

Current Recommendations for Patient-Reported Outcome Measures Assessing Domains of Quality of Life in Neurofibromatosis Clinical Trials

Pamela L. Wolters, PhD, Ana-Maria Vranceanu, PhD, Heather L. Thompson, PhD, CCC-SLP, Staci Martin, PhD, Vanessa L. Merker, PhD, Andrea Baldwin, CRNP, Carolina Barnett, MD, PhD, Kimberley S. Koetsier, MD, Cynthia M. Hingtgen, MD, PhD, Christopher J. Funes, MS, James H. Tongsgard, MD, Elizabeth K. Schorry, MD, Taryn Allen, PhD, Taylor Smith, PhD, Barbara Franklin, BS, and Stephanie Reeve, BSN, RN, on behalf of the REiNS International Collaboration

Correspondence

Dr. Wolters
woltersp@mail.nih.gov

Neurology® 2021;97:S50-S63. doi:10.1212/WNL.00000000000012421

Abstract

Objective

To review and recommend patient-reported outcome (PRO) measures assessing multidimensional domains of quality of life (QoL) to use as clinical endpoints in medical and psychosocial trials for children and adults with neurofibromatosis (NF) type 1, NF2, and schwannomatosis.

Methods

The PRO working group of the Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) International Collaboration used systematic methods to review, rate, and recommend existing self-report and parent-report PRO measures of generic and disease-specific QoL for NF clinical trials. Recommendations were based on 4 main criteria: patient characteristics, item content, psychometric properties, and feasibility.

Results

The highest-rated generic measures were (1) the Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales for NF clinical trials for children or for children through adults, (2) the Functional Assessment of Cancer Therapy–General for adult medical trials, and (3) the World Health Organization Quality of Life–BREF for adult psychosocial trials. The highest-rated disease-specific measures were (1) the PedsQL NF1 Module for NF1 trials, (2) the NF2 Impact on Quality of Life Scale for NF2 trials, and (3) the Penn Acoustic Neuroma Quality of Life Scale for NF2 trials targeting vestibular schwannomas. To date, there are no disease-specific tools assessing multidimensional domains of QoL for schwannomatosis.

Conclusions

The REiNS Collaboration currently recommends these generic and disease-specific PRO measures to assess multidimensional domains of QoL for NF clinical trials. Additional research is needed to further evaluate the use of these measures in both medical and psychosocial trials.

From the Pediatric Oncology Branch (P.L.W., S.M.), National Cancer Institute, NIH, Bethesda, MD; Integrated Brain Health Clinical and Research Program (A.-M.V., C.J.F.), Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston; Department of Communication Sciences and Disorders (H.L.T.), California State University, Sacramento; Center for Healthcare Organization and Implementation Research (V.L.M.), Edith Nourse Rogers Memorial Veterans Hospital, Bedford; Clinical Monitoring Research Program Directorate (A.B., T.A.), Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD; Division of Neurology (C.B.), Department of Medicine, University Health Network and University of Toronto, Ontario, Canada; Department of Otolaryngology/Head and Neck Surgery (K.S.K.), Leiden University Medical Center, the Netherlands; Department of Clinical Neurosciences (C.M.H.), Spectrum Health Medical Group and College of Human Medicine, Michigan State University, East Lansing; University of Chicago Pritzker School of Medicine (J.H.T.), IL; Division of Human Genetics (E.K.S.), Cincinnati Children's Hospital, OH; Department of Psychology and Child Development (T.S.), California Polytechnic State University, San Luis Obispo; and REiNS International Collaboration Patient Representative (B.F., S.R.).

Go to [Neurology.org/N](https://www.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

FACT-G = Functional Assessment of Cancer Therapy–General; **MCID** = minimally clinically important difference; **NF** = neurofibromatosis; **NFTI-QoL** = Neurofibromatosis 2 Impact on Quality of Life; **PANQoL** = Penn Acoustic Neuroma Quality of Life; **PedsQL** = Pediatric Quality of Life Inventory; **PedsQL GCS** = PedsQL Generic Core Scales; **PedsQL NF1** = Pediatric Quality of Life Inventory Neurofibromatosis Type 1 Module; **PRO** = patient-reported outcome; **QoL** = quality of life; **REiNS** = Response Evaluation in Neurofibromatosis and Schwannomatosis; **WHOQoL-BREF** = World Health Organization Quality of Life–BREF.

Neurofibromatosis (NF) 1, NF2, and schwannomatosis are a group of neurogenetic disorders that exhibits the predisposition to develop nerve sheath tumors, but each displays its own distinct characteristics. NF1 is associated with plexiform neurofibromas, cutaneous manifestations, gliomas, bone abnormalities, pain, and learning disabilities.¹ NF2 is characterized by bilateral vestibular schwannomas and other benign tumors of the nervous system, as well as complications such as hearing loss or deafness, tinnitus, balance problems, facial paresis, ophthalmic manifestations, and skin lesions.² Schwannomatosis is characterized by peripheral schwannomas and associated chronic disabling pain.² These histologically benign tumors and other complications can cause significant morbidity and negatively affect quality of life (QoL) regardless of age and type of NF.^{3,4}

Clinical trials are underway to evaluate the efficacy of medical therapies to treat the tumors and other clinical manifestations associated with NF,^{5,6} as well as psychosocial interventions to help patients cope with NF-related symptoms and to improve QoL.^{7,8} For individuals with chronic and progressive diseases like NF, it is critically important to assess the effects of the treatment on the patients' symptoms, functioning, and well-being in clinical trials.⁹ While objective trial endpoints (e.g., laboratory tests or imaging analyses) document changes in physiologic disease severity, assessing the patients' perspective about aspects of their functioning and QoL using patient-reported outcome (PRO) measures provides a unique indicator of the beneficial or detrimental effects of treatment beyond biomedical outcomes.^{9,10}

The Federal Drug Administration, European Medicines Agency, and several professional oncology societies support the use of PRO measures in clinical trials and the development of standardized approaches and consensus-based guidance.^{9,10} Harmonization of clinical outcome measures facilitates implementing multicenter studies, comparing findings across similar treatment trials, and pooling data when appropriate,¹¹ which are particularly important in rare diseases like NF. The Response Evaluation in Neurofibromatosis and Schwannomatosis (REiNS) International Collaboration is an interdisciplinary group tasked with reaching a consensus on appropriate endpoints for NF clinical trials. More specifically, the REiNS PRO group is working to provide guidance on the most appropriate PRO measures for medical and psychosocial NF trials. The PRO group previously developed a systematic method for

evaluating available PRO measures for use in NF clinical trials, which was used to formulate recommendations for assessing pain and physical function.^{12,13} In the current study, this method was applied to generic and disease-specific PRO measures assessing multidimensional domains of QoL.

QoL is a multidimensional construct comprising a range of domains such as physical, social, emotional, and role functioning.¹⁴ PRO measures assessing aspects of QoL vary widely in their content and may consist of items related to the perception of one's own functioning, disability, health, and satisfaction with life.¹⁵ Two types of PRO measures evaluate multidimensional domains of QoL: generic and disease specific. Generic measures assess a broad range of concepts in the general population, including individuals with or without chronic illness.¹⁶ Generic scales allow comparison of results between a specific disease and the general public and across different diseases such as when individuals with NF1 and NF2 participate in a single psychosocial trial. However, generic measures often do not capture unique issues specific to a particular disease such as skin conditions, which are important in NF1, or hearing and balance, which are relevant for NF2.

Disease-specific measures, in contrast, are developed for a particular patient population¹⁶ and assess the symptoms and physical functions typically affected by a specific medical condition. Thus, disease-specific measures are often more sensitive to capturing change after interventions aimed to improve specific signs and symptoms.¹⁶ However, disease-specific measures do not allow comparing or combining various diseases (e.g., NF1, NF2) within a psychosocial study.

With many multidimensional PRO measures available, NF researchers have used a number of different existing scales^{3,4} that evaluate varied domains and item content, assess various age ranges, and may not be developed specifically for NF. However, there is a current need to assess clinical outcomes in NF trials and to use consistent measures when feasible to facilitate comparisons across studies. Thus, the primary aim of this study was to identify existing PRO measures assessing multidimensional domains of QoL that are most appropriate to evaluate changes in clinical outcomes in NF trials. This article presents the results of the REiNS PRO working group's efforts examining generic and disease-specific measures using published literature and provides the current REiNS consensus recommendations in these domains to serve as a guide for NF researchers.

