reliable information than the traditional expert reviews. Level 2 contains synopses that summarize the results of systematic reviews in combination with the best current primary literature. The apex of the pyramid (level 1) contains the most valid evidence and comprises information systems that integrate and summarize all relevant and important research regarding a specific clinical topic.

The validity of a research article is based on how close the study results are to the truth. This can be determined by assessing the study design in the Methods section regarding the patient-selection process, reference standard, internal and external biases, and limitations. A level of evidence or quality scores can assist in determining the better quality research studies. An example is the Quality Assessment of Diagnostic Accuracy Studies appraisal tool,[3] which can be used to evaluate research studies concerning the diagnostic accuracy of a test. The second part of the appraisal process is to assess the strength of the research findings by analyzing the Results section. The sensitivity and specificity, confidence intervals, positive and negative predictive values, and likelihood ratios are all used to assess the ability of a diagnostic test to reliably differentiate between disease and healthy status. The confidence interval around the sensitivity and specificity gives an idea of how close to the truth these results may actually be.

The application of the best evidence into clinical practice requires a transition from thinking about the sensitivity and specificity of a diagnostic test to the likelihood or probability of the patient having the disease. The "pretest probability" is the clinician's estimate of the patient's probability of having the disease, given all the available data. The "posttest probability" is simply defined as the pretest probability updated by the test results. If the pretest probability is above the clinician's inclusion threshold for disease diagnosis or below the exclusion threshold, then no further diagnostic testing is necessary because there is a reasonable degree of certainty that the patient does or does not have the disease. However, between these thresholds is the uncertain area that warrants further diagnostic testing to move the patient's probability of disease either above the inclusion or below the exclusion thresholds. The greater the strength a diagnostic test has with high sensitivity and specificity, the less influence the clinician's pretest probability has in determining the patient's disease status.

There are several limitations and barriers to implementing evidence-based radiology in practice. Getting started may be overwhelming because the critical thinking skill set required is relatively new to radiologists, and many do not have prior experience or training. Training courses, Web-based tutorials, and textbooks are available to learn more about evidence-based medicine. Another option to getting started is to work with a librarian. The expertise of a librarian is a valuable resource in performing a comprehensive search of the literature. Getting a librarian to also search your question can assist in identifying your knowledge gaps and the limitations in your search strategy, improving your skills.

Once you have overcome this obstacle of getting started, time limitation remains the main barrier to performing evidence-based radiology in practice. A valuable short-cut is to seek evidence from as high in the evidence pyramid as possible, such as the evidence-based reviews, systematic reviews, and meta-analyses. These structured reviews provide reliable information by using strict methodology designed to limit bias. The relevant research is critically appraised, and the best evidence is summarized for you in these structured reviews. However, at times, the best evidence may not be readily available because there is a tendency not to publish "negative" studies in the literature. Last, case reports are considered as the lowest evidence in the pyramid; however, these reports may provide valuable information in specific clinical scenarios.

In summary, "evidence-based practice" is defined as "the integration of best research evidence with clinical expertise and patient values." This requires the art of balancing the scientific evidence, clinical expertise, and judgment. When there is strong scientific evidence (at the apex of the pyramid) with information systems that summarize and integrate all relevant research about a clinical topic, then practice guidelines can be developed, and with time, these guidelines are implemented as standard practice. However, when only weak evidence is available, then clinical expertise and judgment become a major component guiding our medical decisions. Judgment is particularly important when the evidence is inconclusive because we rely on our judgment to detect differences between observations in research and to understand their significance in clinical practice.

Many times conclusive evidence is not available at the time a medical decision needs to be made because acquiring strong evidence is time-consuming and costly and may lack research interest. However, at the point-of-care level, a decision needs to be made regardless of the lack of knowledge and evidence available. Therefore, as difficult as it may be, sound clinical judgment may be most valuable in guiding patient care when only weak or inconclusive evidence is available.

