SWA generally agreed on classifying those who changed, the activity monitors disagreed with PA changes measured by the CHAMPS. This discrepancy may have been due to limitations inherent of PA questionnaires such as recall-bias.[48,119]

A weakness of this study is the small sample size. Although this study has a relatively small sample size, it is unlikely that the non-significant results for internal responsiveness were due to type 2 error because the changes in PA were all very small. As discussed above, these small changes were likely the results of difficulties in changing subjects' lifestyle along with high variability in change scores. Therefore, to properly assess the responsiveness of PA measures when there is no warranty that the construct of PA will change, future studies in older adults with arthritis of the lower extremity could be performed in a controlled environment. A recent study assessed the responsiveness of the ACT in children and adolescents (6-16 year of age) who performed low- and high-intensity activities in a simulated free-living setting.[148] Their results showed that the ACT had a high degree of responsiveness to capture changes from low- to high-intensity activities (SRM≥1.1).

3.6 CONCLUSION

Our study showed that, although the ACT, SWA and CHAMPS had small degrees of responsiveness, the activity monitors (ACT and SWA) agreed on detecting changes in PA when measurement error was used as a threshold. Therefore, when investigating changes in PA behavior in individuals with arthritis of the lower extremities, clinicians and researchers should consider the alternative to interpret their results based on changes that occurred beyond measurement error, since those changes may not be detectable at a group-level.

4.0 CHAPTER 4 - SIGNIFICANCE AND DIRECTION OF FUTURE RESEARCH

The focus of this dissertation was to investigate and concurrently compare the psychometric properties of three instruments to measure PA in older adults with knee osteoarthritis who underwent TKA. Chapter 1 presented the gaps in knowledge, relevance of our investigation, and methods and procedures proposed in this dissertation. Chapter 2 described the study that investigated the reliability of the ACT, SWA and CHAMPS questionnaire in measuring PA in free-living condition. Results indicated that the three instruments were highly reliable. However, the magnitude of PA measures from the ACT were significantly lower compared to the SWA across intensity categories; the CHAMPS showed systematic bias in measures of light-intensity PA; and the SWA showed better precision amongst the three PA measurement tools. Also, this study provided information on the reliability of the activity monitors in measuring PA during different timeframes. Measuring PA during wake-hours (i.e., time between getting out of bed in the morning to going back to bed at night) seems to be the most appropriate timeframe to measure PA. Measures of PA taken during wake-hours are highly reliable, provide a good representation of daily activities in comparison to 24-hours timeframe, and may help reduce the burden on subjects when wearing activity monitors.

Chapter 3 described the study that assessed the responsiveness of the ACT, SWA and CHAMPS after exercise programs in subjects who underwent TKA. Results from this study indicated that the three instruments have similarly low degrees of responsiveness. Our findings showed that assessing responsiveness at the group-level is less than ideal since changing PA

behavior is very difficult and the variability in the data for change scores is very large, both of which result in small effect sizes. Also, we found that using an anchor based on self-reported changes in PA was not feasible since its scores did not correlate with changes assessed by the PA measurement tools. In that study, we proposed an alternative to investigate changes in PA at the individual-level by using the standard error of the measurement as a threshold for changes in PA. This approach is practical as it can be used to identify changes in PA for individual patients.

Validity is also an important aspect of the psychometrics of a measurement instrument. Although not part of my dissertation, during my PhD training I investigated the concurrent criterion-related validity of the ACT and SWA in comparison to the criterion method of indirect calorimetry. [84] Twenty-one older adults who underwent TKA participated in an 80-minute protocol that included 9 increasingly demanding activities in a laboratory setting. Energy expenditure was concurrently estimated by the activity monitors and the indirect calorimetry throughout the protocol. Our results indicated that the ACT underestimated the values of energy expenditure by 40% to 100% in comparison to the criterion method. The differences between the SWA and the criterion were small, not greater than 19%. Also, the SWA was able to accurately assess non-walking activities (i.e., lying down, sit and read, computer-work, stand and talk, stand and load shelves, and mopping), while the ACT was not. Overall, the SWA showed better criterion-related validity than the ACT to measure PA in older adults with knee osteoarthritis who underwent TKA.

