

Improvement in the Capacity for Activity Versus Improvement in Performance of Activity in Daily Life During Outpatient Rehabilitation

Catherine E. Lang, PT, PhD, FAPTA, Carey L. Holleran, PT, DPT, DHS, Michael J Strube, PhD, Terry D. Ellis, PT, PhD, FAPTA, Caitlin A. Newman, OTR/L, Meghan Fahey, PT, DPT, Tamara R. DeAngelis, PT, DPT, Timothy J. Nordahl, PT, DPT, Darcy S. Reisman, PT, PhD, FAPTA, Gammon M. Earhart, PT, PhD, FAPTA, Keith R. Lohse, PhD, and Marghuretta D. Bland, PT, DPT, MSC

Background and Purpose: We addressed questions about the potential discrepancy between improvements in activity capacity and improvements in activity performance in daily life. We asked whether this discrepancy is:

1. Common in routine, outpatient care, or an artifact of intervention studies?
2. Unique to upper limb (UL) rehabilitation, or is it seen in walking rehabilitation too?
3. Only seen in persons with stroke, or a broader neurorehabilitation problem?

Methods: A longitudinal, observational cohort of 156 participants with stroke or Parkinson disease (PD) receiving outpatient rehabilitation at 5 clinics was assessed around admission and monthly

during their episode of care. Individual, longitudinal capacity (Action Research Arm Test or walking speed) and performance (wearable sensor measurements of use ratio or steps/day) data were modeled to extract predicted change scores. Simulation methods were used to determine whether an individual's change was greater than 1 standard error. Participants were classified into categories based on whether or not they improved in capacity (C+ or C-) and/or performance (P+ or P-).

Results: A majority (59%) were classified as C+P-. Smaller portions of the sample were classified as C+P+ (20%) and C-P- (21%), with 1 participant (<1%) classified as C-P+. The proportions in the C+P- category were significantly larger in the stroke ($\chi^2 = 48.7, P < 0.0001$) and PD ($\chi^2 = 24.3, P < 0.0001$) walking subgroups than in the stroke UL subgroup.

Discussion and Conclusions: The discrepancy between improvements in capacity and performance is a problem in routine, outpatient neurorehabilitation. If performance information were available, patients and clinicians could act to address it.

Video Abstract available for more insights from the authors (see the Video, Supplemental Digital Content 1, available at: <http://links.lww.com/JNPT/A396>).

Key words: *measurement, outcomes, Parkinson disease, stroke, upper limb*

(*JNPT* 2023;47: 16–25)

Program in Physical Therapy (C.E.L., C.L.H., G.M.E., K.R.L., M.D.B.) and Program in Occupational Therapy (C.E.L., M.D.B.), Washington University School of Medicine, St Louis, Missouri; Departments of Neurology (C.E.L., C.L.H., G.M.E., K.R.L., M.D.B.) and Neuroscience (G.M.E.), Washington University School of Medicine, St Louis, Missouri; Department of Brain and Psychological Sciences, Washington University, St Louis, Missouri (M.J.S.); Department of Physical Therapy, Boston University, Boston, Massachusetts (T.D.E., T.R.D., T.J.N.); Shirley Ryan Ability Lab, Chicago, Illinois (C.A.N., M.F.); and Department of Physical Therapy, University of Delaware, Newark (D.S.R.).

Sources of funding: NIH R01HD068290.

Earlier versions of the data set were part of invited presentations at the APTA Combined Sections Meeting February 2021 and the ANPT Annual Meeting October 2021.

The authors declare no conflict of interest.

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.jnpt.org).

Correspondence: Catherine E. Lang, PT, PhD, FAPTA, Washington University School of Medicine, 4444 Forest Park, MSC: 8502-66-1101, St Louis MO 63122 (langc@wustl.edu).

Copyright © 2022 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of Academy of Neurologic Physical Therapy, APTA.

ISSN: 1557-0576/23/4701-0016

DOI: 10.1097/NPT.0000000000000413

INTRODUCTION

The World Health Organization separates the activity domain into the capacity for activity versus performance of activity in daily life.¹ Capacity (or functional capacity) is what someone is capable of doing, assessed by standardized tests in structured, clinical, or laboratory settings. Performance is the activity that someone actually does in the unstructured, free-living environment. Capacity and performance of activity are separate from participation, defined as an individual's involvement in life situations, which requires someone to perform a collection of activities in daily life.² Despite rehabilitation professional efforts focused on capacity measures,³ patients are referred to and seek out rehabilitation services to improve performance of activity in daily life, as indicated by

self-reported rehabilitation goals.⁴ Advancements in wearable motion sensors now allow for direct measurement of movement performance in daily life. Sensor-derived variables can provide relevant information specific to patient goals that can be used for clinical decision-making.⁵

Clinicians measuring improvements in the capacity for activity generally infer that performance of activity in daily life also improves. We previously found, however, that moderate capacity gains (average improvement >6 points on the Action Research Arm Test [ARAT]⁶) in a randomized controlled trial designed to improve upper limb (UL) capacity post-stroke did not translate to improved UL performance in daily life, as measured by bilateral, wrist-worn accelerometers.⁷ This was true for all accelerometer-derived performance variables and for all participants, regardless of the magnitude of capacity change.⁷ Data suggesting that UL capacity improvements may not readily translate to performance improvements⁷⁻⁹ are largely from research studies where interventions were delivered by specialized research teams, separate from routine clinical care. The data generate critical questions about whether the discrepancies between improvements in capacity and improvements in performance are a function of research interventions, the type of rehabilitation (ie, UL services), and/or the patient population (ie, stroke), or whether the discrepancies are a general phenomenon of routine, clinical neurorehabilitation service.

Using a longitudinal, prospective, observational cohort of persons receiving physical and occupational therapy services at multiple sites, this study asked 3 key questions:

1. Is the discrepancy common in routine outpatient care, not just in research studies?
2. Is the discrepancy unique to UL rehabilitation post-stroke, or is it present in walking rehabilitation too?
3. Is this discrepancy only seen in persons with stroke, or is it a broader problem across neurorehabilitation?

