Validating Techniques for Measurement of Cutaneous Neurofibromas

Recommendations for Clinical Trials

Raquel D. Thalheimer, BA, Vanessa L. Merker, PhD, K. Ina Ly, MD, Amanda Champlain, MD,
Jennifer Sawaya, MD, Naomi L. Askenazi, Hamilton P. Herr, BA, Jennifer L.W. Da, BA, Justin T. Jordan, MD, MPH,
Alona Muzikansky, MA, Elizabeth Morehouse Pearce, BA, Fernanda H. Sakamoto, MD, PhD,
Jaishri O. Blakeley, MD, R. Rox Anderson, MD, and Scott R. Plotkin, MD, PhD,
on behalf of the REiNS International Collaboration

Correspondence
Dr. Plotkin
splotkin@partners.org

Abstract

Objective
To assess the reliability and variability of digital calipers, 3D photography, and high-frequency ultrasound (HFUS) for measurement of cutaneous neurofibromas (cNF) in patients with neurofibromatosis type 1 (NF1).

Background
cNF affect virtually all patients with NF1 and are a major source of morbidity. Reliable techniques for measuring cNF are needed to develop therapies for these tumors.

Methods
Adults with NF1 were recruited. For each participant, 6 cNF were assessed independently by 3 different examiners at 5 different time points using digital calipers, 3D photography, and HFUS. The intraclass correlation coefficient (ICC) was used to assess intrarater and interrater reliability of linear and volumetric measurements for each technique, with ICC values >0.90 defined as excellent reliability. The coefficient of variation (CV) was used to estimate the minimal detectable difference (MDD) for each technique.

Results
Fifty-seven cNF across 10 participants were evaluated. The ICC for image acquisition and measurement was >0.97 within and across examiners for HFUS and 3D photography. ICC for digital calipers was 0.62–0.88. CV varied by measurement tool, linear vs volumetric measurement, and tumor size.

Conclusions
HFUS and 3D photography demonstrate excellent reliability whereas digital calipers have good to excellent reliability in measuring cNF. The MDD for each technique was used to create tables of proposed thresholds for investigators to use as guides for clinical trials focused on cNF size. These criteria should be updated as the performance of these end points is evaluated.

Glossary

cNF = cutaneous neurofibroma; **CV** = coefficient of variation; **HFUS** = high-frequency ultrasound; **ICC** = intraclass correlation coefficient; **MDD** = minimal detectable difference; **NF1** = neurofibromatosis type 1; **REiNS** = Response Evaluation in Neurofibromatosis and Schwannomatosis.

Neurofibromatosis type 1 (NF1) is an autosomal dominant genetic disorder characterized by multiple nervous system tumors, including cutaneous neurofibromas (cNF), which affect >90% of adults with NF1.[1] Although these tumors do not have malignant potential, they often cause pain, itching, and disfigurement, and are associated with impaired quality of life.[2,3] Most cNF are small (typically 1–30 mm) but they can number in the thousands, involving all areas of skin and resulting in life-altering disfigurement.[4] They develop throughout life, and there are no known effective, nonsurgical therapies to treat or prevent these lesions. The mainstay of treatment is either surveillance or surgical excision of individual tumors. Beyond surgery, modalities that have been used for treatment include electrodessication, lasers, radiofrequency ablation, and investigational topical and systemic drug therapies.[5-12] However, none of these modalities have been assessed in multicenter studies, and there is no standard approach to response evaluation for cNF.[13,14]

A major limitation in developing new treatments for cNF is the inability to reliably measure these tumors. Previous studies have explored the reliability of calipers to assess the size of plaster models of cNF and of high-frequency ultrasound (HFUS) to identify ultrasonographic features of cNF.[15,16] Digital photography has been used to assess cNF improvement in clinical trials (NCT01031901). No data on the reliability of these techniques in persons with NF1 have been published. This study assessed the feasibility, reliability, and variability of digital calipers, 3D photography, and HFUS to measure cNF in clinical trials.

Methods

Study Design

Participants were identified and recruited through the Neurofibromatosis Clinic at Massachusetts General Hospital in Boston. This study was approved by the Massachusetts General Hospital institutional review board. Inclusion criteria included a diagnosis of NF1, age ≥18 years, presence of at least 6 visible cutaneous neurofibromas, and ability to give informed consent in English. Exclusion criteria included inability to tolerate imaging procedures. The primary end point was intrarater and interrater reliability of HFUS for cNF. Secondary end points included ability of HFUS to detect change in cNF size over time, feasibility of HFUS, intrarater and interrater reliability of digital calipers and 3D photography, and patient-reported outcomes.

