

Motor assessment using the NIH Toolbox

David B. Reuben, MD
Susan Magasi, PhD
Heather E. McCreath,
PhD
Richard W. Bohannon,
EdD, PT
Ying-Chih Wang, PhD,
OT
Deborah J. Bubela, PhD,
PT
William Z. Rymer, MD,
PhD
Jennifer Beaumont, MS
Rose Marie Rine, PT,
PhD
Jin-Shei Lai, PhD,
OTR/L
Richard C. Gershon, PhD

Correspondence to
Dr. Reuben:
dreuben@mednet.ucla.edu

ABSTRACT

Motor function involves complex physiologic processes and requires the integration of multiple systems, including neuromuscular, musculoskeletal, and cardiopulmonary, and neural motor and sensory-perceptual systems. Motor-functional status is indicative of current physical health status, burden of disease, and long-term health outcomes, and is integrally related to daily functioning and quality of life. Given its importance to overall neurologic health and function, motor function was identified as a key domain for inclusion in the NIH Toolbox for Assessment of Neurological and Behavioral Function (NIH Toolbox). We engaged in a 3-stage developmental process to: 1) identify key subdomains and candidate measures for inclusion in the NIH Toolbox, 2) pretest candidate measures for feasibility across the age span of people aged 3 to 85 years, and 3) validate candidate measures against criterion measures in a sample of healthy individuals aged 3 to 85 years ($n = 340$). Based on extensive literature review and input from content experts, the 5 subdomains of dexterity, strength, balance, locomotion, and endurance were recommended for inclusion in the NIH Toolbox motor battery. Based on our validation testing, valid and reliable measures that are simultaneously low-cost and portable have been recommended to assess each subdomain, including the 9-hole peg board for dexterity, grip dynamometry for upper-extremity strength, standing balance test, 4-m walk test for gait speed, and a 2-minute walk test for endurance. *Neurology*® 2013;80 (Suppl 3):S65–S75

GLOSSARY

FRSTST = 5-repetition sit-to-stand test; **ICC** = intraclass correlation coefficient; **RFI** = request for information; **6MWT** = 6-minute walk test.

Motor function, the ability to use and control muscles and movements, is integrally related to daily functioning and quality of life. Motor function is a complex physiologic process and requires the integration of multiple inputs and systems, including the neuromuscular, neurosensory, musculoskeletal, and cardiopulmonary systems. Impairments in motor function are often indicative of disorders of the central or peripheral nervous systems and can lead to increased risk of activity limitations, participation restrictions, and mortality in people across the lifespan.^{1,2} Accordingly, assessment of motor function was included as a major domain of the NIH Toolbox Assessment of Neurological and Behavioral Function (NIH Toolbox)—an initiative of the NIH Blueprint for Neuroscience Research that seeks to develop a brief but comprehensive set of standard measures of motor, cognitive, sensory, and emotional function across the age span from 3 to 85 years³ for use in epidemiologic, longitudinal, and clinical research.

Assessing change in motor function across the lifespan is a complex measurement task. Precise measures of motor function frequently require specialized testing environments and costly equipment.⁴ Performance-based measures of motor function have demonstrated prognostic and diagnostic value at the individual and population levels^{1,2,5}; however, there is a lack of consistency in how motor function is measured across studies and across the age span that limits the generalizability of research findings. Hence, it is essential to develop a brief, valid, and reliable assessment tool that can be used

From the Division of Geriatrics (D.B.R., H.E.M.), David Geffen School of Medicine, University of California, Los Angeles, CA; Department of Medical Social Sciences (S.M., W.Z.R., J.B., J.-S.L., R.C.G.), Northwestern University Feinberg School of Medicine, Chicago, IL; Program in Physical Therapy (R.W.B., D.J.B.), Department of Kinesiology, University of Connecticut, Storrs, CT; Department of Occupational Science and Technology (Y.-C.W.), University of Wisconsin, Milwaukee; Sensory Motor Processing Program (W.Z.R.), Rehabilitation Institute of Chicago, IL; and Specialty Therapy Source, LLC (R.M.R.), Jacksonville, FL.

Go to Neurology.org for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

across diverse study designs and populations and can also characterize developmental and lifespan trajectories.

The development of the motor domain of the NIH Toolbox project occurred in 3 stages. First, an inventory of specific areas of motor function (subdomains) was created and rated by experts for importance and relevance. Second, a pretesting phase was conducted to evaluate the feasibility of candidate measures across the age span using a convenience sample of healthy individuals aged 3 to 85 years, with oversampling in the oldest and youngest age cohorts. Finally, a validation phase was conducted during which the most promising candidate measures in each subdomain were compared against criterion measures in a convenience sample of healthy individuals, stratified by age.

METHODS **Standard protocol approvals, registrations, and patient consents.** IRB approval was obtained at NorthShore University HealthSystem; University of California, Los Angeles; University of Connecticut; and Northwestern University. All participants or their guardians were consented using age-appropriate procedures.

Phase 1: Identifying subdomains. The subdomain structure was developed based on extensive literature reviews; results from an online request for information (RFI) field survey of 147 researchers with experience in conducting large-cohort preclinical and clinical studies (described in more detail within this supplement), 62 of whom self-identified as having sufficient knowledge of motor function to provide rankings; and in-depth interviews with 9 motor-function experts. Through this process, 8 subdomains were identified as the most relevant areas of motor functioning: locomotion (84% by RFI respondents and 92% by experts, respectively), strength (65%, 67%), coordination (60%, 36%), balance (56%, 50%), endurance (53%, 58%), dexterity (47%, 46%), hand and upper-extremity function (53%, 92%), and flexibility (29%, 8%). At a follow-up consensus meeting involving NIH staff, an external group of consultants, and the NIH Toolbox Steering Committee, participants reviewed the RFI results to identify and rank key components/constructs to include for measurement of motor function both conceptual and clinical perspectives. The following subdomains were identified:

Dexterity: the ability to coordinate the fingers and manipulate objects in a timely manner.

Strength: the capacity of a muscle to produce the force and power necessary for maintaining posture, initiating movement, or controlling movement during conditions of loading the musculoskeletal system.

