Research Article

Development and Validation of Person-Centered Cut-Points for the Figure-of-8-Walk Test of Mobility in Community-Dwelling Older Adults

Peter C. Coyle, PT, DPT, PhD,^{1,2,*} Subashan Perera, PhD,³ Valerie Shuman, PT, DPT,¹ Jessie VanSwearingen, PT, PhD, FAPTA,¹ and Jennifer S. Brach, PT, PhD, FAPTA^{1,*}

¹Department of Physical Therapy, University of Pittsburgh, Pennsylvania. ²Department of Physical Therapy, University of Delaware, Newark. ³Department of Medicine, Division of Geriatric Medicine, University of Pittsburgh, Pennsylvania.

*Address correspondence to: Jennifer S. Brach, PT, PhD, FAPTA, Department of Physical Therapy, University of Pittsburgh, Bridgeside Point 1, 100 Technology Dr, Suite 210, Pittsburgh, PA. E-mail: jbrach@pitt.edu

Received: July 30, 2019; Editorial Decision Date: January 26, 2020

Decision Editor: Jay Magaziner, MSHyg, PhD

Abstract

Background: The Figure-of-8-Walk test (F8WT) is a performance measure of the motor skill of walking. Unlike walking speed over a straight path, it captures curved path walking, which is essential to real-world activity, but meaningful cut-points have yet to be developed for the F8WT.

Methods: A secondary analysis of 421 community-dwelling older adults (mean age 80.7 ± 7.8), who participated in a community-based exercise clinical trial, was performed. Area under receiver operating characteristic curves (AUROCC) were calculated using baseline data, with F8WT performance discriminating different self-reported global mobility and balance dichotomies. Cut-points for the F8WT were chosen to optimize sensitivity and specificity. For validation, F8WT cut-points were applied to postintervention F8WT data. Participants were called monthly for 12 months after intervention completion to record self-reported incident falls, emergency department visits, and hospitalizations; risks of the outcomes were compared between those who performed well and poorly on the F8WT.

Results: F8WT performance times of ≤ 9.09 seconds and ≤ 9.27 seconds can discriminate those with excellent (sensitivity = 0.647; specificity = 0.654) and excellent/very good global mobility (sensitivity = 0.649; specificity = 0.648), respectively. A total number of steps ≤ 17 on the F8WT can discriminate those with excellent/very good/good global balance (sensitivity = 0.646; specificity = 0.608). Compared to those who performed poorly, those who performed well had a lower incidence of negative outcomes: F8WT time ≤ 9.09 seconds = 46%–59% lower; F8WT time ≤ 9.27 seconds = 46%–56% lower; F8WT steps ≤ 17 = 44%–50% lower.

Conclusions: Clinicians may consider these preliminary cut-points to aid in their clinical decision making, but further study is needed for definitive recommendations.

Keywords: Walking, Curved path, Performance, Psychometric properties, Receiver operating characteristic curve

Among older adults, usual walking speed is widely considered the hallmark performance measure of mobility (1–3). Usual walking speed predicts disability, institutionalization, and mortality (4–8). Typically, usual walking speed is measured over a straight path; yet real-world walking requires the navigation of, and continual adaptation to, complex and variable environments (9). The Figure-of-8 Walk test (F8WT), a test of curved path walking, was recently introduced as a performance measure of the motor skill of walking, capturing the directional and postural challenges essential to real-world

walking (10). Prior work has shown that the F8WT relies on cognitive and physiologic mechanisms less associated with straight-path walking speed (11,12), indicating the F8WT captures an aspect of performance-based mobility that is conceptually distinct from walking speed measured over a straight path.

While worse performance on the F8WT was associated with a history of falling within the previous year (13), much is still unknown regarding the interpretation of the F8WT. To our knowledge, little work (14) has been done to identify or validate meaningful

cut-points for the F8WT in the community-dwelling older adult population. Cut-points are crucial to inform clinical decision making about which patients may be at risk for poor health outcomes and, thus, may benefit from intervention. In this secondary analysis of data from a large cluster-randomized controlled trial (15,16), we aimed to: (i) develop person-centered cut-points for the F8WT among community-dwelling older adults; and (ii) determine whether the classifications defined by those cut-points are meaningful with regard to prognosis for poor health outcomes. Specifically, self-reported falls, emergency department (ED) visitation, and hospitalizations were measured monthly in the year following F8WT assessment.

Methods

Overview and Study Design

The original study design, methods, and primary outcomes of the parent randomized controlled trial are described elsewhere in detail (15,16). The parent study was a cluster-randomized, single-blind intervention trial designed to compare the effectiveness of two community-based group exercise interventions. Both exercise programs were held for 50 minutes per session, twice per week, over 12 weeks. Participants were assessed on-site at baseline and 12 weeks (ie, postintervention); after postintervention assessment, participants were followed over 12 months via monthly automated phone calls. The parent study was registered on clinicaltrials.gov (NCT01986647). All policies and procedures were developed in accordance with the Helsinki Declaration of the World Medical Association, and were approved by the University of Pittsburgh Institutional Review Board. Signed informed consent was obtained from all participants.

Participants

Community-dwelling adults aged ≥ 65 years in the greater Pittsburgh, Pennsylvania area were recruited for this study if they attended participating senior community centers or if they resided in participating independent living facilities or senior housing. Recruitment was conducted between April 2014 and January 2016. Participants were included if they could ambulate independently (with or without a cane) for household distances and had a usual walking speed of ≥ 0.60 m/s. Individuals were excluded if they were non-English speaking; had cognitive impairments (ie, unable to follow a two-step command or complete the informed consent process); had a progressive neurological disorder; or if they had an unstable, acute medical condition/illness (15,16).

