ARTICLE

Reliability of Handheld Dynamometry to Measure Focal Muscle Weakness in Neurofibromatosis Types 1 and 2

Srivandana Akshintala, MBBS, MPH, Nashwa Khalil, PT, DPT, Kaleb Yohay, MD, Alona Muzikansky, MA, Jeffrey Allen, MD, Anna Yaffe, BS, Andrea M. Gross, MD, Michael J. Fisher, MD, Jaishri O. Blakeley, MD, Beverly Oberlander, MSMC, Miriam Pudel, CPNP, Celia Engelson, MS, FNP, Jaime Obletz, MA OTR/L, CIMI, Carole Mitchell, MS, RN, Brigitte C. Widemann, MD, David A. Stevenson, MD,* and Scott R. Plotkin, MD, PhD*

Correspondence
Srivandana Akshintala
srivandana.akshintala@nih.gov

Neurology® 2021;97:S99-S110. doi:10.1212/WNL.0000000000012439

Abstract

Objective

To determine a suitable outcome measure for assessing muscle strength in neurofibromatosis (NF) type 1 and NF2 clinical trials, we evaluated the intraobserver reliability of handheld dynamometry (HHD) and developed consensus recommendations for its use in NF clinical trials.

Methods

Patients ≥5 years of age with weakness in at least 1 muscle group by manual muscle testing (MMT) were eligible. Maximal isometric muscle strength of a weak muscle group and the biceps of the dominant arm was measured by HHD. An average of 3 repetitions per session was used as an observation, and 3 sessions with rest period between each were performed on the same day by a single observer. Intrasession and intersession intraclass correlation coefficients (ICCs) and coefficients of variation (CVs) were calculated to assess reliability and measurement error.

Results

Twenty patients with NF1 and 13 with NF2 were enrolled; median age was 12 years (interquartile range [IQR] 9–17 years) and 29 years (IQR 22–38 years), respectively. By MMT, weak muscle strength ranged from 2−/5 to 4+/5. Biceps strength was 5/5 in all patients. Intersession ICCs for the weak muscles were 0.98 and 0.99 in the NF1 and NF2 cohorts, respectively, and for biceps were 0.97 and 0.97, respectively. The median CVs for average session strength were 5.4% (IQR 2.6%–7.3%) and 2.9% (IQR 2.0%–6.2%) for weak muscles and biceps, respectively.

Conclusion

HHD performed by a trained examiner with a well-defined protocol is a reliable technique to measure muscle strength in NF1 and NF2. Recommendations for strength testing in NF1 and NF2 trials are provided.

*These authors contributed equally to this work as senior authors.

From New York University (NYU) School of Medicine and NYU Langone Health (S.A., N.K., K.Y., J.A., A.Y., M.P., C.E., J.O., C.M.), New York; Pediatric Oncology Branch (S.A., A.M.G., B.C.W.), Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD; Massachusetts General Hospital (A.M.), Boston; Division of Oncology (M.J.F.), The Children's Hospital of Philadelphia, PA; Department of Neurology (J.O.B.), Johns Hopkins University, Baltimore, MD; Neurofibromatosis Network (B.O.); Department of Pediatrics (D.A.S.), Division of Medical Genetics, Stanford University School of Medicine, Palo Alto, CA; and Cancer Center and Department of Neurology (S.R.P.), Massachusetts General Hospital, Boston.

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

Glossary

ALS = amyotrophic lateral sclerosis; **CV** = coefficient of variation; **HHD** = handheld dynamometry; **ICC** = intraclass correlation coefficient; **IQR** = interquartile range; **MMT** = manual muscle testing; **MRC** = Medical Research Council; **NF** = neurofibromatosis; **NYU** = New York University; **REiNS** = Response Evaluation in Neurofibromatosis and Schwannomatosis.

Neurofibromatosis (NF) type 1 and NF2 are genetic disorders with multisystem manifestations that include development of tumors of the CNS and peripheral nervous system.[1,2] In addition, skeletal abnormalities and cognitive deficits may be seen in NF1.[1,3] Generalized or focal muscle weakness secondary to tumor or nontumor manifestations is a significant concern in both NF1 and NF2[2,4-6] and may negatively affect typical motor development of children, physical capability at work and school, and quality of life. Recent clinical trials targeting peripheral nerve tumors in children with NF1 have shown improvement in muscle weakness with the 0 to 5 Medical Research Council (MRC) scale.[7] However, more reliable and sensitive measures of muscle strength are necessary to systematically assess clinical benefit of interventions on muscle strength in NF1 and NF2 clinical trials.

Handheld dynamometry (HHD) is a convenient technique that provides a quantitative measurement of isometric muscle strength and has been used as an outcome measure in trials for other neurologic conditions such as amyotrophic lateral sclerosis (ALS).[8,9] Studies comparing HHD to isokinetic measurement as a reference standard have concluded that HHD can be considered a valid method for strength testing in a clinical setting.[10,11] The reliability of strength measurements using HHD has been studied in children and adults who are healthy and those with various disease conditions. Reported reliability has varied according to patient population and muscle group tested.[10] The objectives of this study were to assess the reliability of HHD in measuring strength in patients with NF and to evaluate its utility as an outcome measure in NF clinical trials.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

The protocol was approved by the institutional review board at New York University (NYU) Langone Health, and all participants or their legal guardians provided voluntary written informed consent before participating in the study.

The goal of the functional outcomes working group of the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) International Collaboration is to identify standardized functional measures appropriate for use as endpoints in NF clinical trials.[12] Working group members from various disciplines therefore developed a single-institution prospective study to assess the reliability of HHD in patients with NF1 and NF2. From the results of the study, the group subsequently developed consensus recommendations on use of HHD in NF clinical trials.

Outcome Measurement Selection for the Reliability Study

The following different methods of measuring muscle strength were considered as potential outcome measures by the working group: manual muscle testing (MMT) using the 0 to 5 MRC scale, isokinetic dynamometers, handgrip strength, and HHD. MMT using the 0 to 5 ordinal MRC scale is routinely used in clinical practice to assess strength of different muscle groups. However, published literature suggests it may not provide a sensitive or reliable measure, especially for muscles that are rated between 4/5 and 5/5 in strength; therefore, this outcome measure was not selected.[13] Isokinetic dynamometers are computerized machines that measure multiple elements of muscle strength such as the peak force, endurance, power, and angle of maximal force and generate strength curves.[11] These dynamometers are reliable and valid for testing muscle force and are often used as a reference standard. However, the need for complex and expensive equipment with instrument cost of approximately $40,000 or more is a major limitation for many centers, and thus isokinetic dynamometers were not considered practical for NF clinical trials.[11] Handgrip strength using a grip strength dynamometer measures the maximal voluntary muscular force of muscles of the hand, and it has been used as a measure of generalized weakness such as myopathy. It has been widely studied in different patient populations, including NF1, and is relatively inexpensive and easy to use in a clinical setting.[6,14] Handgrip strength was not recommended for measurement of tumor-related weakness because it cannot measure focal weakness of specific muscle groups. HHD is a convenient technique that provides a quantitative measurement of force generated by an individual muscle.[8] HHD can measure strength in various muscle groups, is relatively inexpensive with an instrument cost of approximately $1,000, and can be used in a clinic setting.[10,11] Testing of each muscle group takes only a few minutes, and HHD can measure muscle strength without testing until exhaustion of the muscle.[4] Strength testing by HHD is therefore less tiring for patients and allows repeat measurements after a short period of rest. Limitations of HHD include the increased variability in measurements when the participant overcomes the strength of examiner (ceiling effect), as well as dependence of the measurement on patient effort and consistency of testing procedure.[10,15] On the basis of the need for outcome measures to be relatively quick and easy to perform in an outpatient clinic

setting and the ability to measure focal weakness, strength testing by HHD was selected for further evaluation.

