Neuro-Oncology
19(5), 625–635, 2017 | doi:10.1093/neuonc/nox029

The Neurologic Assessment in Neuro-Oncology (NANO) scale: a tool to assess neurologic function for integration into the Response Assessment in Neuro-Oncology (RANO) criteria

Lakshmi Nayak, Lisa M. DeAngelis, Alba A. Brandes, David M. Peereboom, Evanthia Galanis, Nancy U. Lin, Riccardo Soffietti, David R. Macdonald, Marc Chamberlain, James Perry, Kurt Jaeckle, Minesh Mehta, Roger Stupp, Alona Muzikansky, Elena Pentsova, Timothy Cloughesy, Fabio M. Iwamoto, Joerg-Christian Tonn, Michael A. Vogelbaum, Patrick Y. Wen, Martin J. van den Bent, and David A. Reardon

Center for Neuro-Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts (L.N., P.Y.W., D.A.R.); Department of Neurology, Memorial Sloan-Kettering Cancer Center, New York, New York (L.M.D., E.P.); Department of Medical Oncology, Azienda USL-IRCCS Institute of Neurological Science, Bologna, Italy (A.A.B.); Rose Ella Burkhardt Brain Tumor and Neuro-Oncology Center, Cleveland Clinic, Cleveland, Ohio (D.M.P., M.A.V.); Department of Oncology, Mayo Clinic, Rochester, Minnesota (E.G.); Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts (N.U.L.); Department of Neurology and Neuro-Oncology, University of Turin, Turin, Italy (R.S.); Department of Oncology, London Health Sciences Centre, University of Western Ontario, London, Ontario, Canada (D.R.M.); Department of Neurology, Fred Hutchinson Cancer Research Center and Seattle Cancer Care Alliance, University of Washington School of Medicine, Seattle, Washington (M.C.); Division of Neurology, Sunnybrook Health Science Center, University of Toronto, Toronto, Ontario, Canada (J.P.); Department of Neurology and Hematology/Oncology, Mayo Clinic, Jacksonville, Florida (K.J.); Department of Radiation Oncology, Miami Cancer Institute, Miami, Florida (M.M.); Department of Oncology, University of Zurich, Zurich, Switzerland (R.S.); Biostatistics Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts (A.M.); Neuro-Oncology Program, Ronald Reagan UCLA Medical Center, University of California Los Angeles, Los Angeles, California (T.C.); Division of Neuro-Oncology, Neurological Institute of New York College of Physicians and Surgeons, Columbia University, New York, New York (F.M.I.); Department of Neurosurgery, LMU University of Munich, Munich, Germany (J.-C.T.); Neuro-Oncology Unit, Erasmus MC Cancer Center, Erasmus MC, Rotterdam, the Netherlands (M.J.B.)

Corresponding Author: Lakshmi Nayak, MD, Center for Neuro-Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, MA 02215, USA (Lakshmi_Nayak@dfci.harvard.edu).

Abstract

Background. The Macdonald criteria and the Response Assessment in Neuro-Oncology (RANO) criteria define radiologic parameters to classify therapeutic outcome among patients with malignant glioma and specify that clinical status must be incorporated and prioritized for overall assessment. But neither provides specific parameters to do so. We hypothesized that a standardized metric to measure neurologic function will permit more effective overall response assessment in neuro-oncology.

Methods. An international group of physicians including neurologists, medical oncologists, radiation oncologists, and neurosurgeons with expertise in neuro-oncology drafted the Neurologic Assessment in Neuro-Oncology (NANO) scale as an objective and quantifiable metric of neurologic function evaluable during a routine office examination. The scale was subsequently tested in a multicenter study to determine its overall reliability, inter-observer variability, and feasibility.

Results. The NANO scale is a quantifiable evaluation of 9 relevant neurologic domains based on direct observation and testing conducted during routine office visits. The score defines overall response criteria. A prospective, multinational study noted a >90% inter-observer agreement rate with kappa statistic ranging from 0.35 to 0.83 (fair to almost perfect agreement), and a median assessment time of 4 minutes (interquartile range, 3–5).

Conclusion. The NANO scale provides an objective clinician-reported outcome of neurologic function with high inter-observer agreement. It is designed to combine with radiographic assessment to provide an overall assessment of outcome for neuro-oncology patients in clinical trials and in daily practice. Furthermore, it complements existing patient-reported outcomes and cognition testing to combine for a global clinical outcome assessment of well-being among brain tumor patients.

Key words

brain tumor | neurologic function | outcome | response criteria

Improved understanding of brain tumor biology along with advances in drug development over the past decade have led to a substantial increase in the evaluation of novel treatments through clinical trials for neuro-oncology patients. Although an improvement in overall survival is considered the gold standard for oncology clinical trials, evaluating clinical benefit also constitutes a valuable endpoint.[1,2] However, the definition of *clinical benefit* may vary between different stakeholders, including physicians, regulatory agencies, the pharmaceutical industry, and most importantly, patients and their families. For patients with brain tumors, loss of neurologic integrity markedly compromises quality of life and was recently identified as a key priority regarding expectations of therapy benefit in a survey of 1851 brain tumor patients conducted by the Jumpstarting Brain Tumor Drug Development Coalition.[3,4] Ultimately, maintaining neurologic function is a paramount endpoint to all stakeholders.

Current Response Assessment

Outcome assessment based solely on radiographic criteria, which has been the gold standard for the assessment of treatment efficacy, can be an insufficient surrogate for survival. Furthermore, this can be particularly challenging in neuro-oncology, as imaging findings may be misleading and may not translate into clinical benefit. For example, some patients may worsen neurologically while their radiographic findings remain stable; alternatively, imaging can worsen as patients improve clinically. The Response Assessment in Neuro-Oncology (RANO) criteria were developed to address and standardize critical radiographic parameters used to assess therapeutic outcome among patients with high-grade glioma,[5] low-grade glioma,[6] or brain metastases more reliably.[7–9] Although the RANO criteria and the preceding Macdonald criteria[10] specify that clinical status be incorporated into the assessment of overall response, neither provide specific parameters to do so, and instead recommend that neurologic status be categorized as simply the same, better, or worse. Neurologic assessment in this context is subjective and nonspecific. Furthermore, different observers may have disparate standards, which may increase the likelihood of inaccurate and inconsistent overall response assessment.