Balance: the ability to orient body parts in space and maintain an upright posture under both static and dynamic conditions, and following internal or external disturbances. Methods and results from balance testing and validation are reported by Rine et al. in this supplement.

Locomotion: the act of moving the body from one place to another, reflecting ambulation ability including walking distance, velocity, and quality of gait with different environments and ground surfaces. The NIH Toolbox focused on walking gait speed across a flat surface.

Endurance: the ability to sustain effort that requires conjoint work capacities from cardiopulmonary, biomechanical, and neuromuscular function. This measure focuses on overall physical fitness instead of individual muscle endurance.

Phase 2: Pretesting phase. Several versions of the subdomain tasks were pretested for feasibility, with emphasis on evaluating performance among children and older adults (table 1). Tests of dexterity and strength were conducted at one site (the Rehabilitation Institute of Chicago) including 51 participants aged 4 to 78 years (mean = 30, SD = 23), whereas locomotion and endurance were pretested at another site (University of California, Los Angeles) with 70 participants aged 3 to 85 years. For locomotion and endurance, English and Spanish instruction sets were evaluated.

Dexterity. Five pegboard tests were evaluated: 1) Rolyan 9-hole pegboard (Sammons Preston Rolyan, Bolingbrook, IL); 2) 9-hole wooden pegboard; 3) Lafayette 25-hole pegboard (Lafayette Instrument Company, Lafayette, IN); 4) Lafayette 25-hole grooved pegboard; and 5) a custom-built, 9-hole electronic pegboard (designed at the Rehabilitation Institute of Chicago).

Strength: upper extremity. Four grip dynamometers were evaluated: 1) Jamar hydraulic hand dynamometer (Sammons Preston Rolyan); 2) J-Tech Commander GripTrack hand dynamometer (J-Tech Medical, Salt Lake City, UT); 3) Lafayette Instrument's 100-kg/220-lb Hand Grip Dynamometer; and 4) Lafayette Instrument's 50-kg/110-lb Hand Grip Dynamometer.

Strength: lower extremity. Two tasks were evaluated: 1) knee-extension strength measured by handheld dynamometer (Lafayette Manual Muscle Testing System); and 2) 5-repetition sit-to-stand test (FRSTST). Several versions of knee-extension strength were evaluated, including the traditional method with the administrator opposing the extension force and a variety of methods with the dynamometer externally stabilized by belts or cables.

Locomotion. Three forms of gait speed were evaluated: 1) 20-ft walk using a straight course, 2) 4-m walk using a straight course, and 3) 20-ft walk using an out-and-back course (U-turn at 10-ft distance).

Endurance. Four forms of endurance tasks were evaluated: 1) 6-minute walk (100-ft course), 2) 6-minute walk (50-ft course), 3) 2-minute walk (50-ft course), and 4) 3-minute step test.

Most tasks were deemed feasible for participants to perform (table 1). In addition to information provided by administrators, results were reviewed with an expert panel to select a subset of these instruments for the validation phase.

Two pegboards, wooden and electronic, presented problems with standardization and reliability, respectively. Although both the 25-hole pegboard and 25-hole grooved pegboard performed similarly, the grooved version was selected for validation because of the additional challenge posed by the unique shape of the pegs enabling observation of the influence of participants' perceptual and fine motor coordination abilities on dexterity. Upon recommendation of the pediatric experts, the 9-hole pegboard was also included in validation because it is considered to be adequately challenging for children.

The Jamar Plus Digital Grip Dynamometer (Sammons Preston Rolyan) was recommended for validation testing because it provided the best balance of price and performance for upper-extremity strength, and the digital display facilitated quick and accurate readings of peak force, reducing the risk of administrator errors. To measure lower-extremity strength, knee-extension force using a belt-stabilized, handheld dynamometer was deemed the best combination of price and performance. The MicroFET2 handheld dynamometer (Hoggan Health Industries, West Jordan, UT) was recommended for validation testing because its low-profile ergonomic design made it easier to stabilize the device during testing. Although concerns were raised that

Table 1 Candidate motor measures by subdomain and level of evaluation

Dexterity						
Measure	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure	
Roylan 9-hole pegboard	Time in seconds	8 Children	Feasible across the age span	1 Practice plus 2 timed trials per hand	Children (3-13 y): Bruininks-Oseretsky Test of Motor Proficiency II-Dexterity Module	
		10 Adults	May lack sensitivity (proceed to validation)		Adults (14-85 y): Purdue Pegboard	
9-Hole pegboard (wooden)	Time in seconds	3 Children	Results similar to Roylan pegboard but lacks standardized equipment placement (do not proceed to validation)			
		3 Adults				
9-Hole pegboard (electronic)	Time in seconds	3 Children				
		3 Adults	Device lacked reliability (do not proceed to validation)			
Grooved pegboard	Time in seconds	6 Children	Feasible across age span; too difficult for youngest children (proceed to validation)	1 Practice plus 2 timed trials per hand	Children (3-13 y): Bruininks-Oseretsky Test of Motor Proficiency II-Dexterity Module	
		10 Adults		Children 3-9 y: complete 2 rows (10 pegs)	Adults (14-85 y): Purdue Pegboard	
					10-85 y: complete 5 rows (25 pegs)	
25-Hole pegboard	Time in seconds	8 Children	Feasible across age span; results similar to grooved pegboard (do not proceed to validation)			
		10 Adults				
Strength: Upper extremity						
Measure	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure	
Grip dynamometer (Jamar Plus)	Grip force in lbs	5 Children	Feasible across age span	1 Submaximal practice and 2 test trials per hand	Children (3-13 y): Bruininks-Oseretsky Test of Motor Proficiency II-Strength Module	
		10 Adults	Digital display reduces administrator error (proceed to validation)		Adults (14-85 y): Biodex System 3 Isokinetic Dynamometer	
Griptrack dynamometer (J-Tech Commander)	Grip force in lbs	6 Children	Feasible in all but youngest children			
		8 Adults	Cost incompatible with NIH Toolbox mandates (do not proceed to validation)			
Grip dynamometer (Lafayette 78010)	Grip force in lbs	10 Adults	Not appropriate for use in children, separate devices would be required for use in adults making cost incompatible with NIH Toolbox mandates (do not proceed to validation)			
Grip dynamometer (Lafayette 78011)	Grip force in lbs	8 Children	Maximum grip force of 50 lbs, separate devices required for children and adults making cost incompatible with NIH Toolbox mandates (do not proceed to validation)			
		10 Adults				
Strength: Lower extremity						
Instrument	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure	
Isometric knee extension via belt-stabilized, handheld dynamometer (MicroFET2)	Isometric knee-extension force in lbs	3 Children	Feasible across the age span, design appropriate for belt stabilization	1 Submaximal practice and 2 test trials per lower leg	Children (3-13 y): Bruininks-Oseretsky Test of Motor Proficiency II-Strength Module	
		3 Adults	Digital display (proceed to validation)		Adults (14-85 y): Biodex System 3 Isokinetic Dynamometer	
Isometric knee extension via belt-stabilized, handheld dynamometer (Lafayette)	Isometric knee-extension force in lbs	3 Children	High profile and location of testing pad make it inappropriate for belt-stabilized testing (do not proceed to validation)			
		3 Adults				

