

UCLA

UCLA Previously Published Works

Title

Radiographic read paradigms and the roles of the central imaging laboratory in neuro-oncology clinical trials

Permalink

<https://escholarship.org/uc/item/6ng1h7pp>

Journal

Neuro-Oncology, 23(2)

ISSN

1522-8517

Authors

Ellingson, Benjamin M
Brown, Matthew S
Boxerman, Jerrold L
et al.

Publication Date

2021-02-25

DOI

10.1093/neuonc/noaa253

Peer reviewed

Radiographic read paradigms and the roles of the central imaging laboratory in neuro-oncology clinical trials

Benjamin M. Ellingson^o, Matthew S. Brown, Jerrold L. Boxerman^o, Elizabeth R. Gerstner, Timothy J. Kaufmann, Patricia E. Cole, Jeffrey A. Bacha, David Leung, Amy Barone, Howard Colman, Martin J. van den Bent, Patrick Y. Wen, W. K. Alfred Yung, Timothy F. Cloughesy, and Jonathan G. Goldin

UCLA Brain Tumor Imaging Laboratory (BTIL), Center for Computer Vision and Imaging Biomarkers, University of California Los Angeles, Los Angeles, California, USA (B.M.E., M.S.B., J.G.G.); Department of Radiological Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E., M.S.B., J.G.G.); Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (B.M.E.); UCLA Neuro-Oncology Program, University of California Los Angeles, Los Angeles, California, USA (B.M.E., T.F.C.); Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA (T.F.C.); Center for Neuro-Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA (P.Y.W.); The University of Texas MD Anderson Cancer Center, Houston, Texas, USA (W.K.A.Y.); Department of Diagnostic Imaging, Warren Alpert Medical School, Brown University, Providence, Rhode Island, USA (J.L.B.); Department of Neuro-Oncology, Erasmus Medical Center Cancer Institute, Rotterdam, Netherlands (M.J.v.d.B.); Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA (E.R.G.); Edison Oncology Holding Corp, London, UK (J.A.B.); Bayer Pharmaceuticals, Whippany, New Jersey, USA (P.E.C.); Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah, USA (H.C.); Bristol Myers Squibb, New York, New York, USA (D.L.); MedQIA, Los Angeles, California, USA (M.S.B., J.G.G.); US Food and Drug Administration, Silver Spring, Maryland, USA (A.B.); Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA (T.J.K.)

Corresponding Author: Benjamin M. Ellingson, Ph.D., Director, UCLA Brain Tumor Imaging Laboratory (BTIL), Professor of Radiology and Psychiatry, Departments of Radiological Sciences and Psychiatry, David Geffen School of Medicine, University of California, Los Angeles, 924 Westwood Blvd., Suite 615, Los Angeles, CA 90024 (bellingson@mednet.ucla.edu).

Abstract

Determination of therapeutic benefit in intracranial tumors is intimately dependent on serial assessment of radiographic images. The Response Assessment in Neuro-Oncology (RANO) criteria were established in 2010 to provide an updated framework to better characterize tumor response to contemporary treatments. Since this initial update a number of RANO criteria have provided some basic principles for the interpretation of changes on MR images; however, the details of how to operationalize RANO and other criteria for use in clinical trials are ambiguous and not standardized. In this review article designed for the neuro-oncologist or treating clinician, we outline essential steps for performing radiographic assessments by highlighting primary features of the *Imaging Charter* (referred to as the Charter for the remainder of this article), a document that describes the clinical trial imaging methodology and methods to ensure operationalization of the Charter into the workings of a clinical trial. Lastly, we provide recommendations for specific changes to optimize this methodology for neuro-oncology, including image registration, requirement of growing tumor for eligibility in trials of recurrent tumor, standardized image acquisition guidelines, and hybrid reader paradigms that allow for both unbiased measurements and more comprehensive interpretation.

Keywords

clinical trials | Imaging Charter | imaging endpoints | neuro-oncology | RANO

Determination of therapeutic benefit in intracranial tumors is intimately dependent on serial assessment of radiographic images and clinical status. While a number of RANO criteria provide some basic principles for the interpretation of changes on these images, there are a number of choices that need to be made to implement or operationalize radiographic endpoints (Figure 1). The process of radiographic response assessment starts with image acquisition at the individual sites. After images have been acquired and sent to the independent radiologic facility or imaging contract research organization (CRO), a particular radiographic read paradigm is employed by the independent readers, which involves particular rules about selection of images to review, order to display them, rules of adjudication, etc. Readers then make measurements on the images according to the read paradigm and in accordance with specific measurement guidelines (eg, bidirectional measurements, volumetric segmentation). From these measurements, particular response criteria including RANO and other endpoints (eg, growth modeling) can be applied.

For trials intended to support regulatory decision making, the FDA recommends development of an Imaging Charter in order to specify and standardize these imaging procedures among the key clinical trial stakeholders, including imaging acquisition, image interpretation, and imaging data management. The Charter should be authored with the same rigor as a protocol and serve as the document against which the imaging aspects of a trial can be audited. The Charter should describe the clinical trial imaging methodology, including modality-specific technical details, what data get transferred to and from the sites and sponsor, details on personnel involved in the entire process, image interpretation, image archiving procedures, and tumor response assessment. In this review article, we outline essential steps for performing radiographic assessments by highlighting primary aspects of the Charter and methods to ensure its operationalization

into the workings of a clinical trial. Additional guidance for use of imaging endpoints in clinical trials can be found at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-trial-imaging-endpoint-process-standards-guidance-industry>