Current Neurologic Assessment Tools

Characterizing neurologic status of brain tumor patients currently relies on the assessment of symptoms, quality of life, and performance status, as well as neurocognitive tests and the neurologic examination (Table 1). These measures are highly relevant for neuro-oncology patients but they are subject to limitations.

Measures of symptom burden and quality of life, such as the MD Anderson Symptom Inventory Brain Tumor module (MDASI-BT), the 30-item European Organisation for Research and Treatment of Cancer Quality of Life Core Questionnaire (EORTC QLQ-C30), the 20-item EORTC QLQ for Brain Neoplasm (-BN20) (specifically for brain tumor patients), and the Functional Assessment of Cancer Therapy –Brain (FACT-Br) broadly assess important aspects of day-to-day physical, social, and emotional well-being among brain tumor patients but are inherently subjective.[6,11–14] The MDASI-BT captures not only symptom severity but also interference with daily life which can predict tumor progression.[13] While such health-related quality of life assessment tests have been validated, they may be impacted by practical issues, including variations in patient compliance, response shift, as well as missing data.[8,11,14] Similarly, functional rating scales, such as the KPS and the Eastern Cooperative Oncology Group (ECOG) scales—which evaluate the ability of patients to care for themselves, work, and carry on normal activities—are also subjective.[15,16] Although these scales may predict prognosis and represent global assessments of functional status, they lack reproducibility[17] and fail to capture meaningful changes in neurologic function. Impairment in neurocognitive function is commonly seen in patients with brain tumors and its assessment is of immense value. The Mini-Mental State Examination is useful as a simple and brief screen of general neurocognitive function but lacks sensitivity and fails to detail memory, verbal fluency, visual-motor speed, and executive function, which are often impaired in brain tumor patients.[18,19] Accordingly, the Wechsler Adult Intelligence Scale–Revised, the Hopkins Verbal Learning Test–Revised, the Trail Making Tests, and the Controlled Oral Word Association are highly valuable, objective, and comprehensive measures of cognition and have also been reported to be predictive of progression among brain tumor patients.[19,20] Nonetheless, neurocognitive testing does not address other domains of neurologic function, and requires specialized expertise and dedicated time to administer.[6] Neurocognitive testing may also be impacted by drop out, timing, and frequency of tests and data interpretation.

While measurements of symptoms, quality of life, and global function, as well as objective neurocognitive testing, provide critical and unequivocal value for outcome assessment, none were designed to objectively assess

Table 1 Current neurologic outcome assessment tools

Endpoint	Applied Tools
Performance status	• Karnofsky performance status (KPS) • Eastern Cooperative Oncology Group (ECOG)
Symptom assessment	• MD Anderson Symptom Inventory Brain Tumor module (MDASI-BT)
Quality of life assessment	• European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30) • EORTC QLQ-BN20 (specifically for brain tumor patients) • Functional Assessment of Cancer Therapy–Brain (FACT-Br)
Neurocognitive assessment	• Mini-Mental State Examination (MMSE) • Wechsler Adult Intelligence Scale–Revised (WAIS-R) • Hopkins Verbal Learning Test–Revised (HVLT-R) • Trail Making Tests (TMT) • Controlled Oral Word Association (COWA)

neurologic function. As an additional concern, the etiology of neurologic deficits among neuro-oncology patients is often complex and may be due to treatment-related changes, comorbid events, changes in concurrent medications, and underlying tumor activity.

Detailed neurologic assessment scales for other neurologic subspecialties such as stroke (National Institutes of Health Stroke Scale [NIHSS]), multiple sclerosis (Expanded Disability Status Scale [EDSS]), Parkinson disease (Unified Parkinson Disease Rating Scale), ataxia (Scale for Assessment and Rating of Ataxia), myopathy (Kendall muscle scale), and amyotrophic lateral sclerosis (Amyotrophic Lateral Sclerosis Functional Rating Scale–Revised [ALSFRS-R]) have expanded in recent years and are widely utilized in daily practice as well as in the assessment of patients enrolled in clinical trials.[21] In contrast, a measurement scale of neurologic function has never been developed specifically for brain tumor patients.

We sought to develop a disease-specific, clinician-reported outcome (ClinRO) assessment tool to measure neurologic function across the multiple neurologic domains routinely assessed during an office examination that will provide objective and quantifiable data with adequate inter-observer agreement to provide a measure of neurologic outcome. In addition, such a ClinRO is envisioned to complement highly valuable existing patient-reported outcome (PRO) tools and the assessment of cognition and thereby provide, in aggregate, a comprehensive clinical outcome assessment (COA) of well-being among brain tumor patients.[3]

Development of the NANO Scale

We sought to develop a disease-specific, ClinRO assessment tool to measure neurologic function across multiple neurologic domains routinely assessed during an office examination that will provide objective and quantifiable data with adequate inter-observer agreement to score neurologic outcome.

An international, multidisciplinary working committee comprising neurologists, medical oncologists, radiation oncologists, and neurosurgeons with neuro-oncology expertise was convened and subsequently conducted biweekly teleconferences and semiannual meetings for 18 months to develop the Neurologic Assessment in Neuro-Oncology (NANO) scale. Similar to the RANO working group, the NANO working group includes leaders from major neuro-oncology institutions and brain tumor cooperative groups in the United States, Canada, and Europe. Formal progress reports were presented and additional volunteer committee members were solicited at the annual American Society of Clinical Oncology and Society for Neuro-Oncology meetings.