Methods

The REiNS PRO group consists of an interdisciplinary group of clinical researchers with expertise in NF and several patient representatives, including adults with NF and caregivers of children with NF. We used a systematic process to review and rate existing PRO measures, as previously described in detail.^{12,13} To summarize, we first discussed and selected the most promising multidimensional generic and disease-specific PRO measures published before May 1, 2019, on the basis of literature reviews. Next, we initially reviewed and rated these measures according to 6 established criteria using our Patient Reported Outcomes–Rating and Acceptance Tool for Endpoints system.¹² These criteria include (1) patient characteristics (age range, normative data), (2) use in published studies (validation, descriptive, clinical trials, and NF clinical trials), (3) domain and item content (content of domains and items, wording, format), (4) scores available (raw, standardized, item, domain), (5) psychometric data (reliability, validity, sensitivity to change, factor analysis), and (6) feasibility (cost, languages, length, ease of administration and scoring, recall period). Each criterion was rated on a scale of 0 (no to poor data or information) to 3 (solid published data and information supporting use in NF trials), which were averaged to produce an overall group rating. To make our final ratings and selections, we predominantly considered 4 of the criteria: patient characteristics, domains and item content, psychometric data, and feasibility for NF clinical trials. When ≥ 2 of the PRO tools in a particular domain were rated the highest with close total scores, we conducted a detailed side-by-side comparison to rereview, rate, and discuss the strengths and limitations of each measure.

We focused our discussions, ratings, and recommendations specific to identifying PRO measures to evaluate change in clinical trials for individuals with NF rather than for descriptive studies or for general chronic illness. Thus, under these 4 main criteria, we took several issues into account when identifying and rating measures. NF trials involving young children or even older individuals with learning disabilities require self-report PRO measures that are easy to understand in addition to parallel parent-report forms to obtain observer-reported outcomes. For trials enrolling children through adults, measures are needed that assess a wide age range with similar domains and items. Of utmost importance, we carefully considered each measure for item content most relevant to individuals with NF and that may demonstrate change with treatment in clinical trials. To be used as trial endpoints, good psychometric properties are imperative. Finally, multicenter and international NF trials are ongoing that necessitate PRO measures in multiple languages.

Results

Table 1 lists all 23 PRO measures evaluated by our group and their strengths and limitations specifically for use in NF

clinical trials according to our extensive reviews and Patient Reported Outcomes–Rating and Acceptance Tool for Endpoints criteria. Below we provide information about the multidimensional generic and disease-specific PRO measures rated the highest and the type of trials and study populations for which they were considered. In addition, tables 2 through 6 present detailed information about each of the PRO measures described below and their final group ratings.

Generic Measures

Due to the wide age range of individuals to be enrolled and the different domains to be assessed when evaluating medical and psychosocial treatments for NF, we present the 3 highest-rated generic measures our group evaluated for adult-only, pediatric-only, and adult-through-pediatric trials.

Adult-Only Clinical Trials

Our group identified 5 adult-only and 5 pediatric-through-adult generic measures assessing domains of QoL as potential tools suitable for use in adult NF clinical trials. After our initial reviews, the 2 highest-rated measures for adult trials were the Functional Assessment of Cancer Therapy–General (FACT-G)¹⁷ and World Health Organization Quality of Life–BREF (WHOQoL-BREF)¹⁴ scales. Side-by-side comparisons of these 2 measures resulted in very similar ratings that informed our recommendations.

Adult Medical Trials

The FACT-G was the highest-rated generic PRO measure assessing multidimensional domains of QoL for use in medical clinical trials for adults with NF (table 2). This measure assesses 4 domains (physical, emotional, social/family, and functional well-being) and consists of 27 items rated on a 5-point Likert scale¹⁷ that were deemed relevant to evaluating meaningful changes in NF treatment trials. Scores are computed for the 4 subscales and summed to produce a total score. The FACT-G has been used extensively in medical trials,¹⁸ including 1 trial in NF1.¹⁹ Moreover, it has strong reliability²⁰ and is valid when administered by paper, interviewer, or electronically.²¹ This scale takes 5 to 10 minutes to complete, is easy to read, has a simple format, and is freely available in >70 languages. Limitations our group identified for using this measure in NF trials are the lack of items assessing cognitive function and the absence of normative and psychometric data for patients with NF.

Adult Psychosocial Trials

The WHOQoL-BREF¹⁴ was the highest-rated generic PRO measure to evaluate clinical changes in psychosocial trials for adults with NF (table 3). It was developed from the parent measure WHOQoL-100 through field testing in 25 countries. The WHOQoL-BREF comprises 4 domains of QoL: physical health, psychological health, social relationships, and environment. In addition, 2 items measure overall QoL and satisfaction with general health. All items are answered on a 5-point Likert scale,²² and within each domain, items are

Table 1 List of PRO Measures Reviewed to Assess Multidimensional Domains of QoL in NF Clinical Trials

Type and name of measure	Age range available	Strengths for use in NF clinical trials ^b	Limitations for use in NF clinical trials ^b
Generic			
Adult-only			
EORTC-QLQ-C30	≥18 y SR	Well studied in Europe and the United States, including in clinical trials; most item content relevant for NF trials, generally good psychometric properties; brief, simple format; multiple languages available	No child or adolescent forms for clinical trials enrolling adult and younger ages combined; not used in NF studies or trials; some items combine topics, some missing items relevant to NF
FACT-G^a	≥18 y SR	Well studied with good psychometric properties; appropriate domain/item content for NF medical trials; widely used in clinical trials (1 in NF); brief, simple format and wording; free; multiple languages available	No child or adolescent forms for clinical trials enrolling both adults and younger individuals; no cognitive items, few items on activities of daily living
SF-12	≥18 y SR	Well studied with good psychometric properties; used in numerous studies and clinical trials; very brief; multiple languages available	Only an adult form; lack of use in NF populations and trials; instructions are complex, items are wordy, and answer format varies across items; not free
SF-36	≥18 y SR	Well studied with good psychometric properties; used in NF studies; general population normative data and norm-based T scores; brief; multiple languages available	No NF normative data or use in NF trials; limited to no coverage of social and cognitive function; varying response format and recall periods across questions may be confusing
WHOQoL-BREF^a	≥18 y SR	Well studied, good psychometrics; widely used in clinical trials including NF; good content for the physical, social, and emotional domains; brief, free, easy to administer; multiple languages available	Some items have difficult wording for patients with learning disabilities or adolescents; no child form for clinical trials enrolling both adults and younger individuals; some items combine depression and anxiety
Pediatric-Through-Adult EuroQoL			
	≥8 y SR ^c ; 4–7 y PR	Well studied in a variety of populations with good psychometric properties; used in numerous studies including clinical trials; short, simple; available in multiple languages	Only 1 item per domain; some items assess 2 different activities or symptoms (e.g., anxiety and depression); lacking items in important areas for NF (e.g., cognitive, social)
PedsQL GCS^{a,d}	≥5 y SR ^c ; ≥2 y PR ^c	Wide age range; well studied in a variety of populations; can be used with the PedsQL Infant Scale; used in clinical trials including NF; assesses areas of functioning relevant to NF; items worded clearly; simple format; good psychometric properties; available in multiple languages	Potentially long process to obtain license for use at some institutions and significant fee for commercial use; minimal research using the adult form (≥26 y)
PROMIS and Neuro-QoL	≥8 y SR ^c ; 5–17 y PR	Used in many populations; 1 article on use of specific domains in children with NF1; has standardized T scores; computerized adaptive testing allows assessment with fewer items; good psychometric data; excellent feasibility (brief, free); available in multiple languages	No PR for younger ages; varied recall periods on some forms; different items from pediatric to adult scales; needs more psychometric data using these measures in NF, with computerized adaptive testing, and in clinical trials
PROMIS Global Health	≥8 y SR ^c ; 5–17 y PR	Used in multiple conditions, good age range and normative samples; standardized scores; good psychometrics; multiple languages available; brief; free	Complicated wording; child, parent, and adult forms have inconsistent formats and an item combines thinking and mood; misses younger age ranges; few studies in children; limited use in NF
TNO AZL QoL	>8 y SR ^c ; 6 mo–15 y PR ^c	Wide age range of self and PR forms, used in various pediatric populations with one cross-sectional study in NF1; parallel child and parent items; good psychometric data for child form; multiple languages available	No longitudinal studies or use in clinical trials in NF; different items from pediatric to adult forms; some items not as relevant for NF clinical trials; limited psychometric data of the adult and preschool parent forms; vague recall period; somewhat long
Pediatric-Only Child Health Questionnaire^d			
	10–18 y SR; 5–18 y PR ^c	Satisfactory age range; mostly parallel with the ITQoL so could be used together for younger children; has been a primary outcome in clinical trials (not NF); discriminant and concurrent validity adequate for parent measure; multiple languages available	Self-report (45-item SF) a bit long; some items not relevant for NF; items vary in recall period; possible problems with ceiling effects, poor internal consistency for some subscales; no summary scores on child long form; sensitivity to change questionable for the parent SF
DISABKIDS	4–18 y SR ^c /PR ^c	Items content developed with qualitative research; good psychometric properties of SR form across a range of chronic medical conditions; used in 1 NF cross-sectional study; brief, simple format, easy to administer; multiple languages available	Limited published data on 4- to 7-y-olds, parent-proxy measure, and use in longitudinal studies, clinical trials, and the United States; some wording or content is perceived as being challenging for younger kids and those with learning disabilities
ITQoL^d	2 mo–5 y PR	Mostly parallel with Child Health Questionnaire so could be used together in clinical trial for wider age range; internal consistency adequate to good; discriminant validity good; available in multiple languages	Content of some items not as relevant for NF trials; recall periods vary across questions; some ceiling effects observed; no factor analysis; cost for use; a bit long