To address the first question, we examined data from all persons recruited into the cohort. To address the second question, we compared capacity and performance outcomes in persons with stroke undergoing UL versus walking rehabilitation. We hypothesized that the discrepancy would be greater for the UL, since activities of daily living can be completed unilaterally in daily life, but walking requires the use of 2 limbs. To address the third question, we compared capacity and performance outcomes in persons with stroke versus Parkinson disease (PD) undergoing walking interventions. PD was chosen as the comparator group because it is the second-largest population seen for motor neurorehabilitation, behind stroke. Both stroke and PD populations regularly demonstrate improvements in capacity over the course of neurorehabilitation research studies, despite the fact the PD is a neurodegenerative disease.^{10,11} While the neurobiological underpinnings of the 2 conditions are very different, common factors that may hamper translation of in-clinic, capacity gains to gains in performance in daily life are most likely to be multifactorial with personal and environmental contributions.¹² Given that people are referred to or seek out rehabilitation services to improve performance of activities in daily life, the answers

to these questions are important for evaluating the effectiveness of current services and potentially improving future services.

METHODS

The study design was a prospective, longitudinal cohort of people participating in outpatient services at 5 clinics in the United States. Outpatient service delivery was chosen over inpatient service delivery as the study recruitment environment because people receiving outpatient care are, or have returned to, living in their own homes. This provides the opportunity for motor skills gained in therapy sessions to be applied throughout the day at home. Outpatient services also see a broad range of neurorehabilitation conditions over a longer time span than is possible during inpatient rehabilitation, at least in the United States. Participants were recruited from 5 clinics initially: (1) 2 academic physical therapy clinics (St Louis, Missouri, and Boston, Massachusetts); (2) 2 rehabilitation clinics that provide physical, occupational, and speech therapy services affiliated with but not directly considered academic clinics (St Louis, Missouri, and Chicago Illinois); and (3) 1 community outpatient rehabilitation clinic that provides physical, occupational, and speech therapy services (St Louis, Missouri). Midway through the data collection period, 1 academic physical therapy clinic merged with one of the academic-affiliated clinics, such that data collection was coming from only 4 clinics. The study was approved by the local institutional review boards at each site and all participants provided written informed consent.

Participants

Persons with stroke or PD were recruited if they met all of the following inclusion criteria: (1) neurologist diagnosis of stroke or idiopathic PD (Hoehn-Yahr score 2-3), but not both diagnoses in the same individual; (2) referral for outpatient physical or occupational therapy; (3) anticipated to receive rehabilitation services for at least 1 month; (4) documented therapy goal(s) to improve UL limb function or walking mobility; (5) able to follow 2-step commands and participate in testing; and (6) for persons with PD, stable dose of PD medication more than 2 weeks prior to enrollment and no medication changes anticipated during the time of therapy services. Potential participants were excluded if they met one or more of the following criteria: (1) other neurological or psychiatric conditions, including deep brain stimulator implants; (2) other orthopedic conditions that limit UL capacity or mobility (eg, amputation, severe arthritis, and significant pain); (3) other comorbid conditions such that the physician or therapy documentation indicates minimal chance for improvement in function (eg, end-stage cancer diagnosis); (4) UL or walking capacity that is already near normal (as indicated by ARAT scores ≥ 52 or self-selected gait speeds ≥ 1.2 m/s). Participants were enrolled in 1 of 3 subgroups—stroke UL, stroke walking, or PD walking—such that each subgroup could be considered statistically independent from one another. If a participant met criteria for more than one subgroup, they were assigned to the subgroup that had the smallest number of participants at the time.

Prior to data collection, a power analysis employing a 2-group comparison for each hypothesis, a Wald test with 1 degree of freedom, a 15% dropout rate, a 2-tailed α of 0.05, and a power of 0.8 to detect differences between subgroups of 10% to 15% indicated that we would need a target sample ranging from 132 to 231 participants. This size sample would detect 10% to 15% differences in frequency of outcome classification (see the Data Management and Analyses subsection) between the 3 subgroups. Enrollment in the cohort was paused in March 2020 and then restarted in September 2020 due to the COVID-19 global pandemic.

Measures

Study measures were collected within 1 week of starting therapy and monthly (± 10 days) thereafter for the duration of services. Monthly assessments allowed ample time for change and matched most insurer reporting requirements. Demographic and descriptive data collected at the initial visit included age, gender, race, ethnicity, comorbidities, current medications, time since injury and type (stroke) or time since diagnosis (PD), concordance (dominant limb = paretic limb, for UL subgroup only), and self-report of rehabilitation services to date (eg, did or did not go to inpatient rehabilitation). PD severity was assessed with the Movement Disorders Society Unified Parkinson's Disease Rating Scale (UPDRS).¹³ The Montreal Cognitive Assessment (MoCA)¹⁴⁻¹⁶ screened general cognitive abilities. Participants received medical and rehabilitation services in accordance with their overall plan of care. We recorded information about but did not interact nor interfere with the routine rehabilitation services delivered to participants. Information from the study measures was not fed back to the treating clinicians or patients at 4 of the 5 sites. At the fifth site, the personnel doing the assessments were the treating clinicians. Performance data from the previous assessment were therefore accessible in the research database, but not in the clinical record. From the therapy service records, we collected type of therapy, goals for therapy, and the duration and frequency of therapy services. Initial training, periodic study team meetings, and video-taping and rescored across sites ensured that administration of study assessments was uniform over the 4-year duration of data collection.

Capacity Measures

UL capacity was quantified by the ARAT, a test that quantifies the ability to reach, grasp, manipulate, and release a variety of everyday objects (higher scores are better, range = 0-57). The ARAT was chosen because: (1) it is an established clinical measure; (2) it has consistently strong psychometric properties, including sensitivity to change in people with stroke¹⁷⁻²⁵; (3) the time to administer is short compared with other, similar measures²⁶; and 4) it is widely used in UL rehabilitation research and clinical practice around the world.

Walking capacity was quantified by gait speed on the 10-m walk test. This test is valid, reliable, sensitive to change, and quick to administer and is the current gold standard for measuring walking capacity for stroke, PD, and older adults in research and clinical practice.²⁷⁻³¹ Gait speed was collected with instructions to walk at a "comfortable" speed (3 trials) and "as fast as possible" (3 trials). The average of the fast

trials was used as the primary measure to indicate walking capacity. Fast walk speed was chosen over self-selected walk speed because it is the most rigorous way to quantify what the individual was capable of achieving. Within this sample, fast and self-selected speeds were highly correlated ($r = 0.91$, $P < 0.0001$), and produced similar overall statistical conclusions.