Measures

Baseline descriptive measures

Participants reported their age, sex, race, education level, fear of falling, and history of falls within the previous year (yes/no), and body mass index (from self-reported height and weight) at baseline assessment. In addition, the Duke Comorbidity Index, a measure of the number of body systems affected by disease (eg, cardiovascular, musculoskeletal, neurological), was used to quantify comorbid burden (17); scores ranged 0–8 with lower scores indicating less comorbidity burden.

Performance measures

At baseline and 12 weeks, participants underwent mobility performance assessments of F8WT performance and usual walking speed over a straight path.

The F8WT performance was measured according to previously published procedures (10–13). Briefly, two cones were placed 1.52 m (five feet) apart. From a standstill position, participants started the walk from the midpoint between the cones and walked in a figure-of-8 pattern around the cones at their usual speed. The total time and number of steps taken to complete one figure-of-8 pattern were recorded, as time and steps appear to be uniquely related to different gait and cognitive characteristics (10,12). One trial was completed, unless the participant made contact with the cone, in which case the trial was repeated. Faster times and fewer steps indicated better walking skill. The F8WT has excellent test–retest (intraclass correlation coefficient [ICC] = 0.84 and 0.82 for time and steps, respectively) and interrater reliability (ICC = 0.90 and 0.92 for time and steps, respectively) (10).

Usual walking speed was measured using a computerized, instrumented walkway (Zeno Walkway, Protokinetics, Havertown, PA). Participants completed six passes at their usual walking speed. The test–retest reliability of walking speed measured by a computerized walkway is excellent (ICC = 0.98) (18).

Anchor measures

Single-item, global state questions were used to ascertain self-perceived mobility and balance at baseline and 12 weeks. Items and responses were developed following the Medical Outcomes Study SF-36 question of current health status as a model (19). Participants were asked, “Would you say your level of mobility in general is excellent, very good, good, fair, or poor?” The question was repeated regarding balance.

Falls, hospitalization, and ED visitation measures

Participants were called monthly for 12 months following postintervention assessment via an interactive voice response automated system. Participants that withdrew from the study or were lost to follow up during the intervention were not called. Automated calls were conducted two times per day for the 4 days before and after the due date. Data were considered missing if participants did not complete the call within the timeframe. If participants did not answer in a given month, future calls were still attempted in the following month(s).

During the automated call, participants were asked if they had fallen (ie, “landed on the floor or the ground and could not stop yourself”), were admitted to the hospital for at least one night for any reason, and visited the ED for any reason in the previous month (ie, yes/no). The interactive voice response system has been found to be reliable, valid, and acceptable by community-dwelling older adults (20). The number of months in which a participant experienced a fall, hospitalization, or ED visit in the year following the intervention were counted (0 = no months, 12 = every month).

Data Analysis

Developing person-centered cut-points using self-reported global mobility and balance anchors

Initially, we considered four possible dichotomized operational definitions (excellent vs very good/good/fair/poor; excellent/very good vs good/fair/poor; excellent/very good/good vs fair/poor; and excellent/very good/good/fair vs poor) of each anchor as meaningful person-centered classifications based on face validity. However, due to a low frequency of participants who rated themselves “poor,” we were only able to pursue analyses for three dichotomies: excellent versus very good/good/fair/poor; excellent/very good versus good/fair/poor; and

excellent/very good/good versus fair/poor. We fitted a series of logistic regression models with cross-sectional data from the baseline time point with each dichotomized anchor as the dependent variable, and either F8WT time or number of steps, as the sole independent variable. An area under the receiver operator characteristic curve (AUROC) of 0.675 was considered to provide a sufficiently strong association between the F8WT and the self-reported anchor for cut-point determination. A cut-point that equally maximized both sensitivity and specificity was chosen. Robustness of the cut-point against the alternative strategy of minimizing the index union was also considered (21).

Evaluating the person-centered performance cut-point classifications

First, to explore the validity of the person-centered F8WT cut-point classifications, we created an analytic subsample that included only participants who completed 12-week performance tests, 12-week anchor measures, and at least one monthly follow-up phone call. To ensure the psychometric test properties of the newly defined cut-points were intact, we applied the cut-points to 12-week cross-sectional data (F8WT performance and anchors) and re-examined their sensitivity and specificity using 2×2 tables (ie, above versus below cut-point compared to better versus worse self-reported global mobility or global balance dichotomy). Next, we examined the predictive validity of the cut-point classifications longitudinally in the analytic subsample, by analyzing the associations between 12-week F8WT cut-point classifications and adverse monthly count outcomes (ie, fall, ED visit, and hospitalization) in the ensuing year. We report descriptive statistics of incident rate per 1,000 person months of follow-up. Finally, we fitted generalized estimating equations models with each count outcome as the dependent variable; a negative binomial distribution and a logarithmic link function for the outcome; person-centered F8WT cut-point classification as the independent factor; and an exchangeable working correlation structure to account for clustering due to facility. Estimated incident rate ratios (IRRs), 95% confidence intervals, and *p*-values are reported.