Continued

Table 1 List of PRO Measures Reviewed to Assess Multidimensional Domains of QoL in NF Clinical Trials (continued)

Type and name of measure	Age range available	Strengths for use in NF clinical trials ^b	Limitations for use in NF clinical trials ^b
Kidscreen	8–18 y SR/PR	Developed in several countries with large validation studies in different populations; excellent internal consistency; normative values available; child and parent forms; simple format; multiple languages available	Lacks certain items relevant for NF and includes some items not as relevant to NF trials; test-retest reliability coefficients are low; limited data on responsiveness; no child or parent forms for younger children
KINDL-R	4–17 y SR ^c /PR ^c	Broad age range for SR including young children and parallel PR forms; good psychometric data; simple format, brief, easy to administer; multiple languages available	Content of some items not relevant for NF1 trials, in particular the family items; categories of scoring are broad; the words “seldom” and “quarreled” could potentially be confusing; cost for use
PedsQL Infant^{a,d}	1–24 mo PR ^c	Compliments PedsQL GCS by extending age range to infants and toddlers; used in numerous studies and clinical trials; good internal consistency and discriminant validity; simple format; multiple languages available	Items limited by lack of specificity to infants and toddlers; no longitudinal studies or use in clinical trials with NF; five-factor model provides poor-questionable fit of the data; ceiling effects for some subscales, test-retest data not available; cost for use
NF1 disease-specific			
INF1-QoL	≥18 y SR	Covers multiple domains relevant to NF1; developed with qualitative research; good preliminary psychometric data; brief; no cost	No child/adolescent or parent-proxy forms; only 1 item per domain; some items assess >1 symptom; lengthy items may be confusing; limited use and psychometric data to date; only in English
PedsQL NF1 Module^a	≥5 y PR ^c ; ≥5 y SR ^c	Wide age range; qualitative research yielded excellent item content with all domains applicable to NF1 with most relevant to NF trials; simple wording and format; good psychometric properties	Large number of items; cost per study licensing fees for use; available only in English
Skindex	≥12 y SR ^c	Developed on a clear theoretical framework that guided domains and item selection; covers a lot of NF issues; many shared items between adult and teen forms; extensive psychometric testing; Skindex-29 has been used in many studies, including NF, and may be seen as having normative data for NF; Skindex-Teen form for younger ages (but fewer data on this form)	Designed for skin conditions, not NF; some different items between adult and teen forms, some domains vary slightly in content, and domain scores not comparable across measures; some items relevant to NF (cognitive, functional) are not adequately covered; wording (skin condition) may not be the best for some patients with NF; cost for funded studies
NF2 disease-specific			
Neary NF2 questionnaire	≥16 y SR	Comprehensive set of items assessing NF2 symptoms; brief, free, and easy to administer	Currently appears to be a clinical tool; no scores to obtain; no psychometric data reported
NF2 QoL survey	≥16 y SR	Multiple domains relevant to NF2; item selection based on EORTC modules allows comparability of some items to general population and oncology norms; free	Very limited psychometric data; concerns with domain scoring, depth of coverage, and grouping of items; lengthy
NFTI-QoL^a	≥16 y SR	Specifically designed for NF2 through qualitative research in individuals with NF2 ≥16 y of age; used in several NF2 studies; good preliminary psychometric properties; short; feasible for use in trials	Not validated in youth <16 y of age; limited in scope with only 8 items; some items combine symptoms; somewhat limited 4-point Likert-type response scale; different recall periods on some items; limited languages
PANQoL^a	≥18 y SR	Well researched and feasible questionnaire for trials focusing on patients with vestibular schwannomas; relevant item content; good reliability and validity; multiple languages available	Not validated in patients with NF2 or children; questions limited to vestibular schwannoma-related complaints and lack some NF2 specific symptomology; recall period not specified; limited data on sensitivity to change

Abbreviations: EORT-QLQ-C30 = The European Organisation for Research and Treatment of Cancer Quality of Life questionnaire (30 items); EuroQoL = European Quality of Life; FACT-G = Functional Assessment of Cancer Therapy–General; INF1-QoL = Impact of NF1 on Quality of Life; ITQoL = Infant Toddler Quality of Life questionnaire; NF = neurofibromatosis; NFTI-QoL = Neurofibromatosis 2 Impact on Quality of Life; PANQoL = Penn Acoustic Neuroma Quality of Life; PedsQL = Pediatric Quality of Life Inventory; PedsQL GCS = PedsQL Generic Core Scales; PR = parent report; PRO = patient-reported outcome; PROMIS = Patient-Reported Outcomes Measurement Information System; QoL = quality of life; SF = Short Form; SR = self-report; TZO-AZL QoL = Netherlands Organisation for Applied Scientific Research Academic Medical Centre Children’s Quality of Life Questionnaire; WHOQoL-BREF = World Health Organization Quality of Life–BREF.

^a Measures currently recommended by the Response Evaluation in Neurofibromatosis and Schwannomatosis (REINS) International Collaboration for use in clinical trials specifically for individuals with NF.

^b Strengths and limitations of these PRO measures were specifically focused on relevance for individuals with NF and use in NF clinical trials.

^c Different forms for various age groups.

^d A separate PRO measure for a younger age group was designed to be used with another separate measure for older children, which have different names, to extend the age range.

summed and transformed into domain scores. This scale can be administered in ≈5 minutes via paper and is validated for administration by phone or electronic devices.²³ It has been

translated into 19 languages and used in hundreds of psychosocial clinical trials, including NF.^{7,8,24} Limitations include no normative data as yet for patients with NF. In addition,

Table 2 Generic PRO Measure for Adults on Medical NF Trials

Criteria (ratings)	FACT-G characteristics Overall group rating = 2.74
Patient characteristics (2.75)	Age range is > 18 y Self-report form No observer-report form Normative data on several groups, including oncology and the general population
Domains/item content (2.44)	27 Items assess 4 domains: physical, social/family, emotional, and functional well-being Item content includes experience of physical symptoms, difficulties in functioning, and satisfaction with aspects of life Does not include a cognitive subscale and limited items on activities of daily living
Psychometric data (2.75)	Adequate to good reliability with a coefficients typically in the 0.70s–0.80s ²⁰ Strong construct validity ⁵⁴ and known group validity ⁵⁵ Sensitivity to change is documented ⁵⁶
Feasibility (3.0)	5–10 min to complete Recall period is “the past 7 d” Simple format and wording (fourth-grade reading level) ¹⁷ Valid paper and electronic versions available ²¹ Freely available at the FACIT Group website Available in > 50 languages

Abbreviations: FACIT = Functional Assessment of Chronic Illness Therapy; FACT-G = Functional Assessment of Cancer Therapy–General; NF = neurofibromatosis; PRO = patient-reported outcome.

some of the wording may be difficult for individuals with NF1 who have learning disabilities, and the Environment domain may not be relevant or sensitive to change in NF trials, although it does not have to be administered.

Pediatric-Only or Pediatric-Through-Adult Clinical Trials

Our group identified 6 pediatric and 5 pediatric-through-adult generic PRO measures that assess multidimensional domains of QoL that we deemed appropriate for further review for NF trials. The top 3 rated measures for pediatric-only trials included the Pediatric Quality of Life Inventory (PedsQL) version 4.0 Generic Core Scales (GCS),²⁵ the KINDL-R questionnaire,²⁶ and the DISABKIDS questionnaire.²⁷ The 2 highest-rated generic measures for trials enrolling children through adults were the PedsQL GCS²⁵ and Patient-Reported Outcomes Measurement Information System/Neuro-QoL measures.²⁸ In our side-by-side comparisons, the PedsQL GCS received the highest overall group rating.

Table 3 Generic PRO Measure for Adults on Psychosocial NF Trials

Criteria (ratings)	WHOQoL–BREF characteristics Overall group rating = 2.73
Patient characteristics (2.75)	Age range is ≥ 18 y Self-report form No observer-report form Used with adolescents but not children General norms and effect sizes available for adults, by age and sex ⁵⁷
Domains/item content (2.40)	26 Items assess 4 domains: physical health, psychological health, social relationships, and environmental Item content primarily includes satisfaction with various aspects of one’s functioning, environment, and QoL 2 Additional items assess overall QoL and satisfaction with health
Psychometric data (2.75)	Internal consistency: is good with Cronbach α >0.70 for overall domains; high reliability (>0.80) for all individual domains except social relationships (0.68), which has only 3 items and thus is not concerning ²² Test-retest reliability is adequate (>0.70) across studies ²² Strong validity (construct, discriminant, criterion, concurrent) and sensitivity to change after treatment, in particular for physical health and psychological domains ²² Factor analyses confirmed subscale factor structure ²²
Feasibility (3.0)	5 min to complete High compliance, minimal missing data Measure, manual, and scoring syntax (SPSS) freely available online on the WHOQoL–BREF website Valid for multiple recall periods but 2 wk is standard Valid administration by paper, phone, or electronically ²³ Validated in 19 languages

Abbreviations: NF = neurofibromatosis; PRO = patient-reported outcome; QoL = quality of life; SPSS = Statistical Package for the Social Sciences; WHOQoL–BREF = World Health Organization Quality of Life–BREF.