Performance Measures

UL performance was captured with accelerometers, an established, valid, and reliable methodology in nondisabled adults and adults with stroke.³²⁻³⁵ Consistent with previous work,^{9,36-41} GT9X Link accelerometers (Actigraph Inc, Pensacola, Florida) were worn on both wrists for 3 days (including sleep and bathing/showering) while people went about their normal, daily routines. Data were collected at 30 Hz, downloaded, and then processed in Actigraph software and in MATLAB to calculate numerous variables. The use ratio was chosen as the primary measure of UL performance because it is the most highly consistent of the accelerometer variables in healthy adults and children, it is responsive to change in people with stroke, and it is highly correlated with other accelerometer variables.^{41,42} The use ratio is the hours of paretic limb activity divided by the hours of nonparetic limb activity and quantifies the contribution of the paretic limb relative to the nonparetic limb over the course of the wearing period.³⁹ Healthy, neurologically intact adults have a use ratio of 0.95 ± 0.06 , indicating nearly equal durations of UL movement during daily activity.³⁹

Walking performance was captured by StepWatch Activity Monitors (SW1002, Modus Health). Step activity monitors provide valid and reliable measures of walking performance in daily life, and are easy to use with adults from a variety of patient populations.⁴³⁻⁵⁰ Step detection accuracy exceeds 98%, even for shuffling, dyskinetic PD gait.^{51,52} Step activity monitors were worn on the less affected leg for 7 days,⁴⁶ and calibrated per standard procedures.^{44,45} The longer wearing period for walking was due to the fact that day-to-day walking performance can be highly variable.⁵³ The primary variable for walking performance was steps per day because it the most common variable used across populations, with established psychometric properties.^{5,45}

Data Management and Analyses

Data were collected and managed via a REDCap database.²⁸ Data across sites were checked on entry and audited quarterly. Analyses were done in R version 4.1.2,⁵⁴ employing nonlinear, longitudinal, multilevel modeling with the lme4 package.⁵⁵ Longitudinal, multilevel analyses (measures nested within people) are the preferred method for these data, given it does not require the same number of assessments across participants, can account for missing data, and can minimize noise in the clinical measures.⁵⁵⁻⁵⁷

Determination of whether or not change occurred is difficult in rehabilitation research and clinical practice. We rejected the option of individual change scores greater than a published minimal detectable change (MDC) or minimal clinically important difference (MCID) because: (1) MDC and MCID values are estimates from research samples with specific inclusion/exclusion criteria that may not be a good match

with patients seen in routine outpatient services; (2) studies of MDC and MCID use a variety of anchors (including arbitrary percentage of a scale) to determine whether change occurred⁵⁸⁻⁶⁰; (3) MCID values may change with time^{21,61} or severity, which would differentially affect individual participants in our heterogeneous sample; and (4) several of our measures (use ratio, steps/day) have minimal information available to make MCID estimates.⁵ Additionally, individuals exhibit varying amounts of stability in their measures over time; individually determined standard errors provide a better basis of classifying change than a single standard error that is applied to all participants. We therefore employed simulation methods to obtain individual-level standard errors and used those to generate model-based probabilistic estimates of improvement over time for each individual, as follows.

First, individual participant trajectories for the capacity and performance data were modeled (Figure 1A) using polynomial curve fitting with time (centered on baseline). While both linear and quadratic models were adequate fits with the data, the quadratic model was chosen to obtain the most accurate fit possible for the subsequent analytic steps. The random effects part of this model was as complex as the fitting algorithm allowed (a random intercept and typically a random slope).

Second, this base model was used to generate a new set of model-consistent outcome values. A new set of regression coefficients was created, sampled from the multivariate normal distribution with the original model coefficients as the mean vector and a variance-covariance matrix equal to the original coefficient variance-covariance matrix. New level 2 random effects were generated from a multivariate normal distribution with mean vector corresponding to the individual-specific random effect vector from the original analysis and variance-covariance matrix equal to the individual-specific

uncertainty from the original analysis. New level 1 residuals were generated from a normal distribution with mean of 0 and standard deviation equal to level 1 residual standard error from the original analysis. These new vectors were then used along with the original model fixed effect design matrix and random effect design matrix, z , to produce a new vector of model-consistent outcome values. These outcome values, along with the original predictor values, were then analyzed to get new model-predicted outcome values from which a new change score was calculated for each participant.

Third, the previous step was repeated 1000 times to create, for each person, a distribution of model-consistent change scores, from which an individual-level change standard error was calculated. At the individual level, the predicted change scores from the original model were transformed into z scores using the individual-specific standard error from the simulated distributions, allowing for comparison of changes across measurements and measurement levels with different metrics (Figure 1B).

Fourth, each person was categorized into 1 of 4 groups (Figure 1C) based on their z scores: (1) improved capacity and improved performance (C+P+); (2) improved capacity and unimproved performance (C+P-); (3) unimproved capacity and improved performance (C-P+); and (4) unimproved capacity and unimproved performance (C-P-). We considered several thresholds for z scores. A z -score threshold of 1.0 was chosen as the cut-off for classification into the improved category, such that an individual was considered to have improved if their change score was more than 1 standard error of their change score distribution (corresponds to a 1-tailed $P \sim 0.15$ relative to 0). A threshold of 1.0 was selected over the more rigorous $z = 1.645$ (corresponds to a 1-tailed $P < 0.05$ relative to 0) because these were individual, not group estimates.