As a first step, the committee defined the purpose of the scale to be the objective measurement of neurologic function *relative* to underlying tumor activity. A comprehensive review of existing scales of neurologic function for other classes of neurologic, non-oncologic disorders was then performed to determine whether any of these scales could be appropriately adapted for the purpose of evaluating neurologic function in neuro-oncology patients. However, these scales were deemed suboptimal and insufficiently applicable for this purpose because they rely too heavily on the performance of a detailed neurologic examination completed by a trained neurologist or neurology subspecialist, are too time-consuming, or may be difficult to interpret. Some scales, such as the NIHSS, were in fact designed for non-neurologists, including emergency room nurses and physicians; however, this scale is not suited for neuro-oncology patients because it measures acute changes due to infarction that are based on vascular territories.[22] Furthermore, this scale is not intended for the assessment of sub-acute changes as are seen in progressive and invasive brain tumors. Scales such as EDSS for multiple sclerosis[23] and ALSFRS-R for amyotrophic lateral sclerosis[24] evaluate gradual progression in these chronic diseases but include specific elements that correlate with level of disability for these diseases and may be insensitive to clinical change more commonly observed among brain tumor patients.

Thus, the committee reasoned that a novel, standard neurologic examination scale was required to provide an objective and quantifiable measure of neurologic function. Furthermore, in order to be of optimal practical value and widely utilized, the committee agreed that the tool should be able to be rapidly and readily completed by neurologists and non-neurologists alike in the context of a routine office visit.

With these goals in mind, the NANO scale was organized into 9 relevant domains of neurologic function likely to be impacted by supratentorial, infratentorial, and brainstem lesions (Fig. 1). These domains were selected on the basis of the most common clinical features identified in patients with brain tumors.[4,25,26] Some aspects of neurologic function, such as visual acuity and cognition, were deliberately excluded because their effective assessment requires expertise and time that are beyond the scope of a routine neuro-oncology office visit. Relevant and discrete levels of

Scoring assessment is based on direct observation and testing performed during clinical evaluation and is not based on historical information or reported symptoms. Please check 1 answer per domain. Please check "Not assessed" if testing for that domain is not done. Please check "Not evaluable" if a given domain cannot be scored accurately due to pre-existing conditions, co-morbid events and/or concurrent medications.

Patient Identifier: _____

Date Assessment Performed (day/month/year): _____

Study time point (i.e. baseline, cycle 1, day 1, etc):_____

Assessment performed by (please print name): _____

Domains	Key Considerations
Gait 0 ☐ Normal 1 ☐ Abnormal but walks without assistance 2 ☐ Abnormal and requires assistance (companion, cane, walker, etc.) 3 ☐ Unable to walk ☐ Not assessed ☐ Not evaluable	• Walking is ideally assessed by at least 10 steps
Strength 0 ☐ Normal 1 ☐ Movement present but decreased against resistance 2 ☐ Movement present but none against resistance 3 ☐ No movement ☐ Not assessed ☐ Not evaluable	• Test each limb separately • Recommend assess proximal (above knee or elbow) and distal (below knee or elbow) major muscle groups • Score should reflect worst performing area • Patients with baseline level 3 function in one major muscle group/limb can be scored based on assessment of other major muscle groups/limb
Ataxia (upper extremity) 0 ☐ Able to finger to nose touch without difficulty 1 ☐ Able to finger to nose touch but difficult 2 ☐ Unable to finger to nose touch ☐ Not assessed ☐ Not evaluable	• Non-evaluable if strength is compromised • Trunk/lower extremities assessed by gait domain • Particularly important for patients with brainstem and cerebellar tumors • Score based on best response of at least 3 attempts
Sensation 0 ☐ Normal 1 ☐ Decreased but aware of sensory modality 2 ☐ Unaware of sensory modality ☐ Not assessed ☐ Not evaluable	• Recommend evaluating major body areas separately (face, limbs and trunk) • Score should reflect worst performing area • Sensory modality includes but not limited to light touch, pinprick, temperature and proprioception • Patients with baseline level 2 function in one major body area can be scored based on assessment of other major body areas

Fig. 1 Neurologic Assessment in Neuro-Oncology (NANO) scale.

function were then defined for each domain. The levels of function between normal and most severe for each domain were carefully defined so that they could be readily ascertained without ambiguity or overlap. Language to address scoring changes in neurologic function relative to preexisting deficits was incorporated. A parallel but separate effort was initiated to evaluate neurologic function of patients impacted by leptomeningeal tumor dissemination and will be reported elsewhere.[27,28]

As described below, the committee then defined how observed changes in neurologic function would be scored within and across the established domains to categorize response, progression, and stable disease. The agreed upon scoring approach was specifically defined to reflect clinically meaningful and readily measured changes in neurologic function. In order to ensure that observed changes reflect underlying tumor activity, a score of "non-evaluable" was included to address changes in neurologic function felt to be due to concurrent medication changes, side effects of therapy, or other comorbid events. Finally, in order to gain initial insight into whether the NANO scale and scoring system were adequately designed, an inter-observer variability study was conducted.

Scoring assessment is based on direct observation and testing performed during clinical evaluation and is not based on historical information or reported symptoms. Please check 1 answer per domain. Please check "Not assessed" if testing for that domain is not done. Please check "Not evaluable" if a given domain cannot be scored accurately due to pre-existing conditions, co-morbid events and/or concurrent medications.