Study Participants

Patients with clinically confirmed NF1 using the NIH consensus criteria,[16] NF2 (NIH or Manchester criteria), or a known germline *NF1* or *NF2* pathogenic variant were eligible to participate and were recruited from the NF clinic at NYU Langone Health. Additional eligibility criteria included age ≥5 years, ability to follow instructions and to cooperate with strength examination, and weakness on neurologic examination (defined as <5/5 strength on the 0 to 5 MRC scale by MMT) in at least 1 of the 11 muscle groups listed in table 1. Patients were excluded if they had undergone any orthopedic procedure or other major surgery that could influence extremity strength in the 6 months before enrollment and if they had any history of tibial dysplasia.

Data Availability

Anonymized data will be shared by request from any qualified investigator.

Study Design

Patients were enrolled in the study into 2 cohorts: cohort 1 for patients with NF1 and cohort 2 for patients with NF2. Muscle strength was assessed by MMT either as part of routine clinical examination or as part of this study to identify weak muscle groups. For each patient, 1 weak muscle group was identified for testing reliability by HHD (referred to as the weak muscle hereafter). Potential muscle groups to be studied included wrist extensors (extensor carpi ulnaris/radialis), biceps, triceps, shoulder external rotators, deltoid, gastrocnemius, hamstring, quadriceps, iliopsoas, gluteus medius, and gluteus maximus (table 1). For participants with >1 weak muscle group, 1 muscle group was randomly selected for the study. In addition, to systematically measure a common muscle group in all patients, the strength of the biceps muscle on the dominant side was also measured (referred to as the dominant biceps hereafter). The dominant biceps would be measured regardless of strength by MRC scale, and if the biceps on the dominant side was the only weak muscle, then only the biceps on that side would be evaluated.

Testing was performed by a single trained physical therapist using a standardized testing protocol (table 1). Each muscle group strength was measured in 3 sessions, and each session consisted of 3 repetitions of strength testing (figure 1). Patients who completed at least 1 session for at least 1 muscle group were included in the analysis. Reasons for testing failure were noted such as if testing was unable to be completed due to patient inability to comply with instructions or if the examiner was unable to overcome the strength of the participant and keep the dynamometer stationary.

Clinical data were extracted from patients' medical records, including participant age, sex, weight, height, muscle strength by modified MRC scale,[17] and history of NF manifestations such as presence of spinal or peripheral nerve tumors, CNS manifestations, attention-deficit/hyperactivity disorder, skeletal deformities, and prior surgeries. Each patient was given a unique identification, and deidentified clinical and strength testing data were entered into a secure REDCap database.

Strength Testing Protocol

Muscle strength was measured in Newtons with the Ametek Chatillon DFE2 handheld dynamometer (Carrollton, TX). The standardized testing protocol defined the correct positioning of patient, handheld dynamometer placement, and stabilization of the limb for each potential muscle group to be tested (table 1). For muscle strengths <3/5 per the modified MRC scale, when the patient would be unable to maintain the limb and joint position against gravity without support, modified participant/joint positions (gravity-eliminated positions) were developed. A "make test" in which the patient was asked to exert maximal force while the examiner held the handheld dynamometer stationary was used to ensure an isometric muscle contraction. Hand dominance was determined by asking the patient directly or by asking them to write down their name or to draw a picture (for younger pediatric patients).

Statistical Analysis

Intraobserver reliability of HHD was estimated by calculating the intraclass correlation coefficient (ICC) in the NF1 and NF2 cohorts. Intrasession and intersession ICCs were calculated for the weak muscle group and the dominant biceps group separately. For calculating the intersession ICC, the average strength from the 3 repetitions within a session (average session strength) was calculated and used as 1 observation, and the average strength measurements from the 3 sessions were used for the analysis. The ω^2 was estimated to assess the partial variation accounted for by the repeat measurements within a patient compared to the variability between patients. In addition, the coefficient of variation (CV) was calculated as an estimate of measurement variability. Percent difference between the average session strength measurements for each muscle was also calculated. For descriptive statistics, unless otherwise indicated, percentages, median, interquartile ranges (IQRs), and ranges are provided. The study planned to enroll 20 patients each in the NF1 and NF2 cohorts because a sample size of 20 participants with 3 observations per participant would achieve a 90% power to detect an ICC of 0.91 under the alternative hypothesis when the ICC under the null hypothesis is 0.75 using an F test with a significance level of 0.05.

Results

Between October 8, 2018, and December 2, 2019, 20 patients with NF1 and 13 patients with NF2 were enrolled. Clinical characteristics of the enrolled patients are summarized in table 2. Common NF1-associated clinical manifestations in the 20 enrolled patients included learning difficulties (n = 15, 75%),

Table 1 Standardized Testing Protocol

Muscle Group	Position; Limb/Joint Position	Modified Participant and Limb/Joint Position (Gravity-Eliminated Position if Needed for MMT <3/5)	Dynamometer Placement
Wrist extensors (extensor carpi ulnaris/radialis)	Sitting; elbow 90° flexion, forearm resting on support surface and fully pronated, wrist in neutral position	Sitting; elbow 90° flexion, forearm and wrist resting on support surface and in neutral position	Just proximal to third metacarpal head
Elbow flexors (biceps brachii)	Supine; shoulder 0°, elbow 90° flexion, forearm in full supination	Side-lying (lying on side contralateral side of muscle being tested); shoulder 0°, elbow 90° flexion, forearm in full supination	Most distal on flexor surface of forearm, just proximal to wrist
Elbow extensors (triceps)	Supine; shoulder 0°, elbow 90° flexion, forearm in full supination	Side-lying (lying on side contralateral side of muscle being tested); shoulder 0°, elbow 90° flexion, forearm in neutral position	Most distal on extensor surface of forearm, just proximal to wrist
Shoulder external rotators	Supine; shoulder 0°, elbow flexed 90°, forearm in neutral position	Side-lying (lying on side contralateral side of muscle being tested); shoulder 0°, elbow 90° flexion, forearm in neutral position	Most distal on posterior surface of forearm, just proximal to wrist
Shoulder abductors (deltoid–middle)	Supine; shoulder and elbow 0°, forearm in neutral	NA	Most distal on lateral surface of arm, just proximal to lateral epicondyle of humerus
Ankle plantarflexors (gastrocnemius)	Supine; hip and knee extended 0°, ankle in neutral and resting off the end of the mat/examination table	NA	Plantar aspect of foot, just proximal to the first metatarsal head
Knee flexors (hamstrings)	Sitting; knee flexed 90°, hip flexed 90°, trunk straight; feet on floor	Side-lying (lying on side contralateral side of muscle being tested); knee flexed 90°, hip flexed 90° with leg resting on supportive surface	Most distal on posterior surface of leg, just proximal to malleoli
Knee extensors (quadriceps femoris)	Sitting; knee flexed 90°, hip flexed 90°, trunk straight; feet on floor	Side-lying (lying on side contralateral side of muscle being tested); knee flexed 90°, hip flexed 90° with leg resting on supportive surface	Most distal on anterior surface of leg, just proximal to malleoli
Hip flexors (iliopsoas)	Supine; hip and knee 90° with leg resting on supportive surface	Side-lying (lying on side contralateral side of muscle being tested); hip and knee 90° with leg resting on supportive surface	Anterior aspect of thigh, most distal, just proximal to knee joint
Hip abductors (gluteus medius)	Supine; hip and knee 0°, contralateral limb stabilized on table with foot flat on surface	NA	Most distal on lateral surface of thigh, on lateral femoral epicondyle
Hip extensors (gluteus maximus)	Supine; hip and knee 90° with leg resting on supportive surface	Side-lying (lying on side contralateral side of muscle being tested); hip and knee 90° with leg resting on supportive surface	Posterior aspect of thigh, most distal, just proximal to knee joint