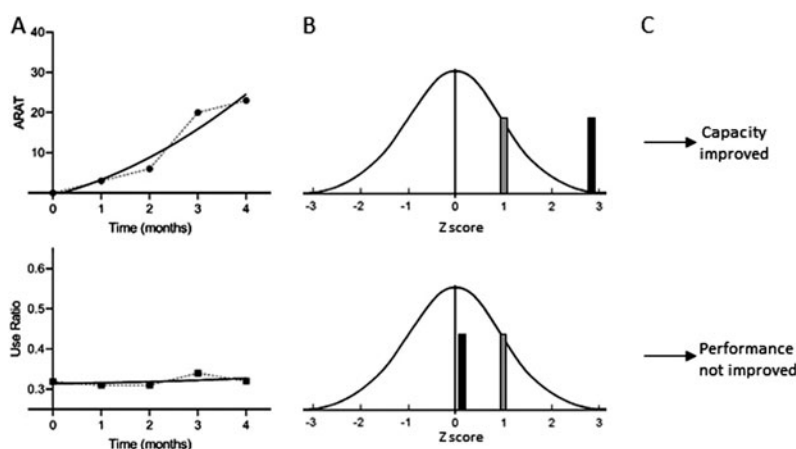


Figure 1. Illustration of analytic process to determine whether improvement occurred. (A) Participant in the stroke UL subgroup capacity (top) and performance (bottom) measurements from onset to discharge from outpatient rehabilitation services. Symbols are the measurements and the thick lines are the individual models from those measurements. (B) Individual models were used to simulate distributions of change scores using model coefficients, uncertainties, and covariance estimates. The gray bar = +1 SE in the simulated z distribution. The black bar marks the predicted change z score from the actual model coefficients. (C) Capacity (top) is judged as improved because the z score is larger than 1 SE, while performance (bottom) is judged as unchanged because the z score is smaller than 1 SE. This participant was classified as C+P-. ARAT, Action Research Arm Test; SE, standard error; UL, upper limb.

Thus, these procedures resulted in probabilistic decisions (85% probability) that an improvement is real for each person for each measurement level (Figure 1C). The probabilistic decisions provide *no* information about whether or not any improvement was meaningful to the patient or treating clinician.

Descriptive statistics on demographics and other characteristics were compared via analyses of variance (ANOVAs) to evaluate differences between subgroups. Statistical significance for this and subsequent analyses was set at $\alpha < 0.05$. To address the first question regarding improvements in routine outpatient care, we report the classification distribution from the entire sample. To address the second question regarding a unique problem of UL rehabilitation, we compared the classification distributions from the stroke UL versus stroke walking subgroups. To address the third question regarding a unique stroke rehabilitation problem, we compared classification distributions from the stroke walking versus PD walking subgroups. Comparisons between distributions were done via χ^2 tests. Additionally, we explored how age, number of assessments (a proxy for duration of services), time since stroke or PD diagnosis, MoCA score (cognition), and concordance (within stroke UL group only) might differ across the classifications using ANOVAs in the full sample. Numbers in each cell were too small to do this within the subgroups. We ran the entire analyses twice, first with the participants who enrolled and completed data collection before the global pandemic ($n = 138$), and then with everyone enrolled ($n = 156$), because the pandemic could have influenced performance data from daily life. These 2 analyses produced the same statistical conclusions; we therefore report data from all 156 participants.

RESULTS

The flow diagram for recruitment and enrollment from outpatient physical and occupational therapy services is shown in Figure 2. Overall, the sample (Table 1) was a hetero-



Figure 2. Flow diagram of participants into the observational cohort. Mercy: Mercy Outpatient Therapy Services, St Louis, Missouri; SRAL: Shirley Ryan Ability Lab, Chicago, Illinois; TRISL: The Rehabilitation Institute of Saint Louis, St Louis, Missouri.

geneous group of individuals with stroke or PD that resembles the population seen in outpatient rehabilitation services. As expected, the PD walking subgroup was older, capable of walking faster, and took more steps/day than the stroke mobility subgroup ($P < 0.05$). Examples of classifications for each of the 3 subgroups are provided in Figure 3 to show the variety in the participants and how individual trajectories resulted in

Table 1. Sample Characteristics^a

	Stroke UL n = 51	Stroke Walking n = 48	PD Walking n = 57
Age, y	60 ± 12	62 ± 12	71 ± 7 ^b
Sex, female	37%	40%	44%
Race			
White	56%	58%	92%
Black	42%	40%	4%
Asian	2%	2%	4%
Ethnicity, Hispanic/Latinx	2%	0%	0%
Time post-stroke, mo	1.7 (1.3, 15)	2.3 (1, 7)	...
Time since PD diagnosis, y	5 (2, 8)
Number of assessments	3.5 ± 1.5	2.8 ± 1.0	3.1 ± 1.1
MoCA	23 (20, 26.5)	23 (21, 26)	24 (20, 27)
Upper limb concordance	51%
UPDRS (iii_total)	37 (26, 44)
UL capacity (ARAT)	23 (4, 45)
UL performance (use ratio)	0.53 (0.35, 0.71)
Walking capacity (fast walk speed)	...	0.83 (0.67, 1.05)	1.26 (1.05, 1.51) ^c
Walking performance, steps/d	...	4996 (3330, 7888)	7220 (4476, 10264) ^c

Abbreviations: ARAT, Action Research Arm Test; MoCA, Montreal Cognitive Assessment; PD, Parkinson disease; UL, upper limb; UPDRS, United Parkinson's Disease Rating Scale.

^aValues are mean ± SD, percentage, or median (first quarter, third quarter). Concordance = dominant upper limb = paretic limb, reported just for stroke UL subgroup.

^bSignificantly different from the stroke UL and stroke walking subgroups.

^cSignificantly different from the stroke walking subgroup.