Table 1 Continued

Strength: Lower extremity					
Instrument	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure
5-Repetition sit-to-stand test	Time in seconds	2 Children	Feasible for adults	1 Practice plus 2 timed trials	Children (3–13 y): Bruininks-Oseretsky Test of Motor Proficiency II–Strength Module
		8 Adults	Children require more training to understand task (proceed to validation)		Adults (14–85 y): Biodex System 3 Isokinetic Dynamometer
Locomotion					
Instrument	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure
Fast walking gait speed (20-ft course)	Time in seconds	10 Children	Feasible across the age span (proceed to validation)	1 Practice and 2 test trials	No clear criterion measure
		12 Adults			Comparative validity and reliability of candidate measures evaluated
Fast walking gait speed (4-m course)	Time in seconds	14 Children	Feasible across the age span (proceed to validation)	1 Practice and 2 test trials	No clear criterion measure
		10 Adults			Comparative validity and reliability of candidate measures evaluated
Fast walking gait speed (20-ft out-and-back course)	Time in seconds	10 Children	Feasible across the age span (proceed to validation)	1 Practice and 2 test trials	No clear criterion measure
		12 Adults			Comparative validity and reliability of candidate measures evaluated
Endurance					
Instrument	Outcome measure	Pretesting sample size	Pretesting results (decision)	Validation testing conditions	Validation criterion measure
6-min walk test (100-ft course)	Distance walked in 6 min	10 Children	Feasible across the age span		
		9 Adults			
6-min walk test (50-ft course)	Distance walked in 6 min	8 Children	Feasible across the age span (proceed to validation)	Single trial, distance recorded at 2-, 4-, and 6-min intervals	6-min walk test
		9 Adults			2-, 4-min intervals were evaluated
2-min walk test (50-ft course)	Distance walked in 2 min	9 Children	Feasible across the age span (proceed to validation)	Single trial, distance recorded at 2 min	6-min walk test (first 2-min interval)
		10 Adults (English and Spanish)			Comparative validity and reliability were evaluated
3-min step test	1-min recovery heart rate (total beats)	8 Children	Feasible across the age span (proceed to validation)	Cadence: <13 and >65 y: 88 steps/min, 14–64 y: 96 steps/min; Step height: 8 in.; people <57 in. tall, 12 in.; people ≥58 in. tall	6-min walk test
		8 Adults			

timed-chair rise involved more dimensions than lower-extremity strength, the experts encouraged the retention of FRSTST for validation because of the low cost, common use by clinicians (particularly in geriatrics), and functional relevance.⁶

All locomotion tasks were retained as candidate measures, even though performance was similar, because the 20-ft distance is favored in US studies and the 4-m distance is favored in non-US studies. For endurance, the 100-ft course was deemed too long for the variety of settings for which the NIH Toolbox battery is designed. A hybrid version of the 6-minute, 50-ft course was developed and tested to assess performance at each 2-minute interval. The 3-minute step was feasible for most participants

and had the advantage of prior research validating it with traditional exercise stress tests.^{7,8}

Phase 3: Validation phase. A convenience sample of English-speaking participants who did not use assistive devices for walking was recruited for the validation study according to the age (years) bands 3–4, 5–6, 7–10, 11–13, 14–20, 21–30, 31–45, 46–65, 66–75, and 76–85. For all age bands (except 76–85 years), 30 participants were recruited, with some age bands having more than 30. Recruitment targeted those in the youngest and oldest age ranges to allow adequate evaluation of tasks with those most likely to demonstrate difficulty with tasks.

Tasks were organized by subdomain, and order of administration within each subdomain was randomized. Subdomains were ordered from fine to gross motor with the exception of endurance. Endurance in final NIH Toolbox testing will be performed last because it imposes the greatest energy demand on the participant and may be expected to affect performance in other motor tasks, particularly among people with lower motor-functional capacity. Tasks, test conditions, outcomes, and validity measures are presented in table 1.

Reliability assessment. A subset of participants (approximately 5 per age band) were invited for a second visit within 7 to 21 days of the first visit. At the second visit, participants completed all of the candidate measures but not the validation measures.

Statistical methods. Distributional characteristics were examined for each measure (mean, SD, range) overall and by age band. To determine comparative validity of the measures, Pearson correlation coefficients (r) were computed between all candidate and validation tasks within a subdomain. For test-retest reliability, intraclass correlation coefficients (ICCs) were calculated for values from visits 1 and 2 for each measure. A priori criteria for validity and test-retest were established based on standards in the field of motor-functional testing as $r \geq 0.70$ and $ICC > 0.80$, respectively.^{9,10}

RESULTS A total of 340 participants were enrolled in the validation study across 3 testing sites. Sociodemographic characteristics are provided in table 2. The mean age of the sample was 30.7 years, 19% were 3 to 6 years, 23% were 7 to 13 years, and 19% were 66 to 85 years; 54% were female; 36% were non-white; and 29% had an annual household income of less than \$40,000.