Based on the combined results from my studies on the psychometric properties (reliability, criterion-related validity, and responsiveness) of the ACT, SWA and CHAMPS, it seems that in general the activity monitors are a better choice than the CHAMPS questionnaire to measure PA in older adults with arthritis of the lower extremities. Although the ACT and SWA are costly in comparison to the CHAMPS, the questionnaire was susceptible to recall bias and did not agree with

the activity monitors in the identification of patients who changed versus those who did not change PA participation. However, it is important to mention that the recall bias occurred for light-intensity PA only, and that the reliability of measures of light-to-moderate and moderate-intensity PA from the CHAMPS was high and comparable to the reliability of the SWA in those PA categories. Therefore, the CHAMPS may be useful to assess light-to-moderate and moderate-intensity PA in cross-sectional studies since its responsiveness to interventions seems to be limited. Amongst the activity monitors, the SWA seems to be the best option. The SWA showed better accuracy and reliability than the ACT. In particular, the SWA has been shown to be capable of accurately and reliably capture light-intensity activities, such as slow walking and tasks that do not involve ambulation, which are usually performed by individuals with arthritis of the lower extremities. The significance of these series of studies is that they provide evidence for a well-educated choice of measurement tools to be used when assessing PA in this population and will help to interpret data on changes in PA in future trials targeting PA promotion to increase activity participation.

From the time when we proposed these studies to date, newer accelerometry-based activity monitors have been released in the market. These devices are much cheaper and smaller than the ones used in this study. Some examples are the BodyMedia FIT, Fitbit, Jawbone Up, NikeFuel band, iFit, Runtastic Orbit, Garmin Vivosmart, and more. Since these activity monitors are usually wrist-mounted and a few are waterproof, they can decrease patients' burden when measuring their daily PA. In addition, numerous applications for smartphones have been developed and refined over the years. Although all these new activity monitors and mobile apps have been developed to attract the general public, future research studies should test their psychometric properties to provide insights on appropriateness to specific populations.

By conducting these studies I acquired comprehensive and in depth understanding of measures of PA and confirmed with my own data that changing activity behavior in older adults is challenging. I plan to use this knowledge to transition into a different but related area of research that entails the development of interventions to promote PA. For many years the world health organization has highlighted the need for developing interventions to increase PA participation. While recent systematic reviews and meta-analysis identified several types of PA promotion interventions, they showed only small treatment effects, [149,150] which warrants refinement of those interventions. The development of innovative and effective interventions to promote PA is currently the focus of several funding institutions such as the National Institute of Health and several Foundations. The world health organization is targeting for a reduction of 10% in physical inactivity by 2025. Therefore, my plan is to take advantage of this momentum in research to develop innovative interventions to promote PA that may impact health in individuals with arthritis.

5.0 BIBLIOGRAPHY

1. Naal FD, Impellizzeri FM. How active are patients undergoing total joint arthroplasty?: A systematic review. *Clin Orthop Relat Res.* 2010;468(7):1891-1904.
2. Namba RS, Paxton L, Fithian DC, Stone ML. Obesity and perioperative morbidity in total hip and total knee arthroplasty patients. *J Arthroplasty.* 2005;20(7 Suppl 3):46-50.
3. Dickstein R, Heffes Y, Shabtai EI, Markowitz E. Total knee arthroplasty in the elderly: patients' self-appraisal 6 and 12 months postoperatively. *Gerontology.* 1998;44(4):204-210.
4. Hawker G, Wright J, Coyte P, et al. Health-related quality of life after knee replacement. *J Bone Joint Surg Am.* 1998;80(2):163-173.
5. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. *J Rheumatol.* 2000;27(7):1745-1752.
6. American Education Research Association APA, & National Council on Measurement in Education. *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association; 1999.
7. Guyatt G, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40(2):171-178.
8. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis.* 1985;38(1):27-36.

9. Murphy CO, Davidshofer KR. *Psychological testing: principles and applications*. 6th ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2005.

10. Taylor JR. *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. University Science Books; 1999.

11. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2(14):i-iv, 1-74.

12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.

13. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976)*. 2000;25(24):3192-3199.

14. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459-468.

15. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28(7):632-642.

16. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract*. 2000;6(1):39-49.

17. Verbunt JA, Huijnen IP, Koke A. Assessment of physical activity in daily life in patients with musculoskeletal pain. *Eur J Pain*. 2009;13(3):231-242.

