

Responsiveness of Physical Activity Measures Following Exercise Programs after Total Knee Arthroplasty

Gustavo J. Almeida^{1*}, Lauren Terhorst², James J. Irrgang^{1,3}, G. Kelley Fitzgerald¹, John M. Jakicic⁴ and Sara R. Piva¹

¹ Department of Physical Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh. Address: 100 Technology Dr., Suite 210. Pittsburgh, PA 15219. USA

² Department of Occupational Therapy, School of Health and Rehabilitation Sciences, University of Pittsburgh. Address: 5017 Forbes Tower, Pittsburgh, PA 15260. USA

³ Department of Orthopedic Surgery, School of Medicine, University of Pittsburgh. Address: 3471 Fifth Ave., Suite 1010, Pittsburgh, PA 15213. USA

⁴ Department of Health and Physical Activity, School of Education, University of Pittsburgh. Address: 128 Oak Hill Dr., Pittsburgh, PA 15213. USA

Received: November 09, 2017; Accepted: November 29, 2017; Published: December 06, 2017

*Corresponding author: Gustavo J. Almeida, PT, PhD, Department of Physical Therapy, University of Pittsburgh, 100 Technology Dr, Suite 210, Pittsburgh, PA 15219; Tel : (412) 648-2215; Fax: (412) 648-5970; E-mail: gja4@pitt.edu

Abstract

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. Evidence for responsiveness of these instruments after an intervention is limited.

Objectives: o estimate and compare the responsiveness of two activity monitors and one questionnaire in assessing PA after an intervention following total knee Arthroplasty.

Methods: This one-group pretest-posttest, repeated-measures study analyzed changes in duration of daily PA and the standardized response mean (SRM) to assess internal responsiveness that were compared across instruments. Correlations between changes in PA measured by the proposed instruments and the global rating of change were used to test external responsiveness. Agreement between PA instruments on identifying individuals who changed their PA based on measurement error was assessed using weighted-Kappa (K).

Results: Thirty subjects, mean age 67(6) and 73% female, were analyzed. Changes in PA measured by each instrument were small ($p > 0.05$), resulting in a small degree of responsiveness ($SRM < 0.30$). Global rating of change scores did not correlate with changes in PA ($\rho = 0.13-0.28$, $p > 0.05$). The activity monitors agreed on identifying changes in moderate-intensity PA ($K = 0.60$) and number of steps ($K = 0.63$), but did not agree with scores from questionnaire ($K \leq 0.22$).

Conclusion: Analyzing group-based changes in PA is challenging due to high-variability in the outcome. Investigating changes in PA at the individual-level may be a more viable alternative.

Keywords: Physical activity; Psychometrics; Knee; Arthroplasty; Osteoarthritis;

counterparts,[2, 3]and therefore, more susceptible to comorbidities.[4] Interventions to increase PA and prevent comorbidities in individuals after TKA have been developed lately. [5, 6]To test the effectiveness of these interventions, researchers need measurement tools responsive to changes in PA behavior. However, information on the responsiveness of PA measurement tools is limited.

Responsiveness evaluates the ability of a measurement tool to accurately detect changes in the concept being measured when change has occurred, which can be determined by internal and external responsiveness methods.[7-9]There are numerous PA measurement tools available, but evidence of their responsiveness to change is limited. Two activity monitors (Actigraph [ACT; Actigraph LLC, Pensacola, FL] and Sensewear Armband [SWA; Body media, Pittsburgh, PA]) and one self-reported questionnaire (Community Health Activities Model Program for Seniors [CHAMPS]) have been validated in older adults, including those with knee osteoarthritis.[10-12]and are commonly used in research in this population.[13-15]These three instruments have the advantage to measure light-intensity PA (i.e., household chores and slow-walking) that are mostly performed by older adults who undergo TKA.[16-21]While activity monitors are costly and need several days of data collection, questionnaires have low cost and it takes 15-20 minutes to complete. In contrast, the monitors measure PA in real time, which eliminates problems with recall bias commonly seeing in self-report PA measures.[21, 22]

To our knowledge, no studies 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 such as those undergoing TKA.[12, 23-25]Additionally, studies have not concurrently compared responsiveness across these three measurement tools, which will provide information

Background

The benefits of regular physical activity (PA) to improve general health are well known.[1] Individuals who undergo total knee Arthroplasty (TKA) for end-stage knee osteoarthritis are typically older adults who are less active than their healthier

for evidence-based instrument selection to assess changes in PA behavior over-time in individuals with osteoarthritis of the lower extremities. Therefore, this study aims to estimate and compare the responsiveness of the ACT, SWA, and CHAMPS in assessing light-, moderate- and vigorous-intensity PA following a rehabilitation program after TKA.