Pediatric-Only and PEDIATRIC-Through-Adult Medical and Psychosocial Trials

The PedsQL GCS was the generic measure that we rated the highest for both medical and psychosocial trials targeting only children and adolescents, as well as trials including children through adults (table 4). The PedsQL GCS has self-report forms for children ages 5 to 18 years (preschool, child, teen), young adults (18–25 years), and adults (>26 years) and parallel parent-report forms for ages ≥2 years. In addition, the PedsQL has parent-report forms for infants ages 1 through 24 months, thus allowing similar measures to be used across all ages.^{29–31}

Table 4 Generic PRO Measure for Pediatric-Only or Pediatric-Through-Adult NF Trials

Criteria (ratings)	PedsQL GCS and Infant Scales characteristics Overall group rating = 2.79
Patient characteristics (2.9)	SR forms for children 5–7, 8–12, and 13–18 y of age and adults 18–25 and ≥26 y of age PR forms for infants 1–12 and 13–24 mo of age, children 2–4, 5–7, 8–12, and 13–18 y of age, and adults 18–25 and ≥26 y of age Normative data collected for 8,591 children 5–16 y of age ⁵⁸ and 683 infants 1–24 mo of age ²⁹
Domains/item content (2.75)	Infant Scales PR forms (1–12 and 13–24 mo) have 36 and 45 items, respectively; Child Scales PR toddler form (2–4 y old) has 21 items; SR (≥5 y) and PR (≥5 y) child and adult forms have 23 items Child, Adolescent, and Adult Scales domains assess physical, emotional, social, and work/school functioning Infant Scale domains assess physical symptoms and physical, emotional, social, and cognitive functioning Similar content and wording of items across all forms; item content assesses primarily experience of physical symptoms and difficulties in different areas of functioning
Psychometric data (2.75)	Reliability is adequate to excellent: internal consistency for the PedsQL GCS total scores of the young child to adolescent total sample is SR = 0.89 and PR = 0.92 (individual scales SR = 0.68–0.83 and PR = 0.75–0.88) ²⁵ and for the Infant total scores of the total sample PR = 0.92 (individual scales PR = 0.72–0.89) ²⁹ Good construct validity and discriminant validity (between healthy infants and children and those with chronic health conditions) ^{29,58} Parent-child agreement ranges from fair to strong across subscales ⁵⁸ Factor analysis indicates that 52% of the total variance is explained for the Child Scales for the 5-factor solution ³⁰ ; poor to questionable fit of the 5-factor structure for the Infant Scales ²⁹ Large validation samples of healthy children and children with chronic illness ³⁰
Feasibility (2.75)	Short (4–10 min) and easy to administer and complete Items use simple language and are easy to understand Acute (past 7 d) or standard (past 1 mo) recall periods Forms available in >100 languages Free for nonfunded academic research, annual license fee, or fee per study for commercial or noncommercial research; more information available on the PedsQL website

Abbreviations: NF = neurofibromatosis; PedsQL GCS = Pediatric Quality of Life Inventory Generic Core Scales; PR = parent report; PRO = patient-reported outcome; SR = self-report.

Although the PedsQL has a self-report version for adults (>26 years), there currently is scant research using this form, so we decided to recommend separate adult-only measures for trials that do not include children or adolescents at this time.

The PedsQL Infant Scales assess physical symptoms, physical functioning, emotional functioning, social functioning, and cognitive functioning,²⁹ while the PedsQL GCS child, adolescent, young adult, and adult forms assess physical functioning, emotional functioning, social functioning, and work/school functioning.³⁰ Items are rated on 5-point Likert scales that are transformed to a 0 to 100 scale and yield domain and total scores. The PedsQL GCS has good psychometric properties in a variety of patient populations,³⁰ is brief, and is available in >100 languages. The PedsQL GCS was used to evaluate generic QoL in clinical trials with individuals with NF1,^{5,19} and additional NF trials using the PedsQL GCS are underway. The main limitations of these PedsQL measures include the poor to questionable fit of the 5-factor structure of the Infant Scales²⁹ and the lack of research on the adult form.

Disease-Specific Measures

NF1 Disease-Specific QoL

Our group identified and reviewed 3 published disease-specific measures used for the evaluation of patients with NF1: Impact of NF1 on Quality of Life questionnaire,³² Pediatric Quality of Life Inventory Neurofibromatosis Type 1 Module (PedsQL NF1),^{33,34} and Skindex, including the Skindex-29³⁵ and Skindex-Teen.³⁶

In our review process, the PedsQL NF1 Module^{33,37} was rated the highest to assess multidimensional domains of disease-specific QoL in trials for individuals with NF1 (table 5). This measure was developed through a collaboration between experts in the field of NF and the author of the PedsQL using a process that was similar to that of other PedsQL modules.³³ The 104-item scale assesses 18 domains determined to be affected in NF1 through rigorous qualitative research.³⁷ It yields domain and total scores and has acceptable to excellent psychometric properties.^{33,34} Furthermore, the adolescent

Table 5 Disease-Specific PRO Measure for NF1 Trials

Criteria (ratings)	PedsQL NF1 Module characteristics
	Overall group rating = 2.5
Patient characteristics (2.75)	SR forms for children 5–7, 8–12, and 13–18 y of age and adults 18–25 and ≥26 y of age with NF1
	PR forms for children 5–7, 8–12, and 13–18 y of age and adults 18–25 and ≥26 y of age with NF1
	No PR measure for infants or children <5 y of age
	Normative data for 323 pediatric participants 5–25 y of age and 335 parents of pediatric participants, ³⁴ as well as 134 adults with NF1 between 20 and 71 y ³³
Domains/item content (2.75)	Large number of items (104) that assess 18 domains known to be of concern to patients with NF1 according to qualitative research
	Parallel forms assess the following domains: skin itch bother, skin sensations, pain, pain impact, pain management, cognitive functioning, speech, fine motor, balance, vision, perceived physical appearance, communication, worry, treatment anxiety, medicines, stomach discomfort, constipation, and diarrhea ³⁴
	Similar content and wording of items across all forms; item content assesses primarily physical symptoms and difficulties in functioning related to NF1
Psychometric data (2.5)	Reliability in pediatric and parent forms is good to excellent ³⁴ : total scale scores (patient and parent reports were each 0.98) and individual scales (self-report 0.71–0.96, parent report 0.73–0.98)
	Good construct validity across ages and discriminant validity in adults ^{33,34}
	Parent-child agreement ranges from fair to strong across subscales
	Significant ceiling effects on some subscales
	Fewer data available for parent-report form
Feasibility (2.0)	Easy to administer; excellent feasibility ³⁴
	Acute (past 7 d) or standard (past 1 mo) recall periods
	Long length may make it difficult for some people with NF to complete repeatedly in a clinical trial (104 items for children 5 y of age through adults)
	Available only in English currently
	Free for nonfunded academic research, annual license fee or fee per study for commercial or noncommercial research with costs exceeding US \$1,000 for supported trials; more information available on the PedsQL website

Abbreviations: NF = neurofibromatosis; PedsQL = Pediatric Quality of Life Inventory; PR = parent report; PRO = patient-reported outcome; SR = self-report.

and adult modules were used in 2 recent clinical trials that documented changes in several domains scores with treatment to reduce plexiform neurofibromas.^{6,38} Limitations of this measure include the lack of a parent-report form for children <5 years of age, long length (particularly for repeated administration in trials), availability in only English, and lack of clinical trial data in children to date.

NF2 Disease-Specific QoL

Our group identified 4 measures of NF2-related QoL to be reviewed and rated. Of these, the Neurofibromatosis 2 Impact on Quality of Life (NFTI-QoL)³⁹ and Penn Acoustic Neuroma Quality of Life (PANQoL)⁴⁰ scales were the highest-rated measures with close side-by-side ratings (table 6).

The NFTI-QoL is freely available, brief, and easy to administer. It consists of only 8 items but addresses the complications relevant to NF2 based on qualitative research.³⁹ The items are rated on 4-point Likert scales, and a total score (ranging from 0–24) is calculated as the sum of all items.