To assess intrarater and interrater reliability, 3 examiners assessed 6 cNF in 10 participants at 3 time points at least 15 minutes apart on the same day using digital calipers, 3D photography, and HFUS. The same 6 cNF were measured to assess longitudinal changes in cNF size across 4 study visits over the course of 12 months: months 0 (baseline), 4, 8, and 12. Participants completed quality of life surveys including Skindex-29,[17] Impact of NF1 on Quality of Life,[18] and 5-D Pruritus Scale at baseline and month 12. The analysis of reliability and variability in the baseline assessment of cNF is presented in this article. The feasibility of techniques was assessed by recording the time spent performing HFUS, 3D camera image analysis, and digital caliper assessment. These times were averaged for each patient (6 tumors per patient) and for each tumor, including setup. Cost of acquiring hardware and software was used as a metric of accessibility of techniques.

Image Acquisition

Prior to imaging, 6 nonpedunculated cNF were selected on the participant's arms to minimize artifact from movement and for patient convenience. Two methods were used to track individual cNF across study visits: (1) a permanent map was created by marking the location of cNF on a plastic sheet placed over the participant's arm and (2) a 2D photograph of the anatomic region with the marked cNF targets in relation to other anatomical landmarks—for example, a nevus, birthmark, or scar.

High-Frequency Ultrasound

HFUS imaging was performed using the Vevo3100 System (Fujifilm VisualSonics). To image cNF, 12–25 mL of Aquasonic ultrasound gel (Cardinal Health) were placed over the tumor, and the transducer (range, 22–55 MHz; center frequency 30 MHz) was positioned directly above the tumor using a tripod for stabilization. All images were reviewed manually for quality prior to inclusion in the analysis.

3D Photography

3D photographs were obtained using Vectra H1 3D camera (Canfield Scientific) according to the manufacturer's recommendation.

Digital Calipers

cNF were measured with digital calipers (World Precision Instruments). The width, length, and height of each tumor were measured and recorded on a paper form. To assess tumor width, the caliper was adjusted to measure the maximum distance parallel to the participant's wrist. The caliper was zeroed, and the forks were adjusted so that they included the visible tumor, but not the surrounding skin. For elevated tumors, the caliper was placed flush to the skin and adjusted

until the forks touched the outer edges but did not distort tumor shape. The same process was repeated for the length measurement, which was the maximum distance perpendicular to the width. To assess height, a tongue depressor was positioned so that the base touched the surrounding unaffected skin and the side was even against the tumor. The maximum height of the tumor was marked against the tongue depressor and measured using digital calipers.

Image Analysis

Ultrasound images and 3D photographs were analyzed offline using standard software according to the manufacturer's recommendations (FujiFilm Visual Sonics for ultrasound and Vectra H1 software for 3D photographs). For HFUS, volume, width, and depth of cNF were measured; for 3D photographs, volume, surface area, width, and length were measured. Linear measures (length, width) for 3D photography were calculated using both automated and manual measurements to assess maximum cross width and a corresponding perpendicular length (table 1). For digital calipers, tumor width, length, and height were measured, and an approximate ellipsoid volume was calculated for each cNF. An example of measurement techniques is shown in figure 1.

Statistical Assessment

The Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) International Collaboration has previously recommended percent change from baseline (rather than absolute change from baseline) to assess imaging response in NF.[19] One consequence of this approach is that a given change in size represents a larger percent change for small cNF than for large cNF. In order to study this issue, we used a statistical definition to define small and large tumors. The cutoff value was defined as the median tumor diameter rounded up to the nearest integer (in millimeters). Large tumors were defined as ≥5 mm in maximum diameter and small tumors were defined as <5 mm in maximum diameter, as measured by calipers.

The intrarater and interrater reliability of HFUS, 3D photography, and digital calipers for the first 10 participants was assessed using intraclass correlation coefficients (ICCs). ICCs range from 0 to 1 with 0.9–1.0 considered excellent, 0.75–0.89 good, 0.5–0.74 moderate, and <0.5 poor reliability. The ICC was calculated for 2 possible sources of variability in the measurements: image acquisition and tumor measurement. This approach yielded 4 ICC measures per technique (i.e., intrarater and interrater reliability for image acquisition and tumor measurement). Digital calipers did not have ICC values for image acquisition because acquisition and analysis occur simultaneously.

The coefficient of variation (CV) was used to compare the variability of 1D (linear), 2D (area), and 3D (volume) measurement of HFUS, 3D photography, and digital calipers. We defined the minimal detectable difference (MDD) for each technique as at least twice the CV to detect the smallest change that is not the result of measurement variability. The MDD was used to create tables of proposed thresholds for investigators to use in clinical trials directed at evaluating cNF response for therapeutic interventions.

Data Availability

Nonidentifying clinical data will be uploaded to the Synapse online data-sharing platform (synpase.org) maintained by Sage Bionetworks.

Results

Study Cohort

The median age of the 10 participants was 53 years (range, 36–67 years). Half of participants were female, and the majority were non-Hispanic and White (table 2). In total, 57 tumors were included in this analysis (the first patient had 2 tumors that were believed to represent scar on image analysis). The median tumor diameter was 4.2 mm (range 2.4–11.4 mm). There were 24 (42%) large tumors (≥5 mm in maximum diameter) and 33 (58%) small tumors (<5 mm in maximum diameter) included in the analysis.