The following numbers of participants (in parentheses) experienced difficulty complying with administration guidelines and/or completing the task demands for the following measures: grooved pegboard (6), timed chair rise (6), 3-minute step test (50), and 6-minute walk test (6MWT) (6).

Test-retest reliability. Test-retest reliability of all candidate measures (table 3) was good with all measures meeting the predefined criterion for $ICC \geq 0.80$, except for the grooved pegboard (left hand) and the 3 locomotion measures. Within each domain, there were no differences among measures/instruments in reliability.

Validity. Validity of the candidate measures was generally good and met the predefined criterion ($r > 0.70$) of Pearson correlations with criterion measures (table 4) for all measures except for grooved pegboard as a measure of dexterity in children younger than 14 years, timed chair rise as a measure of strength across age groups, and the 3-minute step test as a measure of endurance across all ages. The 2-minute walk test and 4-minute walk test correlated exceptionally highly ($r > 0.96$) with the 6MWT.

DISCUSSION This study evaluated the conceptual and clinical relevance of key aspects of motor function (subdomain) and established the preliminary validity and reliability of candidate measures of motor function

Table 2 Demographic characteristics of validation sample (n = 340)

	No. (%)
Site	
RIC	121 (35.6)
UConn	81 (23.8)
UCLA	138 (40.6)
Sex (n = 2 missing)	
Female	182 (53.8)
Age bands, y (n = 6 missing)	
3-6	64 (19.2)
7-13	75 (22.5)
14-20	33 (9.9)
21-65	99 (29.6)
66-85	63 (18.9)
Hispanic or Latino? (n = 2 missing)	
Yes	40 (11.8)
Race	
White	217 (63.8)
Asian	51 (15.0)
Black	29 (8.5)
Other	43 (12.6)
Household income (n = 19 missing)	
<\$20,000	35 (10.9)
\$20,000-\$39,999	57 (17.8)
\$40,000-\$99,999	120 (37.4)
≥\$100,000	93 (29.0)
Don't know	16 (5.0)

Abbreviations: RIC = Rehabilitation Institute of Chicago; UCLA = University of California, Los Angeles; UConn = University of Connecticut.

recommended for inclusion in the motor battery of the NIH Toolbox. The 5 subdomains identified (dexterity, strength, balance, gait speed, and endurance) have public health and scientific importance.

Dexterity is a central component of hand function and relates to both the speed and accuracy of hand movements during the manipulation of objects in the environment.¹¹ Dexterity in children is correlated with school performance and is a predictor of quality of handwriting,¹² which, in turn, is associated with upper-limb coordination and some cognitive functions (e.g., visual memory, spatial relation, visual sequential memory, visual-motor integration, and sustained attention¹³). Fine-motor-skill deficits are correlated with global psychosocial functioning in both children and adults.¹⁴⁻¹⁷ Decline in manual dexterity is a common phenomenon in older adults and is associated with performance of activities of daily living and independent living.^{18,19}

Table 3 Test-retest reliability statistics

Domain	Measure/instrument	No.	ICC (95% CI)
Dexterity	9-Hole pegboard, right	54	0.87 (0.79-0.92)
	9-Hole pegboard, left	54	0.83 (0.73-0.90)
	Grooved pegboard, right	50	0.83 (0.72-0.90)
	Grooved pegboard, left	52	0.78 (0.65-0.87)
Strength, upper extremity	Grip strength, right	54	0.88 (0.80-0.93)
	Grip strength, left	54	0.98 (0.97-0.99)
Strength, lower extremity	Knee extension, right	53	0.92 (0.87-0.95)
	Knee extension, left	53	0.93 (0.88-0.96)
	Timed chair rise	52	0.84 (0.74-0.90)
Locomotion	20-ft walk test	53	0.78 (0.65-0.87)
	4-m walk test	54	0.64 (0.45-0.77)
	10-ft out-and-back walk test	52	0.72 (0.56-0.83)
Endurance	2-min walk test	52	0.89 (0.82-0.93)
	4-min walk test	53	0.92 (0.87-0.95)
	6-min walk test	52	0.92 (0.87-0.95)
	3-min step test	53	0.84 (0.74-0.90)

Abbreviations: CI = confidence interval; ICC = intraclass correlation coefficient.

Muscle strength is an essential element for humans to move against gravity and provide sufficient force to perform movements within the full range of motion. Although grip strength has been used to characterize total body strength and predict mortality, postsurgical complications, and future disability,²⁰ some research suggests that it may be more accurate to measure a representative muscle group for each limb (e.g., lower extremity as well as grip strength) if the strength of that limb is of interest.²¹ For example, knee extensor muscles have been identified as representative of lower-extremity strength because of their central role in key functional tasks, such as walking, standing from a chair, and climbing stairs.²²⁻²⁴ Muscle strength of the limbs and trunk declines with age and is associated with an increased risk of falls,²⁵ hip fractures,²⁶ loss of bone mineral density,²⁷⁻³⁰ and long-term survival in severe congestive heart failure,³¹ functional dependence in people aged 75 years,³² and loss of functional status in hospitalized patients.³³

Balance allows humans to be able to orient the body in space, maintain an upright posture under static and dynamic conditions, and move without falling. The ability to respond to internal and external disturbance, realign body segments, as well as protect oneself from falling is essential and inherent in every daily task. Examination of balance is important because it predicts people's ability to safely and independently function in a variety of environments. Maintaining stance stability under varying sensory environments is an essential function for the elderly to avoid falling and among patients for better

functional outcomes.³⁴ Several studies have found that changes in balance ability correlate significantly with changes in function.³⁵⁻³⁹

Gait speed as a measure of bipedal locomotion is both a good way to summarize the overall burden of disease as well as a generic indicator of health status, prognosis, and the comorbid burden of disease in older persons.⁴⁰ The speed at which older individuals walk is relevant to their functioning in the community.^{41,42} Moreover, gait speed is an important predictor of outcomes such as length of stay and discharge disposition of patients admitted for acute rehabilitation after stroke,³⁵ mortality,^{36,37} incident ischemic stroke,³⁸ and incident dementia.³⁹