18. Melanson EL, Jr., Freedson PS. Physical activity assessment: a review of methods. *Crit Rev Food Sci Nutr*. 1996;36(5):385-396.

19. Harada ND, Chiu V, King AC, Stewart AL. An evaluation of three self-report physical activity instruments for older adults. *Med Sci Sports Exerc*. 2001;33(6):962-970.

20. Bassett DR, Jr., Ainsworth BE, Leggett SR, et al. Accuracy of five electronic pedometers for measuring distance walked. *Med Sci Sports Exerc.* 1996;28(8):1071-1077.

21. Cyarto EV, Myers A, Tudor-Locke C. Pedometer accuracy in nursing home and community-dwelling older adults. *Med Sci Sports Exerc.* 2004;36(2):205-209.

22. Macko RF, DeSouza CA, Tretter LD, et al. Treadmill aerobic exercise training reduces the energy expenditure and cardiovascular demands of hemiparetic gait in chronic stroke patients. A preliminary report. *Stroke.* 1997;28(2):326-330.

23. Motl RW, McAuley E, Snook EM, Scott JA. Accuracy of two electronic pedometers for measuring steps taken under controlled conditions among ambulatory individuals with multiple sclerosis. *Mult Scler.* 2005;11(3):343-345.

24. Schmalzried TP, Szuszczewicz ES, Northfield MR, et al. Quantitative assessment of walking activity after total hip or knee replacement. *J Bone Joint Surg Am.* 1998;80(1):54-59.

25. Swartz AM, Bassett DR, Jr., Moore JB, Thompson DL, Strath SJ. Effects of body mass index on the accuracy of an electronic pedometer. *Int J Sports Med.* 2003;24(8):588-592.

26. Ainsworth BE, Bassett DR, Jr., Strath SJ, et al. Comparison of three methods for measuring the time spent in physical activity. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S457-464.

27. Bassett DRJ, Strath S. *Use of pedometers to assess physical activity.* Champaign, IL: Human Kinetics; 2002.

28. Melanson EL, Jr., Freedson PS. Validity of the Computer Science and Applications, Inc. (CSA) activity monitor. *Med Sci Sports Exerc.* 1995;27(6):934-940.

29. Lee IM, Paffenbarger RS, Jr., Hsieh C. Physical activity and risk of developing colorectal cancer among college alumni. *J Natl Cancer Inst.* 1991;83(18):1324-1329.

30. LaPorte RE, Montoye HJ, Caspersen CJ. Assessment of physical activity in epidemiologic research: problems and prospects. *Public Health Rep.* 1985;100(2):131-146.

31. Bouten CV, Sauren AA, Verduin M, Janssen JD. Effects of placement and orientation of body-fixed accelerometers on the assessment of energy expenditure during walking. *Med Biol Eng Comput.* 1997;35(1):50-56.

32. Services USDoHaH. 2008 Physical Activity Guidelines for Americans. 2008; http://www.health.gov/paguidelines/.

33. Hendelman D, Miller K, Baggett C, Debold E, Freedson P. Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S442-449.

34. Welk GJ, Schaben JA, Morrow JR, Jr. Reliability of accelerometry-based activity monitors: a generalizability study. *Med Sci Sports Exerc.* 2004;36(9):1637-1645.

35. Welk GJ. *Use of accelerometry based activity monitors to assess physical activity.* Champaign, IL Human Kinetics; 2002.

36. Katz P., O'Grady M., Davis G., et al. Exercise prescription for older adults with osteoarthritis pain: consensus practice recommendations. A supplement to the AGS Clinical Practice Guidelines on the management of chronic pain in older adults. *J Am Geriatr Soc.* 2001;49(6):808-823.

37. Blair SN, Dowda M, Pate RR, et al. Reliability of long-term recall of participation in physical activity by middle-aged men and women. *Am J Epidemiol.* 1991;133(3):266-275.

38. Jacobs DR, Jr., Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. *Med Sci Sports Exerc.* 1993;25(1):81-91.

39. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. *J Clin Epidemiol.* 1993;46(2):153-162.

40. Washburn RA, McAuley E, Katula J, Mihalko SL, Boileau RA. The physical activity scale for the elderly (PASE): evidence for validity. *J Clin Epidemiol.* 1999;52(7):643-651.