Visual Fields

- 0 ☐ Normal
- 1 ☐ Inconsistent or equivocal partial hemianopsia (≥quadrantanopsia)
- 2 ☐ Consistent or unequivocal partial hemianopsia (≥quadrantanopsia)
- 3 ☐ Complete hemianopsia
- ☐ Not assessed
- ☐ Not evaluable

- Patients who require corrective lenses should be evaluated while wearing corrective lenses
- Each eye should be evaluated and score should reflect the worst performing eye

Facial Strength

- 0 ☐ Normal
- 1 ☐ Mild/moderate weakness
- 2 ☐ Severe facial weakness
- ☐ Not assessed
- ☐ Not evaluable

- Particularly important for brainstem tumors
- Weakness includes nasolabial fold flattening, asymmetric smile and difficulty elevating eyebrows

Language

- 0 ☐ Normal
- 1 ☐ Abnormal but easily conveys meaning to examiner
- 2 ☐ Abnormal and difficulty conveying meaning to examiner
- 3 ☐ Abnormal. If verbal, unable to convey meaning to examiner. OR non-verbal (mute/global aphasia)
- ☐ Not assessed
- ☐ Not evaluable

- Assess based on spoken speech. Non-verbal cues or writing should not be included.
- **Level 1**: Includes word finding difficulty; few paraphasic errors/neologisms/word substitutions; but able to form sentences (full/broken)
- **Level 2**: Includes inability to form sentences (<4 words per phrase/sentence); limited word output; fluent but "empty" speech.

Level of Consciousness

- 0 ☐ Normal
- 1 ☐ Drowsy (easily arousable)
- 2 ☐ Somnolent (difficult to arouse)
- 3 ☐ Unarousable/coma
- ☐ Not assessed
- ☐ Not evaluable

- None

Behavior

- 0 ☐ Normal
- 1 ☐ Mild/moderate alteration
- 2 ☐ Severe alteration
- ☐ Not assessed
- ☐ Not evaluable

- Particularly important for frontal lobe tumors
- Alteration includes but is not limited to apathy, disinhibition and confusion
- Consider subclinical seizures for significant alteration

Fig. 1 *Continued*

NANO Scale

Purpose: The primary purpose of the NANO scale is to objectively define clinical parameters of response and progression related to underlying tumor activity among neuro-oncology patients based on a clinician-performed neurologic examination. It is specifically designed to measure neurologic integrity as a ClinRO metric in day-to-day practice and to serve as a valid clinical trial endpoint. A secondary purpose of NANO is to complement existing subjective PRO measures and cognition testing to provide a comprehensive COA of well-being among neuro-oncology patients.

Salient Features: The NANO scale is designed to be a simple, clinician-friendly, real-time assessment of neurologic function readily performed within the time frame of routine office visits by non-neurologists and neurologists alike. In order to preserve objectivity, the ratings are based on direct observation and testing of the patient by the clinician and are not based on history or patient-reported symptoms. Additionally, evaluation of symptoms such as headaches, weakness, and seizures, which are frequent in brain tumor patients, was excluded, as they are adequately assessed in existing validated symptom burden inventories and quality of life assessment tools.

Domains: The NANO scale evaluates 9 major domains of neurologic function that are most relevant to patients with supratentorial, infratentorial, and brainstem tumors (Fig. 1), including gait, strength, upper extremity ataxia, sensation, visual fields, facial strength, language, level of consciousness, and behavior. As designed, the gait domain includes assessment of lower extremity ataxia.

Score: Each domain is subdivided into 3 or 4 levels of function with scores based on discrete quantifiable measures. Thus, levels of function for each domain range from 0 to 2 or 0 to 3. A score of 0 indicates normal function, while the highest score indicates the most severe level of deficit for that domain. Levels of function are distinguished by significant and measurable differences in order to avoid misinterpretation of subtle or nonspecific changes. A deliberate and systematic attempt was made to eliminate ambiguity between levels of function for each domain in order to keep the scale as objective as possible.

Additional Considerations

- Most domains are accompanied by user-friendly scoring guidelines to reinforce proper performance of the examination and maintain standardization.
- Premorbid neurologic deficits are documented at baseline visits.
- Domains with a preexisting most severe level of deficit at baseline may be evaluated at follow-up for assessment of improvement at that site or worsening at other anatomic locations assessed by that domain. For example, a patient with no movement of the left arm at baseline should still undergo evaluation of strength at subsequent visits, but the assessment score at these time points should reflect function of the remaining major muscle groups/limbs or improvement of the left arm.
- The sensation domain is intended only to assess loss or impairment of sensation from the underlying tumor rather than peripheral nerve lesions.

Definition of Neurologic Response: The NANO scale is intended to be performed at baseline and at follow-up visits, especially those where neuroimaging is obtained. An overall NANO score will be determined following assessment of each domain and will include one of 5 possible outcomes: neurologic response, neurologic progression, neurologic stability, not assessed, and non-evaluable.

Neurologic response is defined as a ≥2 level improvement in at least one domain without worsening in other domains from baseline or best level of function that is not attributable to change in concurrent medications or recovery from a comorbid event.

Neurologic stability indicates a score of neurologic function that does not meet criteria for neurologic response, neurologic progression, non-evaluable, or not assessed.

Neurologic progression is defined as a ≥2 level worsening from baseline or best level of function within ≥1 domain or worsening to the highest score within ≥1 domain that is felt to be related to underlying tumor progression and not attributable to a comorbid event or change in concurrent medication. Of note, an assessment of neurologic progression does not require evaluation of a minimum number of domains of the NANO scale if any of these conditions is met.

Non-evaluable should be selected if it is more likely than not that a factor other than underlying tumor activity contributed to an observed change in neurologic function. Such factors may include changes in a concurrent medication, such as corticosteroids, sedatives, narcotics, or anti-epileptic agents; acute or chronic adverse events related to therapeutic interventions; or a comorbid event such as a toxic-metabolic encephalopathy, post-ictal state, stroke, etc. Non-evaluable could also be selected if measurement of a given domain is not feasible due to an alteration of another domain. For example, assessment of upper extremity ataxia may not be possible if weakness of the extremity limits mobility. In this case, the strength domain should be assigned a numeric score but the upper extremity ataxia domain would be scored as non-evaluable.