Types of Radiographic Reads

There are typically 2 types of radiographic reads performed by a centralized imaging core lab in neuro-oncology—eligibility reads and response reads (Table 1). (Note that safety reads are less common in neuro-oncology.) The FDA recommends central blinded review in situations where clinical site image interpretation is variable and results of image measurements are important for eligibility determination, safety, or response endpoints. This recommendation is based on the agency's assertion that the centralized process can better provide verifiable and uniform reader training, as well as ongoing management of reader performance, ensuring the process produces high data quality and integrity, and that bias, interreader and intrareader variability are minimized. Eligibility reads, as the name implies, involve standardized methods for determining whether image-based study inclusion and/or exclusion criteria are met across all sites and screened patients. This might include minimum or maximum size restrictions, like those required to classify “measurable disease” by RANO criteria (ie, ≥ 1 cm \times 1 cm on at least 2 slices that are 5 mm or less thick with no interslice gap, otherwise the minimum size is twice the slice thickness plus the interslice gap), or assessments to verify growing tumor as an entry criterion in studies for recurrent disease. Other examples of eligibility reads include presence or absence of enhancement in lower-grade glioma studies, where large regions of

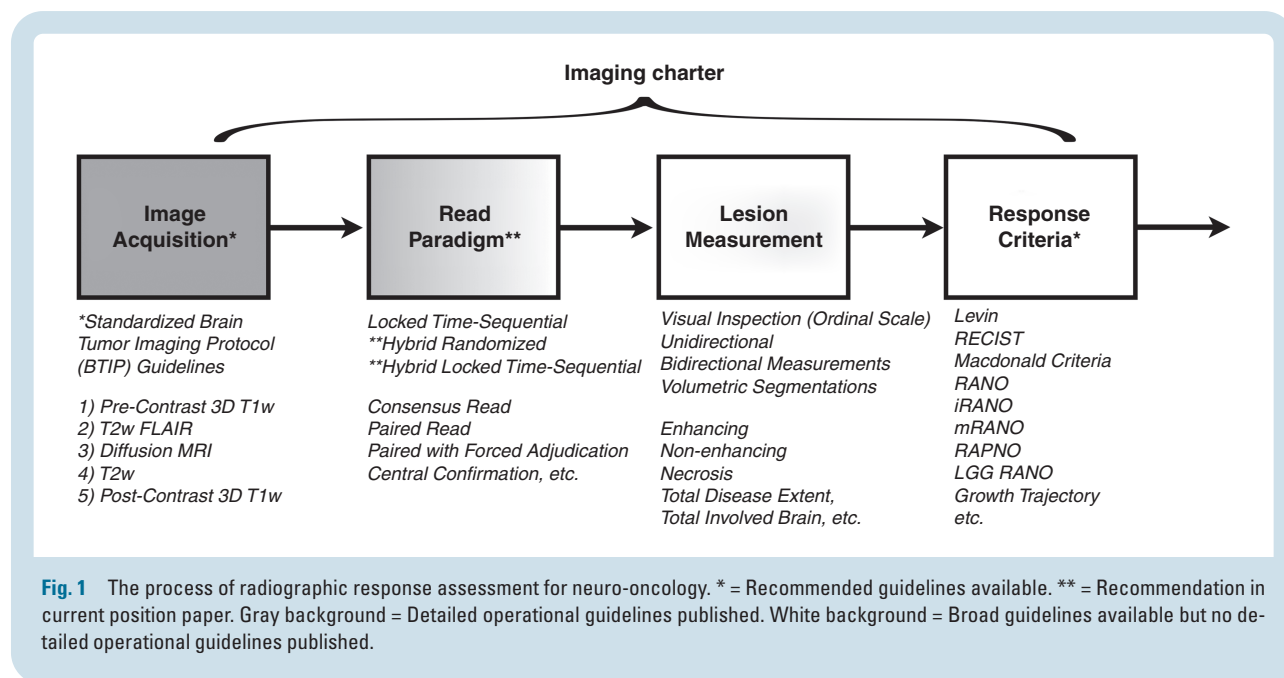


Table 1 Types of radiographic reads

Type of Read	Purpose	Example
Eligibility Read	To centrally confirm radiographic inclusion and/or exclusion criteria for a study (lesion size or features, etc.)	Confirming “measurable disease” (>1 cm × 1 cm) in recurrent disease with response as an endpoint
Response Read	To evaluate radiographic response and/or confirm date of disease progression after experimental treatment. Most common category of radiographic evaluation in neuro-oncology.	Radiographic response (PR, CR, SD, PD) Date of radiographic progression (PFS) Date of best response

Abbreviations: PR, partial response; CR, complete response; SD, stable disease; PD, progressive disease.

enhancement might suggest a more aggressive phenotype with atypical sensitivity to the particular treatment.

Radiographic reads for response are the most common task performed by neuroradiologists as part of a clinical trial that incorporates evaluation by CRO (Table 1). Response reads require the neuroradiologist to evaluate a series of images and determine whether the tumor is getting smaller or larger while on the study. Neuroradiologists would typically use their measurements, or properly authenticated measurements made by a technologist, at specified time points for application of the appropriate response criteria. Traditionally, overall response would then be computed through integration of radiographic and clinical information by a neuro-oncologist; however, we recommend that overall response be computed by standardized software that integrates steroid and neurological status. Below, we will discuss the challenges and importance of consistent algorithmic implementation of response criteria, highlighting the need for standardization and automation. The goal of the response read is to evaluate for response and/or disease progression in order to generate data supporting treatment response endpoints including objective durable response and date of best response, and to determine the date of progression for estimating time-to-progression, progression-free survival (PFS), and duration of clinical benefit (DOCB), defined as the time from first confirmed response or disease stabilization until radiographic disease progression or death.

In addition to eligibility reads and response reads that support study endpoints, adaptive trial designs may require central confirmation of disease progression before switching to an alternate therapy or salvage pathway. These reads need to be performed as scans are received, necessitating read paradigms that include a review of all serial images obtained up until the date of review. Careful consideration must be paid to the interaction between on-site, local reads and central reads in the context of such trials. As such, these progression confirmation reads are often performed with a single reader, independent of

the primary paradigm used for formal radiographic study endpoints.