41. Bonnefoy M, Normand S, Pachiaudi C, Lacour JR, Laville M, Kostka T. Simultaneous validation of ten physical activity questionnaires in older men: a doubly labeled water study. *J Am Geriatr Soc.* 2001;49(1):28-35.

42. Dipietro L, Caspersen CJ, Ostfeld AM, Nadel ER. A survey for assessing physical activity among older adults. *Med Sci Sports Exerc.* 1993;25(5):628-642.

43. Colbert LH, Matthews CE, Havighurst TC, Kim K, Schoeller DA. Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc.* 2011;43(5):867-876.

44. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Med Sci Sports Exerc.* 2001;33(7):1126-1141.

45. Stewart AL, Verboncoeur CJ, McLellan BY, et al. Physical activity outcomes of CHAMPS II: a physical activity promotion program for older adults. *J Gerontol A Biol Sci Med Sci.* 2001;56(8):M465-470.

46. Giles K, Marshall AL. Repeatability and accuracy of CHAMPS as a measure of physical activity in a community sample of older Australian adults. *J Phys Act Health.* 2009;6(2):221-229.

47. Kaleth AS, Ang DC, Chakr R, Tong Y. Validity and reliability of community health activities model program for seniors and short-form international physical activity questionnaire as

physical activity assessment tools in patients with fibromyalgia. *Disabil Rehabil.* 2010;32(5):353-359.

48. Hekler EB, Buman MP, Haskell WL, et al. Reliability and validity of CHAMPS self-reported sedentary-to-vigorous intensity physical activity in older adults. *J Phys Act Health.* 2012;9(2):225-236.

49. Castro CM, Pruitt LA, Buman MP, King AC. Physical activity program delivery by professionals versus volunteers: the TEAM randomized trial. *Health Psychol.* 2011;30(3):285-294.

50. Pruitt LA, Glynn NW, King AC, et al. Use of accelerometry to measure physical activity in older adults at risk for mobility disability. *J Aging Phys Act.* 2008;16(4):416-434.

51. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. *Am J Prev Med.* 2008;34(2):138-142.

52. Stewart AL, Gillis D, Grossman M, et al. Diffusing a research-based physical activity promotion program for seniors into diverse communities: CHAMPS III. *Prev Chronic Dis.* 2006;3(2):A51.

53. Feldman LS, Kaneva P, Demyttenaere S, Carli F, Fried GM, Mayo NE. Validation of a physical activity questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy. *Surgery.* 2009;146(1):31-39.

54. Wilcox S, Dowda M, Leviton LC, et al. Active for life: final results from the translation of two physical activity programs. *Am J Prev Med.* 2008;35(4):340-351.

55. Berntsen S, Hageberg R, Aandstad A, et al. Validity of physical activity monitors in adults participating in free-living activities. *Br J Sports Med.* 2010;44(9):657-664.

56. Bassett DR, Jr., Ainsworth BE, Swartz AM, Strath SJ, O'Brien WL, King GA. Validity of four motion sensors in measuring moderate intensity physical activity. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S471-480.

57. Welk GJ, Blair SN, Wood K, Jones S, Thompson RW. A comparative evaluation of three accelerometry-based physical activity monitors. *Med Sci Sports Exerc.* 2000;32(9 Suppl):S489-497.

58. King GA, Torres N, Potter C, Brooks TJ, Coleman KJ. Comparison of activity monitors to estimate energy cost of treadmill exercise. *Med Sci Sports Exerc.* 2004;36(7):1244-1251.

59. Crouter SE, Schneider PL, Karabulut M, Bassett DR, Jr. Validity of 10 electronic pedometers for measuring steps, distance, and energy cost. *Med Sci Sports Exerc.* 2003;35(8):1455-1460.

60. Bassett DR, Jr. Validity and reliability issues in objective monitoring of physical activity. *Res Q Exerc Sport.* 2000;71(2 Suppl):S30-36.

61. Bassett DR, Jr., Cureton AL, Ainsworth BE. Measurement of daily walking distance-questionnaire versus pedometer. *Med Sci Sports Exerc.* 2000;32(5):1018-1023.