Not assessed should be scored if the clinician omits evaluation of that particular domain during his/her examination. If a particular domain is marked not assessed at baseline, then that domain cannot thereafter be considered for progression or response. In general, assessment and scoring of all domains is encouraged.

An alternative scoring method based on a composite score of all 9 domains was also considered but deemed suboptimal because major changes affecting one or more domains were felt to more likely reflect underlying tumor activity than modest or small changes across multiple domains. Given that different manifestations of involvement of a certain part of the nervous system can be interrelated and impairment of more than one domain can result from a lesion in a particular region of the brain, a total score was felt to more likely overestimate tumor progression or response and it was the committee's consensus that a "significant" change in any one domain would be most appropriate to utilize rather than a composite score.

Inter-Observer Variability Study

For an objective scale of neurologic function to be useful as an outcome measure, key requirements include that it should: (i) readily detect changes in neurologic function in response to treatment and disease progression and (ii) exhibit adequate inter-observer agreement. Demonstration of acceptable inter-observer variability also provides reassurance that the levels of function specified for each domain are clearly defined by the scale. A prospective, international multicenter, multidisciplinary study was therefore conducted to determine the inter-observer variability of scoring each NANO scale domain.

Patients, Methods, and Study Design

This study protocol was approved by the institutional review boards of all participating centers. Eligible patients had a diagnosis of primary or secondary brain tumors, were ≥18 years of age, and were able to provide informed consent. Each patient underwent neurologic evaluation by 2 providers (physician, physician assistant, or nurse practitioner) on the same day during a scheduled, routine clinic visit that occurred in a neuro-oncology, medical oncology, or radiation oncology outpatient clinic setting (Fig. 2). At

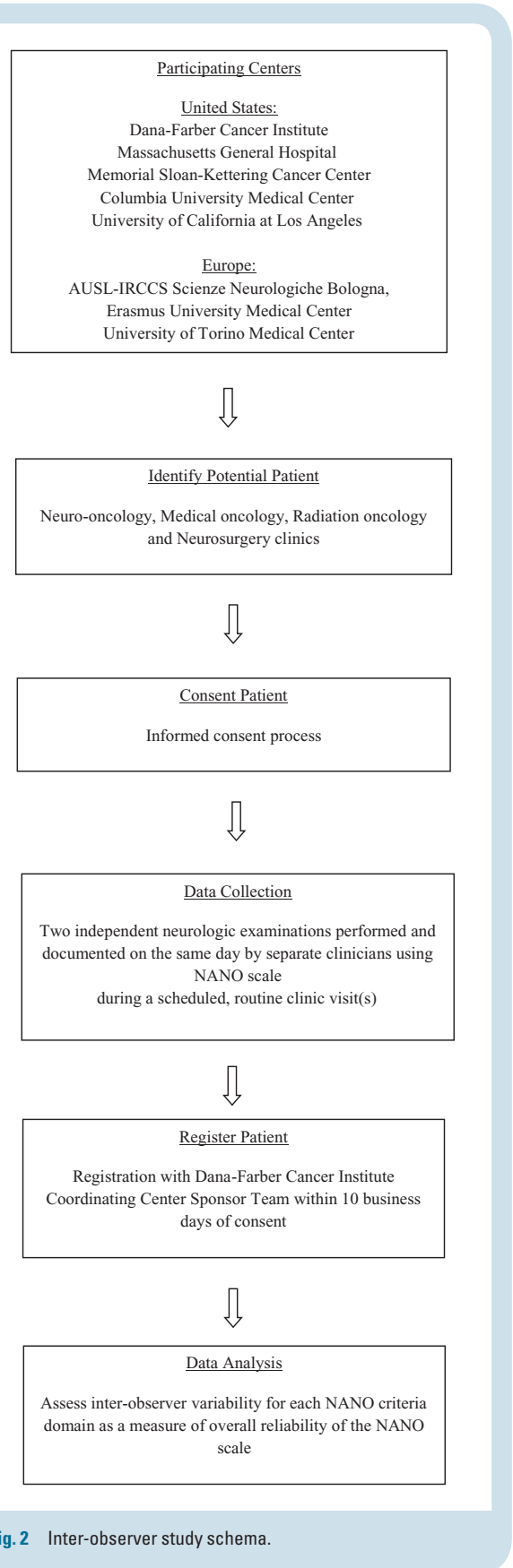

Fig. 2 Inter-observer study schema.

least one of the 2 providers was a physician. The primary provider (clinician A) confirmed eligibility; documented basic demographics, including age, sex, date of assessment, type of brain tumor, type of ongoing treatment if any; and performed an initial NANO neurologic examination. Subsequently, a second provider (clinician B), who was blinded to the findings determined by clinician A, performed a second neurologic NANO examination on the same patient. Each clinician independently documented his/her score using the NANO scale. Both clinicians also recorded the time required to complete the NANO assessment. Only an English version of the NANO scale was used.

Statistical Analysis

The primary endpoint was to evaluate the inter-observer variability of each NANO scale neurologic domain. We evaluated inter-observer agreement for each domain using the kappa statistic as well as the percent agreement.[29] The kappa statistic measures agreement between observers corrected for chance and ranged from –1 for no agreement to +1 for perfect agreement between observers, with 0 equating to agreement by chance.[30,31] The kappa measure can lower the estimate of agreement because it is also affected by the prevalence of the finding under consideration. Thus, for rare findings, the kappa method may be less reliable. Percent agreement is the proportion of evaluations in which observers report identical findings (in this case, scoring for each domain). Percent agreement does not account for chance but is reliable when raters are well trained and guessing is not expected. Percent agreement was calculated by the number of domains that received the same rating by both observers divided by the total number of domains rated by both observers per patient.