Reliability of Measurement Techniques

Figure 2 summarizes the intrarater and interrater ICCs of the 3 measurement techniques. All ICC values for HFUS were consistently excellent (ICC ≥0.9). Similarly, nearly all ICC

Table 1 Comparison of Measures Determined by Digital Calipers, 3D Photography, and High-Frequency Ultrasound (HFUS)

Dimension	Caliper	Photography (manual measurement)	Photography (automated measurement)	HFUS
Linear (1D)	Width	Width	Width	Width
	Length	Length	Length	—
	Height (above skin)	—	—	Depth (above and below skin)
Surface area (2D)	—	—	Surface area	—
Volume (3D)	Ellipsoid volume	Volume	—	Volume

Figure 1 Representative Images for Measurement of Cutaneous Neurofibroma (cNF)
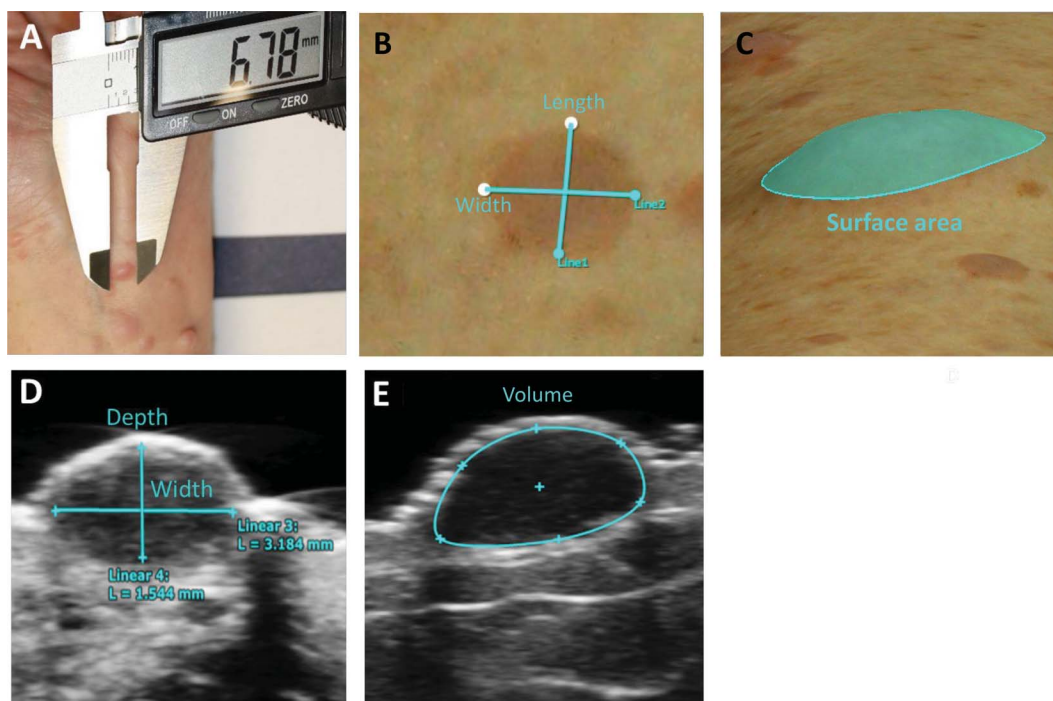
(A) Measurement of cNF using digital calipers. (B) Manual length and width using 3D photography. (C) Automated length, width, and surface area using 3D photography. (D) Depth and width using high-frequency ultrasound (HFUS). (E) Volume using HFUS. Hypoechoic (dark) areas on HFUS correspond to cNF.

Table 2 Demographics and Tumor Characteristics of Participants

Feature	Value
Participants, n	10
Age, y	53 (36–67)
Sex (female)	5 (50)
Ethnicity	
Non-Hispanic	8 (80)
Hispanic	1 (10)
Declined	1 (10)
Race	
White	9 (90)
Asian	1 (10)
Cutaneous neurofibromas measured, n	57
Diameter, mm	4.2 (2.4–11.4)
<5	33 (58)
≥5	24 (42)

Values are median (range) or n (%).

values for 3D photography were excellent except for interrater ICC of image acquisition using automated length measurements (ICC 0.87; good). The lowest ICCs were observed for digital calipers. Because acquisition and analysis for digital calipers occur simultaneously, only analysis reliability is reported for this method. The intrarater ICC values for digital caliper were excellent (ICC ≥0.9) except for height (ICC 0.81; good). By contrast, the interrater ICC values were lower; good (ICC 0.85–0.88) except for height (ICC 0.62; moderate).

Minimal Detectable Difference for Each Technique

The CV and MDD for the 3 measurement techniques are listed in figure 3. In general, linear measures had smaller CV than volumetric measures for all 3 techniques, ranging from 5.2% to 13.4% depending on technique, tumor size, and type of linear measurement. These smaller CV values produce smaller MDD values as well. The CV for linear measures was smaller for 3D photography and HFUS than for digital calipers and was smaller for measuring cNF ≥5 mm than for cNF <5 mm.