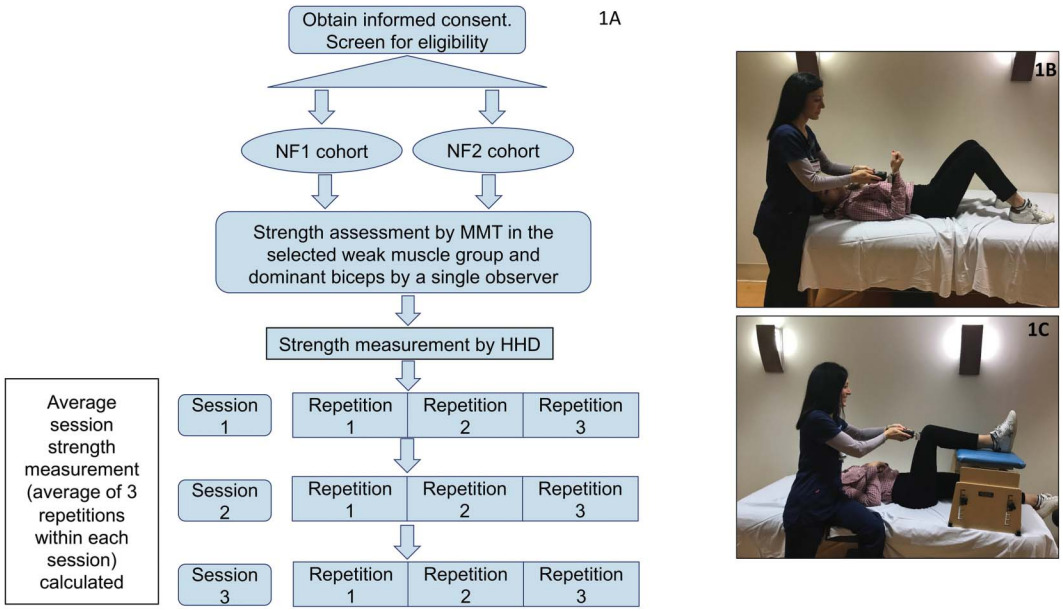
Abbreviations: MMT = manual muscle testing; NA = not applicable.
Verbal/written instructions provided before testing and verbal encouragement during testing were also standardized. Subjects were instructed to avoid explosive contractions and asked to gradually increase their effort with verbal cueing "3, 2, 1, go". Each contraction was held for 5 seconds followed by a 60-second rest time between repetitions. For patients with hearing impairment in the neurofibromatosis type 2 cohort, cueing was modified using hand gestures (and/or mouth reading) indicating "3, 2, 1, go".

plexiform neurofibromas (n = 11, 55%), hypotonia (n = 10, 50%), and scoliosis (n = 8, 40%). One patient reported preexisting pain on the day of the evaluation. All of the patients with NF2 had bilateral vestibular schwannomas and other additional schwannomas, and a majority had other NF2-associated tumors. Gait problems were reported in 10 patients, and 12 had history of surgery.

In the NF1 cohort, the weak muscles tested included deltoid (n = 1), gluteus medius (n = 10), iliopsoas (n = 3), quadriceps (n = 3), shoulder external rotators (n = 2), and wrist extensors (n = 1). In the NF2 cohort, the weak muscles tested included biceps (n = 1), deltoid (n = 1), iliopsoas (n = 3), quadriceps (n = 5), and shoulder external rotators (n = 3). The strength in the weak muscles ranged from 2−/5 to 4+/5 on the modified MRC scale by MMT, with 28 of 33 patients having 4−/5 to 4+/5 strength. The strength in the dominant biceps was 5/5 by MMT in all patients tested (table 3).

A session testing 2 muscle groups typically lasted <10 minutes, and none of the patients requested testing to be terminated secondary to pain or fatigue. In some patients with NF1, increased time for testing was required due to higher rates of attention and behavioral issues, including impulsivity and decreased ability following commands. Children often required increased redirection to task. Accommodations for visual instructions or sign language interpretation for patients with NF2 with hearing impairment were also necessary. Despite implementation of alternative test positions for muscles that do not have full antigravity strength (strength <3/5), some compensatory use of proximal musculature during testing was appreciated. In addition, participants with muscles

Figure 1 Study Design



(A) Study schema. After 1 to 2 practice trials with the same muscle group of the contralateral side, strength of the weak muscle and the dominant biceps was measured with handheld dynamometry (HHD) in 3 sessions (sessions 1, 2, and 3, respectively). Each session consisted of 3 repetitions of strength testing with a minimum of 1 minute of rest time provided between the 3 repetitions. Repeat sessions (sessions 2 and 3) were performed on the same day after a minimum of 15 minutes of rest between sessions. The average of the 3 repetitions within a session was calculated and used as the strength measurement for that session. The same observer performed all 3 sessions, with the 3 repetitions per session using the same standardized protocol. (B) Example of testing biceps strength. (C) Example of testing iliopsoas strength. MMT = manual muscle testing; NF = neurofibromatosis.

with <3/5 strength demonstrated increased fatigability and decreased endurance with repeat measurements.

Overall, excellent (>0.9) intrasession and intersession ICCs were observed for the weak muscles group and the dominant biceps group in both the NF1 and NF2 cohorts (table 4). The high reliability and similar ICCs noted in both cohorts allowed reliability to be analyzed in both cohorts together, and the study was therefore stopped after 13 patients with NF2 were enrolled. For the weak muscle group, the partial ω^2 attributed to repeat measurements within a patient was 0.10 (90% confidence limits 0.03–0.22), whereas that attributed to variation between patients was 0.99 (90% confidence limits 0.98–0.99), suggesting that most of the variation in the data came from variability noted between patients with very little variability contributed by the repeat measurements within a patient. Similarly, for the biceps muscle, the partial ω^2 attributed to repeat measurements within a patient was 0.03 (90% confidence limits 0.00–0.12), whereas that attributed to variation between patients was 0.98 (90% confidence limits 0.97–0.98). For the weak muscle group, in the NF1 cohort, the median CV between the repeat measurements taken during the 3 sessions was 5.4% (IQR 2.8%–7.1%, maximum 15.4%). The median CV in the NF2 cohort was 5.9% (IQR 2.3%–7.3%, maximum 17.4%). For the biceps, the median CV was 2.7% (IQR 1.7%–6.1%, maximum 22.1%) in the NF1 cohort and 3.5% (IQR 2.8–6.1%, maximum 6.8%) in the NF2 cohort (table 3). The CV was ≤10% in 59 of 64 (92%) of the muscle groups tested. The corresponding difference in the average session strength measurement between the highest and lowest reading for each muscle group was <25% in 60 of 64 (94%) muscle groups tested (table 3). The CV did not vary on the basis of the measured strength (figures 2, A and B).