Methods

Design and Subjects

This ancillary study used a one-group pretest-posttest, repeated-measures design. Baseline and 6-month data on PA from a randomized trial that investigated the effect of rehabilitation approaches to improve physical function in individuals following TKA were analyzed. This study took place in the Department of Physical Therapy, University of Pittsburgh from October'11 to August'13. All subjects signed a consent form approved by the University's Institutional Review Board prior to participation.

Inclusion and exclusion criteria followed that of the parent study: individuals at least 50 years old who had unilateral TKA done 3-6 months prior to study participation were included; and those who had any medical conditions that precluded safe exercise participation were excluded.[6] Specific to this ancillary study, individuals were included only if they were willing to wear the activity monitors for 7 days during baseline and follow-up assessments.

Study Protocol

Subjects attended two testing visits: one prior to the rehabilitation program (baseline), and one at the end of the program (6-month follow-up). At baseline, subjects completed questionnaires of demographics, pain (11-point numeric pain scale) and physical function (17-item Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function subscale). Subjects' height and weight were also measured. At the end of the testing-session, subjects were fitted with the ACT and SWA and instructed to wear the monitors for 7 days, except during water activities because they are not waterproof. After the 7-day period, subjects returned to our facility to download data from the monitors and to complete the CHAMPS. The CHAMPS questionnaire queries about PA participation in the past week, which corresponds to the time monitors collected data. Only subjects who had useful PA data from the monitors (i.e., ≥ 5 days with 10-hours of PA data/day)[26, 27] were analyzed.

After completing PA measures, subjects were randomized into two exercise groups.[6] Both exercise programs included endurance and strengthening for the lower extremities. The experimental-group performed more intense training than the control and it was exposed to balance and functional exercises along with PA promotion. Exercise programs consisted of 12-supervised sessions delivered within 3 months followed by a home-exercise program for another 3 months. For the purpose of this ancillary study, data from both groups were combined to provide a wide range of change in PA to test responsiveness.

PA measures were repeated at follow-up. Additionally, subjects rated their perceived changes in PA from baseline to follow-up, using a modified global rating of change scale that consists of 15 points ranging from -7 ("a very great deal less active") to 0 ("about the same") to +7 ("a very great deal more active"). Subjects were classified as 'more active' if ratings were $\geq +3$ ("somewhat more active") or higher. Subjects were classified as 'not changed' if ratings were between +2 ("a little bit more active") and -2 ("a little bit less active"). Subjects were classified as 'less active' if ratings were ≤ -3 ("somewhat less active") or lower.

Measures of Physical Activity

The ACT is a small triaxial accelerometer-based monitor worn around the waist at the hip-bone level, over the right anterior superior iliac spine. The ACT GT3X and the actilife 5 software (Actigraph LLC, Pensacola, FL) were used. This device generates data on activity counts per minute (CPM) and number of steps. Duration of daily activities in minutes per day (min/day) were categorized by the software using the following CPM cut-points: 760-1951 CPM for light, 1952-5724 CPM for moderate, and >5724 CPM for vigorous intensities.[28] The software calculates non-wear periods following manufacturer's recommendation, which were also visually analyzed. The ACT has demonstrated moderate accuracy in comparison to doubly labeled water (reference standard) to measure PA in older adults and good reliability in individuals after TKA.[10, 29]

The SWA is a small multi-sensor device worn on the right arm over the triceps muscle at midpoint between shoulder and elbow. The SWA Pro-3 and the Professional software v6.1 (Body media, Pittsburgh, PA) were used. Information from biaxial accelerometer, heat flux, galvanic signal (i.e., sweat rate) and skin temperature is integrated and processed by the software using proprietary algorithms that account for subjects' characteristics (gender, age, height and weight). The SWA was set to provide data on duration of PA (min/day) in light- (2-2.9 metabolic equivalents [mets]), moderate- (3-5.9 mets) and vigorous-intensity (≥ 6 mets), as well as number of steps. The SWA turns off when not in use, enabling the software to recognize non-wear periods. Data were also visually screened for non-wear periods. The SWA showed good accuracy compared to doubly labeled water to measure PA in older adults and good reliability in individuals after TKA.[10, 29]