Clinical studies generally report good reliability and validity.^{41–44} However, some studies have demonstrated limited construct validity for psychological functioning and emotional well-being.^{44,45} Our group noted some limitations of this measure. The wording of response options and recall periods are not consistent across all items. The measure does not address potentially important facets of NF2-related QoL, including speech/communication, motor function, eating/swallowing, bowel/bladder issues, and potential meningioma symptoms such as fatigue and cognitive difficulties. In addition, it has only 1 question per domain, with some items addressing multiple related domains, leading to concerns regarding sensitivity to change and interpretability of changes over time. However, the NFTI-QoL reflected improvement in response to treatment with bevacizumab during routine clinical care,⁴³ and it is being used in current NF2 clinical trials, so additional longitudinal data will be forthcoming.

The PANQoL is a 26-item scale that assesses 7 QoL domains affected by vestibular schwannomas⁴⁰ and includes more

Table 6 Disease-Specific PRO Measures for NF2 Trials

Criteria (ratings)	NFTI-QoL characteristics	PANQoL characteristics
	Overall group rating = 2.19	Overall group rating = 2.0
Patient characteristics		
NFTI-QoL (2.25) PANQoL (1.75)	Specific NF2 questionnaire	VS-specific questionnaire: no patients with NF2 included
	SR form developed for ≥16-y-olds, although it has been used in studies for those ≥10 y of age	SR form ≥18 y
	No pediatric SR forms or PR forms	No pediatric SR forms or PR forms
	Validated in patients with NF2, patients with sporadic VS, and healthy controls	Validated in patients with sporadic VS, control patients without hearing or balance-related symptoms, and general population
Domains/item content		
NFTI-QoL (1.75) PANQoL (2.25)	Items assess appropriate content for NF2	Items assess appropriate content for symptoms of VS
	8 Items covering 8 domains: hearing, dizziness/balance, facial weakness, pain, anxiety/depression, outlook on life, mobility, and sight	26 Items covering 7 domains: hearing, balance, facial dysfunction, pain, anxiety, general health, and energy
	Item content assesses how much NF2 symptoms affect daily functioning	Item content assesses difficulties in functioning related to VS
	Does not address important domains of speech/communication, motor function, eating/swallowing, and bowel/bladder	Items do not address other important NF2 domains so adequate only for a trial targeting VS in NF2
Psychometric data		
NFTI-QoL (2.5) PANQoL (1.75)	Good internal consistency (Cronbach α = 0.85) and test-retest reliability (r = 0.84) ⁴¹	Moderate to good internal consistency (Cronbach α = 0.71–0.88) and test-retest reliability (r = 0.80 for total score, r = 0.63–0.76 for domain scores) ⁴⁰
	Demonstrated construct validity with generic QoL measures (r = –0.81 with sum of SF-36 scales; r = –0.61 to –0.70 for physical health scales; r = 0.71 with EuroQoL), ^{39,44} but weak correlations of relevant items to patient-reported psychological health/emotional well-being and objective hearing measures ^{43–45}	Demonstrated construct validity to established generic QoL measures (e.g., SF-36) and domain-specific QoL scales such as emotional intelligence ⁴⁷ and facial dysfunction ⁴⁶
	Discriminative validity by disease type (unilateral VS vs NF2) and NF2 severity ^{42,44}	Known group validity between patients and controls ⁴⁰
	Sensitive to changes with bevacizumab ⁴³	Limited data on sensitivity to change in clinical trials
Feasibility		
NFTI-QoL (2.25)	Brief (\leq 3 min), easy to administer	Brief, easy to administer
	Validated in English; translated but not yet validated in Swedish ⁵⁹ and Dutch (personal communication, K.S. Koetsier, April 2, 2020)	Validated in English, Spanish, French, German, Dutch, and Japanese (article lists all translations/references) ⁶⁰
PANQoL (2.25)	Freely available	Freely available

Abbreviations: EuroQoL = European Quality of Life; NF = neurofibromatosis; NFTI-QoL = Neurofibromatosis 2 Impact on Quality of Life; PANQoL = Penn Acoustic Neuroma Quality of Life; PedsQL = Pediatric Quality of Life Inventory; PR = parent report; PRO = patient-reported outcome; QoL = quality of life; SF-36 = Short Form 36 items; SR = self-report; VS = vestibular schwannoma.

items on emotional well-being than the NFTI-QoL. Items are rated on a 5-point Likert scale; scores are then summed and transformed to a 0 to 100 scale. An overall score and 7 domain scores are available. The measure has moderate to good internal consistency and test-retest reliability.⁴⁰ Convergent validity has been demonstrated across multiple domains (fatigue, facial paresis, and emotional intelligence).^{46–48} The PANQoL was recommended to be used only as an exploratory outcome measure because patients with NF2 were excluded during the development of this scale and in all currently published studies. Therefore, it does not address all

aspects of NF2-related QoL and does not account for all potential etiologies of NF2-related symptoms (e.g., questions address difficulties in walking due to balance issues but not neuropathies or muscle atrophy).

Schwannomatosis Disease-Specific QoL

Our group was not able to identify any disease-specific measures developed for individuals with schwannomatosis. Prior systematic reviews have documented a relative paucity of published studies assessing QoL in patients with schwannomatosis, highlighting the need for additional investigation in this area.^{3,49}

Table 7 Current REiNS Recommendations for Generic and Disease-Specific PRO Measures for NF Clinical Trials^a

Trial age range	Generic domains of QoL	NF1 disease-specific domains of QoL	NF2 disease-specific domains of QoL
Pediatric	PedsQL GCS and Infant Scales (SR 5–18; PR 1–18)	PedsQL NF1 Module (SR 5–18; PR 5–18)	NFTI-QoL (SR 16–17)
Pediatric-adult	PedsQL GCS and Infant Scales (SR ≥ 5; PR ≥ 1)	PedsQL NF1 Module (SR ≥ 5; PR ≥ 5)	NFTI-QoL (SR ≥ 16)
Adult	FACT-G (SR ≥ 18; medical trials)	PedsQL NF1 Module (SR ≥ 18)	NFTI-QoL (SR ≥ 18)
	WHOQoL-BREF (SR ≥ 18; psychosocial trials)		PANQoL (SR ≥ 18) ^b

Abbreviations: FACT-G = Functional Assessment of Cancer Therapy–General; NF = neurofibromatosis; NFTI-QoL = Neurofibromatosis 2 Index-Quality of Life; PANQoL = Penn Acoustic Neuroma Quality of Life; PedsQL = Pediatric Quality of Life Inventory; PedsQL GCS = PedsQL Generic Core Scales; PR = parent report; PRO = patient-reported outcome; QoL = quality of life; REiNS = Response Evaluation in Neurofibromatosis and Schwannomatosis; SR = self-report; WHOQoL-BREF = World Health Organization Quality of Life–BREF.

^a Measures currently recommended by the REiNS International Collaboration for use in clinical trials specifically for individuals with NF.

^b Exploratory measure for adult NF2 trials targeting vestibular schwannomas.

Discussion

PRO measures provide unique and critical data to evaluate treatment effects in medical and psychosocial trials by assessing patients' subjective experiences. In this study, the REiNS PRO group, including the patient representatives, used their established methodology¹² to review, rate, and recommend existing generic and disease-specific PRO measures to assess multidimensional domains of QoL in individuals with NF1, NF2, and schwannomatosis who are participating in clinical trials. Our ratings and recommendations target the specific, unique needs of PRO assessments for clinical trials involving individuals with NF and may not reflect the appropriateness of these measures for other types of NF research or patient populations.

For the domain of generic PRO measures, we currently recommend 3 measures due to the wide age range of individuals who may participate in NF clinical trials and the different domains and items that are important to assess in medical vs psychosocial trials. For adult-only medical trials, we recommend the FACT-G because it covers a range of domains and items important for evaluating the impact of treatments on symptoms, functioning, and aspects of QoL for adults with NF, is easy to complete, and is available in multiple languages. Additional strengths include its excellent psychometric properties in the adult cancer population and frequent use in clinical trials and regulatory submissions,¹⁸ including 1 NF1 trial evaluating a drug treatment for plexiform neurofibromas.¹⁹ However, further research is needed to examine the use of the FACT-G in adult NF clinical trials and to obtain additional measurement data from the various types of NF. For medical trials seeking drug approval, the Food and Drug Administration prefers PRO assessments of proximal concepts such as symptoms rather than more distal concepts like generic QoL, while the European Medicines Agency is more accepting of QoL data.^{18,50} However, because the domains in multidimensional PRO measures yield scores from the same normative sample,⁹ separate analyses of the individual domains (e.g., functional well-being) can be performed to provide a focused evaluation of a particular area when needed. Furthermore, specific symptoms or complications may influence

functioning and QoL,⁵¹ thus supporting the use of generic measures in clinical trials.