Discussion

There is a great need for developing clinical and functional outcome measures for patients with NF1 and NF2 that can assist in the study of the disease natural history and the effect of interventions in clinical trials. These measures are needed in clinical trials for noncancerous conditions in which tumor measurement responses alone are insufficient for registration and there is a need to document clinical benefit.[12] Muscle weakness is a well-described morbidity in both NF1 and NF2, and early clinical trials targeting plexiform neurofibromas in children with NF1 have shown improvement in muscle strength with the use of the MRC 0 to 5 scale by MMT.[7] MMT, however, is not a sensitive or reliable measure, especially for muscles with strength rated as 4/5 to 5/5.[13] Therefore, outcome measures that can reliably and sensitively measure muscle strength are needed for studying this key morbidity in future clinical trials.

An effective outcome measure should be reliable (low intra-participant variation in those whose health status is stable), valid (change detected by the measure should be consistent

Table 2 Clinical Characteristics

Patient	Age, y	Sex	Weight, kg	Height, cm	Clinical History
NF1 (n = 20)					
2	10	M	33.7		OPG, scoliosis, LD, hypotonia, seizures
5	10	M	22.4	131	ADHD, LD
6	12	M	48.8	154	Scoliosis, ADHD, LD
8	19	M	60.3	160	Scoliosis, LD, spinal ganglioglioma, 6/10 pain, Sx
9	6	F	32.2	129	PN, Sx
10	27	M	71.7		PN, other glioma, scoliosis, ADHD, LD, seizure, pseudoaneurysm, Sx
11	45	M	27.9	127	ADHD, LD, hypotonia
12	11	M	27.9	127	PN, OPG, other glioma, LD, Sx
13	9	M	34.7	128	PN, other glioma, LD, moyamoya, sphenoid wing dysplasia, Sx
14	6	M	31.3	124	OPG, ADHD, LD, hypotonia
15	6	F	20.9	118	Hypotonia
16	7	M	32.6	140	ADHD, LD, hypotonia
17	16	M	93.0	175	PN, scoliosis, LD
18	13	F	39.3	142	LD, hypotonia
20	19	F	58.6	161	PN, Sx
21	12	M	40.6	152	PN, ADHD, LD, hypotonia
22	19	M	60.5	168	PN, scoliosis, hypotonia
23	16	M	63.0	175	PN, scoliosis, hydrocephalus, Sx
29	8	F	49.7	147	PN, OPG, LD, hypotonia
31	14	M	53.8	155	PN, scoliosis, LD, hypotonia
Median	12		40.0	144	
Range	6–45		20.9–93.0	118–175	
NF2 (n = 13)					
1	25	F	45.8	157	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, Sx
3	23	M	53.7	166	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, Sx
4	22	F	70.1	160	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, Sx
7	15	F	61.0	160	BLVS, schwannoma (CNS, spinal),
19	39	F	78.8	178	BLVS, schwannoma (CNS, spinal), gait dysfn, Sx
24	29	M	56.4	163	BLVS, schwannoma (CNS, spinal), meningioma, gait dysfn, Sx
25	19	F	57.5	166	BLVS, schwannoma (CNS, spinal), meningioma, gait dysfn, Sx
26	29	M	66.0	178	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, Sx
27	75	F	54.9	163	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, Sx
28	53	M	70.8	191	BLVS, schwannoma (spinal), meningioma, ependymoma, gait dysfn, Sx
30	32	F	47.5	164	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, Sx
32	21	F	41.3	155	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, gait dysfn, pain, Sx
33	38	F	43.0	146	BLVS, schwannoma (CNS, spinal), meningioma, ependymoma, Sx
Median	29		56.4	163	
Range	15–75		41.3–78.8	146–191	

Abbreviations: ADHD = attention-deficit/hyperactivity disorder; BLVS = bilateral vestibular schwannoma; dysfn = dysfunction; LD = learning difficulties; NF = neurofibromatosis; OPG = optic pathway glioma; PN = plexiform neurofibroma; schwannoma = other schwannomas; Sx = history of surgery.

Table 3 Strength Measurements and Measurement Variability

Patient	Weak Muscle	Weak Muscle MMT	Weak Muscle Average Session Strength, N[a]			CV	% Difference[b]	Biceps MMT	Biceps Average Session Strength, N[a]			CV	% Difference[b]
NF1 (n = 20)													
2	R iliopsoas	4/5	70.9	73.5	81.1	7.0	14.3		51.3	75.5	79.4	22.1	54.8
5	L quadriceps	4+/5	75.2	80.4	88.9	8.5	18.3	5/5	69.3	67.6	69.0	1.3	2.5
6	R gluteus medius	4+/5	83.7	82.7	86.0	2.0	4.0	5/5	91.2	98.0	95.4	3.7	7.5
8	R iliopsoas	3+/5	121.7	113.3	117.0	3.6	7.4	5/5	156.7	157.0	144.7	4.6	8.5
9	R gluteus medius	4+/5	59.0	59.7	60.0	0.9	1.7	5/5	51.0	58.0	62.3	10.0	22.2
10	L quadriceps	4/5	119.3	113.7	113.3	2.9	5.3						
11	R gluteus medius	4+/5	157.0	165.7	171.0	4.3	8.9	5/5	170.7	163.7	161.7	2.9	5.6
12	R gluteus medius	4+/5	68.7	70.3	70.0	1.3	2.4	5/5	71.7	72.7	74.3	1.8	3.7
13	R gluteus medius	3−/5	46.3	63.3	56.3	15.4	36.7	5/5	57.0	57.7	58.7	1.5	2.9
14	R gluteus medius	4/5	45.0	42.7	39.3	6.7	14.4	5/5	68.0	67.7	71.0	2.7	4.9
15	R shoulder ext rotators	4/5	37.0	38.0	37.7	1.4	2.7	5/5	60.3	64.0	63.3	3.1	6.1
16	R gluteus medius	4−/5	42.3	48.3	48.3	7.5	14.2	5/5	90.3	89.7	87.0	2.0	3.8
17	L gluteus medius	4+/5	81.3	87.3	84.7	3.6	7.4	5/5	176.7	153.7	186.3	9.7	21.3
18	L deltoid	4+/5	67.0	66.3	73.0	5.3	10.1	5/5	89.7	92.0	87.3	2.6	5.3
20	L wrist extensors	4−/5	76.7	83.3	88.0	6.9	14.8	5/5	98.3	98.3	101.0	1.6	2.7
21	L shoulder ext rotators	4−/5	49.3	56.7	50.7	7.5	14.9	5/5	78.0	78.3	76.3	1.4	2.6
22	R gluteus medius	4−/5	96.3	99.7	101.3	2.6	5.2	5/5	86.7	80.7	99.7	10.9	23.6
23	R gluteus medius	4/5	70.7	61.7	64.7	7.0	14.6	5/5	165.7	166.7	165.3	0.4	0.8
29	R iliopsoas	4/5	105.3	100.3	111.7	5.4	11.3	5/5	87.3	84.3	87.3	2.0	3.6
31	R quadriceps	2/5	28.7	28.0	32.0	7.3	14.3	5/5	74.7	78.7	67.7	7.6	16.3
Median						5.4	10.7					2.7	5.3
Range		2/5–4+/5				0.9–15.4	1.7–36.7	5/5–5/5				0.4–22.1	0.8–54.8
NF2 (N = 13)													
1	R iliopsoas	4+/5	125.5	122.9	126.2	1.4	2.7	5/5	100.4	100.1	95.5	2.8	5.1
3	R quadriceps	4+/5	52.0	46.1	45.8	7.3	13.6	5/5	139.2	132.7	149.7	6.1	12.8
4	L quadriceps	4+/5	238.3	247.8	243.9	2.0	4.0	5/5	191.9	192.8	192.9	0.3	0.5
7	L quadriceps	2−/5	25.7	36.3	30.3	17.4	41.6	5/5	92.7	97.7	97.3	2.9	5.4
19	R shoulder ext rotators	4+/5	84.3	95.3	94.3	6.7	13.0	5/5	118.3	120.7	123.7	2.2	4.5
24	L Shoulder ext rotators	4−/5	64.0	61.0	64.3	2.9	5.5	5/5	108.7	113.7	118.3	4.3	8.9
25	L biceps	4+/5	102.0	103.7	106.7	2.3	4.6	5/5	154.0	137.7	138.7	6.4	11.9
26	R shoulder ext rotators	4+/5	105.7	103.0	104.0	1.3	2.6	5/5	120.0	113.3	122.7	4.1	8.2
27	L quadriceps	4/5	137.3	132.7	152.3	7.3	14.8	5/5	84.0				
28	R iliopsoas	3+/5	53.3	62.7	71.7	14.7	34.4	5/5	163.0	155.0	177.0	6.7	14.2
30	L iliopsoas	4+/5	128.7	111.3	133.0	9.2	19.5	5/5	106.7	110.7	113.0	2.9	5.9