CHAMPS is a self-reported questionnaire that queries type, frequency and duration of 41 activities usually performed by older adults. Duration in hours per week (hr/week) of each activity is selected from a range of <1 hr/week to ≥ 9 hr/week and categorized in two intensity levels according to the CHAMPS activities codebook.[12] Categories are light-to-vigorous PA (≥ 2 mets) and moderate-to-vigorous PA (≥ 3 mets). A light-intensity category was computed (i.e., light-to-vigorous PA minus moderate-to-vigorous PA) to allow direct comparison with the activity monitors. CHAMPS scores in hr/week were converted into min/day. The questionnaire showed small significant association with doubly labeled water to measure PA in older adults, and moderate reliability in people with musculoskeletal disorders and healthy older adults.[10, 29, 30]

The main outcome measure of this study was PA duration in min/day estimated by the ACT, SWA and CHAMPS during waking hours (i.e., from out of bed in the morning to back to bed at night). 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). 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). Number of steps was compared between the ACT and SWA only.

Data Analysis

Descriptive statistics for continuous variables included mean(SD) or median (25-75 percentiles), according to data distribution. Counts and frequencies were used for categorical variables. Histograms depicting the magnitude of changes in PA from pre-post intervention were to visually assess changes in PA. Internal responsiveness (group-based) was estimated by calculating the magnitude of changes in PA and its 95% confidence intervals, as well as the standardized response mean (SRM), which is as an index of responsiveness. The SRM is a ratio of mean change to the SD of change scores.[31]The SRM is interpreted as trivial (<0.20), small (0.20-0.49), moderate (0.50-0.79) and high (≥0.80)degree of responsiveness.[31]Confidence intervals around the SRM were calculated using the bias-corrected and accelerated bootstrap.[32]

To estimate external responsiveness (individual-based), we first explored if the modified global rating of change was suitable as an external anchor to capture perceived changes in PA, i.e., correlations between its scores and changes measured by the PA instruments should be at least moderate(≥0.30). Pearson or Spearman rho correlation coefficients are used according to data distribution. If global rating of change and changes in PA measured by the three instruments were associated, then cut-offs of clinical important change were derived.[9]

To compare responsiveness across the PA instruments one-way repeated measures Analysis of Variance (ANOVA) was performed for each PA intensity category to determine if the magnitude of change obtained from the three instruments were statistically different from each other. Changes in PA at each intensity category measured by each instrument was used as the repeated factor. If anovas indicated significant differences, post-hoc comparisons between PA measures were performed applying Bonferroni adjustments with alpha level set at 0.016 to account for multiple comparisons. Bland and Altman plots were used to visually compare change scores across instruments at each intensity category.[33]Internal responsiveness was also compared across instruments by examining the 95% confidence intervals around the SRMs. Non-overlapping confidence intervals indicate that responsiveness between instruments was significantly different.[34]

Additionally, weighted-Kappa (linear weighing scheme) was used to investigate the agreement between instruments on identifying subjects who were less, more, or similarly active after the intervention, based on the standard error of the

measurements previously published.[29]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).[31]Analyses were performed using the IBMSPSS Statistics 21 (IBM Corporation, Armonk, NY)and Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA).

Results

Forty-two subjects completed the randomized trial, of which 30 had complete data in the three PA instruments and were included in the responsiveness analyses. Amongst the 12 subjects not entered in the analyses, 3 refused wearing the monitors at follow-up, and the remaining 9 subjects had no data in one of the devices due to equipment failure. The demographic and biomedical characteristics between subjects included and those excluded from the responsiveness analyses were similar (Table-1). Subjects included in the analyses were on average 67(6) years old, predominantly females (73.3%) and obese (body mass index= 30(4) kg/m²).Monitors wear time was similar at baseline and follow-up: 15(2) hr/day.

Table 1: Baseline characteristics of study sample. Data represents mean (SD) or Frequency (%), unless otherwise indicated.

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 25 th ; Q75: quartile 75 th ; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index-Physical Function sub-scale;		

(Figure-1)depicts the distribution of changes in PA from baseline to follow-up. The graphs indicate that the amount of subjects who became less versus more active was similar. For example, using zero as a threshold for no change at all, measures from the ACT in light-to-moderate PA showed that 18 subjects