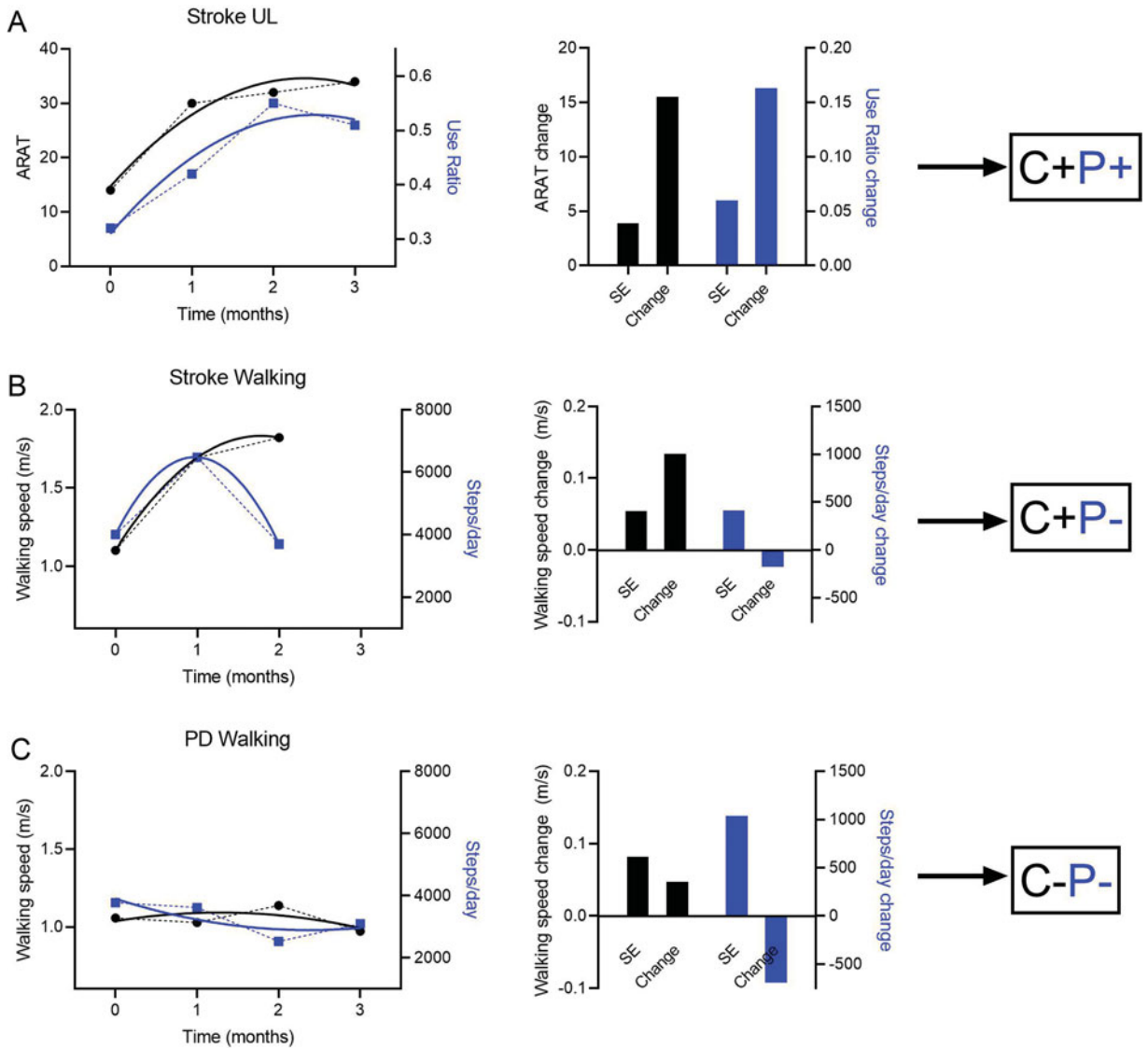


Figure 3. Example participants. Capacity measures are black and scaled by the left y-axis. Performance measures are blue and scaled by the right y-axis. In the left column, symbols are the measurements and the thick, solid lines are the modeled data. In the middle column, SE = 1 standard error from the individual simulated distributions, shown in units of the original scale and change = model predicted change scores, also shown in units of the original scale. (A) Participant from the stroke UL subgroup, classified as C+P+. (B) Participant from the stroke walking subgroup, classified as C+P-. (C) Participant from the Parkinson disease (PD) walking subgroup, classified as C-P-. This figure is available in color online (www.jnpt.org).

model predicted change scores, standard errors in units of the measurement scales, and eventual classifications.

With respect to question 1, discrepancies between improvements in capacity and improvements in performance appear to be common in routine outpatient rehabilitation services. The classification distribution for the full sample is shown in the top part of Table 2. The majority (59%) of the sample improved capacity for activity but did not improve on performance of activity in daily life. Exploration of potential factors influencing classifications revealed that the classification groups had wide ranges for each of the 3 significant variables, with the C+P+ group being slightly younger, and the C-P- group being

more chronic and having more assessments (middle part of Table 2).

With respect to question 2, the discrepancy between capacity and performance improvements is not an issue isolated to UL limb stroke rehabilitation. Classification distributions for the subgroups are shown in the bottom part of Table 2. To facilitate pairwise comparisons among groups, the C-P+ classification (containing only 1 participant) was eliminated. More persons improved in both capacity and performance in the stroke UL group compared with the stroke walking group ($\chi^2 = 48.7, P < 0.0001$). Likewise with respect to question 3, this issue is not isolated to stroke, as indicated by the PD walking subgroup distribution (different from UL subgroup,

Table 2. Classification Distribution for the Full Sample, Factors Influencing Classifications, and Classification Distributions of the 3 Subgroups^a

Full sample (n = 156)	Improved Performance, P+		Unimproved Performance, P–
Improved capacity, C+	20% (31)		59% (92)
Unimproved capacity, C–	<1% (1)		21% (32)
Factors significantly influencing classifications in the full sample (n = 155) ^b	C+P+	C+P–	C–P–
Age, y	60 (28, 87)	65 (29, 86)	70 (45, 88)
Time since stroke or diagnosis, mo	1.6 (<1, 76)	4.1 (<1, 212)	38 (1.4, 263)
Number of assessments	2 (2, 7)	2.5 (2, 7)	4 (2, 7)
Stroke upper limb (n = 51)	Improved Performance, P+		Unimproved Performance, P–
Improved capacity, C+	51% (26)		33% (17)
Unimproved capacity, C–	0% (0)		16% (8)
Stroke walking (n = 48)			
Improved capacity, C+	0% (0)		100% (48)
Unimproved capacity, C–	0% (0)		0% (0)
PD walking (n = 57)			
Improved capacity, C+	9% (5)		47% (27)
Unimproved capacity, C–	2% (1)		42% (24)

^aValues are % (n), or medians (min, max). Values are rounded to the nearest whole percentage; totals may not equal 100% are due to rounding.

^bAnalyses exclude the C–P+ group since it had only 1 participant. Significant differences between the 3 groups as indicated by $P = 0.007$ for age and $P < 0.0001$ for time and number of assessments.

$\chi^2 = 24.3$, $P < 0.0001$; different from stroke walking subgroup, $\chi^2 = 34.5$, $P < 0.0001$). The statistics of the change scores (in units of the measurement scales) are provided in the top of Table 3, confirming that the method of individual probabilistic change judgments was adequate to detect whether a change did or did not occur. Data were further examined to see whether ceiling effects at the time of initial assessment could account for the lack of improvement over time in capacity or performance measures, with this information provided in the bottom of Table 3.