Continued

Table 3 Strength Measurements and Measurement Variability *(continued)*

Patient	Weak Muscle	Weak Muscle MMT	Weak Muscle Average Session Strength, N[a]			CV	% Difference[b]	Biceps MMT	Biceps Average Session Strength, N[a]			CV	% Difference[b]
32	R quadriceps	4−/5	74.0	70.3	72.7	2.6	5.2	5/5	63.0	55.0	60.0	6.8	14.5
33	L deltoid	4/5	74.0	69.0	77.7	5.9	12.6	5/5	125.0	119.0	124.7	2.7	5.0
Median						5.9	12.6					3.5	7.1
Range		2-/5 to 4+/5				1.3–17.4	2.6–41.6	5/5–5/5				0.3–6.8	0.5–14.5

Abbreviations: CV = coefficient of variation; ext = external; MMT = manual muscle testing; N = Newtons; NF = neurofibromatosis.
[a] Average session strength is the average measurement readings from 3 repetitions within a session. CV is for the 3 sessions tested per muscle.
[b] Percent difference in average session strength for the 3 sessions tested per muscle = [(maximum − minimum value)/minimum value] × 100 among the 3 observations (average session strength measurements) per muscle.

with an external standard of change), and responsive to change (able to detect change over time).[18] A systematic review of 17 articles that included 19 studies compared HHD to isokinetic dynamometers as a reference standard and concluded that HHD can be considered a valid measure of strength in a clinical setting.[11] However, reported reliability of HHD has been variable in the literature according to the study population tested and the muscles to be tested. Given the importance of assessing the reliability of a measure within the target population of interest, we decided to conduct this reliability study in patients with NF1 and NF2 who have known muscle weakness in at least 1 muscle group.[10] Muscle weakness in these patients can be secondary to tumors (such as plexiform neurofibromas in NF1 or spinal tumors in NF1 and NF2) or nontumor manifestations (such as primary myopathy or neuropathy in NF1). In addition, patients can have different phenotypes with muscle weakness of varying degrees that can be focal or generalized, depending on the etiology.[2,4-6,14] This contrasts with other neuromuscular conditions in which generalized weakness or a defined pattern of focal weakness is noted. The eligibility requirements and the design of this reliability study were chosen considering this large variation in phenotype in the NF1 and NF2 population.

Table 4 Reliability of HHD

	Intrasession ICC; Sessions 1–3	Intersession ICC
Weak muscle		
NF1 (n = 20)	0.96–0.98	0.98
NF2 (n = 13)	0.97–0.98	0.99
All (n = 33)	0.97–0.98	0.99
Biceps[a]		
NF1 (n = 19)	0.97–0.98	0.97
NF2 (n = 13)	0.93–0.95	0.97
All (n = 32)	0.96–0.97	0.97

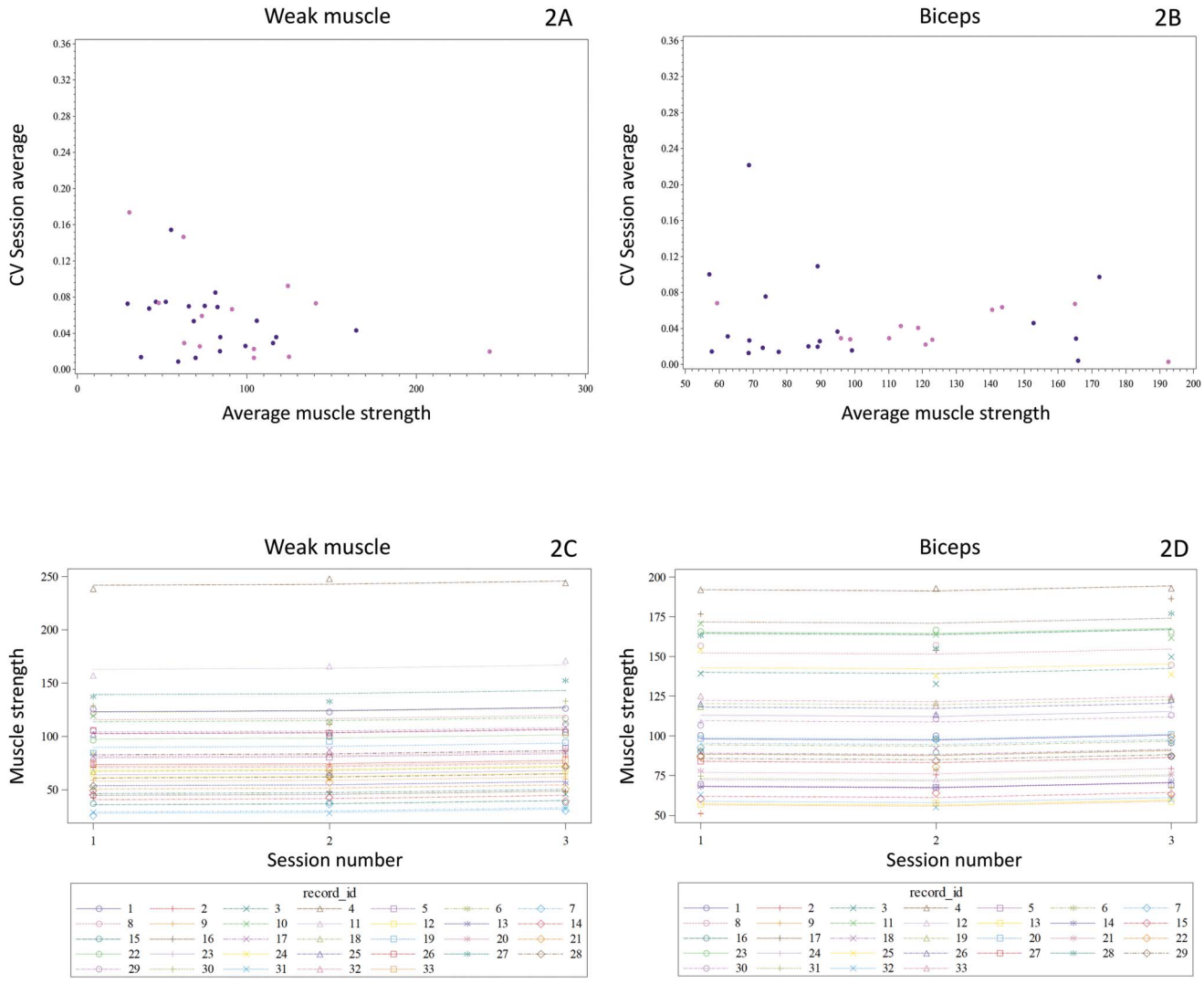
Abbreviations: HHD = handheld dynamometry; ICC = intraclass correlation coefficient; NF = neurofibromatosis.
[a] Biceps strength was unable to be tested in 1 patient with NF1 and was tested for only 1 session in 1 patient with NF2.