Time for Image Acquisition and Image Analysis

For image acquisition, the median time required per patient was 5 minutes for HFUS, less than 30 seconds for 3D photography, and 6 minutes for digital calipers. For image analysis, the median time per patient was as follows: 7 minutes for HFUS

Figure 2 Intraclass Correlation Coefficient (ICC) Results for Caliper Measurements, 3D Photography, and High-Frequency Ultrasound (HFUS)

	Image Acquisition Reliability		Image Analysis Reliability	
	Intra-rater ICC	Inter-rater ICC	Intra-rater ICC	Inter-rater ICC
HFUS				
Volume	0.98	0.97	0.99	0.98
Width	0.98	0.98	0.98	0.99
Depth	0.98	0.97	0.99	0.97
3D Camera				
Volume	0.97	0.95	0.98	0.98
Width (manual)	0.97	0.96	0.97	0.98
Width (automated)	0.96	0.96	0.98	0.97
Length (manual)	0.97	0.96	0.98	0.98
Length (automated)	0.97	0.87	0.91	0.93
Surface Area	0.98	0.97	0.99	0.98
Calipers				
Volume			0.9	0.77
Width			0.96	0.88
Length			0.93	0.85
Height			0.81	0.62

ICC	Reliability
<.5	Poor
0.5-.75	Moderate
0.75-0.9	Good
0.9-1.0	Excellent

Image acquisition refers to the real-time process of acquiring images. Image analysis was performed offline after image acquisition. Reliability was assessed for both steps.

linear analysis (range 5.0–10.5 minutes), 21.5 minutes for HFUS volume analysis (range 16.5–45.5 minutes), 8.0 minutes for 3D photography–based manual linear analysis (range 5.0–11.0 minutes), 10.0 minutes for 3D photography–based automated linear analysis (range 6.0–12.0 minutes), and 9.5 minutes for 3D photography–based volume analysis (range 7.0–15.0 minutes). No time was required for analysis of digital caliper–based measurements. Notably, image analysis for both HFUS and 3D photography yields durable data files with raw and processed images. There are no durable data files for digital calipers.

Cost of Hardware and Software

The cost of the hardware and software used in this study was approximately $225,000 for HFUS, $13,400 for 3D photography, and $100 for digital calipers.

Discussion

We assessed 3 techniques for measurement of the cNF size in adults with NF1 with the goal of establishing metrics to support their use in clinical trials. Each technique has strengths and weaknesses for consideration by investigators. HFUS and 3D photography had excellent reliability for image acquisition and analysis within raters and across different raters. In contrast, caliper measurements showed less reliability with weakest performance when measuring the height of cNF both within and across raters.

We then evaluated the variability of measurements to propose thresholds for each technique that would allow accurate assessment of tumor growth/shrinkage. First, the variability of measurement for small cNF (<5 mm) was greater than for

Figure 3 Variability of Measurement as Estimated by Coefficient of Variation (CV) for Caliper Measurements, 3D Photography, and High-Frequency Ultrasound (HFUS)

	Small cNF CV	Large cNF CV	Proposed threshold for small cNF	Proposed threshold for large cNF
Calipers				
Width	19.0%	10.6%	40%	25%
Length	17.8%	13.0%		
Height	38.7%	25.0%	75%	50%
Volume	63.2%	33.4%	n/a	70%
3D Camera				
Width (manual)	11.80%	7.00%	25%	25%
Width (automated)	10.2%	7.6%		
Length (manual)	10.7%	6.6%		
Length (automated)	10.9%	9.9%		
Surface Area	18.7%	12.2%	40%	25%
Volume	57.2%	18.6%	n/a	40%
HFUS				
Width	13.4%	5.9%	30%	15%
Depth	11.6%	5.2%		

Threshold	
≥35%	
30-34%	
26-30%	
21-25%	
16- 20%	
11-15%	
≤ 11%	

Small cutaneous neurofibroma (cNF) are <5 mm in diameter; large cNF are ≥5 mm in diameter. The minimal detectable difference (MDD) for each technique was defined as at least twice the CV. The MDD was used to propose thresholds for investigators to use in clinical trials directed at evaluating imaging response for interventions.

large cNF (≥5 mm), raising concerns about accurate measurement of small tumors in clinical trials designed to shrink tumors. Second, HFUS and 3D photography have less variability in measurement than digital calipers and are amenable to central review, which can remove the potential for reader bias and promote standardization of response assessment. Third, volume measurements demonstrated greater variation than linear measures across all measurement tools: the overall CV for cNF volume using HFUS, 3D photography, and digital caliper was 29.5%, 43.2%, and 51.8%, as compared with CV for linear measures of 15%, 10%, and 10%. Measuring longest diameter parallel to the skin has the least variability and can be used for linear or bidimensional measurements of cNF.