DISCUSSION

In this longitudinal cohort, discrepancies between improvements in capacity for activity and improvements in

performance of activity in daily life occurred in the majority of people receiving outpatient rehabilitation services. Contrary to our hypothesis, the discrepancy was larger for walking rehabilitation than for UL rehabilitation. The discrepancy in outcomes was not restricted to stroke rehabilitation, but was also present in people with PD. These novel results come from 5 outpatient clinics in United States, making them more generalizable than if they had come from a single clinic or a single region. Overall, these data suggest that measuring effectiveness of rehabilitation services with only capacity measures may be insufficient.

The majority of participants in the overall sample improved on capacity but not performance measures over the course of their rehabilitation episode. Episodes of care often

Table 3. Model Predicted Change Scores According to Improvement Versus No Improvement for Each of the Subgroups, in Units of the Measurement Scales (Means \pm SDs) and Participants Meeting All Study Inclusion Criteria With Initial Assessment Scores at or Near Ceiling by Classification (n)

	Model Predicted Change Scores			
	C+	C–	P+	P–
Stroke UL	6.8 \pm 5.5 points	1.0 \pm 1.0 points	0.05 \pm 0.05 ^a	0.0 \pm 0.03 ^a
Stroke walking	0.12 \pm 0.04 m/s	8 \pm 120 steps/d
PD walking	0.06 \pm 0.03 m/s	0.03 \pm 0.05 m/s	2476 \pm 2049 steps/d	–505 \pm 1187 steps/d
	Numbers of Participants Meeting Study Inclusion Criteria With Initial Scores at a Level That Could Be a Potential Ceiling Effect ^b			
	C+	C–	P+	P–
Stroke UL	2	0	3	2
Stroke walking	8	3
PD walking	15	22	1	11

Abbreviations: C+, improved capacity; C–, no capacity improvement; P+, improved performance; P–, no performance improvement; PD, Parkinson disease; UL, upper limb.

^aRatio scale.

^bSelected cut-offs for potential ceiling effects (ie, a score that a participant might not be able to improve much beyond or a value that is already sufficient for community mobility and/or maintaining overall health status) were: Action Research Arm Test > 50/56 points; use ratio > 0.91 (1 SD below referent value); fast walking speed > 1.2 m/s (estimated speed to cross a busy street⁶⁸) and steps/day > 10 000 (recommended daily stepping activity⁶⁹).

extended for multiple months, with the median episode of care being around 2 months (3 assessments) and the longest being 6 months (7 assessments). In the UL at least, performance tends to change on approximately the same time scale as capacity, or perhaps even slightly earlier, as individuals recover from stroke.⁶² Thus, it would be hard to argue that lack of time is a cause of lack of performance change in this cohort. Given the current rehabilitation focus primarily on capacity measures,³ our data show that patients are improving on what is being measured by clinicians (ie, the capability to execute activity within the clinical setting). These data, along with a few previous reports,⁷⁻⁹ now make it clear that one cannot assume that improvements observed in a clinical setting carry over to improvements outside the clinic. Further, these data do not say that performance cannot change, just that performance often *did not change* in the current delivery model. Measures of performance in daily life are not currently a routine part of clinical care.⁶³ These data open up the opportunity that if performance information were available, then patients and clinicians could act to address and improve it.

The proportion of persons who improved capacity for activity but failed to improve performance of activity in daily life was greater in walking rehabilitation than in UL rehabilitation. This result is likely not a function of the walking subgroups being more severely affected by their stroke or by PD (see values at bottom of Table 1). One possibility for this result is that participants may have been more self-limiting due to fear of serious consequences (eg, fall and fracture) when trying to increase walking outside the clinic compared with trying to increase UL activity (eg, minor consequences such as slower to complete tasks or spills). Another possibility is that people undergoing outpatient therapy may have felt they were getting sufficient walking practice in therapy sessions and did not need to engage in extra walking activity outside of the services. This would be consistent with findings suggesting people compensate for their structured physical activity by doing less activity during the unstructured time.^{64,65} A third possibility is that the daily environment in which people live may place more restrictions on walking performance than UL performance. Larger amounts of UL activity can occur easily indoors, whereas larger amounts of walking activity are influenced by a range of social and environmental factors in the outside environment.⁶⁶ A fourth possibility is that occupational therapists (the primary profession delivering UL rehabilitation services in the United States) may be better at facilitating carryover of gains outside the clinic than physical therapists. This could stem from the professional training programs, where occupational therapy educational programs may be more focused on improving performance of activity in daily life, as that facilitates participation in important life roles. Regardless of which or how much these 4 possibilities contributed to the result, the data make it clear that it is time to reexamine the content of and how outpatient neurorehabilitation services are delivered.⁶⁷

These data are applicable to a range of severities and periods post-stroke. The sample included participants with broad ranges of capacity (ARAT scores 0-51 points, fast walking speed 0-1.3 m/s, self-selected walking speed 0-1.0 m/s) and broad ranges of time post-stroke (<1 month out to 212 months

[>17 years]) at the start of their rehabilitation episodes of care. As can be seen in the middle portion of Table 2, time post-stroke did influence the likelihood of improving performance. Those who did not improve (C–P– category) were, on average, more chronic. While this was statistically significant, it is clear from the minimum and maximum values that each classification included participants from a wide range of times post-stroke. Based on the wide ranges here, one cannot predict who will or will not improve performance based on time post-stroke or PD diagnosis.

Limitations

Several limitations influence the interpretation of our data. First, the capacity and performance-level measures used here are not perfect. Each one is a sample of the underlying construct, not a complete picture. Thus, there may have been changes (for better or worse) in capacity and/or performance that were not captured by the measures chosen here. Second, a small portion of the classifications (primarily PD walking subgroup participants who initially walked more than 10 000 steps/day) could have been influenced by a potential ceiling effect based on initial assessment values. These participants, however, met all study inclusion criteria, including self-selected walking speeds of less than 1.2 m/s, duration of outpatient therapy services for more than 1 month, and documented goals to improve walking. And third, classification frequencies were determined from probabilistic decisions of improvement versus no improvement. We have no knowledge of whether or not improvements were meaningful to the patients who experienced them, making it possible that we have overestimated the proportion of people who were classified as C+ and/or P+.

CONCLUSIONS

Improvements in capacity for activity measured in the clinic setting during outpatient neurorehabilitation episodes of care often do not translate into improvements in activity performance in daily life. Future research is critically needed to: (1) develop more clinically feasible devices and methods to measure and track performance in routine rehabilitation care and (2) determine how best to modify, restructure, or supplement rehabilitation interventions so that the benefits gained during therapy services are realized in the daily life of people who seek those services.

