



Published in final edited form as:

*Arch Phys Med Rehabil.* 2018 September ; 99(9): 1913–1916. doi:10.1016/j.apmr.2017.12.025.

## Comparison of self-report vs. sensor-based methods for measuring the amount of upper limb activity outside the clinic

Kimberly J. Waddell, MS<sup>1</sup> and Catherine E. Lang, PhD<sup>2</sup>

<sup>1</sup>Program in Physical Therapy, Washington University, St. Louis, MO

<sup>2</sup>Program in Physical Therapy, Department of Neurology, Program in Occupational Therapy, Washington University, St. Louis, MO

### Abstract

**Objective**—To compare self-reported with sensor-measured upper limb (UL) performance in daily life for individuals with chronic (> 6 months), UL paresis post-stroke.

**Design**—Secondary analysis of 64 participants enrolled in a Phase II randomized, parallel, dose-response UL movement trial. This analysis compared the accuracy and consistency between self-reported UL performance and sensor-measured UL performance at baseline and immediately post an 8-week intensive UL task-specific intervention.

**Setting**—Outpatient rehabilitation

**Participants**—Community dwelling individuals with chronic (> 6 months) UL paresis post-stroke.

**Intervention**—Not applicable

**Main outcome measures**—Motor Activity Log-Amount of Use Scale and the sensor-derived use ratio from wrist-worn accelerometers

**Results**—There was a high degree of variability between self-reported UL performance and the sensor derived use ratio. Using sensor-based values as a reference, three distinct categories were identified: accurate reporters (reporting difference  $\pm 0.1$ ), over reporters (difference  $> 0.1$ ), and under reporters (difference  $< -0.1$ ). Five of 64 participants accurately self-reported UL performance at baseline and post-intervention. Over half of participants (52%) switched categories from pre- to post-intervention (e.g. moved from under reporting pre-intervention to over reporting post-intervention). For the consistent reporters, no participant characteristics were found to

---

Corresponding author: Catherine E. Lang, PT, PhD, Washington University School of Medicine, Campus Box 8502, 4444 Forest Park Blvd., St. Louis, MO 63108, langc@wustl.edu, (314) 286-1945.

**Conflicts of interest:** The authors have no conflicts of interest to disclose.

**Acknowledgement of previous presentation:** Preliminary results of this analysis were presented at the Translational Science National Meeting in Washington D.C. in April 2017

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

influence whether someone over- or under-reported performance compared to sensor-based assessment.

**Conclusions**—Participants did not consistently or accurately self-report UL performance when compared to the sensor-derived use ratio. While self-report and sensor-based assessments are moderately associated and appear similar conceptually, these results suggest self-reported UL performance is often not consistent with sensor-measured performance and the measures cannot be used interchangeably.

## Keywords

Accelerometry; upper limb paresis; stroke; self-report; adult

---

Individuals are referred for stroke rehabilitation services to improve performance in daily life. Performance, defined as what a person actually does in his/her current environment, outside of a clinic or laboratory,<sup>1</sup> is difficult to measure for the upper limb (UL). Performance is most commonly quantified by amount, with other aspects (e.g. quality, efficiency) being more difficult to measure *during* everyday life. Researchers must choose between self-report measures of UL performance, which provide critical information about patient perception of abilities but are subject to inherent biases such as social desirability and recall bias,<sup>2,3</sup> or sensor-based methods, such as accelerometry. Accelerometry is a valid, reliable, quantitative measure of UL performance in daily life<sup>4,5</sup> and is not subject to the same biases as self-report measures, but cannot determine the specific activities someone performs and captures functional as well as non-functional movements.

A recent physical activity review indicates self-report and sensor quantifications of physical activity vary widely and unsystematically, with correlations ranging from  $-0.7$  to  $0.7$  across studies.<sup>6</sup> The purpose of this brief report was to compare self-report and sensor-based measures of UL performance in daily life in a clinical trial cohort of persons with chronic stroke. While these measures are moderately correlated,<sup>5</sup> it is critical to know the accuracy and consistency between them.

## Methods

This was a secondary analysis from a Phase II, randomized, dose-response trial of intensive, task-specific UL motor training (see<sup>7</sup> for comprehensive assessment battery).<sup>7</sup> Data from baseline and post-intervention assessment time points were used here. The trial was approved by the Washington University Human Research Protection Office and all participants provided informed consent.