Equipment cost, required training, and execution time should be considered when selecting the most appropriate measurement tool and measurement type for a cNF clinical trial. 3D photography was the most feasible technique for participants, as assessed by time required for image acquisition; digital calipers were most feasible for investigators as assessed by time required for image analysis. The cost of the HFUS machine (approximately $225,000) is prohibitive for many research groups. The 3D camera was more moderately priced at $13,400 (including laptop and software), and digital calipers were affordable (approximately $100). In addition, use of HFUS equipment requires substantial training. To assure quality results, researchers must be trained in appropriate ultrasound technique and image optimization, in addition to learning the corresponding Vevo measurement software. However, our results indicate that raters can achieve excellent reliability (ICC ≥0.97) following appropriate training. Image acquisition using the 3D camera requires training as well, but significantly less than for HFUS. The 3D photography software requires training, but researchers can also achieve excellent reliability (ICC ≥0.91) with appropriate training. Digital calipers require minimal training and do not require postvisit analysis. However, while intrarater reliability is in the good to excellent range (ICC ≥0.81), the interrater reliability of image analysis falls within the moderate to good range (ICC ≥0.62).

The goal of developing cNF response criteria is to detect the smallest clearly measurable difference in size of cNF as a sign of activity of an investigational agent or device. Based on these findings and in consultation with the REiNS cNF working group (see Cannon et al. in this supplement), we developed the following recommendations for assessing cNF in clinical trials.

General Recommendations for Measuring cNF

Measurable cNF
Measurable cNF must be surrounded by normal-appearing skin to demarcate the lesion. For example, cNF abutting other cNF or those obscured by rashes or hair are not considered measurable. Pedunculated cNF with a narrow stalk and broad body are not suitable for study given their irregular shape that is difficult to assess by current measurement techniques.[20]

Tracking Individual cNF in Clinical Trials for Accurate Repeat Measures
Given the large number of cNF in most patients with NF1, a tracking system is essential to ensure that the same cNF are measured throughout the study. In our study, the use of 2D photography to document target cNF in relation to cutaneous landmarks in combination with a permanent map created by marking the location of cNF on a plastic sheet template placed over the participant's arm was effective in tracking individual cNF. These techniques, or others as appropriate, should be defined by the study prior to enrollment of participants.

Minimizing Variability
The group does not recommend the use of double baseline measurements for cNF trials but does recommend using the same technician to acquire images or caliper measurements throughout the trial and use of central review for image analysis of 3D photographs and HFUS images.

Confirmatory Measurement
Confirmation of cNF response at least 4 weeks after response is desirable to minimize the rate of false-positive or false-negative responses.

Recommended Outcomes for HFUS
For clinical trial outcomes, the group considered the following measures as end points: width and depth (linear measures) and volume (table 1). The group recommended HFUS for measurement of small and large cNF given the low variability for all cNF sizes. The group recommended use of HFUS linear measures (depth or width) as outcome measures because (1) depth is a clinically meaningful measure of cNF size that may be particularly important to patients; (2) HFUS is the only technique that can measure the amount of cNF beneath the skin; (3) variability of cNF volume was high, limiting the use of this variable in clinical trials; and (4) HFUS images can be reviewed centrally.

Feasibility
The cost of HFUS devices limits the availability of this technique in NF clinics. Furthermore, the technique cannot be easily integrated into a clinic setting given the size of the device. Training of personnel is essential to minimize variability, and measurements should ideally be performed by the same individual throughout the study.

Baseline Evaluations
The working group recommended that a minimum of 5 cNF should be identified as target lesions and recorded and measured at baseline. The maximum depth and width for each lesion should be recorded. Each study protocol should identify whether maximum depth or width will be used as a reference by which to characterize individual cNF responses; the same dimension should be measured over time to determine response. The orthogonal dimension (e.g., maximum width if maximum depth is used as reference) should also be used as a reference for an exploratory outcome in the trial. In addition, the sum of depths or widths for all target lesions will be

calculated and reported as the baseline sum depth for exploratory analyses. The baseline sum longest diameter will be used as the reference by which to characterize the collective tumor response.

Response Criteria

HFUS response criteria are defined in reference to the baseline maximum depth or width of individual cNF at study initiation. Our study suggests that cNF <5 mm at baseline (small cNF) demonstrate more variability in measurement than cNF larger than 5 mm (large cNF). Thus, the group recommended a greater percentage change in maximum depth or width for small cNF (30%) than for large cNF (15%) in order to minimize the rate of false-positive or false-negative responses. In addition, the group recommends using change in the sum of maximum depth or width as an exploratory end point.

cNF shrinkage is defined as a relative decrease in the width or depth of 30% for small cNF and of 15% for large cNF. cNF growth is defined as a relative increase in the longest diameter of 30% for small cNF and of 15% for large cNF. Stable size is defined as all other changes. For studies using the sum of longest diameters, cNF shrinkage or growth is defined as a relative decrease or increase, respectively, of 15% in the sum of longest diameters.