Cardiorespiratory and muscle endurance are important components of physical fitness and contribute to both performance and health status. People with better endurance are able to complete daily tasks, pursue leisure activities, and accomplish higher-intensity workloads. The clinical significance of endurance, as measured by timed walk tests, to morbidity and mortality outcomes has been extensively reported in healthy and clinical populations across the age span.⁴³⁻⁴⁹

Although the findings of this study advance the field of motor assessment, the limitations of the study must be recognized. The sample size was relatively small and included only a few geographical sites. Test administrators were highly trained, and reproducibility of results with nonexpert administrators (as envisioned by the NIH Toolbox) needs to be evaluated. Simplification of administration and clear safety protocols may be required for administrators who lack clinical expertise and judgment. Moreover, healthy volunteers were used for the feasibility and validation testing. The validation data may not apply to persons with specific diseases. The measures of balance and endurance require standing and walking and may not be appropriate for people with mobility limitations; alternate measures are needed for people who are nonambulatory. Similarly, the visual demands of the dexterity measure make it inappropriate for administration to people with severe visual impairments. Given that a major goal of the NIH Toolbox is to develop a brief but comprehensive assessment battery of motor function and performance, compromises had to be made between brevity and specificity. Furthermore, although the motor battery meets NIH Toolbox requirements for portability and low cost, the equipment costs and space demands are not trivial. Finally, the motor battery does not assess all aspects of motor performance. For the sake of time and cost of administration, choices needed to be made about which subdomains could be included.

Final recommendations for inclusion in the NIH Toolbox Motor Domain are described below. For all measures, we have developed administration protocols that include task description, demonstration,

Table 4 Validity statistics^a

Subdomain	Candidate measure	Criterion measure						
Dexterity		BOT-Dexterity Scale				Purdue Pegboard		
		<9 y		9-13 y		14-85 y		
		n	r	n	r	n	r	
		Grooved pegboard, left	75	-0.65	48	-0.31	193	-0.78
		Grooved pegboard, right	80	-0.74	47	-0.39	192	-0.77
			3-13 y		14-85 y			
		9-Hole pegboard, left	132	-0.81			194	-0.72
		9-Hole pegboard, right	134	-0.87			193	-0.72
	Strength		BOT-Strength Scale				BioDex	
			3-13 y		14-85 y			
		n	r	n	r	n	r	
		Grip strength, right	136	0.77			187	0.78
		Grip strength, left	136	0.75			187	0.77
		Knee extension, right	134	0.70			187	0.86
		Knee extension, left	134	0.69			185	0.86
		Timed chair rise	131	-0.45			183	-0.34
Locomotion			20-ft walk time					
			n	r				
		10-ft out-and-back test, walk time	327	-0.79				
		20-ft walk test, walk time	332	1				
Endurance		6-min walk test						
		n	r					
		2-min walk test	328	0.96				
		4-min walk test	328	0.98				
		3-min step test	314	-0.15				

Abbreviation: BOT = Bruininks-Oseretsky Test of Motor Proficiency.

^aAll $p < 0.05$.

and practice before scored trial (available at www.nihtoolbox.org).

Dexterity. The Rolyan 9-hole pegboard, single trial per hand, is recommended as the dexterity measure for inclusion in the NIH Toolbox. The 9-hole pegboard confers the advantages of acceptable reliability and validity, the ability to be completed by all age groups, and minimization of extraneous functional task demands (especially the visual perceptual components). The grooved pegboard has widespread use in neuropsychological testing and evidence of sensitivity. However, it has greater visual perceptual task demands and required modifications and nonstandard testing protocols to ensure completion in very young children. The grooved pegboard was recommended for further validity testing and norming with the aim of possible

inclusion among a wider group of valid instruments that assess motor domain function (the NIH Toolbox Supplemental Measures).

Strength. Grip strength dynamometry with the Jamar Plus Digital Hand Dynamometer, single trial per hand, is recommended as the strength measure for inclusion in the NIH Toolbox. Grip strength dynamometry confers the advantages of excellent reliability and validity, ease of administration, and documented predictive abilities. It is already well established in epidemiologic research.⁵⁰⁻⁵² Analyses of the validation data indicate that in healthy individuals, grip and knee-extension strength reflect a common underlying construct.⁵³ Although the NIH Toolbox requirements for portability and low cost justified selection of a single measure of strength, the correlation between grip and knee-extension

strength decreases with age and possibly in the presence of neurologic impairments. Therefore, isometric knee-extension force measured by a portable device will be included in the NIH Toolbox Supplemental Measures. Equivalency testing between belt-stabilized, handheld dynamometry of knee-extension force measured using the MicroFET2 handheld dynamometer and an integrated tension-based load cell device showed good reliability and validity across the age span.

The FRSTST has a strong legacy of use in geriatrics research and practice as a quick and portable measure of functional lower-extremity strength. Sit-to-stand tests are composite functional measures that integrate lower-extremity strength, balance, and endurance. Validity statistics indicate that the FRSTST is tapping into underlying constructs other than pure lower-extremity strength. Although not recommended for inclusion in the NIH Toolbox, the FRSTST will be normed for possible inclusion in the NIH Toolbox Supplemental Measures for adults.

Balance. The NIH Toolbox Standing Balance Test is the recommended balance measure for inclusion in the NIH Toolbox. The NIH Toolbox Standing Balance Test (described in detail by Rine et al. in this supplement as the Balance Accelerometry Measure) uses accelerometer measures of anterior-ventral postural sway to evaluate balance under 6 testing conditions (feet together, eyes open; feet together, eyes closed; feet together on foam, eyes open; feet together on foam, eyes closed; tandem stance, eyes open; and tandem stance, eyes closed).