Image Acquisition

It is important that standardized image acquisition protocols are used in studies across all sites and time points for a given subject. The Charter should specify the modality or modalities to be used, image acquisition parameters, schedule of assessments, procedures for site qualification and image de-identification, specifications for the upload and secure transfer of images, as well as methods for monitoring these aspects. These aspects should be addressed whenever imaging is to be used in a clinical trial. (Note: Often a separate *Imaging Manual* is used to describe site-specific image acquisition parameters in more detail, with the Charter referring to the Imaging Manual. It is important the language in the study protocol, Charter, Imaging Manual, and the statistical analysis plan are consistent).

Image Acquisition Protocols

Standardized brain tumor imaging protocols (BTIPs) for use in clinical trials have been established for both high-grade gliomas¹ and brain metastases² based on international consensus recommendations. Since 2015, use of BTIP has been required and widely implemented for therapeutic clinical trials in malignant gliomas with few issues. While not identical to the standard of care (SOC) exams at some institutions, most sites have been able and willing to include both the core sequences required for clinical trials as well as the additional site-specific sequences required for local SOC exams. The use of standardized BTIPs for brain tumor clinical trials will permit better harmonization of both image acquisition and image interpretation, as well as potentially facilitate cross-trial comparisons.

While most brain tumor clinical trials require the use of standardized BTIPs, this choice is made by the study sponsor and not the imaging CRO per se. Due to the cost or perceived logistical challenges associated with standardizing image acquisition in a multicenter study, many sponsors choose only to require SOC images, even from unqualified MRI systems and sites. For example, sites may insist a trial accept their SOC images and procedures, but if the radiology team is directly engaged early in the trial, they can often easily accommodate the such requirements. Experience suggests that many trials targeting accelerated approval or full registration may use a standardized BTIP and imaging CRO, but earlier phase trials do not always require this level of rigor. As a consequence, response assessment has been inconsistent and of questionable reliability, with adjudication rates > 40%³⁻⁵ (compared with ~30–40% for non-CNS trials⁶), as differences in image features and interpretation can result from differences in pulse sequence acquisition parameters (echo time, repetition time, inversion time, flip angle, etc), fat saturation details, MRI scanner field strength, and bandwidth or voxel size (Figure 2A–E). While high adjudication rates do not necessarily mean the assessment is

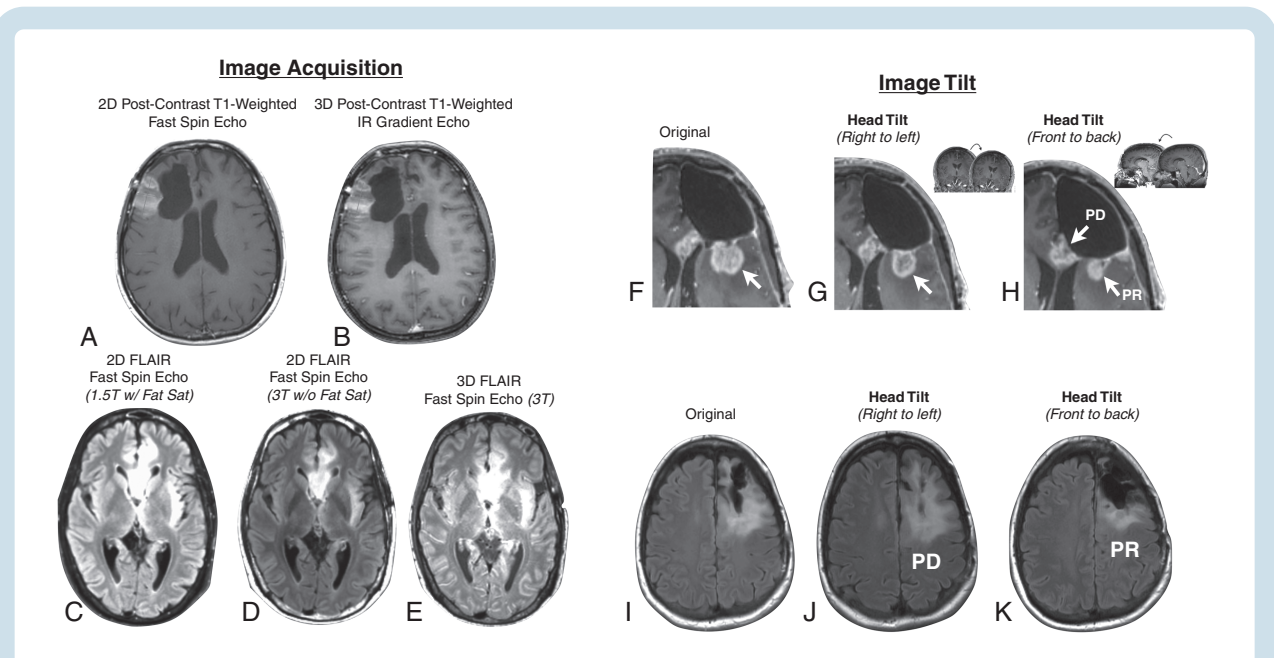


Fig. 2 Confounds to image interpretation including differences in image acquisition and image tilt. (A) 2D post-contrast T1-weighted image of a glioblastoma patient acquired with TE = 13 ms, TR = 560 ms, and slice thickness of 3 mm (bidimensional lesion measurements = 7.53 mm²). (B) Same patient at same time point showing 3D postcontrast T1-weighted images with TE = 3 ms, TR = 10ms, and slice thickness of 1 mm (bidimensional lesion measurements = 7.05 mm²). This resulted in a 6.6% difference in measurements, enough to influence response assessment. (C) 2D T2-weighted FLAIR images of a low-grade glioma scanned on a 1.5T scanner with TI = 2500 ms, TE = 98 ms, TR = 9000 ms, slice thickness = 4 mm, and fat saturation. (D) 2D T2-weighted FLAIR images for the same patient, but scanned on a 3T scanner with TI = 2500 ms, TE = 105 ms, TR = 9000 ms, and no fat saturation. (E) 3D T2-weighted FLAIR images for the same patient, scanned on a 3T scanner, with TI = 2400 ms, TE = 226 ms, TR = 6000 ms, and slice thickness of 1 mm. Note the apparent changes in non-enhancing lesion size and contrast between the lesion and background tissue. (F) Original postcontrast T1-weighted image of a patient with multifocal disease. (G) Same images in the same patient, but rotated 15 degrees to the right. (H) Same images, but also rotated 5 degrees forward, now showing “disease progression,” PD, and “partial response” (PR) via bidimensional measurements, purely an artifact of head tilt. (I) Original T2-weighted FLAIR images of a patient with a low-grade glioma. (J) Same images from the same patient, but rotated left 9 degrees, now showing PD. (K) Same images from the same patient, also rotated backward 10 degrees, now suggesting PR, again purely an artifact of head tilt.