Meaningful classifications for walking speed have been proposed in the literature (1,2). We conducted the same analyses using walking speed as a means of quality control and validation of our larger approach, rather than to draw definitive conclusions regarding walking speed. We used SAS version 9.3 (SAS Institute, Inc., Cary, NC) for all statistical analyses.

See Figure 1 for a conceptual illustration of the analysis plan relative to the study time points used for each component.

Results

Participant Characteristics

Baseline characteristics for the full baseline sample ($n = 421$), those included in the analytic subsample ($n = 316$), and those excluded from the analytic subsample ($n = 105$) are listed in Table 1. Participants were predominantly female, white, well-educated, overweight, and had comorbidities in at least two body systems. Means and proportions for baseline characteristics were relatively similar between the full baseline sample and analytic subsample; compared to participants included in the analytic subsample, those who were excluded did not differ significantly on baseline demographic characteristics, global ratings, or mobility performance ($p > .05$). Participants in the analytic subsample completed a median of 10 monthly phone calls over the course of the year. Approximately 30% reported having a fall within the previous year, and 35% reported a fear of falling.

Figure 1. Conceptual illustration of analysis plan.

Developing Person-Centered Cut-Points Using Global Mobility and Balance Anchors

Areas under the receiver operating characteristic curves for the baseline association of F8WT time and steps with the self-reported global mobility and balance anchors are reported in Table 2. For those that met the criteria (ie, minimum area of 0.675), the associated meaningful person-centered cut-points, sensitivities, and specificities are reported in Table 3. F8WT time discriminated between excellent versus very good/good/fair/poor mobility (cut-point of ≤ 9.09 seconds) and between excellent/very good versus good/fair/poor mobility (cut-point of ≤ 9.72 seconds). F8WT steps discriminated between excellent/very good/good versus fair/poor balance (cut-point of ≤ 17 steps). Walking speed discriminated between the same meaningful mobility and balance classifications (cut-points for the classifications above were ≥ 0.96 , ≥ 0.90 , and ≥ 0.89 m/s, respectively). For all cut-points across all performance metrics, sensitivity ranged 0.646–0.656 and specificity ranged 0.608–0.657. The cut-points did not materially change with the alternative strategy of minimizing the index union.

Evaluating the Person-Centered Performance Cut-Point Classifications

The sensitivities and specificities of the cut-points at the 12-week assessment are also reported in Table 3. Generally, sensitivities and specificities showed only minor changes compared to the baseline sample. For F8WT time, absolute change in sensitivity ranged 0.025–0.065 and specificity ranged 0.044–0.051. For F8WT steps, absolute change in sensitivity was 0.023 and specificity was 0.083. For walking speed, absolute change in sensitivity ranged 0.001–0.105 and specificity ranged 0.005–0.017.

Incidence of, and IRRs for, falls, ED visits, and hospitalizations for each 12-week performance classifications in the 12-week subsample are presented in Table 4. Performing well, as defined by each meaningful performance classification cut-point, on all performance

Table 1. Baseline Characteristics of the Full Sample for Cut-Point Development and 12-wk Subsample for Predictive Validity

	Mean $\pm$ SD or N (%)			<i>p</i> Value ^b
	Baseline Sample (<i>n</i> = 421)	Included in 12-wk Subsample ^a (<i>n</i> = 316)	Excluded from 12-wk Subsample ^a (<i>n</i> = 105)	
Age (years)	80.7 $\pm$ 7.8	80.8 $\pm$ 7.7	80.5 $\pm$ 8.2	.336
Sex (female)	347 (82.4)	263 (83.2)	84 (80.0)	.375
Race (non-white)	72 (17.1)	48 (15.2)	24 (22.9)	.284
Education level (some college or more)	213 (50.8)	162 (51.3)	50 (47.6)	.841
Fear of falling (yes)	147 (34.9)	114 (36.1)	33 (31.4)	.434
Fall within prior year (yes)	126 (29.9)	90 (28.5)	36 (34.3)	.259
BMI (kg/m ²)	28.9 $\pm$ 13.5	28.9 $\pm$ 14.2	28.9 $\pm$ 10.1	.940
Comorbidity Index (0–8)	2.8 $\pm$ 1.4	2.9 $\pm$ 1.4	2.6 $\pm$ 1.3	.051
Global Mobility (excellent/very good)	245 (58.2)	182 (57.6)	63 (60.0)	.605
Global Balance (excellent/very good)	131 (31.1)	103 (32.6)	28 (26.7)	.083
F8WT time (s)	10.45 $\pm$ 3.29	10.38 $\pm$ 3.18	10.67 $\pm$ 3.31	.673
F8WT number of steps	18 $\pm$ 4	18 $\pm$ 4	18 $\pm$ 5	.462
Walking speed (m/s)	0.91 $\pm$ 0.20	0.92 $\pm$ 0.20	0.89 $\pm$ 0.20	.464

Note: BMI = body mass index; F8WT = Figure-of-8 Walk test; SD = standard deviation.

^aCompleted mobility performance tests and global state anchor questions at the 12-wk time point, as well as at least one monthly call during 12-mo period after intervention completion. ^bBetween-group comparison of those included versus excluded from 12-wk analytic subsample.