The results of our study suggest that HHD is a reliable technique for pediatric and adult patients with NF1 and NF2 when measurements are performed by a trained examiner with a well-defined protocol. The intrasession and intersession ICCs were excellent, and the CVs for most muscles tested were ≤10%. Testing for 1 muscle group was brief and did not cause any discomfort or pain. Special considerations for pediatric patients, patients with attention or behavioral issues, those with skeletal issues limiting limb/dynamometer placement, or those with hearing impairment needed to be considered in this population.

Potential limitations of the study include the cross-sectional design, which cannot provide data on the longitudinal changes in the muscle strength in the NF population; thus, further evaluation in longitudinal studies will be needed. In addition, only intraobserver reliability was tested. To minimize observer and participant bias, the instrument was held in a position such that the reading would not be visible until the end of the muscle contraction. However, because a single observer performed all measurements, observer bias could not be completely excluded. In this study, reliability was tested in the selected weak muscle that differed across patients. Therefore, this study cannot provide the reliability estimates for individual muscle groups in the NF population. However, given the variable phenotype in NF and concern for focal weakness, we anticipate different target muscle groups to be affected in different patients in NF1 and NF2 clinical trials. Therefore, the design for our study was chosen to most closely resemble the study population of interest in therapeutic or intervention trials in NF 1 and NF2.

On the basis of the results of this study and review of literature, the REiNS functional group has the following recommendations and considerations for strength testing in NF1 and NF2 trials.

Figure 2 CV and Interaction Plots



(A and B) Coefficients of variation (CVs) for the average session strength measurements (3 strength measurements per muscle tested) plotted vs the average muscle strength. Patients with neurofibromatosis (NF) type 1 are shown in purple; patients with NF2 are shown in pink. (A) Plot for the weak muscle; (B) plot for the dominant biceps. The majority of CVs were ≤0.1 (i.e., ≤10%). (C and D) Interaction plots for measured muscle strength in the (C) weak muscle and (D) dominant biceps. Session number is on the x-axis; muscle strength is on the y-axis. Each line represents 1 patient. Each value represents the mean of the 3 repetitions within a session.

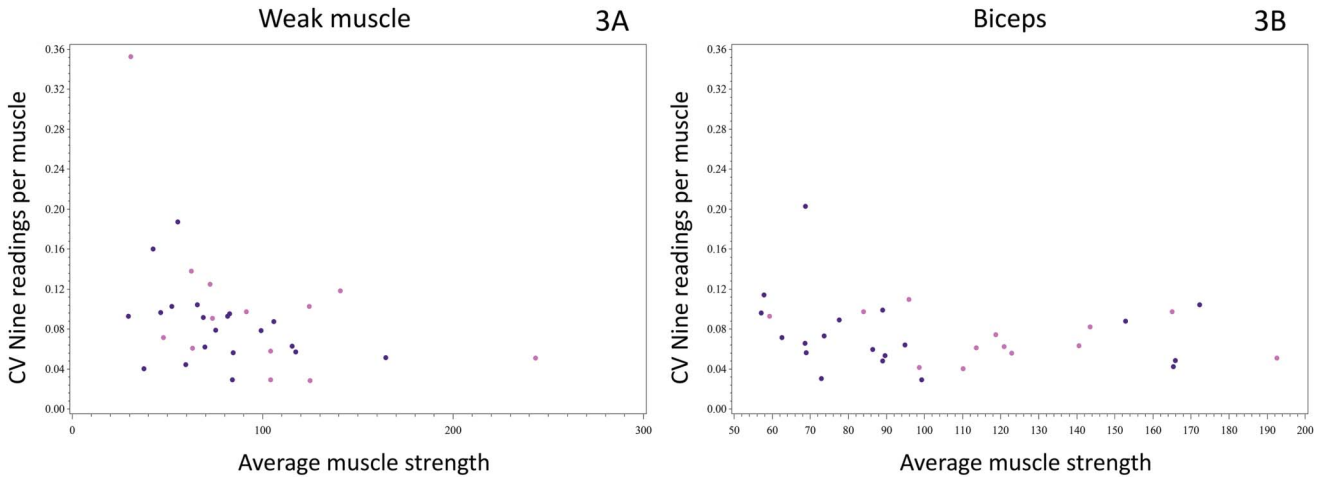
Muscle strength as measured by HHD should be evaluated prospectively as an outcome measure in NF trials. We recommend measuring at least 3 replicates per muscle and using an average of the replicates as the main outcome strength measurement. This is based on reduced variability with average of repetitions within a session compared to individual strength trial measurements (figure 3).

On the basis of the measurement variability observed in this study (CV ≤10% in 92% of the muscle groups tested and the percent change in the average session strength measurements <25% in 94% muscle groups tested), an increase in strength (measured in Newtons) of ≥25% from baseline can be considered a measurable improvement in the strength of an individual muscle group (functional response), and a decrease in strength (measured in Newtons) ≥25% from baseline can be considered measurable worsening in the muscle strength (functional progression).

The number of muscle groups to be tested should depend on the study objectives. In studies in which patients are required to have preexisting weakness in a particular muscle group at baseline for eligibility, change in the strength of the involved muscle group or the patient-reported most clinically significant weak muscle should be followed.