We recommend the WHOQoL-BREF for adult-only psychosocial trials for several reasons. This scale assesses both overall QoL and domain-specific satisfaction with one's physical, social, and psychological health, which are highly modifiable with psychosocial interventions. It is available in many languages and has been used widely in research, including mind-body interventions among adults with NF1, NF2, and schwannomatosis^{8,24} and adolescents with NF1 and NF2.⁷ Future studies should examine the psychometric properties of the WHOQoL-BREF in individuals with NF. Although our group attempted to identify 1 adult measure for both medical and psychosocial trials for all 3 types of NF to facilitate comparability across studies, the items and domains of each of the measures did not seem suitable for both types of trials. Because the 2 measures we recommend are brief, it would be useful to administer both measures in future trials when possible to determine whether only 1 of them is sufficient for both kinds of trials.

For medical and psychosocial NF trials that enroll only children and adolescents or a wider age range from children/adolescents through adults, we recommend the PedsQL, consisting of the GCS and Infant Scales. The main strengths of these measures include that they consist of parallel and simply worded items and assess similar domains of functioning across child to adult age groups with both self-report and parent-report versions. The PedsQL GCS has good psychometric properties and has been used in numerous studies and clinical trials, including those with NF1.^{5,19} This generic measure also may be administered in combination with the PedsQL NF1 disease-specific module to obtain a more comprehensive PRO assessment. However, 1 gap to be addressed is that the parent-report Infant Scales have not been used in an NF trial to date, but they currently are being considered for a pediatric treatment trial being planned for very young children with NF1. Additional research is needed on (1) the general psychometric properties of the PedsQL Adult form, because the lack of data prohibited us from

recommending it for adult-only trials, and (2) the use of the PedsQL GCS and Infant Scales in the NF population. Another limitation is that the new licensing agreements for the PedsQL forms are very complicated for PRO measures, which currently make it difficult and time-consuming for the NIH and other federal government research facilities to obtain the appropriate permissions for its use in studies.

For the NF1 disease-specific QoL measure, we recommend the PedsQL NF1 Module. Strengths include rigorous qualitative research that was conducted to develop the item content so that it assesses the many domains affecting the QoL of individuals with NF1³⁷ and the ability to evaluate children through adults with similar domains and items. Future research on this new scale is needed to develop a related parent-report measure for children <5 years of age because NF medical trials are enrolling younger children. In addition, it will be important to examine further the psychometric properties of the parent-report form, to analyze the factor structure, and to determine the minimally clinically important difference (MCID). Another critical step is to translate the forms into languages other than English. Finally, the PedsQL NF1 Module has been used in 2 adult NF treatment trials,^{6,38} but it has not been used in a pediatric trial to date.

We are recommending 2 measures to assess disease-specific QoL in individuals with NF2: the NFTI-QoL for all NF2 clinical trials and the PANQoL as an exploratory measure for NF2 trials targeting vestibular schwannomas. The NFTI-QoL was developed specifically to assess the domains affected in patients with NF2; it is brief to facilitate repeated evaluations in clinical trials; and it has good reliability and validity.^{39,41,44} Future research should address calculating the MCID, examining its use and sensitivity to change in NF2 clinical trials, and developing measures in additional languages. There also is a need for validated NF2 disease-specific pediatric and parent-proxy forms given recent NF2 trials with lower age limits. The PANQoL was recommended for clinical trials enrolling individuals with NF2 and vestibular schwannomas because it is a well-studied tool with good psychometric data, comprehensively assesses domains important to these patients, and is available in several languages. However, future research is needed to validate the PANQoL in patients with NF2. For both of these NF2 disease-specific tools, there is a lack of items covering social issues (and emotional concerns in the NFTI-QoL), communication/speech, eating/swallowing, cognitive functioning, and bowel/bladder issues. Given recent treatment advances for NF2-related vestibular schwannomas and reduction of related complications, symptoms associated with other manifestations of NF2 (e.g., meningiomas, ependymomas, peripheral neuropathy) may become more apparent; thus, current NF2-specific measures may not be adequate to fully capture the impact of these manifestations on functioning and QoL. Adding a generic QoL or a specific symptom measure in a clinical trial may be needed to thoroughly assess these other domains important to patients with NF2.

There are no published multidimensional PRO measures developed for schwannomatosis. Given that chronic pain is the most common and concerning symptom of schwannomatosis,⁵² we recommend that researchers seeking to assess aspects of QoL in schwannomatosis clinical trials use previously recommended measures for pain intensity and pain interference¹³ in combination with 1 of the age-appropriate generic measures described here.

Table 7 summarizes our current recommendations for the most appropriate PRO measures to assess the multidimensional aspects of QoL in NF trials from the published literature, but researchers still must determine which measures to use in specific trials. During the selection of the most appropriate PRO measures to use, it is important to consider carefully the objectives of the trial, age range of patients, domains that may be affected by NF, wording and content of the items, and areas that might be improved by therapy or possible toxicities that may be related to the treatment. The generic measures can be used for clinical trials enrolling any of the 3 types of NF, unlike the disease-specific tools. In some cases, unidimensional PRO tools that assess disease-specific symptoms such as pain intensity or pain interference¹³ in NF1 or schwannomatosis trials and specific functions such as communication in NF2 trials may need to be administered in addition to a generic measure. In other cases, 1 multidimensional NF1 or NF2 disease-specific tool may be sufficient for measuring the particular domains of interest in a medical trial. For a more comprehensive assessment, a generic and NF disease-specific measure (e.g., PedsQL GSC and NF1 Module) can be administered together. Given the cognitive difficulties in individuals with NF1,⁵³ we recommend administering self-report measures in children starting at 8 years of age. Finally, it is important to acknowledge that some patients enrolling in clinical trials may need assistance with completing PRO self-report measures due to physical or cognitive impairments. Helping patients who have difficulties with reading items, writing down responses, or pushing the desired button on an electronic tablet is appropriate for the administration of PRO measures so that most of the participants in a trial can provide these important clinical endpoint data. REiNS PRO group members and other experts in PRO assessment can be consulted to assist with these choices when needed.

There are limitations to our current recommendations. First, although our systematic method evaluates existing PRO measures on several important criteria¹² and involves NF experts and patient representatives to carefully examine and compare the item content to ensure a fit for the NF population, we did not use the World Health Organization International Classification of Functioning, Disability, and Health to standardize terminology and conceptual frameworks as is being recommended by some PRO researchers.¹⁵ We also did not use a mapping methodology to compare the items and domains between measures. Updates to these recommendations in the future should consider using a specific

methodology such as the International Classification of Functioning, Disability, and Health to define the domains and items assessed to be consistent with these new methods and to evaluate and compare the PRO measures in a more precise manner. Second, the generic measures have limited psychometric and normative data in patients with the different types of NF. We are recommending these PRO measures partially on the basis of their strong psychometric properties in other populations and the item content deemed important for NF clinical trials by our group's expertise and patient input, despite not necessarily being a perfect fit for NF. These generic measures have yet to undergo the rigorous testing needed to determine their reliability, validity, sensitivity to change, and MCID specifically in NF populations, which is critical for PRO measures being used in clinical trials seeking drug approval and for product labeling. However, researchers need PRO measures to assess clinical outcomes in NF trials at this time. In response to this critical need, we made these recommendations for existing tools on the basis of the current literature. As the field gains experience using these tools in patients with NF across the lifespan, collects longitudinal data on trials, and develops new measures, some recommendations may need to be updated. For the NF1 and NF2 disease-specific tools, preliminary psychometric findings are promising, but more work is necessary to further examine the use of these new PRO measures for assessing meaningful change in trial outcomes. Third, most measures do not evaluate young children, although NF clinical trials already are assessing children as young as 3 years of age and new trials are being planned for infants. Thus, parent-report measures or other functional methods of assessment for the youngest children need to be explored. Finally, these recommendations are made to serve as a current guide for the researchers to help promote harmonization in the selection of PRO tools and to suggest areas for further outcomes research in the NF community. However, researchers should consider using other existing PRO measures or developing new tools if needed to match the specific objectives of their clinical trials.

Our REiNS PRO group continues to review existing PRO measures in various domains to provide recommendations for NF clinical trials. Future domains our group may address include disfigurement, fatigue, sleep, tinnitus, skeletal issues, and specific social-emotional domains such as anxiety and depression. In addition, we plan to revisit previous recommendations and to provide updates based on any new data reported in subsequent publications. Thus, our group encourages the NF scientific community to use the recommended PRO measures in studies to collect additional data on their psychometric properties and to further evaluate their appropriateness for both medical and psychosocial trials in NF1, NF2, and schwannomatosis. Researchers also need to continue to develop new measures specifically designed for NF using rigorous methodologies, including qualitative research involving patients, in an effort to increase the pool of reliable and valid PRO measures critically needed as endpoints for various NF clinical trials.