unreliable, as overall adjudication rate is a complex function of the particular reader training or experience, the response criteria used, the particular disease or disease subtype, the drug and/or mechanism under investigation, as well as technical factors. Discordance in image acquisition guidelines and expectations between the sponsor, the imaging CRO, and the individual sites highlights one of the most common challenges and frustrations facing the imaging CRO. Much of this can be mitigated through widespread and proper adherence to standardized BTIPs with reasonable ranges for target acquisition parameters (eg, ~10%), or flagging “off protocol” exams to interpret with caution.

In addition to image acquisition guidelines, the Charter should also specify the methods by which images are de-identified and electronically (or manually) transferred to the imaging CRO for subsequent procedures. While the Health Insurance Portability and Accountability Act in the United States and the Convention for the Protection of Personal Data with Regard to the Automatic Processing of Personal Data in Europe⁷ prohibit the release and use of protected health information (PHI), it does not specify the procedures needed to properly follow these regulations. Since much of the information

stored in the header portion of the DICOM images potentially contains PHI as well as information that is scientifically useful from vendor-specific DICOM fields, the methodology and procedures for de-identifying images is not standardized and can vary trial-to-trial or even site-to-site within a trial. Although sites may have their own regulatory rules in place, standardized guidelines for DICOM de-identification have recently been created in order to minimize the risks associated with release of PHI while maximizing the scientific information that may be stored in the DICOM images.⁸ A similar, unified strategy for de-identifying DICOM images should be considered in prospective brain tumor clinical trials, for which a number of freely available software tools are available.⁹

Site Qualification

Once the sponsor and imaging CRO agree on a suitable image acquisition protocol and the trial begins, qualification of the imaging personnel and equipment at the various clinical sites is performed. Although there is no strict standard method for qualifying medical imaging devices for use in clinical trials, the FDA recommends the use of

imaging equipment that has received FDA marketing authorization or fulfills requirements of 21 CFR Part 812 if exploring an investigational imaging device. (Similarly, a CE mark may be recommended for equipment in Europe.) In addition to the primary imaging system, the Charter should specify all imaging equipment to be used in the trial including contrast agent or drug injectors, gaiting belts, software packages, etc. In standard neuro-oncology trials or trials with fewer than 20 sites, it may be practical to perform the site qualification process on all scanner hardware and software versions, and update those procedures for new scanners or new software that may evolve during the course of the trial. For large multinational studies, this might not be practical particularly for nonprofit or cooperative groups, so initial or annual site qualification for scanners that may be used for specific trials may be sufficient.

While scanner accreditation procedures may differ across imaging CROs or imaging core labs, a 2-stage process is often used for site qualification in neuro-oncology: (i) assessment of what machines (hardware and software) sites will be using and (ii) assessment of images obtained from a test subject or phantom. The latter is more important in protocols requiring more advanced or innovative MRI pulse sequences or for precision measurements for quantitative endpoints, where oversight of acquisition hardware and software is useful to reduce measurement variability.

Site Training

Clinical site personnel responsible for various aspects of imaging exam scheduling, image procurement, and image transfer need to be identified and trained in proper procedures. This should include study coordinators, MR physicists, and MRI technologists who are involved in protocol setup and the scheduling of patients on the study-approved machines for the imaging examination. It is also very important that training be provided to ensure secure anonymized upload and transfer of image data. The basis for training is the Imaging Core Manual (ICM), preferably tailored to the specific MRI scanner(s) at individual sites. Engagement of the MRI technologist is key for ensuring image quality and consistency. Training can be in-person at an investigator meeting or web-based (live or recorded), or at a minimum involve the review and signature of the ICM. This training is typically performed directly by the imaging CRO and the training date and procedures are documented for further regulatory or auditing purposes.

Routine Quality Control

All imaging should undergo ongoing quality control to improve standardization over sites, patients, and time points. This is usually performed by trained modality experts with increasingly automated software. Criteria for classifying an image as uninterpretable based on a technical failure or other considerations that lead to the exclusion of an image from the interpretation process are needed and should be specified prior to data collection. Image quality and

compliance are often scored based on the ability for the exam to be compliant with privacy regulations, use of the correct study ID, and inclusion of the required sequences along with whether they were performed according to protocol. Different weightings of importance may be assigned to different aspects of the exam, including essential sequences or whether the exams were submitted in a timely fashion if a quick or real-time read is required. Additional assessments may include consistency of the exam details compared with previous time points (eg, same physical scanner, same pulse sequence parameters), whether the dates are consistent with expected or scheduled clinical visits, as well as quality and readability of the images, including whether objects or artifacts obscure anatomic features necessary for proper measurements. While the best image quality is obviously desired, a pragmatic approach is to categorize image quality broadly as “optimal,” “suboptimal, but usable,” or “not-usable.” Examples of issues that might deem an exam not usable would include severe image quality issues that do not allow measurement of the lesion or an incomplete exam that is missing slices on a single sequence. All of this information must be documented by the imaging CRO or core lab, along with the date of communication and whether the issues were resolved.