Table 2. AUROCCs for Performance Tests Identifying Global Mobility and Balance Anchors Using Cross-sectional Baseline Data from the Full Sample

Self-Reported Anchor and Operational Definition	Performance Measure		
	F8WT Time	F8WT Number of Steps	Walking Speed
Global Mobility			
Excellent vs Very good/good/fair/poor	0.699*	0.672	0.701*
Excellent/very good vs Good/fair/poor	0.687*	0.658	0.726*
Excellent/very good/good vs Fair/poor	0.671	0.596	0.629
Global Balance			
Excellent vs Very good/good/fair/poor	0.626	0.613	0.642
Excellent/very good vs Good/fair/poor	0.644	0.628	0.675
Excellent/very good/good vs Fair/poor	0.673	0.676*	0.683*

Note: AUROCC = area under the receiver operating characteristic curve; F8WT = figure-of-8 walk test.

*AUROCC $\geq$ 0.675.

Table 3. Sensitivities and Specificities of Newly Developed Cut-Points Using Cross-sectional Baseline Data from the Full Sample, as well as Cross-sectional 12-wk Data

Self-Reported Anchor	Estimated Meaningful Performance Classification	Baseline Data from Full Sample		12-wk Data from Subsample	
		Sensitivity	Specificity	Sensitivity	Specificity
Excellent vs Very good/good/fair/poor mobility	F8WT time $\leq$ 9.09 s	0.647	0.654	0.712	0.610
	Walking speed $\geq$ 0.96 m/s	0.656	0.656	0.761	0.639
Excellent/very good vs Good/fair/poor mobility	F8WT time $\leq$ 9.72 s	0.649	0.648	0.624	0.597
	Walking speed $\geq$ 0.90 m/s	0.655	0.657	0.683	0.662
Excellent/very good/good vs Fair/poor balance	F8WT steps $\leq$ 17 steps	0.646	0.608	0.669	0.691
	Walking speed $\geq$ 0.89 m/s	0.646	0.645	0.647	0.634

Note: F8WT = figure-of-8 walk test.

Table 4. Predictive Validity of Adverse Events in the Following Year Based on Estimated 12-wk Performance Cut-Point Classifications

	Falls			ED Visits			Hospitalizations		
	Incident Rate (per 1,000 person months)	IRR (95% CI)	<i>p</i>	Incident Rate (per 1,000 person months)	IRR (95% CI)	<i>p</i>	Incident Rate (per 1,000 person months)	IRR (95% CI)	<i>p</i>
<i>F8WT time cut-point for excellent mobility</i>									
≤9.09 s	45	0.41 (0.29–0.58)	<.001	30	0.50 (0.33–0.76)	.001	32	0.54 (0.36–0.81)	.003
>9.09 s (ref)	108	1.00	-	64	1.00	-	58	1.00	-
<i>Walking speed cut-point for excellent mobility</i>									
≥0.96 m/s	40	0.38 (0.24–0.60)	<.001	28	0.43 (0.25–0.74)	.002	34	0.61 (0.41–0.90)	.013
<0.96 m/s (ref)	104	1.00	-	64	1.00	-	55	1.00	-
<i>F8WT time cut-point for excellent/very good mobility</i>									
≤9.72 s	51	0.44 (0.28–0.70)	<.001	34	0.54 (0.38–0.76)	<.001	32	0.49 (0.34–0.71)	<.001
>9.72 s (ref)	115	1.00	-	67	1.00	-	65	1.00	-
<i>Walking speed cut-point for excellent / very good mobility</i>									
≥0.90 m/s	46	0.39 (0.27–0.57)	<.001	36	0.60 (0.37–0.97)	.039	40	0.72 (0.50–1.04)	.078
<0.90 m/s (ref)	115	1.00	-	64	1.00	-	55	1.00	-
<i>F8WT steps cut-point for excellent/very good/good balance</i>									
≤17 steps	56	0.50 (0.33–0.76)	.001	37	0.59 (0.42–0.84)	.003	35	0.56 (0.37–0.87)	.010
>17 steps (ref)	113	1.00	-	66	1.00	-	62	1.00	-
<i>Walking speed cut-point for excellent/very good/good balance</i>									
≥0.89 m/s	46	0.39 (0.28–0.55)	<.001	35	0.54 (0.33–0.87)	.011	38	0.66 (0.45–0.97)	.033
<0.89 m/s (ref)	119	1.00	-	67	1.00	-	58	1.00	-

Note: CI = confidence interval; ED = emergency department; F8WT = figure-of-8 walk test; IRR = incidence rate ratio; ref = reference.

metrics (ie, F8WT time, F8WT steps, and walking speed) was generally associated with a significantly lower incidence of all outcomes, compared to performing poorly. For both the F8WT and walking speed, the associations between performing well and lower incidence of outcomes were strongest for falling (50%–62% lower), followed by ED visits (40%–57% lower), and were weakest for hospitalizations (28%–51% lower).

Discussion

The cut-points for the F8WT, based on self-reported global mobility and balance, were useful to classify older adults who had a significantly greater rate of future adverse outcomes. Older adults who performed well on the F8WT were at a substantially reduced risk of falls, ED visits, and hospitalizations at any point within the year following the performance assessment. It is not surprising that the greatest risk reduction was found for falls. Falling is likely more directly related to mobility and balance (the anchors on which our analyses were predicated) compared to health care utilization outcomes, which may be influenced by a host of biopsychosocial factors. Nevertheless, “better” F8WT performance was still associated with a significantly lower incidence of ED visits and hospitalizations. These associations underscore the clinical importance of these cut-point classifications with respect to downstream health and utilization outcomes. Thus, clinicians may find it helpful to use the defined F8WT cut-points as preliminary guides, to aid in determining

who is at-risk for poor health outcomes within the coming year. Furthermore, older adults at risk for adverse health outcomes may benefit from interventions to improve mobility and balance.