Special Notes on High-Frequency Ultrasound

This technique is uniquely able to image tumor beneath the skin and is well suited for nascent cNF that are minimally raised and not easily measured by calipers or 3D photography. Due to its low variability, HFUS is the only technique recommended for use in cNF smaller than 5 mm in longest diameter. Hence, this may be a technique that is particularly valuable in prevention studies. It is unknown whether changes in width or depth are more closely associated with patient benefit. Thus, the proposed response criteria allow individual studies to choose either width or depth to measure as primary outcome for individual cNF.

Recommended Outcomes for 3D Photography

For clinical trial outcomes, the group considered the following measures as end points: manual and automated measurement of length/width (linear measures), surface area, and volume (table 1). The group recommended use of 3D photography to measure surface area as an outcome measure for large cNF (≥5 mm) because (1) surface area is a reliable and clinically meaningful measure of cNF size; (2) 3D photographs are easy to acquire; (3) 3D photographs can be reviewed centrally; and (4) 3D photography can be used to monitor a large number of cNF with minimal effort for participant or examiner. In addition, the group recommended collecting automated length or width (linear measure) and volume of cNF as exploratory end points. Measurement of surface area was preferred over linear measures as the former incorporates elevation above the skin, an important component of tumor visibility, and over volumetric measures given the reduced variability of measuring surface area. For small cNF (<5 mm), the group recommended use of linear measures as an outcome; this

was preferred over surface area or volumetric measures given the reduced variability of measuring linear dimensions.

Feasibility

3D cameras are widely available, of moderate cost for clinical research, and can be easily integrated into a clinic setting. In order to minimize variability, measurements should be performed on cNF ≥5 mm by trained personnel and analyzed by a central facility.

Baseline Evaluation

The working group recommended that a minimum of 5 lesions should be identified as target lesions and recorded and measured at baseline. The surface area for each lesion will be measured using the manufacturer's software and recorded. The individual surface areas will be used as a reference by which to classify individual cNF response. In addition, the longest diameter and volume for each lesion and the sum of surface areas for all target lesions will be calculated. The individual longest diameters, volume, and baseline sum surface area will be used as the reference by which to characterize the collective tumor response for exploratory studies.

Response Criteria

3D photography response criteria are defined in reference to the baseline surface area of individual cNF at study initiation. The group recommends change in surface area as the primary imaging end point (table 2). The group recommends using change in longest diameter for each lesion, in volume, and in the sum of surface area for all cNF as exploratory end points. In addition, the group recommended studying changes in appearance (e.g., coloration) as an outcome measure.

cNF shrinkage is defined relative to baseline as a decrease in surface area of 25%. cNF growth is defined relative to baseline as an increase in the surface area of 25%. Stable size is defined as all other changes. For studies using longest diameter, cNF shrinkage or growth is defined as a relative decrease or increase, respectively, of 25% compared with baseline. For studies using volume, cNF shrinkage or growth is defined as a relative decrease or increase, respectively, of 70% compared with baseline. For studies using the sum of longest diameters, cNF shrinkage or growth is defined as a relative decrease or increase, respectively, of 25% compared with baseline.

Special Notes on 3D Photography

3D photography is performed in real time and generates a permanent record that is ideal for central review, for measuring and counting large numbers of tumors, for identifying new tumors in a large field, and for capturing the overall appearance of tumors. This photographic record can be used to identify novel end points of response such as global assessment of change. However, the technique cannot detect or quantitate the amount of tumor below the skin.

Recommended Outcomes for Digital Calipers

For clinical trial outcomes, the group considered the following measures as end points: length/width (linear measures) and

calculated ellipsoid volume (table 1). The group recommended use of longest diameter (linear measure) as an outcome measure because it (1) is a clinically meaningful measure of cNF size; (2) is widely available at low cost; and (3) is the most used end point in previous studies of cNF. Digital calipers were not recommended for measurement of cNF <5 mm or for measurement of height in any cNF given the increased variability using this technique.

Feasibility

Digital calipers are widely available, and evaluation can be performed routinely in a clinic setting. In order to minimize variability, measurements should be performed by trained personnel, ideally by the same individual throughout the study.

Baseline Evaluation

A minimum of 5 lesions representative of cNF should be identified as target lesions and recorded and measured at baseline. The longest diameters for each lesion should be recorded. The individual longest diameters will be used as a reference by which to characterize individual cNF response. In addition, the sum of the longest diameter for all target lesions will be calculated and reported as the baseline sum longest diameter. The baseline sum longest diameter will be used as the reference by which to characterize the collective tumor response.

Response Criteria

Caliper response criteria are defined in reference to the baseline longest diameter of individual cNF at study initiation. The group recommends change in longest dimension as the primary end point (table 2) and the change in the sum of longest diameter as an exploratory end point.

cNF shrinkage is defined as a relative decrease in the longest diameter of 25%. cNF growth is defined as a relative increase in the longest diameter of 25%. Stable size is defined as all other changes. For studies using the sum of longest diameters, cNF shrinkage or growth is defined as a relative decrease or increase, respectively, in the sum of longest diameters of 25%.