When >1 muscle group per patient is measured, in addition to changes in individual muscle groups, overall response for the patient may need to be analyzed, depending on the study objectives. Prior clinical trials have used different approaches

Figure 3 CV Plots of Individual Strength Trial Measurements



(A and B) Coefficient of variations (CVs) for repeat measurements of the same muscle plotted vs the average muscle strength for the weak muscle and biceps, respectively. Unlike in figure 2, A and B, instead of the average session strength measurements, all 9 readings were used to calculate the CV. Variability is higher compared to CV calculated for the average session strength measurements, suggesting that averaging 3 repetitions decreases the variability in measurement.

to assess changes in muscle strength as an outcome measure for a particular extremity, a body quadrant, or the patient overall such as the phase 2 trial of selumetinib in children with inoperable plexiform neurofibromas (using the 0–5 MRC scale) and phase 3 trials for ALS (using HHD).[7,9]

Strength is expected to increase with age in children.[19] Thus, for studies that measure change in strength over long periods of time such as ≥1 year, the change in strength needs to be interpreted in the context of anticipated changes with age. Age- and sex-adjusted *z* scores can potentially be used to interpret change in strength in children in these situations. In addition, strength measurements with HHD in growing children are anticipated to improve just by virtue of an increase in the distance between the position of the dynamometer placement and the joint as the child grows taller. Measurement of strength and the distance between the position of the dynamometer placement and the joint and estimation of torque are recommended in the pediatric age group when significant changes in height are anticipated during the study follow-up period: torque = force (strength) of the muscle tested by HHD × distance between the joint and the position of the dynamometer.[19,20]

From this study and literature evaluating HHD in children ≥4 years of age,[20] we recommend that HHD can be used in children as young as 5 years of age. However, studies need to account for the inability of very young children or children and adults with cognitive or behavioral issues to provide reliable estimates of muscle strength using HHD because testing does require participant cooperation. These issues may be particularly relevant to the NF1 population. In contrast, hearing impairment, which can be a concern particularly in the NF2 population, was overcome in our study with the help of sign language interpretation and visual cues for encouragement during the testing.

On the basis of the high reliability in measuring the unaffected biceps muscle in both patients with NF1 and those with NF2 in this study and some studies in the literature reporting good reliability even in strong muscles in healthy adults,[10,21] we recommend that HHD can be used to assess strength in affected (strength <5/5 by MMT) and unaffected (no evidence of clinical weakness and strength 5/5 by MMT) extremity muscles. HHD provides a quantitative measure of strength in Newtons, kilogram-force, or pound-force; therefore, there is no upper limit in terms of strength measurement by HHD. Hence, change may be observed in strength even in muscles that are assessed to be 5/5 by MMT at baseline. However, the measurement of strong muscle groups such as hip flexors and knee extensors may be less reliable. Using biomechanically sound setups and procedures, obtaining repeat baseline measures, and excluding observations in which the examiner was unable to overcome the strength of participant should be considered in those instances. Stabilizing belts have also been used in other studies to help overcome the limitations imposed by examiner strength in measuring strength of strong muscles with HHD.[20,22] There is relatively less literature on the measurement of core muscle strength using HHD,[10] and core muscles were not tested in this study.

Testing with HHD should be performed by an experienced examiner because measurements are operator dependent. Testing should require a standardized protocol that should aim to decrease measurement variability and to minimize observer bias. Ideally, each patient's baseline and follow-up measurements should be performed by the same examiner who has received hands-on training in using the dynamometer and the testing protocol and has demonstrated high reliability in their measurements similar to that observed on this study. However, a core group of examiners who have been trained on

the use of the instrument and the testing protocol may need to be considered if a single examiner cannot perform all the required evaluations. In our study, a single experienced examiner performed all the evaluations, and only intraobserver reliability was assessed. Therefore, if future studies include multiple examiners, then agreement on the testing protocol and a baseline assessment of interobserver variability will be crucial to interpret the results because interobserver reliability may be less than intraobserver reliability. A similar approach was used in the ALS trials in which a core group of trained examiners were required to have tested at least 4 healthy volunteers before evaluating patients in the studies to ensure agreement on the testing protocol and to assess interobserver variability.[9]

Each patient should be tested with the same instrument at baseline and follow-up, and appropriate calibration of the instrument should be ensured before use. If multiple instruments are to be used in a study, the instruments should be cross-calibrated to ensure comparable measurements.

Timing of follow-up evaluations should depend on the study objectives and anticipated outcomes. They should also correspond to timing of other evaluations such as imaging, patient-reported outcomes, and other functional measures. Getting a repeat set of baseline measures when no significant change is anticipated (such as screening and cycle 1 day 1), if feasible, should also be strongly considered to assess for baseline test-retest variability.

In trials focusing on reducing long-term morbidity, confirmation of functional response or progression is desirable because that would increase the confidence that an observed response is the result of the intervention and is not due to baseline variability.

We recommend testing muscle strength by MMT using the MRC scale in addition to obtaining measurements with HHD for each of the muscle groups being evaluated to help correlate the values observed by HHD to the commonly assessed clinical evaluations. Depending on the study objectives, testing of other functional measures that assess motor function may also be considered at the time of strength assessment such as grip strength, functional strength measurement score, functional reach, timed up and go, and 10-m walk tests. Correlating measurements obtained by HHD to other functional measures or patient-reported outcomes can help provide additional data to study clinically meaningful improvement in strength.

The definition of duration of functional response and stable disease and the reporting of best response in trials have been described in previous REiNS recommendations.[23] In addition, because these recommended outcomes for strength measurement have not yet been used prospectively in NF clinical trials, the REiNS International Collaboration expects to reassess and potentially revise these recommendations on the basis of additional data obtained from incorporation of HHD into future NF clinical trials.

Acknowledgment

The authors acknowledge the following collaborators for their participation in the REiNS functional outcomes working group: Ann Blanton (SUNY Cortland), Ashley Cannon (University of Alabama at Birmingham), Elizabeth Schorry (Cincinnati Children's Hospital), Brad Welling (Massachusetts Eye and Ear Infirmary), Christopher Moertel (University of Minnesota), Colleen Marchetti (REiNS patient representative), David Viskochil (University of Utah, coordinating role for working group meetings), Dusica Babovic (Mayo Clinic), Fred Barker (Massachusetts General Hospital), D. Gareth Evans (St. Mary's Hospital, Manchester, coordinating role for working group meetings), Heather Thompson (California State University, Sacramento), James Tonsgard (University of Chicago), Kathy Gardner (University of Pittsburgh), Kent Robertson (Indiana University), Rebecca Mullin (St. Thomas' Hospital), Irene Moss (REiNS patient representative), Rosalie Ferner (Guy's Hospital), and Trent Hummel (Cincinnati Children's Hospital). The authors also acknowledge the support of the Children's Tumor Foundation for the REiNS International Collaboration.

Study Funding

The authors report no targeted funding.

Disclosure

S. Akshintala, N. Khalil, K. Yohay, A Muzikansky, J. Allen, A. Yaffe, A.M. Gross, M. J. Fisher, B. Oberlander, M. Pudel, C. Engelson, J. Obletz, C. Mitchell, B.C. Widemann, and D.A. Stevenson report no disclosures. J.O. Blakeley is the director of the Neurofibromatosis Therapeutic Acceleration Program. She has served as a paid consultant for Abbvie and an unpaid consultant for Springworks and AstraZeneca and has received research support from GlaxoSmithKline, Sanofi-Aventis, and Lily Pharmaceuticals. S.R. Plotkin is cofounder of NFlection Therapeutics, Inc and of NF2 Therapeutics, Inc. He is a consultant for AstraZeneca and for SonalaSense. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* October 9, 2020. Accepted in final form May 5, 2021.