Our findings are consistent with the broader literature on the F8WT. Welch and colleagues (13) found that increased completion times (ie, worse performance) on the F8WT were associated with a greater rate of falls in the year prior, which is largely consistent with our findings. These relationships may be due to the underlying cognitive mechanisms that the F8WT emphasizes: Lowry and colleagues (12) found that F8WT performance is related to measures of visual scanning, as well as cognitive flexibility (ie, set-shifting), which are key factors in successfully navigating complex environments; deficits in these areas may contribute to unsteadiness in real-world walking.

Of note, however, Welch and colleagues did not find an association between F8WT performance and self-reported history of fall-related injuries or fall-related hospitalizations, while we found that poor performance was related to all-purpose hospitalizations and ED visitation. Taken together, these findings may indicate that the pathway from poor F8WT performance to hospitalization does not operate through a pathway of falling, but rather through a broader, indirect pathway. Ferrucci and colleagues (3) argues that mobility should be viewed as a summary measure of overall health, as it represents the output from complex interactions between underlying biopsychosocial health factors and the environment. It is possible that poor F8WT performance reflects deficits in a multitude of underlying physiological processes, including cognitive,

musculoskeletal, nervous, and cardiovascular systems, which may require hospital-based care regardless of fall status. On the other hand, falls resulting in injury and/or hospitalization are much rarer than falls themselves and, thus, prior studies may have been underpowered to explore the pathway from poor performance to fall-related injury/hospitalization.

We also explored person-centered cut-points for walking speed in our dataset to compare to the broader literature on performance-based mobility. Middleton and colleagues (2) provided a comprehensive review of studies in which researchers investigated a variety of cut-points for walking speed, across a number of populations (eg, community-dwelling older adults, hospitalized older adults, stroke survivors) and anchor measures (eg, disability, hospitalization, mortality). Perhaps, the most frequently reported cut-point recommendations for walking speed in community-dwelling older adults fall between 0.8 m/s and 1.0 m/s (2,22–28). These values are also echoed in the recommendations made by International Academy on Nutrition and Aging Task Forces (29). The cut-points we have identified for walking speed fall within this range, lending credibility to our broader approach.

It is important to note that there are two distinct methods for developing cut-points. In our study, we elected to use an anchor-based approach. Other studies, however, on walking speed (6,30–32) have used a distribution-based approach, in which a threshold is determined based on the distribution of the data, using quantiles (eg, the threshold between the third and fourth lowest quartiles). While common, the distinct limitation to distribution-based methodology is that threshold designation is largely arbitrary, leading one to question its clinical relevance. Alternatively, in an anchor-based approach, threshold determination is based on some clinically meaningful reference measure (ie, the anchor measure) wherein participants are grouped into different classifications (eg, excellent vs less than excellent self-reported global mobility). Then, analyses are performed to see how well the measure of interest (eg, F8WT performance) can discriminate between the two clinically meaningful classifications (33). For example, in determining meaningful cut-points for walking speed, Shimada and colleagues (34) used an anchor of those who did and did not have personal care need; Salbach and colleagues (35) used an anchor of the maximal walking speed needed to cross an intersection safely; Middleton and colleagues (27) used a physical activity anchor of $\geq 8,000$ steps per day, a threshold that is related to better health outcomes. Similar to these studies, our approach has the distinction that it is person-centered, as we elected to use an anchor based on self-perception (ie, global rating).

To our knowledge, only two other studies have investigated cut-points for F8WT performance. In community-dwelling older adults, Hibbs and colleagues (14) used self-reported physical function dichotomies derived from the Late Life Function and Disability Index; it is worth noting that the cut-point of ≤ 8.72 seconds, identified by Hibbs and colleagues, is comparable to those identified in our study. Further, Wong and colleagues (36) found that stroke survivors could be discriminated from healthy matched individuals using a cut-point of 8.20 seconds. The differences in mobility between individuals who survived a stroke versus healthy individuals are greater than that of older adults with excellent versus very good/good/fair/poor self-reported global mobility. Thus, the differences between our study and that of Wong and colleagues are unsurprising.

In our study, we elected to use cross-sectional data to develop cut-points, which is akin to the diagnostic process (ie, identification of underlying issues that are present at the same point in time as when the measurement is made). Other studies have used a longitudinal

approach, investigating the optimal threshold for identifying the incidence of a specific outcome (ie, prognosis). For example, for usual walking speed, Cesari and colleagues (24) identified a cut-point as it related to future persistent lower extremity limitation, measured over a 5-year period; Stanaway and colleagues (28) identified a cut-point as it related to mortality, measured over a 5-year period. In our study, we developed cut-points cross-sectionally; then, we evaluated the cut-points predictive (ie, prognostic) validity by investigating and comparing the incidence of untoward events. Although we ultimately identified two cut-points for the time to complete the F8WT (rather than one), they correspond to different degrees of self-perception.