## Assessments

Our sensor-based measure of UL performance was derived from bilateral, wrist-worn accelerometers (wGT3X+, Actigraph, Pensacola, FL). Participants wore the accelerometers for 24 hours at baseline and again post-intervention during all daily activities.<sup>8</sup> The variable of interest for this comparison was the use ratio, an established metric of UL performance in daily life.<sup>4,9</sup> The use ratio quantifies the amount of time the paretic UL is active relative to the non-paretic UL and ranges from 0–1. A use ratio value of 1 indicates both UL were

active the same amount of time throughout the recording period. Healthy, non-disabled adults have a use ratio value of 0.95 (SD = 0.06).<sup>9</sup>

The self-report measure of UL performance was the Motor Activity Log (MAL) amount of use (AOU) scale, the only measure that directly queries amount of UL use and has strong psychometric properties.<sup>10</sup> Participants reported how much they used the paretic UL across 28 representative functional activities, with scores from 0 = did not use the paretic UL to 5 = used the paretic UL as often as before the stroke. Because the use ratio is near unity and highly consistent in neurologically-intact adults, then a score of 3 on the AOU scale (used paretic UL half as much as before the stroke) is comparable to a sensor-derived use ratio of 0.5, where the paretic UL is active half as much relative to the non-paretic limb, and a use ratio of 1 is analogous to a 5 on the MAL.<sup>9</sup>

### Statistical Analysis

Correlation analyses examined the association between the use ratio and MAL scores at both time points. Each MAL AOU value was scaled from 0–1 to match the range of the use ratio. The use ratio was subtracted from the scaled AOU values at each time point to create a difference score. Participants whose difference score was  $\pm 0.1$  ( $\pm 10\%$ ) were classified as “accurate” reporters. Participants whose difference score was  $> +0.1$  were classified as “over reporters” and those whose difference score was  $< -0.1$  were “under reporters.” This classification scheme was determined separately for the two time points.

To examine the accuracy and consistency between the two measures, frequencies of the categories were first calculated. Second, consistency of classification was examined by computing frequencies of individuals who were in the same category at both time points (e.g. accurate at baseline and post-intervention) vs. those who switched categories (e.g. accurate at baseline and under-reported post-intervention). Third, of the participants who were consistent, we explored characteristics that may explain their consistency. Because of small, imbalanced group numbers, 95% confidence intervals of the characteristics of age, memory (i.e. Short Blessed Test), depressive symptomatology (i.e. Centers for Epidemiologic Studies Depression Scale [CES-D]), baseline UL motor capacity (i.e. Action Research Arm Test [ARAT] score), and concordance (dominant limb = paretic limb) were examined across the categories.

### Results

Sixty-four of 85 trial participants had available data for this analysis. The 21 excluded participants had either accelerometer recording errors, missing self-report data, or withdrew from the study. Participant demographics, and MAL-AOU and use ratio values (Table 1) were not different from the trial cohort.<sup>7,8</sup> Consistent with previous reports<sup>5</sup>, the MAL and use ratio were moderately associated at baseline ( $r=0.31$ , 95% CI [0.07, 0.51],  $p = 0.014$ ) and post-intervention ( $r = 0.52$ , 95% CI [0.32, 0.68],  $p < 0.001$ ).

Our main finding is the high degree of variability between self-reported and sensor derived UL performance, as indicated by the difference scores (Figure 1A baseline; Figure 1B post-intervention). Thirty-one of 64 participants (48%) were consistent in their category, with the

remaining 33 participants (52%) being inconsistent (i.e. switching categories). Table 2 explores the 31 consistent reporters, divided into consistently accurate, under-reporting, and over-reporting. Only five participants were classified as consistently accurate. In this small sample, we could see no differences in age, cognitive function, depressive symptomatology, UL functional capacity, and concordance (dominant limb = paretic limb) across categories to explain reporting differences.

## Discussion

Despite moderate correlations, we found a high degree of variability and inconsistency between self-reported and sensor-measured UL performance measures post-stroke. Half of our sample consistently reported UL performance from baseline to post-intervention, but only five individuals accurately reported at both time points. Many UL clinical trials include an outcome measure of UL performance in daily life. The high degree of variability and inconsistencies between self-reported and sensor-measured performance indicate these measures cannot be used interchangeably. If a tested intervention is intended to improve self-perceptions of UL performance, then a self-report measure is the better choice. If a tested intervention is intended to improve actual arm use in daily life, then using a quantitative measure of UL performance is the better choice.

The inconsistency across time points seen here is concerning. As with the physical activity literature,<sup>6</sup> reporting inconsistency could compromise the results of research studies testing the efficacy of interventions to improve UL performance post-stroke. If individuals are not consistent and/or accurate in their reporting of UL performance, their change scores will contain a large degree of error and it will be difficult to draw conclusions about whether or not an UL intervention can drive change outside of the clinic or laboratory.

## Limitations

Key limitations of this brief report are the modest sample size and the inability to formally test for differences across categories.