Image Interpretation

Image interpretation generally is performed by trained readers, such as qualified radiology and/or neuro-oncology specialists, who review and interpret, or read, images obtained in the course of a clinical trial. The particular qualifications that are needed for acceptable readers should be specified in the Charter, such as whether a subspecialty-trained neuroradiologist with documented experience in brain tumor imaging and response assessment in clinical trials is required. Documentation showing that readers do not have a financial connection to the drug being tested should be required. The reader training process should be described, emphasizing the training documentation process and the use of any specific training materials (eg, a training manual or training images), image display training sessions, and image read testing process. In addition, the Charter should specify that any performance criteria will be used to qualify readers after training and during the course of the trial.

The Charter should describe the timing of image reads with respect to the clinical trial conduct. In some situations, prompt interpretation of images is important (eg, for determining trial eligibility or confirming disease progression in trial patients). In other situations, images are interpreted only following completion of all subject evaluations. Computer-assisted image interpretation may form an important component of the read process. In general, the extent of computer assistance should be described explicitly within the Charter, including a plan for quality control checks upon any critical software functions. Use of unsupervised artificial intelligence reads without human interpretation or quality checks in order to automate response assessment are not currently recommended,

although they are being researched extensively. Selection of the appropriate response criteria, which images to review, the procedures used for displaying the images to the neuroradiologists, the reading queue procedures, as well as the reader paradigm and adjudication design should all be considered. The Charter may also provide guidance on caliper placement to assess tumor bulk—for example, when the tumor surrounds a surgical cavity.

Selection of Response Criteria, Images to Review, and Display Procedures

There are a number of criteria available for response assessment in neuro-oncology (Table 2). While some of these criteria are only subtly different (eg, RANO,¹⁰ Modified [m]RANO,¹¹ Immunotherapy [i]RANO¹²), others are based on distinct tumor types (eg, RANO–Brain Metastasis,¹³ RANO,¹⁰ Response Assessment in Pediatric Neuro-Oncology [RAPNO]¹⁴) or imaging features (eg, Low-Grade Glioma RANO,¹⁵ RANO¹⁰/mRANO¹¹). The type of response criteria chosen for a particular study will also determine which images are chosen for review. For example, a standard RANO review for high-grade gliomas¹⁰ requires precontrast T1-weighted images to exclude intrinsically hyperintense signal from blood products as well as the postcontrast T1-weighted images and either T2-weighted or T2-weighted fluid attenuated inversion recovery (FLAIR) images in order to both measure target lesions and qualitatively assess nontarget, non-enhancing (T2 hyperintense)

lesion changes. The modified RANO criteria, however, do not place as much emphasis on T2-weighted or FLAIR images, as only the enhancing lesion is tracked for response purposes. The RAPNO criteria¹⁴ build on the standard RANO criteria by offering additional guidance on the use of advanced imaging, including diffusion MRI, into the interpretation of response, highlighting the need to match the images presented to the radiologist with the particular response assessment requirements for the trial.

Considering the multitude of available criteria, more than one of which may be applied in a single trial as primary and exploratory endpoints, along with the complexity of integrating clinical data and confirmation requirements, we recommend using software to compute treatment response based on measurements provided by neuroradiologists. Such an algorithmic implementation highlights challenges in interpreting response criteria, even those which are established and widely adopted. The criteria as published often do not completely specify fringe cases that can occur within trials and lead to discrepancies between local site and central reads and between core labs. For example, the publications often incompletely specify how to compute response when one or more target lesions are not evaluated, steroid or neurological data are not available, or anatomic coverage is incomplete, preventing reliable assessment of new lesions. Industry drafted follow-up documents and white papers have described how such situations are handled in an attempt to fill in these gaps (eg, Response Evaluation Criteria in Solid Tumors [RECIST] 1.1,¹⁶ immune-related RECIST¹⁷). But a complete

Table 2 Selection of response criteria and images to review (gray = desired; black = required)

Response Criteria	Purpose	Images to Review
RANO (Standard Response Assessment in Neuro-Oncology)¹⁰	Therapeutic response assessment in high-grade gliomas	1. PostcontrastT1w 2. T2w orT2w FLAIR 3. PrecontrastT1w
LGG RANO (Low Grade Glioma RANO) ¹⁵	Therapeutic response assessment in low-grade gliomas	1. PostcontrastT1w (<i>to rule out enhancement</i>) 2. T2w orT2w FLAIR 3. PrecontrastT1w
iRANO (Immunotherapy RANO) ¹²	Therapeutic response assessment in high-grade gliomas treated with immunotherapies	1. PostcontrastT1w 2. T2w orT2w FLAIR 3. PrecontrastT1w
mRANO (Modified RANO) ¹¹	Therapeutic response assessment and patient management in high-grade gliomas, agnostic of treatment mechanism.	1. PostcontrastT1w 2. PrecontrastT1w (<i>forT1 subtraction</i>) 3. T2w orT2w FLAIR 4. Diffusion-weighted imaging (<i>per BTIP</i>)
RANO-BM (RANO for Brain Metastases) ¹³	Therapeutic response assessment for brain and general CNS metastases	1. PostcontrastT1w 2. PrecontrastT1w (<i>rule out blood products or intrinsicT1 shortening</i>) 3. T2w orT2w FLAIR (<i>for existing non-target qualitative assessment</i>)
RAPNO (Response Assessment in Pediatric Neuro Oncology) ¹⁴	Therapeutic response assessment for pediatric brain tumors	1. PostcontrastT1w 2. T2w orT2w FLAIR 3. PrecontrastT1w (<i>forT1 subtraction</i>) 4. Diffusion-weighted imaging

algorithmic specification of the main criteria enabling computer implementation would be a very useful technical advance that could be adopted uniformly across trials.

The availability of images to the reviewers, whether on-study or off-study, must be documented in order to fully understand and interpret imaging-based endpoints, particularly PFS or time to treatment failure. In addition to the types of images reviewed, image display procedures including patient de-identification should also be documented and justified. The choice of whether to blind the readers to patient identifying information or leave them unblinded is typically based on factors relating to both the size of the trial and the potential for bias. In small early phase or rare disease trials, de-identifying the patient may be challenging because the readers may remember the cases based on the visual appearance, location, or size of the lesion. This can be overcome if a trial has a pool of available readers. However, this may not be necessary or desired. Having experienced neuroradiologists provide an impression based on their practical knowledge and familiarity with the case can often be very important.