REFERENCES

1. World Health Organization. *Towards a Common Language for Functioning, Disability, and Health: ICF*. Geneva, Switzerland: World Health Organization; 2002.
2. World Health Organization. *International Classification of Functioning, Disability, and Health: ICF*. Geneva, Switzerland: World Health Organization; 2001.
3. Moore JL, Potter K, Blankshain K, Kaplan SL, O'Dwyer LC, Sullivan JE. A core set of outcome measures for adults with neurologic conditions undergoing rehabilitation: a clinical practice guideline. *J Neurol Phys Ther*. 2018;42(3):174-220.
4. Waddell KJ, Birkenmeier RL, Bland MD, Lang CE. An exploratory analysis of the self-reported goals of individuals with chronic upper-extremity paresis following stroke. *Disabil Rehabil*. 2016;38(9):853-857.
5. Lang CE, Barth J, Holleran CL, Konrad JD, Bland MD. Implementation of wearable sensing technology for movement: pushing forward into the

- routine physical rehabilitation care field. *Sensors (Basel)*. 2020;20(20):5744.
6. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the Action Research Arm Test. *Neurorehabil Neural Repair*. 2008;22(1):78-90.
 7. Waddell KJ, Strube MJ, Bailey RR, et al. Does task-specific training improve upper limb performance in daily life poststroke? *Neurorehabil Neural Repair*. 2017;31(3):290-300.
 8. Rand D, Eng JJ. Disparity between functional recovery and daily use of the upper and lower extremities during subacute stroke rehabilitation. *Neurorehabil Neural Repair*. 2012;26(1):76-84.
 9. Doman CA, Waddell KJ, Bailey RR, Moore JL, Lang CE. Changes in upper-extremity functional capacity and daily performance during outpatient occupational therapy for people with stroke. *Am J Occup Ther*. 2016;70(3):7003290040p1-7003290040p11.
 10. Tomlinson CL, Patel S, Meek C, et al. Physiotherapy versus placebo or no intervention in Parkinson's disease. *Cochrane Database Syst Rev*. 2012;(8):CD002817.
 11. Osborne JA, Botkin R, Colon-Semenza C, et al. Physical therapist management of Parkinson disease: a clinical practice guideline from the American Physical Therapy Association. *Phys Ther*. 2021;102(4):pzab302.
 12. Danks KA, Pohlig RT, Roos M, Wright TR, Reisman DS. Relationship between walking capacity, biopsychosocial factors, self-efficacy, and walking activity in persons poststroke. *J Neurol Phys Ther*. 2016;40(4):232-238.
 13. Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord*. 2008;23(15):2129-2170.
 14. Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc*. 2005;53(4):695-699.
 15. Popovic IM, Seric V, Demarin V. Mild cognitive impairment in symptomatic and asymptomatic cerebrovascular disease. *J Neurol Sci*. 2007;257(1/2):185-193.
 16. Zadikoff C, Fox SH, Tang-Wai DF, et al. A comparison of the mini mental state exam to the Montreal cognitive assessment in identifying cognitive deficits in Parkinson's disease. *Mov Disord*. 2008;23(2):297-299.
 17. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res*. 1981;4(4):483-492.
 18. De Weerd W, Harrison MA. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunstrom-Fugl-Meyer test and the Action Research arm test. *Physiother Can*. 1985;37:65-70.
 19. Hsieh CL, Hsueh IP, Chiang FM, Lin PH. Inter-rater reliability and validity of the Action Research Arm Test in stroke patients. *Age Ageing*. 1998;27(2):107-113.
 20. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: properties of the Action Research Arm Test. *Arch Phys Med Rehabil*. 2006;87(12):1605-1610.
 21. Lang CE, Edwards DF, Birkenmeier R, Dromerick AW. Estimating minimal clinically important differences of upper-extremity measures early after stroke. *Arch Phys Med Rehabil*. 2008;89(9):1693-1700.
 22. van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehabil Med*. 2001;33(3):110-113.
 23. van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the Action Research Arm Test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil*. 2001;82(1):14-19.
 24. van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. Improving the Action Research Arm test: a unidimensional hierarchical scale. *Clin Rehabil*. 2002;16(6):646-653.
 25. Beebe JA, Lang CE. Relationships and responsiveness of six upper extremity function tests during the first six months of recovery after stroke. *J Neurol Phys Ther*. 2009;33(2):96-103.
 26. Finch E, Brooks D, Stratford PW, Mayo NE. *Physical Rehabilitation Outcome Measures*. 2nd ed. Hamilton, Canada: BC Decker Inc; 2002.
 27. Lang JT, Kassin TO, Devaney LL, Colon-Semenza C, Joseph MF. Test-retest reliability and minimal detectable change for the 10-meter walk test in older adults with Parkinson's disease. *J Geriatr Phys Ther*. 2016;39(4):165-170.
 28. Lang CE, Bland MD, Connor LT, et al. The brain recovery core: building a system of organized stroke rehabilitation and outcomes assessment across the continuum of care. *J Neurol Phys Ther*. 2011;35(4):194-201.
 29. Winstein CJ, Stein J, Arena R, et al. Guidelines for Adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2016;47(6):e98-e169.
 30. Tyson S, Connell L. The psychometric properties and clinical utility of measures of walking and mobility in neurological conditions: a systematic review. *Clin Rehabil*. 2009;23(11):1018-1033.
 31. Sullivan JE, Crouner BE, Kluding PM, et al. Outcome measures for individuals with stroke: recommendations from the American Physical Therapy Association Neurology Section Task Force. *Phys Ther*. 2013;93(10):1383-1396.
 32. Chen KY, Acra SA, Majchrzak K, et al. Predicting energy expenditure of physical activity using hip- and wrist-worn accelerometers. *Diabetes Technol Ther*. 2003;5(6):1023-1033.
 33. 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.
 34. Gebruers N, Truijens S, Engelborghs S, Nagels G, Brouns R, De Deyn PP. Actigraphic measurement of motor deficits in acute ischemic stroke. *Cerebrovasc Dis*. 2008;26(5):533-540.
 35. van der Pas SC, Verbunt JA, Breukelaar DE, van Woerden R, Seelen HA. Assessment of arm activity using triaxial accelerometry in patients with a stroke. *Arch Phys Med Rehabil*. 2011;92(9):1437-1442.
 36. Bailey RR, Birkenmeier RL, Lang CE. Real-world affected upper limb activity in chronic stroke: an examination of potential modifying factors. *Top Stroke Rehabil*. 2015;22(1):26-33.
 37. Bailey RR, Klaesner JW, Lang CE. An accelerometry-based methodology for assessment of real-world bilateral upper extremity activity. *PLoS One*. 2014;9(7):e103135.
 38. Bailey RR, Klaesner JW, Lang CE. Quantifying real-world upper-limb activity in nondisabled adults and adults with chronic stroke. *Neurorehabil Neural Repair*. 2015;29(10):969-978.
 39. Bailey RR, Lang CE. Upper-limb activity in adults: referent values using accelerometry. *J Rehabil Res Dev*. 2013;50(9):1213-1222.
 40. Urbin MA, Bailey RR, Lang CE. Validity of body-worn sensor acceleration metrics to index upper extremity function in hemiparetic stroke. *J Neurol Phys Ther*. 2015;39(2):111-118.
 41. Urbin MA, Waddell KJ, Lang CE. Acceleration metrics are responsive to change in upper extremity function of stroke survivors. *Arch Phys Med Rehabil*. 2015;96(5):854-861.
 42. Hayward KS, Eng JJ, Boyd LA, Lakhani B, Bernhardt J, Lang CE. Exploring the role of accelerometers in the measurement of real world upper limb use after stroke. *Brain Impairment*. 2016;17(1):16-33.
 43. Danks KA, Roos MA, McCoy D, Reisman DS. A step activity monitoring program improves real world walking activity post stroke. *Disabil Rehabil*. 2014;36(26):2233-2236.
 44. Knarr B, Roos MA, Reisman DS. Sampling frequency impacts measurement of walking activity after stroke. *J Rehabil Res Dev*. 2013;50(8):1107-1112.
 45. Roos MA, Rudolph KS, Reisman DS. The structure of walking activity in people after stroke compared with older adults without disability: a cross-sectional study. *Phys Ther*. 2012;92(9):1141-1147.
 46. Paul SS, Ellis TD, Dibble LE, et al. Obtaining reliable estimates of ambulatory physical activity in people with Parkinson's disease. *J Parkinsons Dis*. 2016;6(2):301-305.
 47. Mudge S, Stott NS, Walt SE. Criterion validity of the StepWatch Activity Monitor as a measure of walking activity in patients after stroke. *Arch Phys Med Rehabil*. 2007;88(12):1710-1715.
 48. Mudge S, Stott NS. Test-retest reliability of the StepWatch Activity Monitor outputs in individuals with chronic stroke. *Clin Rehabil*. 2008;22(10/11):871-877.
 49. Cavanaugh JT, Ellis TD, Earhart GM, Ford MP, Foreman KB, Dibble LE. Capturing ambulatory activity decline in Parkinson's disease. *J Neurol Phys Ther*. 2012;36(2):51-57.