62. St-Onge M, Mignault D, Allison DB, Rabasa-Lhoret R. Evaluation of a portable device to measure daily energy expenditure in free-living adults. *Am J Clin Nutr.* 2007;85(3):742-749.

63. Jakicic JM, Marcus M, Gallagher KI, et al. Evaluation of the SenseWear Pro Armband to assess energy expenditure during exercise. *Med Sci Sports Exerc.* 2004;36(5):897-904.

64. Cole PJ, LeMura LM, Klinger TA, Strohecker K, McConnell TR. Measuring energy expenditure in cardiac patients using the Body Media Armband versus indirect calorimetry. A validation study. *J Sports Med Phys Fitness.* 2004;44(3):262-271.

65. Fruin ML, Rankin JW. Validity of a multi-sensor armband in estimating rest and exercise energy expenditure. *Med Sci Sports Exerc.* 2004;36(6):1063-1069.

66. Papazoglou D, Augello G, Tagliaferri M, et al. Evaluation of a multisensor armband in estimating energy expenditure in obese individuals. *Obesity (Silver Spring).* 2006;14(12):2217-2223.

67. Dwyer TJ, Alison JA, McKeough ZJ, Elkins MR, Bye PT. Evaluation of the SenseWear activity monitor during exercise in cystic fibrosis and in health. *Respir Med.* 2009;103(10):1511-1517.

68. Esliger DW, Tremblay MS. Technical reliability assessment of three accelerometer models in a mechanical setup. *Med Sci Sports Exerc.* 2006;38(12):2173-2181.

69. McClain JJ, Sisson SB, Tudor-Locke C. Actigraph accelerometer interinstrument reliability during free-living in adults. *Med Sci Sports Exerc.* 2007;39(9):1509-1514.

70. Sirard JR, Forsyth A, Oakes JM, Schmitz KH. Accelerometer test-retest reliability by data processing algorithms: results from the Twin Cities Walking Study. *J Phys Act Health.* 2011;8(5):668-674.

71. Polcyn AF, Lipsitz LA, Kerrigan DC, Collins JJ. Age-related changes in the initiation of gait: degradation of central mechanisms for momentum generation. *Arch Phys Med Rehabil.* 1998;79(12):1582-1589.

72. Mandeville D, Osternig LR, Chou LS. The effect of total knee replacement surgery on gait stability. *Gait Posture.* 2008;27(1):103-109.

73. Gage WH, Frank JS, Prentice SD, Stevenson P. Postural responses following a rotational support surface perturbation, following knee joint replacement: frontal plane rotations. *Gait Posture.* 2008;27(2):286-293.

74. Mizner RL, Snyder-Mackler L. Altered loading during walking and sit-to-stand is affected by quadriceps weakness after total knee arthroplasty. *J Orthop Res.* 2005;23(5):1083-1090.

75. Brazeau AS, Karelis AD, Mignault D, Lacroix MJ, Prud'homme D, Rabasa-Lhoret R. Test-retest reliability of a portable monitor to assess energy expenditure. *Appl Physiol Nutr Metab.* 2011;36(3):339-343.

76. Heiermann S, Khalaj Hedayati K, Muller MJ, Dittmar M. Accuracy of a portable multisensor body monitor for predicting resting energy expenditure in older people: a comparison with indirect calorimetry. *Gerontology.* 2011;57(5):473-479.

77. Wafa SW, Talib RA, Hamzaid NH, et al. Randomized controlled trial of a good practice approach to treatment of childhood obesity in Malaysia: Malaysian Childhood Obesity Treatment Trial (MASCOT). *Int J Pediatr Obes.* 2011;6(2-2):e62-69.

78. Dua JS, Cooper AR, Fox KR, Graham Stuart A. Exercise training in adults with congenital heart disease: feasibility and benefits. *Int J Cardiol.* 2010;138(2):196-205.

79. Oliveira J, Ribeiro F, Gomes H. Effects of a home-based cardiac rehabilitation program on the physical activity levels of patients with coronary artery disease. *J Cardiopulm Rehabil Prev.* 2008;28(6):392-396.

80. Mantovani G, Madeddu C, Serpe R. Improvement of physical activity as an alternative objective variable to measure treatment effects of anticachexia therapy in cancer patients. *Curr Opin Support Palliat Care.* 2010;4(4):259-265.