Strengths and Limitations

A major strength of our design is the utilization of participant self-perception as the anchor, which is, by nature, person-centered. Global state questions, while sometimes criticized for their subjectivity and lack of nuance, have been shown to be strong predictors of poor health outcomes (37), underscoring their importance. In addition, we conducted our analyses from a large sample of older adults, who were representative of the geographic area from which they were recruited, which lends generalizability to our findings. Also, we were able to develop and validate our cut-points using two separate time points; we considered robustness of our findings against employing an alternative strategy based on minimizing the index union to identify cut-points. Furthermore, the data set contained 12 months of follow-up data, which provided rich information on health and health services utilization outcomes. Finally, the usual walking speed cut-points identified from our data set are largely consistent with the greater body of literature, indicating that our broader methodological approach is credible.

Despite these strengths, there are some limitations that should be noted. First, the threshold of AUROC = 0.675 we used for a sufficiently strong association between F8WT and anchors is lower than the generally agreed upon value of 0.700 (38). We also note that estimated AUROC is subject to sampling error and a material change in AUROC is around 0.025 (39,40). Therefore, we used the somewhat liberal criterion of 0.675 in this secondary analysis, in which we were limited by the availability of data. Subsequent studies must employ better anchors with stronger associations to establish more definitive thresholds.

While the sample was representative of the community-dwelling older adult population in the geographic area, there were very few participants who had a walking speed associated with healthy aging and unlimited community ambulation (ie, ≥ 1.20 m/s) (8,22,35). Likewise, we had very few individuals who perceived their global mobility and/or balance as “poor.” Ideally, participant representation across the full ranges of mobility and balance is desired to develop the most externally valid estimates. Nevertheless, our study contained participants with largely varying degrees of mobility, giving us confidence that are estimates are likely accurate. In addition, while we excluded individuals with moderate-to-severe cognitive impairment, it is possible that more mild forms of cognitive impairment went undetected. This condition may influence the reliability of self-reported outcomes.

Furthermore, although we posit that global state questions are acceptable and serve as good anchors for these analyses, other anchors may also be appropriate. The choice of anchor must be dependent upon the research question of focus and the desired clinical application of the findings. An anchor that captures the unique aspects of the F8WT may have been desirable, but gold standard dichotomous

references for motor skill in walking are lacking. In addition, a downstream health outcome (eg, falling) may also serve as a suitable anchor. Practically, however, we were limited by available data in the parent study for selecting suitable anchors. Selecting a downstream health outcome would not have only reduced the sample size of our analyses, but also limited our ability to validate our findings, thereby decreasing the generalizability of this study.

The automated phone calling system is subject to self-report bias and lower compliance compared to more objective methods, such as analyzing health record information. Indeed, approximately 25% of the full baseline sample did not have adequate follow-up information on distal health outcomes. That said, it should be noted that those excluded did not differ from those included in the analytic subsample on demographic and health characteristics. In addition, obtaining and analyzing health record information is burdensome and costly. Automated phone calling has been found to be reliable, valid, and acceptable by older adults to capture these outcomes (20), and it is ideal for studies wherein they are not the central focus.

Finally, although the employed strategies of minimizing index union and equally maximizing sensitivity and specificity are reasonable, they do assume equal costs for false negatives and false positives. Other approaches such as determining a high sensitivity cut-point for screening and high specificity cut-point for ruling in/out may be more practically relevant, but such more definitive, extended findings are best based on anchors with stronger associations with F8WT.

Future Directions

Our study is among the first to investigate the psychometric utility of the F8WT. In future studies, it would be interesting to investigate the relative contribution F8WT performance has beyond straight-path walking speed, in predicting adverse health outcomes. As one might expect, there is considerable overlap between the two measures: prior work has shown Pearson R-value correlations between F8WT performance and walking speed in the range of 0.50–0.57 (10); in our study, correlations between F8WT performance and walking speed ranged from 0.64 to 0.75. Yet, previous findings suggest that the F8WT, independent of gait speed, is associated with activity restriction, gait efficacy, gait variability, and navigational planning during walking (10,12). Therefore, it stands to reason that the F8WT may add to the prediction of adverse health outcomes, such as falling, ED visits, and hospitalizations; however, these hypotheses were beyond the scope of this study, as well as the capability of our dataset (ie, lack of sufficient statistical power to adequately investigate these questions). Future work should investigate the potential added value of the F8WT beyond walking speed alone, in well-designed (ie, appropriately powered) prospective cohort studies.

Conclusions

For the F8WT, performance times of ≤ 9.09 and ≤ 9.27 seconds can discriminate those with self-reported excellent and excellent/very good global mobility, respectively. A total number of steps ≤ 17 on the F8WT discriminates those with excellent/very good/good global balance. These performance classifications have predictive validity, as performing well is associated with an approximately 30%–60% lower risk of negative outcomes (ie, falling and health care utilization). Clinicians may use these cut-points as preliminary guides to aid in their clinical decision making to classify patients with mobility and balance limitations that increase the risk of poor outcomes.

Further study is needed, however, to establish more definitive thresholds, based on anchors with stronger associations.

Funding

This work was supported by the Patient-Centered Outcomes Research Institute (award number CE-1304–6301); the National Institute on Aging (grant number K24 AG057728, P30 AG024827).

Conflict of Interest

S.P., a coauthor of this manuscript, would like to disclose that he received salary support in the past from an unrelated osteoporosis grant to the University of Pittsburgh from Eli Lilly & Co; otherwise, the authors have no other disclosures to report.