50. Cavanaugh JT, Ellis TD, Earhart GM, Ford MP, Foreman KB, Dibble LE. Toward understanding ambulatory activity decline in Parkinson disease. *Phys Ther.* 2015;95(8):1142-1150.
51. Speelman AD, van Nimwegen M, Borm GF, Bloem BR, Munneke M. Monitoring of walking in Parkinson's disease: validation of an ambulatory activity monitor. *Parkinsonism Relat Disord.* 2011;17(5):402-404.
52. Schmidt AL, Pennypacker ML, Thrush AH, Leiper CI, Craik RL. Validity of the StepWatch Step Activity Monitor: preliminary findings for use in persons with Parkinson disease and multiple sclerosis. *J Geriatr Phys Ther.* 2011;34(1):41-45.
53. Holleran CL, Bland MD, Reisman DS, Ellis TD, Earhart GM, Lang CE. Day-to-day variability of walking performance measures in individuals poststroke and individuals with Parkinson disease. *J Neurol Phys Ther.* 2020;44(4):241-247.
54. R: A Language and Environment for Statistical Computing [computer program]. Vienna, Austria: R Foundations for Statistical Computing; 2021.
55. Bates D, Maechler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Sof.* 2015;67(1):1-48.
56. Long JD. *Longitudinal Data Analyses for the Behavioral Sciences Using R.* Thousand Oaks, CA: Sage Publications; 2012.
57. Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods.* 2nd ed. Thousand Oaks, CA: Sage; 2002.
58. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol.* 2002;14(2):109-114.
59. Rennard SI. Minimal clinically important difference, clinical perspective: an opinion. *COPD.* 2005;2(1):51-55.
60. Lehman LA, Veloza CA. Ability to detect change in patient function: responsiveness designs and methods of calculation. *J Hand Ther.* 2010; 23(4):361-370; quiz 371.
61. van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. *Stroke.* 1999;30(11): 2369-2375.
62. Lang CE, Waddell KJ, Barth J, Holleran CL, Strube MJ, Bland MD. Upper limb performance in daily life approaches plateau around three to six weeks post-stroke. *Neurorehabil Neural Repair.* 2021;35(10):903-914.
63. Dobkin BH, Martinez C. Wearable sensors to monitor, enable feedback, and measure outcomes of activity and practice. *Curr Neurol Neurosci Rep.* 2018;18(12):87.
64. Ridgers ND, Timperio A, Cerin E, Salmon J. Compensation of physical activity and sedentary time in primary school children. *Med Sci Sports Exerc.* 2014;46(8):1564-1569.
65. Fanning J, Nicklas BJ, Rejeski WJ. Intervening on physical activity and sedentary behavior in older adults. *Exp Gerontol.* 2022;157:111634.
66. Miller A, Pohlig RT, Reisman DS. Social and physical environmental factors in daily stepping activity in those with chronic stroke. *Top Stroke Rehabil.* 2021;28(3):161-169.
67. Dobkin BH. Behavioral self-management strategies for practice and exercise should be included in neurologic rehabilitation trials and care. *Curr Opin Neurol.* 2016;29(6):693-699.
68. Langlois JA, Keyl PM, Guralnik JM, Foley DJ, Marottoli RA, Wallace RB. Characteristics of older pedestrians who have difficulty crossing the street. *Am J Public Health.* 1997;87(3):393-397.
69. Tudor-Locke C, Bassett DR Jr. How many steps/day are enough? Preliminary pedometer indices for public health. *Sports Med.* 2004;34(1): 1-8.