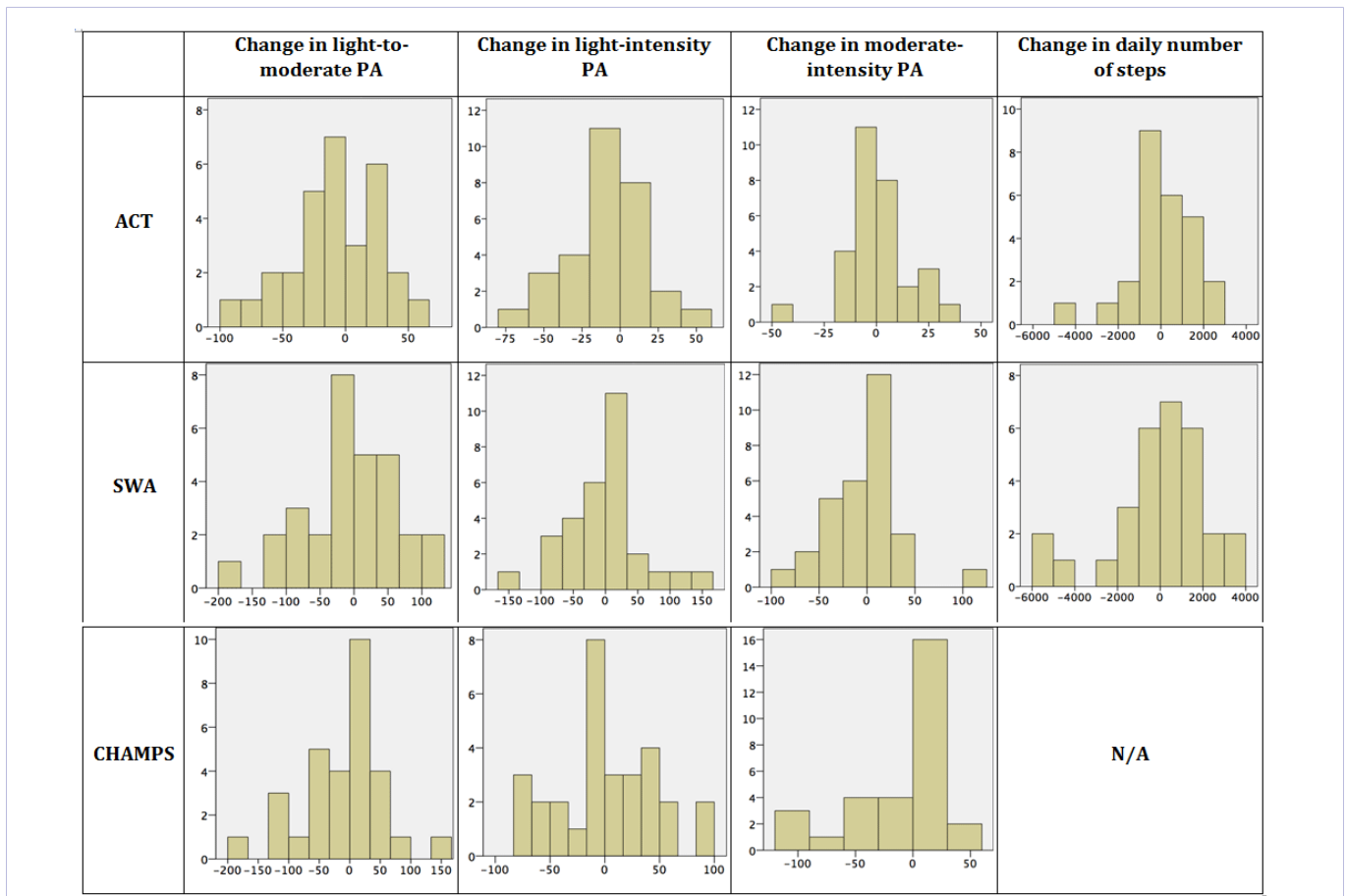


Figure 1: Changes in physical activity (PA) duration measured by the Actigraph (ACT), Sensewear Armband (SWA) and Community Healthy Activities Model Program for Seniors questionnaire (CHAMPS) pre to post intervention across the three PA intensity categories, and number of steps (ACT and SWA only). Numbers on the Y-axis represent frequency and numbers on the X-axis represent PA duration in min/day or daily number of steps. N/A: not applicable since CHAMPS does not measure number of steps.

became less active and 12 more active. As per the SWA, 14 subjects became less active and 16 more active, whereas CHAMPS scores indicated that 16 subjects became less active and 14 more active. By visually analyzing the number of subjects who changed beyond a magnitude that most clinicians would agree to be an important change (i.e., ~20 min/day in light-to-moderate PA), measures from the ACT indicated that 10 subjects became more active, 9 less active and 11 did not change; as per the SWA, 11 subjects became more active, 13 less active and 6 did not change; and based on CHAMPS, 7 subjects became more active, 11 less active and 12 did not change.