Reading Queue and Data Locking Procedures.

The Charter must specify the order in which the reader will receive the images for interpretation along with when these measurements will be locked and unable to be subsequently changed (Table 3). The most common and widely accepted procedure used for general oncology and most late-phase neuro-oncology trials is the locked time-sequential procedure for image presentation. This approach is closest to clinical practice and consists of providing the reader with one time point's scan at a time from baseline to the current time point in chronological order, with the reader blinded to the number of time points for a given subject. Once the reader has made an assessment at that particular time point it is locked, and the reader is shown the next time point's scan, if it exists. Readers are not able to make any changes after locking unless there is a clearly articulated and documented reason to do so. With simultaneous image presentation, the reader has access to all time points simultaneously with no blinding to date or total number of time points. However, allowing the reader to know the number of time points *a priori* may bias lesion selection and interpretation, since the total number of time points may be associated with time to failure as interpreted by the local site. One way to reduce this bias is to use a simultaneous, but randomized temporal image presentation paradigm, where all time points are presented to the reader but the order is randomized and the reader is blinded to the date of exams.

Measuring highly infiltrative or heterogeneous lesions in neuro-oncology is inherently challenging and associated with high adjudication rates, making both the locked time-sequential and the simultaneous, randomized temporal image presentation paradigms problematic. Hybrid designs meant to capture both an unbiased assessment and a holistic interpretation offer a more comprehensive evaluation of therapeutic response in late phase neuro-oncology trials. In a hybrid randomized image presentation paradigm, the time points are first presented similar

Table 3 Reading queue procedures

Reading Queue Procedure	Description
LockedTime-Sequential Presentation*	A patient's complete image set from baseline to current evaluation is presented in the chronological order in which the images were acquired. Unless specified in the Charter, the reader is blinded to the total number of time points per patient. This is the current standard for reading queue procedures in general oncology.
Simultaneous Image Presentation	All of a patient's image set is displayed simultaneously. There is no blinding to total number of time points or date of exams.
Simultaneous, Randomized Temporal Image Presentation	All of the patient's image set is displayed simultaneously, but presented in random order with reader blinded to the date of exam but not to the total number of time points.
Hybrid Randomized Image Presentation**	A patient's image set is presented in random order with reader blinded to the date of exam. Once measurements are locked, the readers are allowed to unlock and review all images in chronological order. Changes from the randomized assessments are tracked.
Hybrid Locked Time-Sequential Image Presentation**	A patient's image set is presented in a locked, time-sequential fashion. The readers are then allowed to unlock and review all images at the same time with no blinding to total number of time points or date of exams. Changes from the randomized assessments are tracked.

*Current standard for general oncology and most neuro-oncology trials.

**Proposed for late-phase neuro-oncology trials.

to the simultaneous, randomized temporal image presentation paradigm. Once the measurements are locked, the readers are allowed to review all time points in chronological order and make changes to their original measurements. All changes from the randomized assessments are then tracked and documented for audit purposes. A hybrid locked time-sequential image presentation paradigm first presents the reader with time points using a traditional locked time-sequential paradigm, and then after locking those measurements, the readers are allowed to review the time points in chronological order and make relevant changes. As always, changes from the original assessments are tracked for audit purposes.

Reader Paradigm and Adjudication Design

The proper reader paradigm and adjudication design chosen for a neuro-oncology trial is dependent on the goals of the trial, size of the trial, complexity of the particular disease, and the potential issues associated with interpretation (Table 4). A consensus read, similar to a tumor board, involves multiple readers or experts working together to discuss each exam, resulting in a single

Table 4 Reader paradigm and adjudication design*

Paradigm	Description	Uses in Neuro-Oncology	Pros/Cons
Consensus Read	3+ readers work together and discuss the exam, coming to a single consensus interpretation.	Phase 0/I/II	<i>Pros:</i> No adjudication or ambiguity. Useful for rare or complex tumors, or small studies. <i>Cons:</i> Logistically difficult to get 3+ readers to discuss a case.
Paired Read with No Adjudication	R1 and R2 perform independent reads. Reader results are averaged or 2 sets of results are provided. Common when reporting both “site determined” and “centrally determined” results.	Phase 0/I	<i>Pros:</i> Efficient and cost-effective. <i>Cons:</i> High discordance rates can cause confusion about results.
Paired Read with Forced Adjudication ^ψ	R1 and R2 perform independent reads. R3 adjudicates any differences between R1 and R2 by choosing the best read, ^ψ R1 or R2.	Phase II/III	<i>Pros:</i> Unbiased <i>Cons:</i> Expensive and time consuming
Paired Read with Open Adjudication	R1 and R2 perform independent reads. R3 adjudicates any differences between R1 and R2 by independently reading the exam or series. Results from R3 are final and can be different from R1 and R2.	Phase II/III	<i>Pros:</i> Unbiased <i>Cons:</i> Expensive and time consuming
Central Confirmation of Local Reads Using Single Read with (Forced or Open) Adjudication	R1 performs independent reads. R2 adjudicates any differences between R1 and the local site reads through either forced or open adjudication.	Phase 0/I/II	<i>Pros:</i> Efficient and cost effective. <i>Cons:</i> Depends heavily on experience of core lab neuroradiologists with disease and treatment mechanism.

*R1 = Reader 1. R2 = Reader 2. R3 = Reader 3.

^ψAdjudication might be based on exact measurements or contours by R1 or R2, response determination (PD, SD, CR, PR) at each time point, or the date of events such as progression or date of best response.

consensus interpretation of the case at each time point. This type of paradigm is most common in smaller trials or trials for rare or complex tumors in which interpretation is difficult and needs a team of experts. A paired read with no adjudication is rarely performed in neuro-oncology, where 2 readers independently read and the results are both reported, except for when reporting both “site determined” and “centrally determined” results. This type of read is efficient and cost-effective, but suffers from potential confusion in the interpretation of results if high discordance rates are observed.