Special Notes on Caliper Measurements

Caliper measurements are easily used in clinical studies. However, they are performed in real time and do not generate a permanent record. Thus, this technique does not permit central review. In addition, the technique cannot detect or qualify the amount of tumor component below the skin, or changes in appearance (e.g., coloration) during treatment. The technique is time-consuming for measurement of large numbers of tumors across the body.

Overall Response Criteria and Duration of Treatment

The REiNS International Collaboration is not recommending overall response criteria given the lack of real-world clinical trial data for cNF assessments. Because the suggested response criteria for cNF trials involve at least 5 cNF for each participant, it is possible that individual cNF will have discordant responses to investigational agents. In cancer trials, overall response is used to determine the duration of treatment: patients with imaging response or stable disease typically continue treatment whereas those with progressive disease discontinue treatment. In contrast to most cancers, growth of individual cNF do not place participants at significantly increased risk of morbidity or mortality because these lesions do not affect critical structures and because surgical removal is an option for virtually all cNF. At this time, the criteria for removal of patients due to tumor growth are not clearly defined. In the absence of clear criteria, trialists should consider the safety profile and existing data on the median time to maximum response of investigational agents. We do not recommend the absolute requirement to remove participants from trials when cNF progress on therapy, particularly in the situation when there is a discordant response (i.e., when some cNF shrink and others grow compared with baseline). Instead, we recommend that trials consider specifying a fixed treatment duration based on best available information about the investigational agent under study.

These recommendations are designed to establish end points for clinical trials that seek to measure the size of cNF in people with NF1. It is hoped that these end points accelerate clinical trials for cNF and facilitate comparison across studies to identify active treatments. The recommended outcomes have not been prospectively used in cNF trials to date and the REiNS cNF Working Group expects to revise these recommendations as data from trials are published.

Study Funding

Disclosure

Publication History

Appendix Authors

Name	Location	Contribution
Raquel D. Thalheimer, BA	Massachusetts General Hospital, Boston	Acquisition of data; design of the study; interpretation of the data; drafting the manuscript for intellectual content
Vanessa L. Merker, PhD	Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA	Acquisition of data; design of the study; interpretation of the data; revising the manuscript for intellectual content
K. Ina Ly, MD	Massachusetts General Hospital, Boston	Acquisition of data; revising the manuscript for intellectual content
Amanda H. Champlain, MD	Massachusetts General Hospital, Boston	Acquisition of data

Appendix (continued)

Name	Location	Contribution
Jennifer Sawaya, MD	Massachusetts General Hospital, Boston	Acquisition of data
Naomi L. Askenazi	Massachusetts General Hospital, Boston	Analysis and interpretation of the data; revising the manuscript for intellectual content
Hamilton Herr, BA	Massachusetts General Hospital, Boston	Acquisition of data; analysis and interpretation of the data; revising the manuscript for intellectual content
Jennifer L.W. Da, BA	Massachusetts General Hospital, Boston	Acquisition of data; analysis and interpretation of the data; revising the manuscript for intellectual content
Justin T. Jordan, MD, MPH	Massachusetts General Hospital, Boston	Design of the study; recruitment of participants; revising the manuscript for intellectual content
Alona Muzikansky, MA	Massachusetts General Hospital, Boston	Design of the study; interpretation of the data; revising the manuscript for intellectual content
Elizabeth Morehouse, BA	Massachusetts General Hospital, Boston	Acquisition of data; revising the manuscript for intellectual content
Fernanda H. Sakamoto, MD, PhD	Massachusetts General Hospital, Boston	Design of the study; interpretation of the data; revising the manuscript for intellectual content
Jaishri O. Blakely, MD	Johns Hopkins School of Medicine, Baltimore, MD	Design and conceptualization of the study; interpretation of the data; revising the manuscript for intellectual content
R. Rox Anderson, MD	Massachusetts General Hospital, Boston	Design and conceptualization of the study; interpretation of the data; revising the manuscript for intellectual content
Scott R. Plotkin, MD, PhD	Massachusetts General Hospital, Boston	Design and conceptualization of the study; interpretation of the data; recruitment of participants; drafting the manuscript for intellectual content

References

1. Blakeley JO, Wolkenstein P, Widemann BC, et al. Creating a comprehensive research strategy for cutaneous neurofibromas. *Neurology.* 2018;91(2 suppl 1):S1–S4.
2. Jouhilahti EM, Peltonen S, Callens T, et al. The development of cutaneous neurofibromas. *Am J Pathol.* 2011;178(2):500–505.
3. Duong TA, Bastuji-Garin S, Valeyrie-Allanore L, Sbidian E, Ferkal S, Wolkenstein P. Evolving pattern with age of cutaneous signs in neurofibromatosis type 1: a cross-sectional study of 728 patients. *Dermatology.* 2011;222(3):269–273.
4. Page PZ, et al.. Impact of neurofibromatosis 1 on quality of life: a cross-sectional study of 176 American cases. *Am J Med Genet A.* 2006;140A:1893–1898.
5. Levine SM, Levine E, Taub PJ, Weinberg H. Electrosurgical excision technique for the treatment of multiple cutaneous lesions in neurofibromatosis type I. *J Plast Reconstr Aesthet Surg.* 2008;61(8):958–962.