References

1. Fritz S, Lusardi M. Whitepaper: “walkingspeed: the sixth vital sign”. *J Geriatr Phys Ther*. 2009;32:46–49. doi: [10.1519/00139143-200932020-00002](https://doi.org/10.1519/00139143-200932020-00002)
2. Middleton A, Fritz SL, Lusardi M. Walking speed: the functional vital sign. *J Aging Phys Act*. 2015;23:314–322. doi: [10.1123/japa.2013-0236](https://doi.org/10.1123/japa.2013-0236)
3. Ferrucci L, Cooper R, Shardell M, Simonsick EM, Schrack JA, Kuh D. Age-related change in mobility: perspectives from life course epidemiology and geroscience. *J Gerontol A Biol Sci Med Sci*. 2016;71:1184–1194. doi: [10.1093/gerona/glw043](https://doi.org/10.1093/gerona/glw043)
4. Guralnik JM, Ferrucci L, Pieper CF, et al. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55:M221–M231. doi: [10.1093/gerona/55.4.m221](https://doi.org/10.1093/gerona/55.4.m221)
5. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995;332:556–561. doi: [10.1056/NEJM199503023320902](https://doi.org/10.1056/NEJM199503023320902)
6. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49:M85–M94. doi: [10.1093/geronj/49.2.m85](https://doi.org/10.1093/geronj/49.2.m85)
7. Perera S, Patel KV, Rosano C, et al. Gait speed predicts incident disability: a pooled analysis. *J Gerontol A Biol Sci Med Sci*. 2016;71:63–71. doi: [10.1093/gerona/glv126](https://doi.org/10.1093/gerona/glv126)
8. Studenski S, Perera S, Patel K, et al. Gait speed and survival in older adults. *JAMA*. 2011;305:50–58. doi: [10.1001/jama.2010.1923](https://doi.org/10.1001/jama.2010.1923)
9. Patla AE, Shumway-Cook A. Dimensions of mobility: defining the complexity and difficulty associated with community mobility. *J Aging Phys Act*. 1999;7:7–19. doi: [10.1123/japa.7.1.7](https://doi.org/10.1123/japa.7.1.7)
10. Hess RJ, Brach JS, Piva SR, VanSwearingen JM. Walking skill can be assessed in older adults: validity of the Figure-of-8 Walk Test. *Phys Ther*. 2010;90:89–99. doi: [10.2522/ptj.20080121](https://doi.org/10.2522/ptj.20080121)
11. Odonkor CA, Thomas JC, Holt N, et al. A comparison of straight- and curved-path walking tests among mobility-limited older adults. *J Gerontol A Biol Sci Med Sci*. 2013;68:1532–1539. doi: [10.1093/gerona/glt060](https://doi.org/10.1093/gerona/glt060)
12. Lowry KA, Brach JS, Nebes RD, Studenski SA, VanSwearingen JM. Contributions of cognitive function to straight- and curved-path walking in older adults. *Arch Phys Med Rehabil*. 2012;93:802–807. doi: [10.1016/j.apmr.2011.12.007](https://doi.org/10.1016/j.apmr.2011.12.007)
13. Welch SA, Ward RE, Kurlinski LA, et al. Straight and curved path walking among older adults in primary care: associations with fall-related outcomes. *PM R*. 2016;8:754–760. doi: [10.1016/j.pmrj.2015.12.004](https://doi.org/10.1016/j.pmrj.2015.12.004)
14. Hibbs R, VanSwearingen J, Studenski S, Brach J. *Curved Path Walking Predicts 12-Month Physical Function in Older Adults Without Mobility Limited Physical Function*. Seattle, WA: American Geriatric Society Annual Meeting; 2012.
15. Brach JS, Perera S, Gilmore S, et al. Effectiveness of a timing and coordination group exercise program to improve mobility in community-dwelling