References

1. Malone DE, Skehan SJ, MacEneaney PM, et al. Evidencebasedradiology.net. www.evidencebasedradiology.net. Accessed September 1, 2010.
2. Sackett DL, Richardson WS, Rosenberg W, et al. *Evidence Based Medicine; How to Practice and Teach EBM.* 2nd ed. Edinburgh, United Kingdom: Churchill Livingstone; 2000
3. Whiting P, Rutjes AW, Reitsma JB, et al. **The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.** *BMC Med Res Methodol* 2003;3:25

P.C. Sanelli
Department of Radiology
New York-Presbyterian Hospital
Weill Cornell Medical College
New York, New York

DOI 10.3174/ajnr.A2390

EDITORIAL

Response Assessment in Neuro-Oncology Criteria: Implementation Challenges in Multicenter Neuro-Oncology Trials

The Response Assessment in Neuro-Oncology Criteria (RANO) Working Group recently published updated guidelines for assessing response to therapy in high-grade gliomas.[1] The goal of the group continues to be the development of a standardized method of response assessment. These guidelines are critical for multicenter trials, particularly when the results are compared against historical controls. Thus in addition to being clinically applicable, the guidelines must address the concerns of regulators with respect to 1) validity, 2) objectivity, 3) reproducibility, and 4) comprehensiveness. RANO replaces the "Macdonald criteria" for response that were originally published in 1990.[2] A major impetus for this update was the widespread adoption of antiangiogenic therapies for the treatment of malignant gliomas. These treatments have a pronounced antipermeability effect, which can cause marked diminishment of tumor enhancement, even when overall tumor size is unchanged or even when tumor growth is apparent.[3] This increased decoupling of enhancement from an accurate measure of tumor burden made the Macdonald criteria, which are based on bidimensional measurements of enhancing lesions only, no longer tenable. The major modification proposed by the RANO guidelines is the assessment of increasing fluid-attenuated inversion recovery (FLAIR) signal intensity change as evidence of tumor progression. In short, the RANO criteria have therefore been referred to as Macdonald + FLAIR.

The incorporation of FLAIR imaging into the RANO guidelines improves the accuracy of determining tumor progression in the setting of antiangiogenic therapy, because it allows the inclusion of nonenhancing tumor into the determination of overall tumor burden. However, there are aspects of the RANO criteria that are not described in sufficient detail to permit these guidelines to be "operationalized." This editorial illustrates some of the challenges in applying the RANO criteria to the real-world scenario of a multicenter drug trial for glioblastoma and outlines some proposed solutions.

Steroids

One of the greatest difficulties in instituting the RANO criteria is the issue of glucocorticoid steroid dosing. Like Macdonald, RANO incorporates steroid dose into the definition of tumor response categories. For example, partial response can only be designated if a patient is on stable or decreasing doses of steroids.

When attempting to apply the RANO guidelines to a patient receiving oral steroids, several questions arise. First, what constitutes an "increase" or "decrease" in steroids? Is there a level of change that can be considered nonsignificant? If so, is that a relative change, an absolute change, or both? Second, over what period preceding the MR imaging (either baseline or follow-up) should steroid data be considered? We have found that acquiring steroid dose for longer periods (>1 week) presents a challenge in terms of consistent data collection. Third, what parameter should be compared: the total (sum) daily dose over the period, or perhaps the average daily dose? For example, operationally one might use the following definition: "an increase in steroids is defined as a ≥10% increase and a ≥4-mg increase in the dexamethasone equivalents for the sum of the daily doses of steroids taken for the 5 days before the current MR imaging compared with the sum of the daily doses of steroids taken for the 5 days before the reference MR imaging." It is likely that the initiation of a steroid dose, even if small, has a greater effect on vessel permeability than changes in already established steroid treatment. Therefore, the initiation of any steroids, even if <4 mg, should be considered an increase in dose. Unfortunately, there is very little in the literature to justify a specific set of parameters that constitute significant changes in steroid treatment,[4] but specific operational definitions for multicenter trials such as those proposed above are necessary. Finally, we must address the question of how to specify what overall response level is to be assigned when the radiographic criteria are met, but the steroid criteria are not. For example: what is the overall response if there is complete tumor resolution by imaging while the patient is on more than a physiologic replacement dose of steroids? Should this be considered partial response?