In terms of external responsiveness, while PA measured by the ACT, SWA and CHAMPS revealed similar number of subjects who became less or more active at follow-up (Figure-1), none of the subjects reported being less active based on the global rating of change in PA scale. According to the scale, 7 subjects (23.3%) had no change in activity level at follow-up and the remaining of the subjects (76.7%) were more active. Consequently, the associations between self-rated and measured changes in PA were below the threshold of $\rho \geq 0.30$ ($\rho = 0.13$ to 0.28 , $p > 0.05$), which precluded calculation of cut-offs of clinical important change.

The comparison of internal responsiveness between PA instruments indicated that the magnitude of changes in power were not different across PA categories ($p \geq 0.12$, data not shown). Moreover, the 95% confidence interval of the SRMs largely overlapped across instruments (Table-2). The Bland-Altman plots also indicated that internal responsiveness was similar across PA instruments (Figure 2): the line of equality (zero) was contained within the lines of 95% confidence interval of the difference between changes in PA scores measured by each instrument across intensity categories.

Using measurement error as a threshold for changes in PA weighted-Kappa (K) indicated moderate agreement between ACT-SWA on identifying subjects who changed their PA beyond the standard error of the measurement in moderate-intensity PA ($K = 0.60$), 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 (Table-3). Agreement was fair between ACT-SWA in measures of light-intensity PA ($K = 0.25$) as well as between ACT-CHAMPS in measures of light-to-moderate PA ($K = 0.22$), which were not statistically significant. There was poor agreement between CHAMPS and the activity monitors in

Table 2: 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 (standard deviation) from an n of 30 unless otherwise indicated.

Instrument	PA categories	Baseline	Follow-up	Change in PA	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	-0.17 (-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	-0.26 (-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	0.04 (-0.33; 0.40)
	Number of Steps (steps/day) *	4676 (2151)	4667 (2109)	-9 (1526) 95% CI: -607; 625	-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	-0.07 (-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	-0.04 (-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	-0.08 (-0.43; 0.29)
	Number of Steps (steps/day)	6003 (3311)	5960 (2995)	-42.8 (2266) 95% CI: -889; 804	-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	-0.18 (-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	-0.03 (-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	-0.20 (-0.52; 0.17)

ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire; PA: physical activity; * n=26 due to monitors' malfunction; 95% CI: 95% confidence interval; SRM: standardized response mean.

Table 3. 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 Categories		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: Sense wear 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; ‡ n=26 due to Actigraph's malfunction; <SEM: less active based on SEM; >SEM: more active based on SEM; No Δ: No changes in PA based on SEM from previous study on reliability [29]

SEM for the ACT: LMPA (16.9 min/day), LPA (13.0 min/day), MPA (5.3 min/day) and number of steps (709 steps/day).
SEM for the SWA: LMPA (29.8 min/day), LPA (22.5 min/day), MPA (10.8 min/day) and number of steps (937 steps/day).
SEM for the CHAMPS: LMPA (16.8 min/day), LPA (12.9 min/day) and MPA (12.2 min/day).