Estimates were obtained using the CI3Cats function of the KappaSize package in R software.[32] A sample size of 98 patients assessed by 2 raters would have a 95% one-sided CI for a lower bound of at least 0.7 for an anticipated kappa of 0.8 for domains with 3 score levels (0, 1, 2) and provided even greater precision for domains with 4 score levels. We expected a low occurrence of abnormalities in some domains, which would potentially lead to a lower kappa or an artificially higher percent agreement for those domains. Due to these concerns, the sample size was empirically doubled to 220 patients in order to further enhance the potential precision of the testing.

A secondary endpoint was to evaluate the time required to perform the NANO scale. Each clinician separately recorded the time taken to perform the NANO neurologic examination and the aggregate data were reported as a mean and median with range. The median time taken by clinicians who were board certified in neurology and those who were not was also calculated.

Results

Two hundred and twenty patients were accrued at 8 hospitals in North America and Europe (Fig. 2 and Table 2). The

percent agreement between observers was >90% per individual domain (Table 3). The kappa statistic ranged from 0.35 (fair agreement) for behavior to 0.83 (near perfect agreement) for language (Table 3). The kappa statistic was moderate to substantially high for the majority of the domains, including ataxia, sensation, facial strength, gait, strength, and visual fields. The kappa statistic was not computed for level of consciousness, as there was only 1 abnormal observation, and all the other patients evaluated were alert with normal level of consciousness. There were 15 instances in 13 patients when a given domain was deemed not assessed or non-evaluable by only one investigator. These scores were included in the statistical analyses.

The median and mean times for assessment of neurologic function using the NANO scale were 4 and 5 minutes, respectively (interquartile range [IQR], 3–5 min) (Table 4). The median time to complete the NANO scale for clinicians with (n = 185) and without board certification in neurology (n = 255) was 5 minutes (IQR, 3–10 min) and 3 minutes (IQR, 3–5 min), respectively.

Discussion

Standardization enhances the inherent value of clinical assessment, especially for challenging patient populations such as those with CNS tumors. The current standard for clinical response assessment in neuro-oncology simply scores patients as "better, worse or unchanged"[5–7] and is therefore insensitive, subjective, and prone to inter-observer variability. Symptom and quality of life assessments provide highly valuable perspective on overall well-being of brain tumor patients but are inherently subjective. Measurement of cognitive dysfunction among brain tumor patients is also of paramount value but requires special expertise and dedicated time that are beyond what is available during a routine outpatient oncology visit.

The NANO scale was developed to provide a standardized and objective ClinRO tool to assess neurologic function among neuro-oncology patients. Nonetheless, clinical assessment is limited without assessment of quality of life and symptoms. Thus, the NANO scale is intended to complement PRO measures and neurocognitive assessment tools which can be combined to provide COA of well-being for brain tumor patients.[3] Importantly, the NANO scale should not be considered to replace existing PRO tools or cognitive testing in clinical trials or daily practice.

Given the increase in clinical trials for neuro-oncology patients, the NANO scale could be integrated into future therapeutic clinical trials and serve as an objective, uniformly applied endpoint. Development of the NANO scale is thus a timely addition to the recently updated RANO criteria for high-grade gliomas,[5] low-grade gliomas,[6] brain metastases,[9] and other CNS tumors as applicable. The NANO criteria are specifically designed to integrate with the radiographic criteria specified by RANO to generate

Table 2 Patient demographics for NANO scale inter-observer variability study

Patient Characteristics	N = 220
Median age, y (range)	54 (22–87)
Sex	
Women	107 (49%)
Men	113 (51%)
Type of brain tumor	
Low-grade glioma	19 (9%)
High-grade glioma	152 (69%)
Brain metastases	25 (11%)
Other primary brain tumors	24 (11%)
Ongoing treatment at the time of NANO assessment	
Chemotherapy	88 (40%)
Radiation	1 (0.5%)
Chemoradiation	18 (8.2%)
Other	31 (14%)
None	82 (37.3%)

Table 3 Inter-observer agreement rate per individual domain

Domain	Inter-Observer Percent Agreement (not chance-adjusted)	Inter-Observer Agreement (kappa statistic*)
Gait	90.8%	0.76
Strength	93.6%	0.80
Ataxia (upper extremity)	90.7%	0.45
Sensation	93.6%	0.50
Visual field	93.0%	0.76
Facial strength	91.7%	0.53
Language	96.4%	0.83
Level of consciousness	99.5%	**
Behavior	95.5%	0.35

*<0.00, poor agreement; 0.00–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; 0.81–1.0, almost perfect agreement.[30]
**Kappa not calculated as too few observations.

Table 4 Time taken for NANO evaluation and scoring in 220 patients*

Time (minutes)	Number of Evaluations (total = 440)	Percent
<3	49	11
3 to <5	215	49
5 to <10	109	25
≥10	67	15

*440 evaluations; median time: 4 minutes.

an objective overall response assessment. Specifically, an overall RANO assessment of response or stable disease mandates that patients be clinically stable or improved. Thus, in order to achieve an overall assessment of response or stable disease, patients would be required to satisfy appropriate imaging and NANO criteria defining these response categories. In contrast, patients can be designated as overall progressive if they satisfy *either* imaging or NANO criteria of progressive disease. Specifically, in this situation, patients would receive an overall assessment of progression if NANO criteria are met, regardless of imaging findings. Importantly, an objective assessment of clinical deterioration attributed to underlying tumor progression represents a standard reason to withdraw patients from therapeutic clinical trials. There is a possibility of introducing bias into the clinical assessment of patients when physicians evaluate the results of the imaging prior to the neurologic evaluation. While this is a limitation, such an assessment reflects real world practice because clinicians routinely review imaging before seeing patients.