Locomotion. The 4-m walk test, 2 trials (1 at usual pace and 1 “as quickly as possible”), are recommended for inclusion in the NIH Toolbox. All of the candidate locomotion measures had acceptable reliability and validity, but based on practical considerations for portability and the emerging international standard of the 4-m walk as a criterion measure of locomotion, we balanced reliability considerations with practical and logistical consideration to recommend the 4-m walk. The 4-m test will reconcile the NIH Toolbox measure of gait speed with international standards and facilitate testing across a variety of clinical and research settings where space might be limited. In an effort to increase the distribution of scores across the age span, the motor domain team made the decision early on to do all tests “as quickly as possible.” However, after consultation with external reviewers and experts in locomotion testing and a subsequent updated review of the literature, the recommendation has been made to add a trial at the usual pace. In prior research, locomotion testing of usual pace was used twice as often as measuring locomotion as quickly as possible.^{54,55}

Endurance. The 2-minute walk test (2MWT), 1 trial on a 50-ft course, is recommended for inclusion in the NIH Toolbox. The 2MWT was highly correlated with the criterion measure of the 6MWT without compromising validity and reliability. The 2MWT was both easy and safe for administration to people across the age span from 3 to 85 years.

General motor function testing guidelines. The recommended test order, proceeding from upper-extremity to lower-extremity function, is dexterity, upper-extremity strength, standing balance, bipedal locomotion, and endurance. Because many of the motor functions build on each other, especially in the lower extremity, this test order was deemed a safe and efficient means of determining whether the person should proceed for testing in subsequent subdomains. For example, if a person is unable to complete the locomotion test, the examiner should evaluate the feasibility of assessing endurance. In our relatively healthy sample, we had few people who were unable to proceed with testing of all subdomains; therefore, precise safety cut-points have not yet been established.

By its nature, motor function testing has inherent safety risks such as falls, physical discomfort, and cardiovascular concerns. Therefore, safety protocols will be developed that are easily interpreted by lay administrators including safety precautions and cut-points for the discontinuation of both individual tests as well as subsequent motor tasks.

The next step in developing the Motor Domain Battery of the NIH Toolbox was a norming of the recommended measures of dexterity, strength, balance, locomotion, and endurance along with cognitive, sensory, and emotional health batteries in a national sample of healthy individuals aged 3 to 85 years. Additional development was conducted to streamline the administration protocols, including simplification of instructions to participants and administrators to increase efficiency and decrease the demands on participants.

In summary, an NIH Toolbox Motor Domain Battery has been created with 5 subdomains (gait speed, balance, dexterity, strength, and endurance) that have public health and scientific importance. The instruments selected to measure these subdomains are feasible and valid, but practical and inexpensive.

AUTHOR CONTRIBUTIONS

D.B. Reuben and S. Magasi: study concept and design, analysis and interpretation, writing of the manuscript, study supervision. H.E. McCreath: study concept and design, acquisition of data, analysis and interpretation, writing of the manuscript, study supervision. R.W. Bohannon and Y.-C. Wang: study concept and design, acquisition of data, analysis and interpretation, writing of the manuscript, study supervision. D.J. Bubela: study concept and design, acquisition of data, analysis and interpretation, writing of the manuscript. W.Z. Rymer: study concept and design, writing of the manuscript, study

supervision. J. Beaumont: study concept and design, analysis and interpretation, writing of the manuscript. R.M. Rine: study concept and design, acquisition of data, analysis and interpretation, writing of the manuscript, study supervision. J.-S. Lai: study concept and design, writing of the manuscript. R.C. Gershon: study concept and design, study supervision.

ACKNOWLEDGMENT

The authors express appreciation to Phoebe Block, Caroline Marchand, and Michael Jesselson at the Rehabilitation Institute of Chicago, Jessica Crocker, at the University of California, Los Angeles, and Alissa Dall, Briana Hoganson, Nicole LaChance, Aaron Morales, and Katherine Taft at the University of Connecticut for data collection assistance. The authors thank the subdomain consultants and select members of the NIH project team, Suzann Campbell, PhD, Christine Chen, ScD, OTR, Jack Guralnik, MD, PhD, Mark Hallett, MD, Jane Clark, PhD, Allen Heinemann, PhD, William Evans, PhD, Dallas Anderson, PhD, Diane Damiano, PhD, PT, Marjorie Garvey, MD, Elizabeth Moberg-Wolff, MD, for expert consultation. Finally, the authors acknowledge Edward Wang, PhD at Northwestern University for early leadership critical to domain development.

STUDY FUNDING

This study is funded in whole or in part with Federal funds from the Blueprint for Neuroscience Research and Office of Behavioral and Social Sciences Research, NIH under contract no. HHS-N-260-2006-00007-C.

DISCLOSURE

D. Reuben is funded by multiple grants from NIH (P30 AG028748, K12AG001004, R21HS017621, R01AG036776, U01AG034661, R01AG023555), the Archstone Foundation, the John A. Hartford Foundation, the Donald W. Reynolds Foundation, the SCAN Foundation, the UniHealth Foundation, the California Breast Cancer Research Program, and the Association of Directors of Geriatric Academic Programs. He is an academic geriatrician and is chair of the Board of Directors of the American Board of Internal Medicine, and lead author of "Geriatrics at Your Fingertips." S. Magasi is funded by multiple grants and contracts from the NIH (HHS-N-260-2006-00007-C, 1U5AR057951-01, U01 AR 052177), the National Institute of Disability and Rehabilitation Research (H133B090024), Agency on Healthcare Research and Quality (RFA-HS-11-001), with minor funding from Forest-Ironwood Pharmaceuticals and Daiichi-Sankyo, Inc. H. McCreath is funded by multiple grants from NIH (P30AG028748, P30AG017265, R01NR010736, RC1AG035664, R03HS018098, RC2MD004770), DHHS/BHPr (2 D01HP00003, HHS230-200432029C), the Archstone Foundation, the John A. Hartford Foundation, the Donald W. Reynolds Foundation, the UniHealth Foundation, the California HIV Research Program, as well as an investigator-initiated study funded by Merck Pharmaceuticals (MK 05-0518). R. Bohannon is a consultant for Hoggan Health Industries, INC Research, and Celtic Pharma Research and Saebo. He received honoraria from Hoggan Health Industries and Saebo. Y.-C. Wang serves as an Assistant Professor at the University of Wisconsin–Milwaukee, an adjunct Assistant Research Professor at the Feinberg School of Medicine, Northwestern University, a senior data analyst in the Focus on Therapeutic Outcomes, Inc., and serves as an editorial board member of the *American Journal of Occupational Therapy* (AJOT), and received research support from NIH as research scientist (HHS-N-260-2006-00007-C) and NIDRR (H133B080031). D. Bubela did not receive any corporate or government funding. Her research has been supported by the University of Connecticut as part of her faculty appointment. W. Rymer serves as a member of the Scientific Advisory Board of Hocoma International, a rehabilitation technology company, and the Neilsen Foundation, a spinal cord injury not-for-profit. Dr. Rymer has served as a consultant for Allergan, a pharmaceutical company. He is a member of the Editorial Board of the *Journal of Neural Engineering and Rehabilitation* and of the *Archives of Physical Medicine and Rehabilitation*. J. Beaumont served as a consultant for NorthShore University HealthSystems, FACIT.org, and Georgia Gastroenterology Group PC. She received funding for travel as an invited speaker at the North American Neuroendocrine Tumor Symposium. R.M. Rine has been an invited speaker for the British Audiological Society, the Royal Society of Medicine, the Hong Kong