6. Lutterodt CG, Mohan A, Kirkpatrick N. The use of electrodessication in the treatment of cutaneous neurofibromatosis: a retrospective patient satisfaction outcome assessment. *J Plast Reconstr Aesthet Surg.* 2016;69(6):765–769.

7. Elwakil TF, Samy NA, Elbasiouny MS. Non-excision treatment of multiple cutaneous neurofibromas by laser photocoagulation. *Lasers Med Sci.* 2008;23(3):301–306.

8. Méni C, Sbidian E, Moreno JC, et al. Treatment of neurofibromas with a carbon dioxide laser: a retrospective cross-sectional study of 106 patients. *Dermatology.* 2015;230(3):263–268.

9. Yoshida Y, Sato N, Furumura M, Nakayama J. Treatment of pigmented lesions of neurofibromatosis 1 with intense pulsed-radio frequency in combination with topical application of vitamin D3 ointment. *J Dermatol.* 2007;34(4):227–230.

10. Slopis JM, Arevalo O, Bell CS, et al. Treatment of disfiguring cutaneous lesions in neurofibromatosis-1 with everolimus: a phase II, open-label, single-arm trial. *Drugs R D.* 2018;18(4):295–302.

11. Kim SH, Roh SG, Lee NH, Yang KM. Radiofrequency ablation and excision of multiple cutaneous lesions in neurofibromatosis type 1. *Arch Plast Surg.* 2013;40(1):57–61.

12. Koenig MK, Bell CS, Hebert AA, et al. Efficacy and safety of topical rapamycin in patients with facial angiofibromas secondary to tuberous sclerosis complex: the TREATMENT randomized clinical trial. *JAMA Dermatol.* 2018;154(7):773–780.

13. Ortonne N, Wolkenstein P, Blakeley JO, et al. Cutaneous neurofibromas: current clinical and pathologic issues. *Neurology.* 2018;91(2 suppl 1):S5–S13.

14. Verma SK, Riccardi VM, Plotkin SR, et al. Considerations for development of therapies for cutaneous neurofibroma. *Neurology.* 2018;91(2 suppl 1):S21–S30.

15. Cannon A, Chen MJ, Li P, et al. Cutaneous neurofibromas in neurofibromatosis type I: a quantitative natural history study. *Orphanet J Rare Dis.* 2018;13(1):31.

16. Raffin D, Zaragoza J, Georgescou G, et al. High-frequency ultrasound imaging for cutaneous neurofibroma in patients with neurofibromatosis type I. *Eur J Dermatol.* 2017;27(3):260–265.

17. Chren MM, Lasek RJ, Flocke SA, et al. Improved discriminative and evaluative capability of a refined version of Skindex, a quality-of-life instrument for patients with skin diseases. *Arch.Dermatol.* 1997;133(11):1433–1440.

18. Ferner RE, Thomas M, Mercer G, et al. Evaluation of quality of life in adults with neurofibromatosis 1 (NF1) using the Impact of NF1 on Quality of Life (INF1-QoL) questionnaire. *Health Qual Life Outcomes.* 2017;15(1):34.

19. Dombi E, Ardern-Holmes SL, Babovic-Vuksanovic D, et al. Recommendations for imaging tumor response in neurofibromatosis clinical trials. *Neurology.* 2013;81(21 suppl 1):S33–S40.

20. Posada JG, Chakmakjian CG. Images in clinical medicine: Von Recklinghausen's disease and breast cancer. *N Engl J Med.* 2005;352(17):1799.

Neurology®

Validating Techniques for Measurement of Cutaneous Neurofibromas: Recommendations for Clinical Trials

Raquel D. Thalheimer, Vanessa L. Merker, K. Ina Ly, et al.

This information is current as of July 6, 2021

Updated Information & Services	including high resolution figures, can be found at: http://n.neurology.org/content/97/7_Supplement_1/S32.full
References	This article cites 20 articles, 4 of which you can access for free at: http://n.neurology.org/content/97/7_Supplement_1/S32.full#ref-list-1
Citations	This article has been cited by 1 HighWire-hosted articles: http://n.neurology.org/content/97/7_Supplement_1/S32.full##otherarticles
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): **Neurofibromatosis** http://n.neurology.org/cgi/collection/neurofibromatosis **Ultrasound** http://n.neurology.org/cgi/collection/ultrasound
Permissions & Licensing	Information about reproducing this article in parts (figures,tables) or in its entirety can be found online at: http://www.neurology.org/about/about_the_journal#permissions
Reprints	Information about ordering reprints can be found online: http://n.neurology.org/subscribers/advertise

AMERICAN ACADEMY OF NEUROLOGY ®