Our inter-observer variability study demonstrated that the NANO scale exhibits adequate reliability for each domain and that its performance during a routine clinic visit is feasible. Specifically, we noted that the percent agreement was over 90% for each domain. The chance-adjusted inter-observer agreement (kappa statistic) was *substantial* for the major domains affected in neuro-oncology patients, including gait, strength, and visual fields, while an *almost perfect score* was achieved for language.[30] As expected, domains that are affected less commonly had a lower chance-adjusted agreement. Specifically, ataxia, sensation, and facial strength, which are more commonly abnormal among patients with less common brainstem and cerebellar tumors, demonstrated *moderate* agreement, while behavior had *fair* agreement.[30] By design, patients with a significant impairment of consciousness were not included in this study, as all patients had to give informed consent. Thus, an assessment of agreement and the kappa was not computed for this domain. Although this deficiency represents a limitation of our inter-observer variability analysis, behavioral alterations can occur among patients with frontal lobe tumors and thus represent an important domain of the NANO scale. Similarly, while impairment in level of consciousness is uncommon in the routine outpatient clinic setting, it is important to assess, particularly among patients with thalamic and brainstem tumors. Although an assessment tool with fewer neurologic domains could improve the chance-adjusted inter-observer kappa agreement score, such a tool would be less comprehensive and thus provide lower overall utility for neuro-oncology patient assessment. Further analysis in a larger patient sample size will be helpful to evaluate the kappa statistic for these specific neurologic domains that are infrequently observed. Moreover, the moderate or higher agreement noted for the majority of domains is comparable to the inter-observer agreement noted for domains currently broadly utilized in various stroke assessment scales, such as NIHSS, the Mathew scale, the Canadian Neurological Scale, and the Stroke Data Bank.[33–36] Similar to the NANO scale, these scales also demonstrate high agreement for domains such as language and motor function (strength) and fair to moderate agreement for domains such as sensation, visual fields, ataxia, facial weakness, and level of consciousness.

We also demonstrated that the NANO scale can be readily performed in the context of a routine clinic visit by both neurologists and non-neurologists. Specifically, the median time to complete the neurologic examination and NANO scoring was under 5 minutes for all clinicians regardless of whether they had dedicated neurology training, indicating that the NANO scale can be incorporated into routine office assessments of brain tumor patients by general medical oncologists and other clinicians without formal neurology training.

The currently designed NANO scale does have limitations. First, it was developed specifically for adults, and the inter-observer variability study was conducted in patients who were ≥18 years of age. Future extension of the NANO scale could include provision for the pediatric population. Second, a separate scale is in development to assess the potentially complex and subtle aspects of neurologic function associated with patients who have leptomeningeal tumor dissemination.[27] Third, the inter-observer variability study was conducted at prominent neuro-oncology centers where clinicians have significant experience caring for brain tumor patients, and thus may not reflect community practice. Future studies could consider further testing of the NANO scale in community centers. Fourth, we did not evaluate learning effect, that is, whether clinicians who participated multiple times in the study decreased their time with repeat performance. Physicians who evaluated more patients with the NANO scale may have been subject to this effect, which in turn may have impacted the time taken and hence the results of the study. Nonetheless, such an effect would actually be of value given that this tool is expected to be routinely used in daily practice. Fifth, we did not assess intra-observer reliability of the NANO scale but rationalized that this was less likely to reflect whether the scale was adequately defined than the inter-observer variability. Finally, we did not correlate the NANO score with imaging data or survival. Future validation studies could aim to prospectively determine whether changes in NANO score predict radiographic outcome as well as survival.

Conclusion

The NANO scale is an objective, relevant, rapid, and simple tool to measure neurologic function among neuro-oncology patients. We confirmed that the NANO scale is associated with a high rate of inter-observer agreement, which provides reassurance that its domain-specific levels of function are effectively and clearly articulated. These results support incorporation of the NANO scale to assess neurologic function in planned clinical trials and routine office assessments of neuro-oncology patients. We postulate that assessment of neurologic outcome by the NANO scale when combined with radiologic assessment as outlined in the RANO criteria will generate a

robust measure of outcome for neuro-oncology patients that surpasses currently used criteria. Furthermore, NANO complements existing highly valuable symptom burden and quality of life PRO measures and cognitive testing. The combination of NANO with these measures represents a comprehensive global COA of well-being among neuro-oncology patients. Finally, NANO provides standardization for the assessment of neurologic function that will lead to more consistent and accurate assessment of this important endpoint in clinical trials. Further efforts will address the translation of the NANO scale as a meaningful and reproducible measure of clinical response in the context of prospective studies, and the degree to which it can supplement assessments by performance status and PRO measures. Specific planned next steps include incorporation of the NANO scale in conjunction with RANO criteria in clinical trials to prospectively assess its validity and utility relative to radiographic and overall outcome.

Acknowledgment

This report has been presented in part at the annual meetings of the Society for Neuro-Oncology and the American Academy of Neurology.

Conflict of interest statement. Authors have no conflict of interest to declare.