Physical Therapy Association, and the American Physical Therapy Association, currently serves on the Scientific Review Committee of the Foundation for Physical Therapy Research, and was funded by NIH grant 5R44DC8022-3. J.-S. Lai's time is supported by the NIH (HHS-N-260-2006-00007-C; PI: Dr. Gershon) and received research support from Pfizer, Inc. R. Gershon has received personal compensation for activities as a speaker and consultant with Sylvan Learning, Rockman, and the American Board of Podiatric Surgery. He is supported through several grants and contracts awarded by NIH: N01-AG-6-0007, 1U5AR057943, HHSN260200600007, U01DK082342, U01NS056 975, K23HL085766, RC2AG036498; and through the NIDRR: H133B090024. Go to Neurology.org for full disclosures.

Received June 6, 2012. Accepted in final form September 26, 2012.

REFERENCES

1. Gill TM, Williams CS, Mendes de Leon CF, Tinetti ME. The role of change in physical performance in determining risk for dependence in activities of daily living among nondisabled community-living elderly persons. *J Clin Epidemiol* 1997;50:765–772.
2. 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.
3. Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, Wagster MV. Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurol* 2010;9:138–139.
4. Ohtake PJ. Field tests of aerobic capacity for children and older adults. *Cardiopulm Phys Ther J* 2005;16:5–11.
5. Reuben DB, Seeman TE, Keeler E, et al. Refining the categorization of physical functional status: the added value of combining self-reported and performance-based measures. *J Gerontol A Biol Sci Med Sci* 2004;59:1056–1061.
6. Bohannon RW. Reference values for the five-repetition sit-to-stand test: a descriptive meta-analysis of data from elders. *Percept Mot Skills* 2006;103:215–222.
7. Liu CM, Lin KF. Estimation of VO_{2max} : a comparative analysis of post-exercise heart rate and physical fitness index from 3-minute step test. *J Exerc Sci Fit* 2007;5:118–123.
8. Santo AS, Golding LA. Predicting maximum oxygen uptake from a modified 3-minute step test. *Res Q Exerc Sport* 2003;74:110–115.
9. Rikli RE, Jones CJ. Development and validation of a functional fitness test for community-residing older adults. *J Aging Phys Act* 1999;7:129–161.
10. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Prentice Hall; 2000.
11. Yancosek KE, Howell D. A narrative review of dexterity assessments. *J Hand Ther* 2009;22:258–270.
12. Volman MJM, van Schendel BM, Jongmans MJ. Handwriting difficulties in primary school children: a search for underlying mechanisms. *Am J Occup Ther* 2006;60:451–460.
13. Tseng MH, Chow SM. Perceptual-motor function of school-age children with slow handwriting speed. *Am J Occup Ther* 2000;54:83–88.
14. Skinner RA, Piek JP. Psychosocial implications of poor motor coordination in children and adolescents. *Hum Mov Sci* 2001;20:73.
15. Bloch MH, Sukhodolsky DG, Leckman JF, Schultz RT. Fine-motor skill deficits in childhood predict adulthood tic severity and global psychosocial functioning in Tourette's syndrome. *J Child Psychol Psychiatry* 2006;47:551–559.