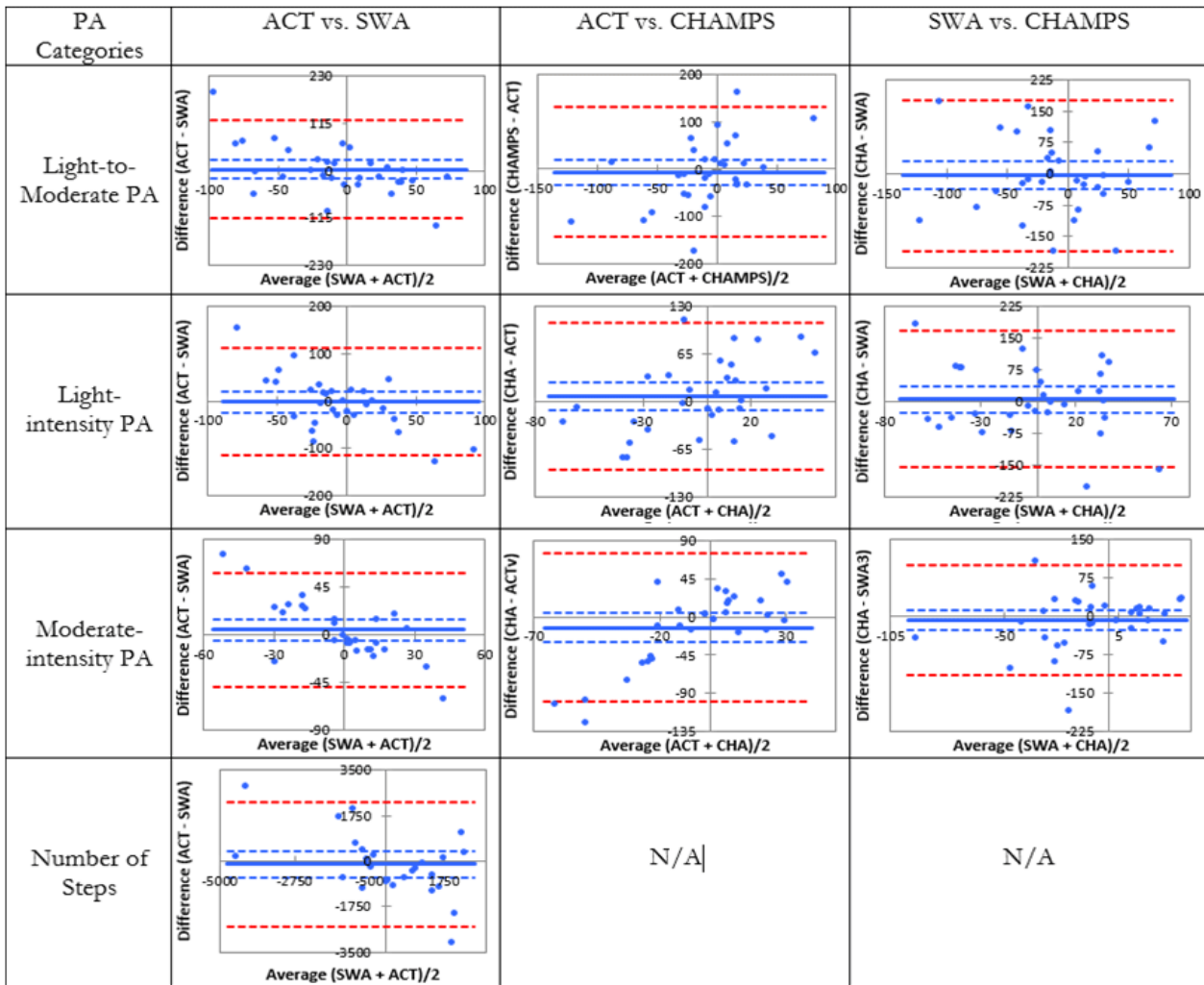


Figure 2: Bland and Altman plots of differences between change scores across physical activity categories. PA: physical activity; ACT: Actigraph; SWA: Sensewear Armband; CHAMPS: Community Healthy Activities Model Program for Seniors questionnaire; N/A: not applicable since CHAMPS does not measure number of steps.

measures of light-to-moderate PA ($K \leq 0.07$) and light-intensity PA ($K \leq 0.12$). Agreement between CHAMPS-SWA was also poor in moderate-intensity PA ($K = 0.12$).

Discussion

This is the first study to estimate and compare the responsiveness of the ACT, SWA, and CHAMPS in assessing changes in light to moderate PA after rehabilitation following TKA. We observed that the high variability inherent in measures of PA is problematic and limits the utilization of any responsiveness index (i.e., SRM and effect-sizes) based on group changes. The inability to detect changes over-time at the group-level in this study seemed to be due to large variability at the subject-level rather than pre-post intervention. Another important finding was that subjects' perception of changes in activity participation appears to be limited due to poor associations between self-rated and measured changes in PA. Study results also demonstrate that

using a threshold to identify changes in paat the individual-level may be useful.

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 instruments.[23-25, 35] Responsiveness indices from studies on measures from ACT ranged from 0.18 to 0.36 in subjects with type-2 diabetes,[23] nonworking older adults,[24] and sedentary healthy adults.[25] One study on the CHAMPS questionnaire in healthy older adults found responsiveness indices of 0.33 and 0.37 in moderate-intensity PA and light-to-moderate PA respectively.[35] The small responsiveness of PA measures across the studies not only highlight the difficulty to change PA behavior, but also support the notion that assessment of changes in PA over-time at the group-level may not be adequate.