Acknowledgment

The authors acknowledge and thank additional members of the REiNS PRO working group who participated in phone calls and discussions to review and rate the PRO measures assessing generic and disease-specific QoL discussed in this article: Dale Berg, BA, Amanda Bergner, MS, Ann Blanton, PhD, CCC-SLP, Alexandra Cellucci, LPN, Kathy L. Gardner, MD, Andrés J. Lessing, MBA, Melissa Reider-Demer, NP, Tena Rosser, MD, and Karin Walsh, PsyD. The authors acknowledge the support of the Children's Tumor Foundation for the REiNS International Collaboration.

Study Funding

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute. Writing of this manuscript was supported by the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Health Services Research, the Center for Healthcare Organization and Implementation Research, and the Edith Nourse Rogers Memorial Veterans Hospital. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government. Writing of this article also was supported by the Department of Defense grants W81XWH-19-1-0184 and W81XWH-17-1-0121.

Disclosures

P. Wolters receives funding from the Neurofibromatosis Therapeutics Acceleration Program. A.M. Vranceanu received funding from the Department of Defense (W81XWH-19-1-0184 and W81XWH-17-1-0121), NF Midwest, NF Northeast, and NF Texas. H. Thompson reports no disclosures relevant to the manuscript. S. Martin receives funding from the Neurofibromatosis Therapeutics Acceleration Program. V. Merker reports consulting income from the Neurofibromatosis Network. A. Baldwin, C. Barnett, K. Koetsier, C. Hingtgen, and C. Funes, report no disclosures relevant to the manuscript. J. Tonsgard receives funding from the Department of Defense Army Research Command as part of the NF Consortium and from Midwest NF, Inc. E. Schorry receives funding from the Department of Defense as a site principal investigator for the NF Consortium and as a coinvestigator for a clinical trial of vitamin D for adults with NF1. T. Allen, T. Smith, B. Franklin, and S. Reeve report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

Publication History

Received by *Neurology* October 9, 2020. Accepted in final form May 13, 2021.

Appendix Authors

Name	Location	Contribution
Pamela L. Wolters, PhD	Pediatric Oncology Branch, National Cancer Institute, NIH, Bethesda, MD	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript

Continued

Appendix (continued)

Name	Location	Contribution
Ana-Maria Vranceanu, PhD	Integrated Brain Health Clinical and Research Program, Massachusetts General Hospital, Harvard Medical School, Boston	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript
Heather L. Thompson, PhD, CCC-SLP	Department of Communication Sciences and Disorders, California State University, Sacramento	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript
Staci Martin, PhD	Pediatric Oncology Branch, National Cancer Institute, NIH, Bethesda, MD	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript
Vanessa L. Merker, PhD	Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript
Andrea Baldwin, CRNP	Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD	Design and conceptualization of the study, acquisition and interpretation of the data, drafting and revising the manuscript
Carolina Barnett, MD, PhD	Division of Neurology, Department of Medicine, University Health Network and University of Toronto, Ontario, Canada	Acquisition and interpretation of the data, drafting and revising the manuscript
Kimberley S. Koetsier, MD	Department of Otolaryngology/Head and Neck Surgery, Leiden University Medical Center, the Netherlands	Acquisition and interpretation of the data, drafting and revising the manuscript
Cynthia M. Hingtgen, MD, PhD	Department of Clinical Neurosciences, Spectrum Health Medical Group and College of Human Medicine, Michigan State University, East Lansing	Design and conceptualization of the study, acquisition and interpretation of the data, revising the manuscript
Christopher J. Funes, MS	Integrated Brain Health Clinical and Research Program, Massachusetts General Hospital, Harvard Medical School, Boston	Acquisition and interpretation of the data, revising the manuscript
James H. Tonsgard, MD	University of Chicago Pritzker School of Medicine, IL	Design and conceptualization of the study, acquisition and interpretation of the data, revising the manuscript
Elizabeth K. Schorry, MD	Division of Human Genetics, Cincinnati Children's Hospital Medical Center, OH	Design and conceptualization of the study, acquisition and interpretation of the data, revising the manuscript
Taryn Allen, PhD	Clinical Monitoring Research Program Directorate, Frederick National Laboratory for Cancer Research, National Cancer Institute, Frederick, MD	Acquisition and interpretation of the data, revising the manuscript

Appendix (continued)

Name	Location	Contribution
Taylor Smith, PhD	Department of Psychology and Child Development, California Polytechnic State University, San Luis Obispo	Acquisition and interpretation of the data, revising the manuscript
Barbara Franklin, BS	REiNS International Collaboration patient representative	Acquisition and interpretation of the data, revising the manuscript
Stephanie Reeve, BSN, RN	REiNS International Collaboration patient representative	Acquisition and interpretation of the data, revising the manuscript

References

- Lu-Emerson C, Plotkin SR. The neurofibromatosis, part 1: NF1. *Rev Neurol Dis*. 2009;6(2):E47-E53.
- Lu-Emerson C, Plotkin SR. The neurofibromatosis, part 2: NF2 and schwannomatosis. *Rev Neurol Dis*. 2009;6(2):E81-E86.
- Vranceanu AM, Merker VL, Park E, Plotkin SR. Quality of life among adult patients with neurofibromatosis 1, neurofibromatosis 2 and schwannomatosis: a systematic review of the literature. *J Neurooncol*. 2013;114(3):257-262.
- Vranceanu AM, Merker VL, Park ER, Plotkin SR. Quality of life among children and adolescents with neurofibromatosis 1: a systematic review of the literature. *J Neurooncol*. 2015;122(2):219-228.
- Gross AM, Wolters PL, Dombi E, et al. Selumetinib in children with inoperable plexiform neurofibromas. *N Engl J Med*. 2020;383(13):1290.
- Weiss BD, Wolters PL, Plotkin SR, et al. NF106: a neurofibromatosis clinical trials consortium phase II trial of the mek inhibitor mirdametinib (PD-0325901) in adolescents and adults with NF1-related plexiform neurofibromas. *J Clin Oncol*. 2021;39(7):797-806.
- Lester E, DiStefano S, Mace R, Macklin E, Plotkin S, Vranceanu AM. Virtual mind-body treatment for geographically diverse youth with neurofibromatosis: a pilot randomized controlled trial. *Gen Hosp Psychiatry*. 2019;62(2):72-78.
- Vranceanu AM, Riklin E, Merker VL, Macklin EA, Park ER, Plotkin SR. Mind-body therapy via videoconferencing in patients with neurofibromatosis: an RCT. *Neurology*. 2016;87(8):806-814.
- Acquadro C, Berzon R, Dubois D, et al. Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration. *Value Health*. 2003;6:S22-S31.
- Mercieca-Berber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas*. 2018;9:353-367.
- Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain*. 2005;113:9-19.
- Wolters PL, Martin S, Merker VL, et al. Patient-reported outcomes in neurofibromatosis and schwannomatosis clinical trials. *Neurology*. 2013;81(2):S6-S14.
- Wolters PL, Martin S, Merker VL, et al. Patient-reported outcomes of pain and physical functioning in neurofibromatosis clinical trials. *Neurology*. 2016;87:S4-S12.
- World Health Organization Quality of Life Assessment (WHOQoL): development and general psychometric properties. *Soc Sci Med*. 1998;46(12):1569-1585.
- Fayed N, de Camargo OK, Kerr E, et al. Generic patient-reported outcomes in child health research: a review of conceptual content using World Health Organization definitions. *Dev Med Child Neurol*. 2012;54(2):1085-1095.
- Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care*. 1989;27(3):S217-S232.
- Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: properties, applications, and interpretation. *Health Qual Life Outcomes*. 2003;1:79.
- Gnanasakthy A, Barrett A, Evans E, D'Alessio D, Romano CD. A review of patient-reported outcomes labeling for oncology drugs approved by the FDA and the EMA (2012-2016). *Value Health*. 2019;22(2):203-209.
- Weiss B, Widemann BC, Wolters P, et al. Sirolimus for non-progressive NF1-associated plexiform neurofibromas: an NF Clinical Trials Consortium phase II study. *Pediatr Blood Cancer*. 2014;61(6):982-986.
- Victorson D, Barocas J, Song J, Cella D. Reliability across studies from the Functional Assessment of Cancer Therapy-General (FACT-G) and its subscales: a reliability generalization. *Qual Life Res*. 2008;17(9):1137-1146.
- Luckett T, King MT, Butow PN, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. *Ann Oncol*. 2011;22(10):2179-2190.
- Skevington SM, Lottly M, O'Connell KA, Group W. The World Health Organization's WHOQoL-BREF quality of life assessment: psychometric properties and results of