81. Pezzino S, Florenty S, Fagour C, Gin H, Rigalleau V. Remedial actions for the physical inactivity of hospitalized patients with type 2 diabetes. *Diabetes Care.* 2010;33(9):1960-1961.

82. Lundholm K, Gunnebo L, Korner U, et al. Effects by daily long term provision of ghrelin to unselected weight-losing cancer patients: a randomized double-blind study. *Cancer.* 2010;116(8):2044-2052.

83. Cazzola M, Segreti A, Stirpe E, Appodia M, Senis L, Matera MG. Energy expenditure and impact of bronchodilators in COPD patients. *Respir Med.* 2010;104(10):1490-1494.

84. Almeida GJ, Wert DM, Brower KS, Piva SR. Validity of physical activity measures in individuals after total knee arthroplasty. *Arch Phys Med Rehabil.* 2015;96(3):524-531.

85. Nichols JF, Morgan CG, Chabot LE, Sallis JF, Calfas KJ. Assessment of physical activity with the Computer Science and Applications, Inc., accelerometer: laboratory versus field validation. *Res Q Exerc Sport.* 2000;71(1):36-43.

86. Brage S, Wedderkopp N, Franks PW, Andersen LB, Froberg K. Reexamination of validity and reliability of the CSA monitor in walking and running. *Med Sci Sports Exerc.* 2003;35(8):1447-1454.

87. Leenders NY, Nelson TE, Sherman WM. Ability of different physical activity monitors to detect movement during treadmill walking. *Int J Sports Med.* 2003;24(1):43-50.

88. Crouter SE, Churilla JR, Bassett DR, Jr. Estimating energy expenditure using accelerometers. *Eur J Appl Physiol.* 2006;98(6):601-612.

89. Catharina B, Gunnevi S, Christel L. Validity of armband measuring energy expenditure in overweight and obese children. *Med Sci Sports Exerc.* 2010;42(6):1154-1161.

90. Johannsen DL, Calabro MA, Stewart J, Franke W, Rood JC, Welk GJ. Accuracy of armband monitors for measuring daily energy expenditure in healthy adults. *Med Sci Sports Exerc.* 2010;42(11):2134-2140.

91. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-415.

92. Mazzeo RS, Tanaka H. Exercise prescription for the elderly: current recommendations. *Sports Med.* 2001;31(11):809-818.

93. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc.* 2007;39(8):1435-1445.

94. Moffet H, Collet JP, Shapiro SH, Paradis G, Marquis F, Roy L. Effectiveness of intensive rehabilitation on functional ability and quality of life after first total knee arthroplasty: A single-blind randomized controlled trial. *Arch Phys Med Rehabil.* 2004;85(4):546-556.

95. Piva SR, Gil AB, Almeida GJ, DiGioia AM, 3rd, Levison TJ, Fitzgerald GK. A balance exercise program appears to improve function for patients with total knee arthroplasty: a randomized clinical trial. *Phys Ther.* 2010;90(6):880-894.

96. Petterson SC, Mizner RL, Stevens JE, et al. Improved function from progressive strengthening interventions after total knee arthroplasty: a randomized clinical trial with an imbedded prospective cohort. *Arthritis Rheum.* 2009;61(2):174-183.

97. Kraus WE, Slentz CA. Exercise training, lipid regulation, and insulin action: a tangled web of cause and effect. *Obesity (Silver Spring).* 2009;17 Suppl 3:S21-26.

98. Slentz CA, Houmard JA, Kraus WE. Exercise, abdominal obesity, skeletal muscle, and metabolic risk: evidence for a dose response. *Obesity (Silver Spring).* 2009;17 Suppl 3:S27-33.

99. Jakicic JM, Wing RR, Butler BA, Robertson RJ. Prescribing exercise in multiple short bouts versus one continuous bout: effects on adherence, cardiorespiratory fitness, and weight loss in overweight women. *Int J Obes Relat Metab Disord.* 1995;19(12):893-901.

100. Jakicic JM, Marcus BH, Gallagher KI, Napolitano M, Lang W. Effect of exercise duration and intensity on weight loss in overweight, sedentary women: a randomized trial. *JAMA*. 2003;290(10):1323-1330.

101. Board NAoSIoMFaN. Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients). In: Agriculture Do, ed: Washington National Academic Press; 2005.