In this study, we attempted to assess external responsiveness using a modified global rating of change scale as an external

anchor. However, associations were poor between the scale and measured changes by the PA instruments. The lack of association between self-rated and measured changes in PA may be partially due to social desirability bias.[36] When completing the global rating of change scale, subjects might have answered the question in a positive waysince none of them reported being less active. Additionally, the poor associations may also be a result of the difficulty that subjects have in adjudicating changes in the amount of activities after a long period of time (baseline to follow-up). This perception of changes in PA may be particularly difficult for older adults with knee osteoarthritis because they usually engage in light-intensity PA, which is very difficult to recall.[37] We are not aware of any studies that attempted to use external responsiveness methods to investigate sensitivity to change of the ACT, SWA or CHAMPS.

A practical step to identify changes in PA at the individual-level is the use of the standard error of the measurement as a threshold, which might be appropriate in situations where patient-perceived changes are unavailable.[38]While measurement error has been used as a threshold in studies investigating changes in physical function,[39, 40] this is the first study to discuss the usability of this method to identify changes in PA. This approach allowed to test the agreement between PA instruments in classifying individuals who changed or not their PA. While the ACT and SWA generally agreed on classifying those who changed, they disagreed with change scores from CHAMPS. This discrepancy may have been due to limitations inherent of PA questionnaires such as recall-bias.[21, 22]

This study is not without weakness. Although the study had a small sample size, it is unlikely that the non-significant results for internal responsiveness were due to type-II error because the changes in PA were all very small. These small changes were likely the result of difficulties in changing subjects' lifestyles along with high variability in change scores among subjects.

Conclusion

In conclusion, our study showed that the high variability in change scores resulted in small degree of responsiveness of PA measured by the ACT, SWA and CHAMPS, which dispute the assessment of change in PA at the group-level. The ACT and SWA seemed to agree on detecting changes in PA using measurement error as a threshold. Therefore, when investigating changes in PA behavior in individuals with arthritis of the lower extremities, clinicians and researchers should consider interpreting their results based on changes at the individual-level rather than at the group-level.

Acknowledgement

This work was supported by the Pepper Center Scholars Pilot Program (P30-AG024827), University of Pittsburgh; The Rehabilitation Institute Pilot Program, University of Pittsburgh Medical Center; and SHRS Research Development Fund, University of Pittsburgh.

Declarations

The authors of this manuscript have nothing to disclose. The University of Pittsburgh Institutional Review Board approved this study and the consent form that was signed by all subjects.

References

1. Paterson DH and DE Warburton. 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.
2. Harding P, Holland AE, Delany C, Hinman RS. Do activity levels increase after total hip and knee Arthroplasty? *Clin Orthop Relat Res.* 2014;472(5):1502-1511. Doi: 10.1007/s11999-013-3427-3
3. 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. Doi: 10.1002/acr.20384
4. Naal FD and FM Impellizzeri . How active are patients undergoing total joint Arthroplasty?: A systematic review. *Clin Orthop Relat Res.* 2010;468(7):1891-1904. Doi: 10.1007/s11999-009-1135-9
5. Pozzi F, L Snyder-Mackler, and J Zeni. Physical exercise after knee Arthroplasty: a systematic review of controlled trials. *Eur J Phys Rehabil Med.* 2013;49(6):877-892.
6. Piva SR, Almeida GJ, Gil AB, DiGioia AM, Hessel DL, Sowa GA. A comprehensive behavioral and exercise intervention improves physical function and activity participation after total knee replacement - a pilot randomized study. *Arthritis Care Res (Hoboken).* 2017;69(12):1855-1862. Doi: 10.1002/acr.23227
7. Guyatt G, S Walter and G Norman. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987;40(2):171-178.
8. Beaton DE, Understanding the relevance of measured change through studies of responsiveness. *Spine (Phila Pa 1976).* 2000;25(24):3192-3199.
9. 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.
10. Colbert LH, et al., Comparative validity of physical activity measures in older adults. *Med Sci Sports Exerc.* 2011; 43(5):867-876.
11. Almeida GJ, Wert DM, Brower KS, Piva SR. et al., Validity of physical activity measures in individuals after total knee Arthroplasty. *Arch Phys Med Rehabil.* 2015;96(3):524-531.
12. 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.
13. Cauley JA, Harrison SL, Cawthon PM, Ensrud KE, Danielson ME, Orwoll E, et al., Objective measures of physical activity, fractures and falls: the osteoporotic fractures in men study. *J Am Geriatr Soc.* 2013;61(7):1080-1088. Doi: 10.1111/jgs.12326
14. Liu SH, Driban JB, Eaton CB, McAlindon TE, Harrold LR, Lapane KL. Objectively Measured Physical Activity and Symptoms Change in Knee Osteoarthritis. *Am J Med.* 2016;129(5):497-505. Doi: 10.1016/j.amjmed.2015.12.029