16. Lehoux C, Everett J, Laplante L, et al. Fine motor dexterity is correlated to social functioning in schizophrenia. *Schizophr Res* 2003;62:269–273.
17. Lee AT, Tsang HW. Validation of the Complete Minnesota Dexterity Test for adults with moderate graded mental retardation in Hong Kong. *Int J Rehabil Res* 2001;24:153–156.
18. Falconer J, Hughes SL, Naughton BJ, Singer R, Chang RW, Sinacore JM. Self report and performance-based hand function tests as correlates of dependency in the elderly. *J Am Geriatr Soc* 1991;39:695–699.
19. Williams ME, Hadler NM, Earp JA. Manual ability as a marker of dependency in geriatric women. *J Chronic Dis* 1982;35:115–122.
20. Bohannon RW. Dynamometer measurements of hand-grip strength predict multiple outcomes. *Percept Mot Skills* 2001;93:323–328.
21. Bohannon RW. Is it legitimate to characterize muscle strength using a limited number of measures? *J Strength Cond Res* 2008;22:166–173.
22. Hughes MA, Myers BS, Schenkman ML. The role of strength in rising from a chair in the functionally impaired elderly. *J Biomech* 1996;29:1509–1513.
23. Hortobágyi T, Mizelle C, Beam S, DeVita P. Old adults perform activities of daily living near their maximal capabilities. *J Gerontol A Biol Sci Med Sci* 2003;58:453–460.
24. Ploutz-Snyder LL, Manini T, Ploutz-Snyder RJ, Wolf DA. Functionally relevant thresholds of quadriceps femoris strength. *J Gerontol A Biol Sci Med Sci* 2002;57:B144–B152.
25. Campbell AJ, Borrie MJ, Spears GF. Risk factors for falls in a community-based prospective study of people 70 years and older. *J Gerontol* 1989;44:112–117.
26. Aniansson A, Zetterberg C, Hedberg M, Henriksson KG. Impaired muscle function with aging: a background factor in the incidence of fractures of the proximal end of the femur. *Clin Orthop Relat Res* 1984;(191):193–201.
27. Sinaki M, McPhee MC, Hodgson SF, Merritt JM, Offord KP. Relationship between bone mineral density of spine and strength of back extensors in healthy postmenopausal women. *Mayo Clin Proc* 1986;61:116–122.
28. Taaffe DR, Pruitt L, Lewis B, Marcus R. Dynamic muscle strength as a predictor of bone mineral density in elderly women. *J Sports Med Phys Fitness* 1995;35:136–142.
29. Snow-Harter C, Boussein M, Lewis B, Charette S, Weinstein P, Marcus R. Muscle strength as a predictor of bone mineral density in young women. *J Bone Miner Res* 1990;5:589–595.
30. Iki M, Saito Y, Kajita E, Nishino H, Kusaka Y. Trunk muscle strength is a strong predictor of bone loss in postmenopausal women. *Clin Orthop Relat Res* 2006;443:66–72.
31. Hülsmann M, Quittan M, Berger R, et al. Muscle strength as a predictor of long-term survival in severe congestive heart failure. *Eur J Heart Fail* 2004;6:101–107.
32. Rantanen T, Avlund K, Suominen H, Schroll M, Frändin K, Pertti E. Muscle strength as a predictor of onset of ADL dependence in people aged 75 years. *Aging Clin Exp Res* 2002;14:10–15.
33. Humphreys J, de la Maza P, Hirsch S, Barrera G, Gattas V, Bunout D. Muscle strength as a predictor of loss of functional status in hospitalized patients. *Nutrition* 2002;18:616–620.
34. Hu MH, Hung YC, Huang YL, Peng CD. Validity of force platform measures for stance stability under varying sensory conditions. *Proc Natl Sci Counc Repub China B* 1996;20:78.
35. Rabadi MH, Blau A. Admission ambulation velocity predicts length of stay and discharge disposition following stroke in an acute rehabilitation hospital. *Neurorehabil Neural Repair* 2005;19:20–26.
36. Ostir GV, Kuo YF, Berges IM, Markides KS, Ottenbacher KJ. Measures of lower body function and risk of mortality over 7 years of follow-up. *Am J Epidemiol* 2007;166:599–605.
37. Rolland Y, Lauwers-Cances V, Cesari M, Vellas B, Pahor M, Grandjean H. Physical performance measures as predictors of mortality in a cohort of community-dwelling older French women. *Eur J Epidemiol* 2006;21:113–122.
38. McGinn AP, Kaplan RC, Verghese J, et al. Walking speed and risk of incident ischemic stroke among postmenopausal women. *Stroke* 2008;39:1233.
39. Waite LM, Grayson DA, Piguot O, Creasey H, Bennett HP, Broe GA. Gait slowing as a predictor of incident dementia: 6-year longitudinal data from the Sydney Older Persons Study. *J Neurol Sci* 2005;229/230:89–93.
40. Studenski S, Perera S, Chandler J, et al. How fast you walk predicts how long you live: pooled data from 34,370 older adults. In: *International Association of Gerontology and Geriatrics XIXth World Congress of Gerontology and Geriatrics*; July 2009; Paris.
41. Robinett CS, Vondran MA. Functional ambulation velocity and distance requirements in rural and urban communities: a clinical report. *Phys Ther* 1988;68:1371–1373.
42. Langlois JA, Wallace RB, Marottoli RA, Foley DJ, Guralnik JM, Keyl PM. Characteristics of older pedestrians who have difficulty crossing the street. *Am J Public Health* 1997;87:393–397.
43. Moriello C, Mayo NE, Feldman L, Carli F. Validating the six-minute walk test as a measure of recovery after elective colon resection surgery. *Arch Phys Med Rehabil* 2008;89:1083–1089.
44. Montgomery PS, Gardner AW. The clinical utility of a six-minute walk test in peripheral arterial occlusive disease patients. *J Am Geriatr Soc* 1998;46:706–711.
45. Pankoff B, Overend T, Lucy D, White K. Validity and responsiveness of the 6 minute walk test for people with fibromyalgia. *J Rheumatol* 2000;27:2666–2670.
46. Schoenmakers MAGC, de Groot JF, Hillaert JLM, Helder PJM, Takken T, Gorter JW. Muscle strength, aerobic capacity and physical activity in independent ambulating children with lumbosacral spina bifida. *Disabil Rehabil* 2009;31:259–266.
47. Kosak M, Smith T. Comparison of the 2-, 6-, and 12-minute walk tests in patients with stroke. *J Rehabil Res Dev* 2005;42:103–108.
48. Jenkins SC. 6-Minute walk test in patients with COPD: clinical applications in pulmonary rehabilitation. *Physiotherapy* 2007;93:175–182.
49. Peeters P, Mets T. The 6-minute walk as an appropriate exercise test in elderly patients with chronic heart failure. *J Gerontol A Biol Sci Med Sci* 1996;51:147–151.
50. Sasaki H, Kasagi F, Yamada M, Fujita S. Grip strength predicts cause-specific mortality in middle-aged and elderly persons. *Am J Med* 2007;120:337–342.
51. Forrest KY, Zmuda JM, Cauley JA. Patterns and correlates of muscle strength loss in older women. *Gerontology* 2007;53:140–147.
52. Rantanen T, Volpato S, Ferrucci L, Heikkinen E, Fried LP, Guralnik JM. Handgrip strength and cause-specific and total mortality in older disabled women: exploring the mechanism. *J Am Geriatr Soc* 2003;51:636–641.

53. Bohannon RW, Magasi S, Bubela DJ, Wang YC, Gershon RC. Grip and knee extension muscle strength reflect a common construct among adults. *Muscle Nerve* 2012;46:555–558.
54. Graham JE, Ostir GV, Kuo YF, Fisher SR, Ottenbacher KJ. Relationship between test methodology and mean velocity in timed walk tests: a review. *Arch Phys Med Rehabil* 2008; 89:865–872.
55. Bohannon RW. Comfortable and maximum walking speed of adults aged 20–79 years: reference values and determinants. *Age Ageing* 1997;26:15–19.