102. Services USDoAaUSDoHaH. Dietary Guidelines for Americans. In: Office USGP, ed. Washington, DC2010.

103. Jakicic JM, Marcus BH, Lang W, Janney C. Effect of exercise on 24-month weight loss maintenance in overweight women. *Arch Intern Med*. 2008;168(14):1550-1559; discussion 1559-1560.

104. Flicker L, McCaul KA, Hankey GJ, et al. Body mass index and survival in men and women aged 70 to 75. *J Am Geriatr Soc*. 2010;58(2):234-241.

105. Perri MG, Nezu AM, McKelvey WF, Shermer RL, Renjilian DA, Viegener BJ. Relapse prevention training and problem-solving therapy in the long-term management of obesity. *J Consult Clin Psychol*. 2001;69(4):722-726.

106. Marlatt GA, George WH. Relapse prevention: introduction and overview of the model. *Br J Addict*. 1984;79(3):261-273.

107. Nunnally JC, Bernstein IH. *Psychometric theory*. New York: McGraw-Hill; 1994.

108. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. Stamford: Appleton & Lange; 1993.

109. Paterson DH, Warburton DE. Physical activity and functional limitations in older adults: a systematic review related to Canada's Physical Activity Guidelines. *Int J Behav Nutr Phys Act.* 2010;7:38.

110. de Groot IB, Bussmann JB, Stam HJ, Verhaar JA. Actual everyday physical activity in patients with end-stage hip or knee osteoarthritis compared with healthy controls. *Osteoarthritis Cartilage.* 2008;16(4):436-442.

111. Brandes M, Ringling M, Winter C, Hillmann A, Rosenbaum D. Changes in physical activity and health-related quality of life during the first year after total knee arthroplasty. *Arthritis Care Res (Hoboken).* 2011;63(3):328-334.

112. Holsgaard-Larsen A, Roos EM. Objectively measured physical activity in patients with end stage knee or hip osteoarthritis. *Eur J Phys Rehabil Med.* 2012;48(4):577-585.

113. Cawthon PM, Blackwell TL, Cauley JA, et al. Objective assessment of activity, energy expenditure, and functional limitations in older men: the Osteoporotic Fractures in Men study. *J Gerontol A Biol Sci Med Sci.* 2013;68(12):1518-1524.

114. Herrmann SD, Barreira TV, Kang M, Ainsworth BE. Impact of accelerometer wear time on physical activity data: a NHANES semisimulation data approach. *Br J Sports Med.* 2014;48(3):278-282.

115. Calabro MA, Lee JM, Saint-Maurice PF, Yoo H, Welk GJ. Validity of physical activity monitors for assessing lower intensity activity in adults. *Int J Behav Nutr Phys Act.* 2014;11(1):119.

116. Van Remoortel H, Raste Y, Louvaris Z, et al. Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS One.* 2012;7(6):e39198.

117. Lee JM, Kim Y, Welk GJ. Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc.* 2014;46(9):1840-1848.

118. Wetten AA, Batterham M, Tan SY, Tapsell L. Relative validity of 3 accelerometer models for estimating energy expenditure during light activity. *J Phys Act Health.* 2014;11(3):638-647.

119. Neilson HK, Robson PJ, Friedenreich CM, Csizmadi I. Estimating activity energy expenditure: how valid are physical activity questionnaires? *Am J Clin Nutr.* 2008;87(2):279-291.

120. Buman MP, Hekler EB, Haskell WL, et al. Objective light-intensity physical activity associations with rated health in older adults. *Am J Epidemiol.* 2010;172(10):1155-1165.

121. Milner CE. Is gait normal after total knee arthroplasty? Systematic review of the literature. *J Orthop Sci.* 2009;14(1):114-120.

122. Almeida GJ, Wasko MC, Jeong K, Moore CG, Piva SR. Physical activity measured by the SenseWear Armband in women with rheumatoid arthritis. *Phys Ther.* 2011;91(9):1367-1376.

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

124. Miller GD, Jakicic JM, Rejeski WJ, et al. Effect of varying accelerometry criteria on physical activity: the look ahead study. *Obesity (Silver Spring).* 2013;21(1):32-44.

125. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15(12):1833-1840.

126. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S531-543.