15. Guirao-Goris, JA, Cabrero-García J, Moreno Pina JP, Muñoz-Mendoza CL. Structured review of physical activity measurement with questionnaires and scales in older adults and the elderly. *Gac Sanit.* 2009;23(4):334 e1-334 e17. Doi: 10.1016/j.gaceta.2009.03.002
16. Harada ND, 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.
17. 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. Doi: 10.1186/s12966-014-0119-7
18. Van Remoortel H, Raste Y, Louvaris Z, Giavedoni S, Burtin C, Langer D, et al., Validity of six activity monitors in chronic obstructive pulmonary disease: a comparison with indirect calorimetry. *PLoS One.* 2012;7(6):e39198. Doi: 10.1371/journal.pone.0039198
19. Lee JM, Y Kim, and GJ Welk, Validity of consumer-based physical activity monitors. *Med Sci Sports Exerc.* 2014; 46(9):1840-1848. Doi: 10.1249/MSS.0000000000000287
20. Wetten AA, Batterham M, Tan SY, Tapsell L, et al., Relative validity of 3 accelerometer models for estimating energy expenditure during light activity. *J Phys Act Health.* 2014;11(3):638-647. Doi: 10.1123/jpah.2011-0167
21. Hecker EB, Buman MP, Haskell WL, Conway TL, Cain KL, Sallis JF, 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.
22. Neilson HK, Paula J Robson, Christine M Friedenreich, and Ilona Csizmad. Estimating activity energy expenditure: how valid are physical activity questionnaires? *Am J Clin Nutr.* 2008;87(2):279-291.
23. 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.* 2015;12(8):1082-1087. Doi: 10.1123/jpah.2013-0265.
24. Gardiner PA, et al., Measuring older adults' sedentary time: reliability, validity, and responsiveness. *Med Sci Sports Exerc.* 2011;43(11):2127-2133. Doi: 10.1249/MSS.0b013e31821b94f7
25. Swartz AM, Rote AE, Cho YI, Welch WA, Strath SJ. Responsiveness of motion sensors to detect change in sedentary and physical activity behavior. *Br J Sports Med.* 2014; 48(13):1043-1047. Doi: 10.1136/bjsports-2014-093520
26. Trost SG, KL McIver, and RR Pate. Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc.* 2005;37(11 Suppl):S531-s543.
27. Almeida GJ, Wasko MC, Jeong K, Moore CG, Piva SR. Physical activity measured by the Sense Wear Armband in women with rheumatoid arthritis. *Phys Ther.* 2011;91(9):1367-1376. Doi: 10.2522/ptj.20100291
28. Freedson PS, E Melanson, and J Sirard. Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc.* 1998;30(5):777-781.
29. Almeida GJ, Irrgang JJ, Fitzgerald GK, Jakicic JM, Piva SR. Reliability of Physical Activity Measures During Free-Living Activities in People After Total Knee Arthroplasty. *Phys Ther.* 2016;96(6):898-907. Doi: 10.2522/ptj.20150407
30. 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. Doi: 10.3109/09638280903166352
31. Portney LG. and MP Watkins. *Foundations of clinical research: applications to practice.* 1993. Stamford: Appleton & Lange.
32. Efron B. Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association.* 1987;82(397):171-185.
33. Bland JM. and DG. Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;1(8476):307-310.
34. Cumming, G. and S. Finch. Inference by eye: confidence intervals and how to read pictures of data. *Am Psychol.* 2005;60(2):170-180.
35. Stewart AL, Verboncoeur CJ, McLellan BY, Gillis DE, Rush S, Mills KM, 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): 465-470.
36. Adams SA, Charles E. Matthews, Cara B. Ebbeling, Charity G. Moore, Joan E. Cunningham, Jeanette Fulton, et al., The effect of social desirability and social approval on self-reports of physical activity. *Am J Epidemiol.* 2005;161(4):389-398.
37. Washburn RA. Assessment of physical activity in older adults. *Res Q Exerc Sport.* 2000;71(2 Suppl):S79-88.
38. 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):102-119. Doi: 10.1016/j.jclinepi.2007.03.012
39. Haley SM. and MA Fragala-Pinkham. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther.* 2006;86(5):735-743. Doi: 10.1093/ptj/86.5.735
40. Copay AG, Subach BR, Glassman SD, Polly DW Jr, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J.* 2007;7(5):541-546.