- older adults: a randomized clinical trial. *JAMA Intern Med.* 2017;177:1437–1444. doi: [10.1001/jamainternmed.2017.3609](https://doi.org/10.1001/jamainternmed.2017.3609)
16. Brach JS, Perera S, Gilmore S, et al. Stakeholder involvement in the design of a patient-centered comparative effectiveness trial of the “On the Move” group exercise program in community-dwelling older adults. *Contemp Clin Trials.* 2016;50:135–142. doi: [10.1016/j.cct.2016.08.003](https://doi.org/10.1016/j.cct.2016.08.003)
 17. Rigler SK, Studenski S, Wallace D, Reker DM, Duncan PW. Co-morbidity adjustment for functional outcomes in community-dwelling older adults. *Clin Rehabil.* 2002;16:420–428. doi: [10.1191/0269215502cr515oa](https://doi.org/10.1191/0269215502cr515oa)
 18. Brach JS, Perera S, Studenski S, Newman AB. The reliability and validity of measures of gait variability in community-dwelling older adults. *Arch Phys Med Rehabil.* 2008;89:2293–2296. doi: [10.1016/j.apmr.2008.06.010](https://doi.org/10.1016/j.apmr.2008.06.010)
 19. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care.* 1992;30:473–483. doi: [10.1097/00005650-199206000-00002](https://doi.org/10.1097/00005650-199206000-00002)
 20. Albert SM, King J, Keene RM. Assessment of an interactive voice response system for identifying falls in a statewide sample of older adults. *Prev Med.* 2015;71:31–36. doi: [10.1016/j.ypmed.2014.12.006](https://doi.org/10.1016/j.ypmed.2014.12.006)
 21. Unal I. Defining an optimal cut-point value in ROC analysis: an alternative approach. *Comput Math Methods Med.* 2017;2017:3762651. doi: [10.1155/2017/3762651](https://doi.org/10.1155/2017/3762651)
 22. Perry J, Garrett M, Gronley JK, Mulroy SJ. Classification of walking handicap in the stroke population. *Stroke.* 1995;26:982–989. doi: [10.1161/01.str.26.6.982](https://doi.org/10.1161/01.str.26.6.982)
 23. Studenski S. Bradypedia: is gait speed ready for clinical use? *J Nutr Health Aging.* 2009;13:878–880. doi: [10.1007/s12603-009-0245-0](https://doi.org/10.1007/s12603-009-0245-0)
 24. Cesari M, Kritchevsky SB, Penninx BW, et al. Prognostic value of usual gait speed in well-functioning older people—results from the Health, Aging and Body Composition Study. *J Am Geriatr Soc.* 2005;53:1675–1680. doi: [10.1111/j.1532-5415.2005.53501.x](https://doi.org/10.1111/j.1532-5415.2005.53501.x)
 25. Studenski S, Perera S, Wallace D, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc.* 2003;51:314–322. doi: [10.1046/j.1532-5415.2003.51104.x](https://doi.org/10.1046/j.1532-5415.2003.51104.x)
 26. Montero-Odasso M, Schapira M, Soriano ER, et al. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *J Gerontol A Biol Sci Med Sci.* 2005;60:1304–1309. doi: [10.1093/gerona/60.10.1304](https://doi.org/10.1093/gerona/60.10.1304)
 27. Middleton A, Fulk GD, Beets MW, Herter TM, Fritz SL. Self-selected walking speed is predictive of daily ambulatory activity in older adults. *J Aging Phys Act.* 2016;24:214–222. doi: [10.1123/japa.2015-0104](https://doi.org/10.1123/japa.2015-0104)
 28. Stanaway FF, Gnjdjic D, Blyth FM, et al. How fast does the Grim Reaper walk? Receiver operating characteristics curve analysis in healthy men aged 70 and over. *BMJ.* 2011;343:d7679. doi: [10.1136/bmj.d7679](https://doi.org/10.1136/bmj.d7679)
 29. Abellan van Kan G, Rolland Y, Andrieu S, et al. Gait speed at usual pace as a predictor of adverse outcomes in community-dwelling older people: an International Academy on Nutrition and Aging (IANA) Task Force. *J Nutr Health Aging.* 2009;13:881–889. doi: [10.1007/s12603-009-0246-z](https://doi.org/10.1007/s12603-009-0246-z)
 30. Abellan van Kan G, Rolland Y, Gillette-Guyonnet S, et al. Gait speed, body composition, and dementia. The EPIDOS-Toulouse cohort. *J Gerontol A Biol Sci Med Sci.* 2012;67:425–432. doi: [10.1093/gerona/glr177](https://doi.org/10.1093/gerona/glr177)
 31. Inzitari M, Newman AB, Yaffe K, et al. Gait speed predicts decline in attention and psychomotor speed in older adults: the health aging and body composition study. *Neuroepidemiology.* 2007;29:156–162. doi: [10.1159/000111577](https://doi.org/10.1159/000111577)
 32. Rantanen T, Guralnik JM, Ferrucci L, et al. Coimpairments as predictors of severe walking disability in older women. *J Am Geriatr Soc.* 2001;49:21–27. doi: [10.1046/j.1532-5415.2001.49005.x](https://doi.org/10.1046/j.1532-5415.2001.49005.x)
 33. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–415. doi: [10.1016/0197-2456\(89\)90005-6](https://doi.org/10.1016/0197-2456(89)90005-6)
 34. Shimada H, Suzuki T, Suzukawa M, et al. Performance-based assessments and demand for personal care in older Japanese people: a cross-sectional study. *BMJ Open.* 2013;3:e002424. doi: [10.1136/bmjopen-2012-002424](https://doi.org/10.1136/bmjopen-2012-002424)
 35. Salbach NM, O’Brien K, Brooks D, et al. Speed and distance requirements for community ambulation: a systematic review. *Arch Phys Med Rehabil.* 2014;95:117–128.e11. doi: [10.1016/j.apmr.2013.06.017](https://doi.org/10.1016/j.apmr.2013.06.017)
 36. Wong SS, Yam MS, Ng SS. The Figure-of-Eight Walk test: reliability and associations with stroke-specific impairments. *Disabil Rehabil.* 2013;35:1896–1902. doi: [10.3109/09638288.2013.766274](https://doi.org/10.3109/09638288.2013.766274)
 37. Schnitker J, Bacak V. The increasing predictive validity of self-rated health. *PLoS One.* 2014;9:e84933. doi: [10.1371/journal.pone.0084933](https://doi.org/10.1371/journal.pone.0084933)
 38. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol.* 2007;60:34–42. doi: [10.1016/j.jclinepi.2006.03.012](https://doi.org/10.1016/j.jclinepi.2006.03.012)
 39. Akobeng AK. Understanding diagnostic tests 3: receiver operating characteristic curves. *Acta Paediatr.* 2007;96:644–647. doi: [10.1111/j.1651-2227.2006.00178.x](https://doi.org/10.1111/j.1651-2227.2006.00178.x)
 40. Apfel CC, Kranke P, Greim CA, Roewer N. What can be expected from risk scores for predicting postoperative nausea and vomiting? *Br J Anaesth.* 2001;86:822–827. doi: [10.1093/bja/86.6.822](https://doi.org/10.1093/bja/86.6.822)