Appendix Authors

Name	Location	Contribution
Srivandana Akshintala, MBBS, MPH	NYU Langone Health, New York; NIH, Bethesda, MD	Designed and conceptualized study; analyzed the data; drafted the manuscript for intellectual content
Nashwa Khalil, PT, DPT	NYU Langone Health, New York	Designed and conceptualized study; major role in the acquisition of data; revised the manuscript for intellectual content
Kaleb Yohay, MD	NYU Langone Health, New York	Designed and conceptualized study; major role in the acquisition of data; revised the manuscript for intellectual content

Continued

Appendix *(continued)*

Name	Location	Contribution
Alona Muzikansky, MA	Massachusetts General Hospital, Boston	Designed and conceptualized study; analyzed the data; performed statistical analysis; drafted the manuscript for intellectual content
Jeffrey Allen, MD	NYU Langone Health, New York	Major role in acquisition of data
Anna Yaffe, BS	NYU Langone Health, New York	Major role in acquisition of data
Andrea M. Gross, MD	NIH, Bethesda, MD	Interpreted the data; revised the manuscript for intellectual content
Michael J. Fisher, MD	The Children's Hospital of Philadelphia, PA	Interpreted the data; revised the manuscript for intellectual content
Jaishri O. Blakeley, MD	Johns Hopkins University, Baltimore, MD	Interpreted the data; revised the manuscript for intellectual content
Beverly Oberlander, MSMC	Neurofibromatosis Network	Interpreted the data; revised the manuscript for intellectual content
Miriam Pudel, CPNP	NYU Langone Health, New York	Major role in acquisition of data
Celia Engelson, MD, FNP	NYU Langone Health, New York	Major role in acquisition of data
Jaime Obletz, MA OTR/L, CIMI	NYU Langone Health, New York	Major role in acquisition of data
Carole Mitchell, MS, RN	NYU Langone Health, New York	Major role in acquisition of data
Brigitte C. Widemann, MD	NIH, Bethesda, MD	Interpreted the data; revised the manuscript for intellectual content
David A. Stevenson, MD	Stanford University, Palo Alto, CA	Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content
Scott R. Plotkin, MD, PhD	Massachusetts General Hospital, Boston	Designed and conceptualized study; interpreted the data; revised the manuscript for intellectual content

References

1. Gutmann DH, Ferner RE, Listernick RH, Korf BR, Wolters PL, Johnson KJ. Neurofibromatosis type 1. *Nat Rev Dis Primers.* 2017;3:17004.
2. Evans DG. Neurofibromatosis type 2. *Handb Clin Neurol.* 2015;132:87-96.
3. Summers MA, Quinlan KG, Payne JM, Little DG, North KN, Schindeler A. Skeletal muscle and motor deficits in neurofibromatosis type 1. *J Musculoskelet Neuronal Interact.* 2015;15(2):161-170.
4. Cornett KM, North KN, Rose KJ, Burns J. Muscle weakness in children with neurofibromatosis type 1. *Dev Med Child Neurol.* 2015;57(8):733-736.
5. Johnson BA, Macwilliams B, Carey JC, Viskochil DH, D'Astous JL, Stevenson DA. Lower extremity strength and hopping and jumping ground reaction forces in children with neurofibromatosis type 1. *Hum Mov Sci.* 2012;31(1):247-254.
6. Souza JF, Passos RL, Guedes AC, Rezende NA, Rodrigues LO. Muscular force is reduced in neurofibromatosis type 1. *J Musculoskelet Neuronal Interact.* 2009;9(1):15-17.
7. Gross AM, Wolters PL, Dombi E, et al. Selumetinib in children with inoperable plexiform neurofibromas. *N Engl J Med.* 2020;382(15):1430-1442.
8. Hebert LJ, Maltais DB, Lepage C, Saulnier J, Crete M, Perron M. Isometric muscle strength in youth assessed by hand-held dynamometry: a feasibility, reliability, and validity study. *Pediatr Phys Ther.* 2011;23(3):289-299.
9. Shefner JM, Liu D, Leitner ML, et al. Quantitative strength testing in ALS clinical trials. *Neurology.* 2016;87(6):617-624.
10. Kolber MJ, ACJ. Strength testing using hand-held dynamometry. *Phys Ther Rev.* 2005;10(2):99-112.
11. Stark T, Walker B, Phillips JK, Fejer R, Beck R. Hand-held dynamometry correlation with the gold standard isokinetic dynamometry: a systematic review. *PM R.* 2011;3(5):472-479.
12. Plotkin SR, Blakely JO, Dombi E, et al. Achieving consensus for clinical trials: the REiNS International Collaboration. *Neurology.* 2013;81(Supp):S1–S5.
13. John J. Grading of muscle power: comparison of MRC and analogue scales by physiotherapists: Medical Research Council. *Int J Rehabil Res.* 1984;7(2):173-181.
14. Stevenson DA, Allen S, Tidyman WE, et al. Peripheral muscle weakness in RASopathies. *Muscle Nerve.* 2012;46(2):394-399.
15. Wikholm JB, Bohannon RW. Hand-held dynamometer measurements: tester strength makes a difference. *J Orthop Sports Phys Ther.* 1991;13(4):191-198.
16. National Institutes of Health Consensus Development Conference Statement: neurofibromatosis: Bethesda, MD,USA, July 13-15, 1987. *Neurofibromatosis.* 1988;1:172-178.
17. Personius KE, Pandya S, King WM, Tawil R, McDermott MP. Facioscapulohumeral dystrophy natural history study: standardization of testing procedures and reliability of measurements: the FSH DY Group. *Phys Ther.* 1994;74(3):253-263.
18. Jerosch-Herold C. An evidence-based approach to choosing outcome measures: a checklist for the critical appraisal of validity, reliability and responsiveness studies. *Br J Occup Ther.* 2005;68(8):347-353.
19. Daloia LMT, Leonardi-Figueiredo MM, Martinez EZ, Mattiello-Sverzut AC. Isometric muscle strength in children and adolescents using handheld dynamometry: reliability and normative data for the Brazilian population. *Braz J Phys Ther.* 2018;22(6):474-483.
20. Hebert LJ, Maltais DB, Lepage C, Saulnier J, Crete M. Hand-held dynamometry isometric torque reference values for children and adolescents. *Pediatr Phys Ther.* 2015;27(4):414-423.
21. Mentiplay BF, Perraton LG, Bower KJ, et al. Assessment of lower limb muscle strength and power using hand-held and fixed dynamometry: a reliability and validity study. *PLoS One.* 2015;10(5):e0140822.
22. Bohannon RW, Kindig J, Sabo G, Duni AE, Cram P. Isometric knee extension force measured using a handheld dynamometer with and without belt-stabilization. *Physiother Theor Pract.* 2012;28(7):562-568.
23. Plotkin SR, Davis SD, Robertson KA, et al. Sleep and pulmonary outcomes for clinical trials of airway plexiform neurofibromas in NF1. *Neurology.* 2016;87(7):S13-S20.