## Conclusions

While self-report and sensor-based assessments are moderately associated and appear similar conceptually,<sup>5</sup> results indicate self-reported UL performance is often not consistent with sensor-measured UL performance. We recommend that clinicians and researchers measure outcomes via self-report if improved perception of UL performance is the primary outcome of interest and measure outcomes via sensors if improvements in actual UL performance is the primary outcome of interest.

## Acknowledgments

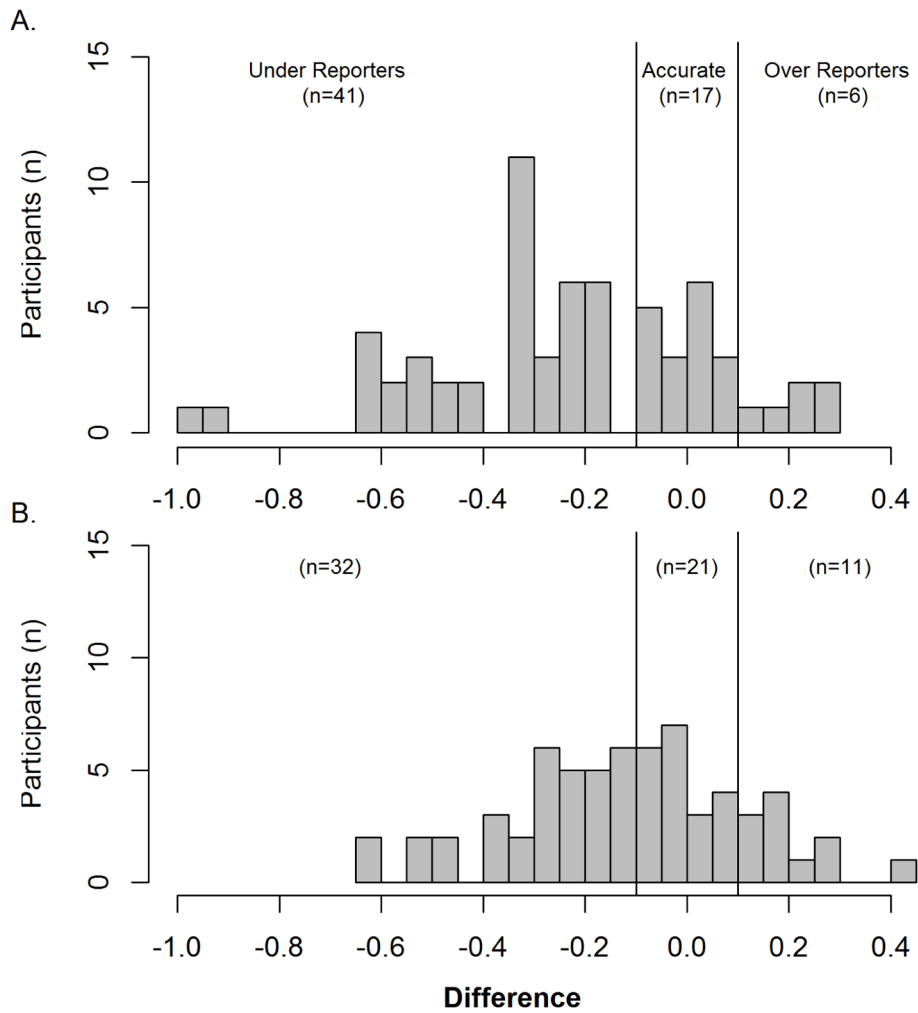
**Funding sources:** This work was supported by the National Institutes of Health R01 HD068290 (CEL) and TL1 TR000449 (KJW)