- the international field trial: a report from the WHOQoL group. *Qual Life Res.* 2004; 13(2):299-310.
23. Shawver Z, Griffith JD, Adams LT, Evans JV, Benchoff B, Sargent R. An examination of the WHOQoL-BREF using four popular data collection methods. *Comput Hum Behav.* 2016;55:446-454.
 24. Funes CJ, Mace RA, Macklin EA, Plotkin SR, Jordan JT, Vranceanu AM. First report of quality of life in adults with neurofibromatosis 2 who are deafened or have significant hearing loss: results of a live-video randomized control trial. *J Neurooncol.* 2019;143(3):505-513.
 25. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul Pediatr.* 2003; 3(6):329-341.
 26. Bullinger M, Brutt AL, Erhart M, Ravens-Sieberer U, Group BS. Psychometric properties of the KINDL-R questionnaire: results of the BELLA study. *Eur Child Adolesc Psychiatry.* 2008;17(suppl 1):125-132.
 27. Schmidt S, Debensason D, Muhlan H, et al. The DISABKIDS generic quality of life instrument showed cross-cultural validity. *J Clin Epidemiol.* 2006;59(6):587-598.
 28. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Qual Life Res.* 2009;18(7):873-880.
 29. Varni JW, Limbers CA, Neighbors K, et al. The PedsQL Infant Scales: feasibility, internal consistency reliability, and validity in healthy and ill infants. *Qual Life Res.* 2011;20(1):45-55.
 30. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 Generic Core Scales in healthy and patient populations. *Med Care.* 2001;39(8):800-812.
 31. Varni JW, Limbers CA. The PedsQL 4.0 Generic Core Scales young adult version: feasibility, reliability and validity in a university student population. *J Health Psychol.* 2009;14(4):611-622.
 32. 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.
 33. Nutakki K, Hingtgen CM, Monahan P, Varni JW, Swigonski NL. Development of the adult PedsQL neurofibromatosis type 1 module: initial feasibility, reliability and validity. *Health Qual Life Outcomes.* 2013;11:21.
 34. Nutakki K, Varni JW, Swigonski NL. PedsQL Neurofibromatosis Type 1 Module for children, adolescents and young adults: feasibility, reliability, and validity. *J Neurooncol.* 2018;137(2):337-347.
 35. Chren MM, Lasek RJ, Quinn LM, Mostow EN, Zyzanski SJ. Skindex, a quality-of-life measure for patients with skin disease: reliability, validity, and responsiveness. *J Invest Dermatol.* 1996;107(5):707-713.
 36. Smidt AC, Lai JS, Cella D, Patel S, Mancini AJ, Chamlin SL. Development and validation of Skindex-Teen, a quality-of-life instrument for adolescents with skin disease. *Arch Dermatol.* 2010;146(8):865-869.
 37. Nutakki K, Varni JW, Steinbrenner S, Draucker CB, Swigonski NL. Development of the Pediatric Quality of Life Inventory Neurofibromatosis Type 1 Module items for children, adolescents and young adults: qualitative methods. *J Neurooncol.* 2017;132(1):135-143.
 38. Fisher MJ, Shih CS, Rhodes SD, et al. Cabozantinib for neurofibromatosis type 1-related plexiform neurofibromas: a phase 2 trial. *Nat Med.* 2021;27(2):165-173.
 39. Hornigold RE, Golding JF, Leschziner G, et al. The NFTI-QoL: a disease-specific quality of life questionnaire for neurofibromatosis 2. *J Neurol Surg B Skull Base.* 2012;73(2):104-111.
 40. Shaffer BT, Cohen MS, Bigelow DC, Ruckenstein MJ. Validation of a disease-specific quality-of-life instrument for acoustic neuroma: the Penn Acoustic Neuroma Quality-of-Life Scale. *Laryngoscope.* 2010;120(8):1646-1654.
 41. Ferner RE, Shaw A, Evans DG, et al. Longitudinal evaluation of quality of life in 288 patients with neurofibromatosis 2. *J Neurol.* 2014;261(5):963-969.
 42. Halliday D, Emmanouil B, Pretorius P, et al. Genetic Severity Score predicts clinical phenotype in NF2. *J Med Genet.* 2017;54(10):657-664.
 43. Morris KA, Golding JF, Axon PR, et al. Bevacizumab in neurofibromatosis type 2 (NF2) related vestibular schwannomas: a nationally coordinated approach to delivery and prospective evaluation. *Neurooncol Pract.* 2016;3(4):281-289.
 44. Shukla A, Hsu FC, Slobogean B, et al. Association between patient-reported outcomes and objective disease indices in people with NF2. *Neurol Clin Pract.* 2019;9(4):322-329.
 45. Quarmby LM, Dalton LJ, Woolrich RA, et al. Screening and intervening: psychological distress in neurofibromatosis type 2 (NF2). *Psychooncology.* 2019;28(7):1583-1587.
 46. Lodder WL, Adan GH, Chean CS, Lesser TH, Leong SC. Validation of the facial dysfunction domain of the Penn Acoustic Neuroma Quality-of-Life (PANQoL) Scale. *Eur Arch Otorhinolaryngol.* 2017;274(6):2437-2442.
 47. van Leeuwen BM, Borst JM, Putter H, Jansen JC, van der Mey AG, Kaptein AA. Emotional intelligence in association with quality of life in patients recently diagnosed with vestibular schwannoma. *Otol Neurotol.* 2014;35(9):1650-1657.
 48. Dhayalan D, Lund-Johansen M, Finnkirk M, Tveiten OV. Fatigue in patients with vestibular schwannoma. *Acta Neurochir (Wien).* 2019;161(9):1809-1816.
 49. Sanagoo A, Jouybari L, Koochi F, Sayehmiri F. Evaluation of QoL in neurofibromatosis patients: a systematic review and meta-analysis study. *BMC Neurol.* 2019;19:123.
 50. *Guidance for Industry FDA: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* Food and Drug Administration; 2009.
 51. Gutmann DH, Ferner RE, Listerneck RH, Korf BR, Wolters PL, Johnson KJ. Neurofibromatosis type 1. *Nat Rev Dis Primers.* 2017;3(2):17004.
 52. Merker VL, Esparza S, Smith MJ, Stemmer-Rachamimov A, Plotkin SR. Clinical features of schwannomatosis: a retrospective analysis of 87 patients. *Oncologist.* 2012; 17(2):1317-1322.
 53. Torres Nupan MM, Velez Van Meerbeke A, Lopez Cabra CA, Herrera Gomez PM. Cognitive and behavioral disorders in children with neurofibromatosis type 1. *Front Pediatr.* 2017;5(10):227.
 54. Yost KJ, Thompson CA, Eton DT, et al. The Functional Assessment of Cancer Therapy - General (FACT-G) is valid for monitoring quality of life in patients with non-Hodgkin lymphoma. *Leuk Lymphoma.* 2013;54(2):290-297.
 55. Teckle P, Peacock S, McTaggart-Cowan H, et al. The ability of cancer-specific and generic preference-based instruments to discriminate across clinical and self-reported measures of cancer severities. *Health Qual Life Outcomes.* 2011;9:106.
 56. Pavel ME, Singh S, Strosberg JR, et al. Health-related quality of life for everolimus versus placebo in patients with advanced, non-functional, well-differentiated gastrointestinal or lung neuroendocrine tumours (RADIANT-4): a multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2017;18(10): 1411-1422.
 57. Hawthorne G, Herrman H, Murphy B. Interpreting the WHOQoL-bref: preliminary population norms and effect sizes. *Soc Indicators Res.* 2006;77:37-59.
 58. Varni JW, Limbers CA, Burwinkle TM. How young can children reliably and validly self-report their health-related quality of life? An analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes.* 2007;5:1.
 59. Lundin K, Stillesjo F, Nyberg G, Rask-Andersen H. Self-reported benefit, sound perception, and quality-of-life in patients with auditory brainstem implants (ABIs). *Acta Otolaryngol.* 2016;136(1):62-67.
 60. Nishiyama T, Oishi N, Kojima T, et al. Validation and multidimensional analysis of the Japanese Penn Acoustic Neuroma Quality-of-Life Scale. *Laryngoscope.* 2020;130(12): 2885-2890.

Neurology®

Current Recommendations for Patient-Reported Outcome Measures Assessing Domains of Quality of Life in Neurofibromatosis Clinical Trials

Pamela L. Wolters, Ana-Maria Vranceanu, Heather L. Thompson, et al.
Neurology 2021;97;S50-S63 Published Online before print July 6, 2021
DOI 10.1212/WNL.00000000000012421

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/S50.full
References	This article cites 59 articles, 6 of which you can access for free at: http://n.neurology.org/content/97/7_Supplement_1/S50.full#ref-list-1
Citations	This article has been cited by 2 HighWire-hosted articles: http://n.neurology.org/content/97/7_Supplement_1/S50.full##otherarticles
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Clinical trials Methodology/study design http://n.neurology.org/cgi/collection/clinical_trials_methodology_study_design Neurofibromatosis http://n.neurology.org/cgi/collection/neurofibromatosis Outcome research http://n.neurology.org/cgi/collection/outcome_research Quality of life http://n.neurology.org/cgi/collection/quality_of_life
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

Neurology® is the official journal of the American Academy of Neurology. Published continuously since 1951, it is now a weekly with 48 issues per year. Copyright © 2021 American Academy of Neurology. All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.