References

1. Booth CM, Tannock I. Reflections on medical oncology: 25 years of clinical trials—where have we come and where are we going? *J Clin Oncol*. 2008;26(1):6–8.
2. Wilson MK, Karakasis K, Oza AM. Outcomes and endpoints in trials of cancer treatment: the past, present, and future. *Lancet Oncol*. 2015;16(1):e32–e42.
3. Helfer JL, Wen PY, Blakeley J, Gilbert MR, Armstrong TS. Report of the Jumpstarting Brain Tumor Drug Development Coalition and FDA clinical trials clinical outcome assessment endpoints workshop (October 15, 2014, Bethesda MD). *Neuro Oncol*. 2016; 18 Suppl 2: ii26–ii36.
4. Armstrong TS, Bishof AM, Brown PD, Klein M, Taphoorn MJ, Theodore-Oklota C. Determining priority signs and symptoms for use as clinical outcomes assessments in trials including patients with malignant gliomas: Panel 1 Report. *Neuro Oncol*. 2016;18 Suppl 2:ii1–ii12.
5. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: Response Assessment in Neuro-Oncology working group. *J Clin Oncol*. 2010;28(11):1963–1972.
6. van den Bent MJ, Wefel JS, Schiff D, et al. Response assessment in neuro-oncology (a report of the RANO group): assessment of outcome in trials of diffuse low-grade gliomas. *Lancet Oncol*. 2011;12(6):583–593.
7. Lin NU, Lee EQ, Aoyama H, et al.; Response Assessment in Neuro-Oncology (RANO) group. Challenges relating to solid tumour brain metastases in clinical trials, part 1: patient population, response, and progression. A report from the RANO group. *Lancet Oncol*. 2013;14(10):e396–e406.
8. Lin NU, Wefel JS, Lee EQ, et al.; Response Assessment in Neuro-Oncology (RANO) group. Challenges relating to solid tumour brain metastases in clinical trials, part 2: neurocognitive, neurological, and quality-of-life outcomes. A report from the RANO group. *Lancet Oncol*. 2013;14(10):e407–e416.
9. Lin NU, Lee EQ, Aoyama H, et al.; Response Assessment in Neuro-Oncology (RANO) group. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol*. 2015;16(6):e270–e278.
10. Macdonald DR, Cascino TL, Schold SC Jr, Cairncross JG. Response criteria for phase II studies of supratentorial malignant glioma. *J Clin Oncol*. 1990;8(7):1277–1280.
11. Armstrong TS, Gilbert MR. Net clinical benefit: functional endpoints in brain tumor clinical trials. *Curr Oncol Rep*. 2007;9(1):60–65.
12. Armstrong TS, Mendoza T, Gning I, et al. Validation of the M.D. Anderson Symptom Inventory Brain Tumor module (MDASI-BT). *J Neurooncol*. 2006;80(1):27–35.
13. Armstrong TS, Vera-Bolanos E, Gning I, et al. The impact of symptom interference using the MD Anderson Symptom Inventory–Brain Tumor module (MDASI-BT) on prediction of recurrence in primary brain tumor patients. *Cancer*. 2011;117(14):3222–3228.
14. Walker M, Brown J, Brown K, Gregor A, Whittle IR, Grant R. Practical problems with the collection and interpretation of serial quality of life assessments in patients with malignant glioma. *J Neurooncol*. 2003;63(2):179–186.
15. Orr ST, Aisner J. Performance status assessment among oncology patients: a review. *Cancer Treat Rep*. 1986;70(12):1423–1429.
16. Roila F, Lupattelli M, Sassi M, et al. Intra and interobserver variability in cancer patients' performance status assessed according to Karnofsky and ECOG scales. *Ann Oncol*. 1991;2(6):437–439.
17. Hutchinson TA, Boyd NF, Feinstein AR, Gonda A, Hollomby D, Rowat B. Scientific problems in clinical scales, as demonstrated in the Karnofsky index of performance status. *J Chronic Dis*. 1979;32(9-10):661–666.
18. Dick JP, Guiloff RJ, Stewart A, et al. Mini-Mental State Examination in neurological patients. *J Neurol Neurosurg Psychiatry*. 1984;47(5):496–499.
19. Meyers CA, Brown PD. Role and relevance of neurocognitive assessment in clinical trials of patients with CNS tumors. *J Clin Oncol*. 2006;24(8):1305–1309.
20. Meyers CA, Hess KR. Multifaceted end points in brain tumor clinical trials: cognitive deterioration precedes MRI progression. *Neuro Oncol*. 2003;5(2):89–95.
21. Herndon RM. *Handbook of Neurologic Rating Scales*. 2nd ed. New York: Demos Medical Publishing; 2006.
22. Brott T, Adams HP Jr, Olinger CP, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke*. 1989;20(7):864–870.
23. Kurtzke JF. Further notes on disability evaluation in multiple sclerosis, with scale modifications. *Neurology*. 1965;15:654–661.
24. Cedarbaum JM, Stambler N, Malta E, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). *J Neurol Sci*. 1999;169(1-2):13–21.
25. Iwamoto FM, Cooper AR, Reiner AS, Nayak L, Abrey LE. Glioblastoma in the elderly: the Memorial Sloan-Kettering Cancer Center experience (1997-2007). *Cancer*. 2009;115(16):3758–3766.
26. Armstrong TS, Cohen MZ, Eriksen LR, Hickey JV. Symptom clusters in oncology patients and implications for symptom research in people with primary brain tumors. *J Nurs Scholarsh*. 2004;36(3):197–206.
27. Chamberlain M, Soffietti R, Raizer J, et al. Leptomeningeal metastasis: a Response Assessment in Neuro-Oncology critical review of endpoints and response criteria of published randomized clinical trials. *Neuro Oncol*. 2014;16(9):1176–1185.

28. Chamberlain M, Junck L, Brandsma D, et al. Leptomeningeal metastases: a RANO proposal for response criteria. *Neuro Oncol*. 2016; 16(Suppl 2): ii56.
29. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276–282.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
31. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Fam Med*. 2005;37(5):360–363.
32. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960; (20):37–46.
33. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol*. 1989;46(6):660–662.
34. Côté R, Battista RN, Wolfson C, Boucher J, Adam J, Hachinski V. The Canadian Neurological Scale: validation and reliability assessment. *Neurology*. 1989;39(5):638–643.
35. D'Olhaberriague L, Litvan I, Mitsias P, Mansbach HH. A reappraisal of reliability and validity studies in stroke. *Stroke*. 1996;27(12):2331–2336.
36. Shinar D, Gross CR, Mohr JP, et al. Interobserver variability in the assessment of neurologic history and examination in the stroke data bank. *Arch Neurol*. 1985;42(6):557–565.