127. Ward DS, Evenson KR, Vaughn A, Rodgers AB, Troiano RP. Accelerometer use in physical activity: best practices and research recommendations. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S582-588.

128. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998;30(5):777-781.

129. Miwa H, Sasahara S, Matsui T. Roll-over detection and sleep quality measurement using a wearable sensor. *Conf Proc IEEE Eng Med Biol Soc.* 2007;2007:1507-1510.

130. Teller A. A platform for wearable physiological computing. *Interacting with Computers.* 2004;16(5):917-937.

131. BaHammam A, Alrajeh M, Albabtain M, Bahammam S, Sharif M. Circadian pattern of sleep, energy expenditure, and body temperature of young healthy men during the intermittent fasting of Ramadan. *Appetite.* 2010;54(2):426-429.

132. Cumming G, Finch S. Inference by eye: confidence intervals and how to read pictures of data. *Am Psychol.* 2005;60(2):170-180.

133. Webber SC, Magill SM, Schafer JL, Wilson KC. GT3X+ accelerometer, Yamax pedometer and SC-StepMX pedometer step count accuracy in community-dwelling older adults. *J Aging Phys Act.* 2014;22(3):334-341.

134. Storti KL, Pettee KK, Brach JS, Talkowski JB, Richardson CR, Kriska AM. Gait speed and step-count monitor accuracy in community-dwelling older adults. *Med Sci Sports Exerc.* 2008;40(1):59-64.

135. Barreira TV, Brouillette RM, Foil HC, Keller JN, Tudor-Locke C. Comparison of older adults' steps per day using NL-1000 pedometer and two GT3X+ accelerometer filters. *J Aging Phys Act.* 2013;21(4):402-416.

136. Pozzi F, Snyder-Mackler L, Zeni J. Physical exercise after knee arthroplasty: a systematic review of controlled trials. *Eur J Phys Rehabil Med.* 2013;49(6):877-892.

137. Piva SR, Catelani MC, Almeida GJ. Comprehensive Behavioral Intervention Compared to Standard of Care Exercise Program after Total Knee Arthroplasty: A Pilot Randomized Trial. 2013.

138. Lee WY, Clark BK, Winkler E, Eakin EG, Reeves MM. Responsiveness to Change of Self-Report and Device-Based Physical Activity Measures in the Living Well With Diabetes Trial. *J Phys Act Health.* 2014.

139. Gardiner PA, Clark BK, Healy GN, Eakin EG, Winkler EA, Owen N. Measuring older adults' sedentary time: reliability, validity, and responsiveness. *Med Sci Sports Exerc.* 2011;43(11):2127-2133.

140. Swartz AM, Rote AE, Cho YI, Welch WA, Strath SJ. Responsiveness of motion sensors to detect change in sedentary and physical activity behaviour. *Br J Sports Med.* 2014;48(13):1043-1047.

141. Almeida GJ, Irrgang JJ, Piva SR. Reliability of physical activity measures during free-living in individuals with knee osteoarthritis. *J Orthop Sports Phys Ther* 2015;45(1):A188.

142. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159-174.

143. Adams SA, Matthews CE, Ebbeling CB, et al. The effect of social desirability and social approval on self-reports of physical activity. *Am J Epidemiol.* 2005;161(4):389-398.

144. Washburn RA. Assessment of physical activity in older adults. *Res Q Exerc Sport.* 2000;71(2 Suppl):S79-88.

145. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102-109.

146. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86(5):735-743.

147. Copay AG, Subach BR, Glassman SD, Polly DW, Jr., Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *The spine journal : official journal of the North American Spine Society.* 2007;7(5):541-546.

148. Montoye AH, Pfeiffer KA, Suton D, Trost SG. Evaluating the Responsiveness of Accelerometry to Detect Change in Physical Activity. *Meas Phys Educ Exerc Sci.* 2014;18(4):273-285.

149. Chase JA. Interventions to Increase Physical Activity Among Older Adults: A Meta-Analysis. *The Gerontologist.* 2015;55(4):706-718.

150. Prince SA, Saunders TJ, Gresty K, Reid RD. A comparison of the effectiveness of physical activity and sedentary behaviour interventions in reducing sedentary time in adults: a systematic review and meta-analysis of controlled trials. *Obesity reviews : an official journal of the International Association for the Study of Obesity.* 2014;15(11):905-919.