## Abbreviations

<b>UL</b>	upper limb
<b>MAL</b>	Motor Activity Log
<b>AOU</b>	Amount of Use

## References

1. International classification of functioning, disability and health (ICF). Geneva: World Health Organization; 2001.
2. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: the impact of memory and inference on surveys. *Science (New York, NY)*. 1987; 236(4798):157–161.
3. Adams SA, Matthews CE, Ebbeling CB, et al. The effect of social desirability and social approval on self-reports of physical activity. *American journal of epidemiology*. 2005; 161(4):389–398. [PubMed: 15692083]
4. Uswatte G, Foo WL, Olmstead H, Lopez K, Holand A, Simms LB. Ambulatory monitoring of arm movement using accelerometry: an objective measure of upper-extremity rehabilitation in persons with chronic stroke. *Arch Phys Med Rehabil*. 2005; 86(7):1498–1501. [PubMed: 16003690]
5. Uswatte G, Giuliani C, Winstein C, Zeringue A, Hobbs L, Wolf SL. Validity of accelerometry for monitoring real-world arm activity in patients with subacute stroke: evidence from the extremity constraint-induced therapy evaluation trial. *Arch Phys Med Rehabil*. 2006; 87(10):1340–1345. [PubMed: 17023243]
6. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *The international journal of behavioral nutrition and physical activity*. 2008; 5:56. [PubMed: 18990237]
7. Lang CE, Strube MJ, Bland MD, et al. Dose response of task-specific upper limb training in people at least 6 months poststroke: A phase II, single-blind, randomized, controlled trial. *Ann Neurol*. 2016; 80(3):342–354. [PubMed: 27447365]
8. Waddell KJ, Strube MJ, Bailey RR, et al. Does Task-Specific Training Improve Upper Limb Performance in Daily Life Poststroke? *Neurorehabilitation and neural repair*. 2017; 31(3):290–300. [PubMed: 27909071]
9. Bailey RR, Lang CE. Upper-limb activity in adults: referent values using accelerometry. *Journal of rehabilitation research and development*. 2013; 50(9):1213–1222. [PubMed: 24458962]
10. Uswatte G, Taub E, Morris D, Light K, Thompson PA. The Motor Activity Log-28: assessing daily use of the hemiparetic arm after stroke. *Neurology*. 2006; 67(7):1189–1194. [PubMed: 17030751]



**Figure 1.** Reporting differences (MAL-Use Ratio) across the sample for (A) baseline and (B) post-intervention. Accurate reporters were defined as a difference value of  $\pm 0.1$ ; under-reporters were  $< -0.1$  and over-reporters were  $> 0.1$ . There was a high degree of variability between self-reported and sensor-derived performance at both time points, with the majority of participants underreporting. We did not control for weekday vs. weekend wearing schedules for participants, as previous work has shown that both days are similar in terms of functional demands for this population.

**Table 1**

Participant demographics. Values are means  $\pm$  SD or median (min, max).

	Entire sample (n=64)
Age	61.2 $\pm$ 11.1
Gender	22 F, 42 M
Race	37 Caucasian 26 Afr American 1 Multi-race
Type of stroke	47 ischemic 6 hemorrhagic 11 unknown
Months post-stroke	11.5 (6, 180)
Affected side	35 R, 29 L
% Concordance <sup>a</sup>	52%
% Independent with ADL	81%
Baseline ARAT <sup>b</sup> score	32 $\pm$ 10.9
Baseline Use Ratio	0.66 $\pm$ 0.2
Baseline MAL AOU <sup>c</sup>	2.73 $\pm$ 0.9

<sup>a</sup> = Concordance = dominant side is paretic side; value here indicates the percentage of the sample who identified their dominant UL as the paretic UL

<sup>b</sup> = Action Research Arm Test; scores range from 0–57 points with higher scores indicating more normal movement. Here, participants had mild to moderate UL paresis at baseline.

<sup>c</sup> = Motor Activity Log, Amount of Use scale; scores range from 0–5 where 0= did not use the paretic UL to 5= used the paretic UL as often as before the stroke.

**Table 2**

Potential modifiers of reporting accuracy across *consistent* reporters. Values are means  $\pm$  SD (95% CI) or n (%)

Modifier	Accurate (n = 5)	Under reporters (n=24)	Over reporters (n=2)
Age	63.8 $\pm$ 7.3	60.7 $\pm$ 10.2	56 $\pm$ 11.3
Cognition (Short Blessed Test) <sup>a</sup>	3.2 $\pm$ 4.1 (0, 6.8)	3.4 $\pm$ 6.1 (0.9, 5.8)	5 $\pm$ 7.1 (0, 14.8)
Depressive Symptomatology (CES-D) <sup>b</sup>	11.4 $\pm$ 9 (3.5, 19.3)	16.5 $\pm$ 11.9 (11.7, 21.3)	11 $\pm$ 9.9 (0, 24.7)
UL functional capacity (ARAT) <sup>c</sup>	33.8 $\pm$ 16 (19.8, 47.8)	31.4 $\pm$ 10.5 (27.2, 35.6)	37 $\pm$ 1.4 (35, 38.7)
Concordance <sup>d</sup> (n (%))	2 (40%)	12 (50%)	2 (100%)

<sup>a</sup> = Cognitive screen for memory, orientation, and concentration with scores ranging from 0–28; lower scores indicate better cognitive function (0–4 = normal cognition)

<sup>b</sup> = Center for Epidemiologic Studies Depression Scale; Scores range from 0–60 with higher scores indicating greater depressive symptomatology

<sup>c</sup> = Action Research Arm Test; Scores range from 0–57 with higher scores indicating better function. All individuals across groups had moderate UL paresis.

<sup>d</sup> = Concordance is when the dominant limb is the affected, paretic limb