The most common reader paradigms and adjudication design for late-phase trials include paired reads with 2 independent readers, with either a forced or open adjudication by a third radiologist or reader. This type of read is largely unbiased, although it can be expensive and relatively time-consuming because of the number of readers involved. In both of these paradigms, 2 readers start by independently reviewing all the cases using one of the reading queue and data locking procedures outlined in Table 3. For forced adjudication, a third reader adjudicates any differences between the first 2 readers by choosing the best read between the 2 independent readers. An open adjudication procedure is similar to forced adjudication but is quite uncommon and requires the third reader to adjudicate through a third independent read of the case. This adjudicator can choose either result from the 2 independent readers or may choose to provide a result that differs from these 2 readers. In all accounts, results from this third

reader are final and therefore more weight is placed on this reader’s interpretation. Adjudication for all these cases may be based on the measurements or contours made by and signed off on by the readers, the response determination at each subject’s time point (eg, progression, stable disease, or response at each time point), or the date of specific events including the date of progression or date of best response.

Another common reader paradigm and adjudication design in neuro-oncology involves central confirmation of local reads in early phase trials. In this type of paradigm, the goal is to confirm local measurements or response determination by an independent reader at a central core lab facility. In this scenario, a single reader first performs an independent read and a second reader then adjudicates any differences between the independent reader and the local site through either forced adjudication (choosing either the site or independent reader) or open adjudication (the adjudicating radiologist independently and conclusively determines the patient response). This type of trial design is efficient and cost-effective, particularly for cooperative groups without a large imaging budget. However, this approach depends heavily on the experience of the core lab personnel, including radiologists, who may disagree with sites based on lack of experience with specific disease subtypes, treatment mechanisms, or clinical factors (eg, seizures) that could obscure interpretation, which can be further complicated by differences in computation of response assessment criteria as described previously.

Proposed Improvements to Optimize Radiologic Read Paradigm for Utilizing Conventional RANO Eligibility Outcome Measures

Currently, eligibility criteria for recurrent or progressive disease trials involve having “measurable” lesions and evidence of disease progression to act as a therapeutic target for monitoring response. However, these trials do not require that there be documented evidence of lesion growth over time prior to enrollment in the trial. Without requiring documented tumor growth coming into the trial, a proportion of these patients could simply be experiencing transient treatment-related changes or could have their disease under control. Similar to the procedures used in recent studies,¹⁸ trials involving recurrent or progressive disease patients should require historic or pre-trial datasets be submitted to the central imaging facility in order to confirm that the patients have both growing tumor and measurable disease prior to enrollment. The time frame and conditions for these pre-trial datasets will depend largely on tumor subtype and disease stage, and should be specified in the Charter.

For outcome assessment, a locked sequential read involves the reader identifying target and nontarget lesions on the baseline study and making bidimensional measurements on the slice with greatest lesion diameter (or whatever the specific rules are for the particular criteria used). Once the baseline is signed, it is locked and the next time point is opened, on which the reader finds the same lesions, makes dimensional measurements on slices with maximum diameter, and looks for new lesions. This approach can be optimized by the following practical upgrades to the read paradigm.

The first is to provide the reader with pertinent clinical history including anatomic location and date of treatments for all surgery and radiation. Additional information about radiation ports would also be useful. The challenge is ensuring that verified source data is collected in a timely fashion and that the reader can access this information in a user-friendly presentation.

Secondly, the plane of the tumor being measured (eg, axial) should be the same over all time points correcting for any tilt of the base of skull at scan acquisition. The brain is a relatively motion-insensitive and rigid organ, so the images for a single patient can be registered and aligned over time. This is practical both because of new high-resolution (1–1.5 mm³) image acquisition guidelines (eg, BTIP¹) as well as the commercial availability of multiple regulatory approved registration algorithms. It is important to note that only a 6-degrees-of-freedom rigid body registration algorithm should be used to avoid geometric scaling or skew that can result from higher order (eg, affine) registration algorithms.

Thirdly, the reader should have a hanging protocol that allows the current study to be read with access to the prior time points with the system highlighting baseline and nadir studies so that the reader can easily visually compare the lesions. This is particularly important for slower growing, lower-grade tumors that may demonstrate indolent progression over a long period of time or multiple

scans.¹⁹ Additionally, the system should support a view of merged images superimposed to confirm the integrity of lesion contours or points of measurement. This will help confirm that the tumors and sites of measurement are corrected for axis alignment and will allow the influences of different imaging parameters and contrast timing to be assessed in a consolidated view.

Fourthly, the reader should be provided with a verification check prior to signing off the cases. This should at very least involve quantifying the measured change between relevant time points and whether an event (eg, progression or response) is triggered. This could be coupled with a visual scoring system that asks the reader to assess tumor change using reader gestalt to encourage the comparison of the calculated RANO disease progression by the system with their clinical opinion as a validation check. If discrepancies arise between quantitative measurements and clinical opinion, the reader should be allowed to re-review the scans to look for any inconsistencies in their measurements that might explain this discordance. This feedback loop will significantly improve the validity of the reader’s measurements.

Finally, although the locked sequential prospective read is the preferred paradigm given the complexity of lesion contours in the setting of recurrent high-grade gliomas, the influence of radiation and surgical scarring, as well as lesion margin conspicuity due to different image noise features and contrast timing, suggests a retrospective review of all time points should also be considered. This allows readers to take advantage of temporal change to confirm lesion selection, lesion contours and measurements so that a more confident determination of date of response or progression can be made. This would improve confidence in the imaging endpoints in the current setting.

While a neuro-oncologist or expert is typically employed for integration of radiologic and clinical data, we recommend that response criteria be computer calculated from neuroradiologist tumor measurements and integrate clinical data (eg, steroid dose, neurological status) and confirmation requirements for progression/response. To facilitate this, publication of complete and algorithmic specifications of standard neuro-oncology criteria are desirable. The Charter should also specify whether/when the clinical data will be provided to the imaging core and policies regarding changes or updates to clinical data as it is cleaned during the course of the study.

There are currently guidelines for reader paradigms that allow for re-reads before adjudication, but the locking or blinding rules have not been properly specified. We recommend the Charter specify the rules around editing tumor measurements, for example, limiting changes based on identification of new active tumor margins or changes meant to ensure temporal consistency across time points. In addition, the use of the audit trail to evaluate the impact of the review and extent of updates performed is important.

Conclusions

The determination of therapeutic benefit from new therapies to treat brain tumors is closely tied to the radiologic

interpretation of images obtained from multicenter trials. Procedures aimed at reducing reader bias while considering practical issues specific to neuro-oncology are important for designing new trials and interpreting results from completed trials. Although many of these procedures have been successfully adopted from general oncology, we recommend specific changes to optimize the methodology for neuro-oncology, including image registration, requirement of growing tumor for eligibility in trials of recurrent tumor, standardized image acquisition guidelines, and hybrid reader paradigms that allow for both unbiased measurements and more comprehensive interpretation.

Funding

This work was supported by NIH/NCI 1P50CA211015-01A1 (Ellingson, Cloughesy), NIH/NCI 2P50CA165962-06A1 (Wen), American Cancer Society RSG-15-003-01-CCE (Ellingson), National Brain Tumor Society (Ellingson, Cloughesy).

Acknowledgments

This paper stems from the efforts of the National Brain Tumor Society's (NBTS) Imaging Workgroup, a group formed to advance recommendations arising from the 2019 NBTS Research Roundtable on imaging endpoints. The authors are grateful to NBTS for its leadership in catalyzing activities designed to push the field forward to improve options for brain tumor patients. We would also like to thank Daniel Kraniak, PhD and Mitchell Anscher, MD from the US Food and Drug Administration for their excellent guidance and insight for this manuscript.

Conflict of interest statement. Ellingson: Consultant for MedQIA, Imaging Endpoints, and Siemens. Brown: Consultant for MedQIA. Boxerman: None. Gerstner: None. Kaufmann: None. Cole: Employee for Bayer. Bacha: Employee for Edison Oncology. Leung: Employee for Bristol Myers Squibb. Barone: None. Colman: None. van den Bent: None. Wen: None. Yung: None. Cloughesy: Consultant for MedQIA. Goldin: Consultant for MedQIA.

References

1. Ellingson BM, Bendszus M, Boxerman J, et al; Jumpstarting Brain Tumor Drug Development Coalition Imaging Standardization Steering Committee. Consensus recommendations for a standardized brain tumor imaging protocol in clinical trials. *Neuro Oncol.* 2015; 17(9):1188–1198.
2. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases (BTIP-BM). *Neuro Oncol.* 2020; 22(6):757–772.
3. Ford RR, O'Neal M, Moskowitz SC, Fraunberger J. Adjudication rates between readers in blinded independent central review of oncology studies. *J Clin Trials.* 2016; 6(5):289.
4. Boxerman JL, Zhang Z, Safriel Y, et al. Early post-bevacizumab progression on contrast-enhanced MRI as a prognostic marker for overall survival in recurrent glioblastoma: results from the ACRIN 6677/RTOG 0625 Central Reader Study. *Neuro Oncol.* 2013; 15(7):945–954.
5. Pope WB, Hessel C. Response Assessment in Neuro-Oncology criteria: implementation challenges in multicenter neuro-oncology trials. *AJNR Am J Neuroradiol.* 2011; 32(5):794–797.
6. Dodd LE, Korn EL, Freidlin B, et al. Blinded independent central review of progression-free survival in phase III clinical trials: important design element or unnecessary expense? *J Clin Oncol.* 2008; 26(22):3791–3796.
7. Levin A, Nicholson MJ. Privacy law in the United States, the EU and Canada: the allure of the middle ground. *Univ Ott Law Technol J.* 2005; 2(2):357–395.
8. Moore SM, Maffitt DR, Smith KE, et al. De-identification of medical images with retention of scientific research value. *Radiographics.* 2015; 35(3):727–735.
9. Aryanto KY, Oudkerk M, van Ooijen PM. Free DICOM de-identification tools in clinical research: functioning and safety of patient privacy. *Eur Radiol.* 2015; 25(12):3685–3695.
10. 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.
11. Ellingson BM, Wen PY, Cloughesy TF. Modified criteria for radiographic response assessment in glioblastoma clinical trials. *Neurotherapeutics.* 2017; 14(2):307–320.
12. Okada H, Weller M, Huang R, et al. Immunotherapy response assessment in neuro-oncology: a report of the RANO working group. *Lancet Oncol.* 2015; 16(15):e534–e542.
13. 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.
14. Erker C, Tamrazi B, Poussaint TY, et al. Response assessment in paediatric high-grade glioma: recommendations from the Response Assessment in Pediatric Neuro-Oncology (RAPNO) working group. *Lancet Oncol.* 2020; 21(6):e317–e329.
15. 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.
16. Parexel. *Recommendations for consistent application of RECIST 1.1 to specific trial indications.* 2014; https://www.parexel.com/application/files_previous/2014/0310/4328/PXL_RECIST_1.1_White_Paper.pdf.
17. Bohnsack O, Ludajic K, Hoos A. *Adaptation of the immune-related response criteria: irRECIST.* 2014; https://www.parexel.com/application/files_previous/7214/2313/4150/Adaptation_of_the_Immune_Related_Response_Criteria_irRECIST_online.pdf.
18. Mellingshoff IK, Ellingson BM, Touat M, et al. Ivosidenib in isocitrate dehydrogenase 1-mutated advanced glioma. *J Clin Oncol.* 2020:JCO1903327.
19. Johnson DR, Guerin JB, Ruff MW, et al. Glioma response assessment: classic pitfalls, novel confounders, and emerging imaging tools. *Br J Radiol.* 2019; 92(1094):20180730.