Nonenhancing Disease

As with steroid dose, there are some ambiguities in the treatment of measurable and nonmeasurable disease. Nonmeasurable disease includes small lesions, lesions that are not enhancing, and lesions with poorly defined margins. As with Macdonald, the RANO guidelines advocate only well-defined enhancing lesions of a minimum size be qualified as "measurable disease." The partial response category is based only on reduction of enhancing tumor measurements. This leads to some difficulties. For example, when tumor changes from enhancing to nonenhancing after antiangiogenic therapy, the criteria for partial response are met, even if there is no change in tumor size. For lesions that are a mixture of enhancing and nonenhancing components, RANO guidelines state that "comparative analysis of changes in the area of both enhancing and nonenhancing component should be performed." But at baseline, measurements do not include the nonenhancing component. Thus, it is unclear what reduction in size of the lesion would qualify as partial response or progression, and whether this is based on the enhancing component, the nonenhancing component, or both. In addition for FLAIR signal intensity change, RANO does not specify what degree of increase qualifies as tumor progression, only that it must be "significant" or "unequivocal." Thus, it could be the case that although a 25% increase in bidimensional measurements for enhancing tumor is categorized as progression, a smaller amount of change in size of abnormal FLAIR signal intensity also may be defined as progression. This discrepancy should be addressed. The elimination of nonenhancing tumor as measurable disease is particularly problematic for grade III tumors that often do not enhance. Because change in nonmeasurable disease cannot be used as evidence of partial response, this would be a significant limitation in the assessment of drug response in trials of grade III tumors.

Nonenhancing tumors can be more difficult to detect and to measure accurately compared with enhancing tumors. However, this does not exclude accurate measurements for all nonenhancing tumors. Accuracy often is a function of the border of the tumor: tumors with distinct margins, whether they are enhancing or not, can be reproducibly measured. Therefore, we would advocate the inclusion of nonenhancing tumor as measurable disease, where possible, based on the ability to define tumor margins. Specifically, we recommend that when disease (either enhancing to nonenhancing) has ≥50% distinct margins, measurements should be made. For lesions with both enhancing and nonenhancing tumors, measurements should encompass both components and standard size changes to designate partial response and progression subsequently applied.

RANO also advocates retrospective analysis of FLAIR signal intensity change and backdating time of progression to the first time point in which progressive nonenhancing tumor is suspected. The introduction of a retrospective component will probably increase sensitivity for progression in comparison with past methods, again making comparison with historical controls potentially problematic. We recommend that for all scans the location of questionable new nonmeasurable disease is marked at the time the scan is first read; then once progressive disease is established, scans are reviewed to determine whether questionable areas did indeed develop into definitive areas of measurable or nonmeasurable disease. If so, progression should then be backdated to the time at which a suspicious area was first identified, if nonmeasurable, or if measurable, to the time when 25% increase in lesion size was demonstrated. This maintains a component of prospective analysis, which should help reduce bias. It will be of interest to quantify in what percentage of cases does the addition of this combination prospective-retrospective approach alter the date of disease progression. It also will be important to determine which method, when there is a discrepancy in the date of progression, most closely predicts the true end point of survival.

Inclusion Criteria

Approximately 20%–30% of patients have "pseudoprogression," defined as increased enhancement on the first scan after surgery and radiation therapy that subsequently abates without further treatment. As a result, the RANO study group advocates the exclusion of patients from drug trials who progress within 3 months of the completion of radiation therapy, unless the tumor is outside the 80% isodose line. However, this may act to exclude the most malignant tumors, which would be more likely to recur rapidly. Because these patients have not been excluded from many previous trials, there could be an element of bias when the treatment effect of new drugs is compared with historical controls if a trial is operating under the new guidelines. In addition, the effects of these new drugs on the fastest growing tumors would probably remain unknown. It is also important to note that pseudoprogression is not uncommon in the 3–6-month postradiation timeframe and can occur at even later dates. Clearly, an improved ability to distinguish tumor recurrence from pseudoprogression is needed. So, we must weigh the risk of excluding early progressors against the risk of including pseudoprogressors. In larger trials, subgroup analyses could be helpful in determining whether patients with early postradiation enrollment had significantly better response rates to address the possible impact of pseudoprogression.

Issues Causing Discordant Reads

As mentioned, one major issue presenting in the assessment of response in multicenter trials is the difficulty in measuring nonenhancing tumor. This is a particular challenge because nonenhancing tumors can be more subtle, and thus harder to detect, and they also are less well defined than enhancing tumors. It seems likely that this is a major cause of discrepancy between reviewers when establishing a date of tumor progression. Such a discrepancy typically requires adjudication by a third reader. This discrepancy or "adjudication rate" can be as high as 50% in trials of antiangiogenic therapy. Moreover, in addition to the radiographic read, usually in the setting of an independent review facility (IRF), the investigator treating the patient often makes an independent determination of progression. The most common discrepancy is when investigators call progression after the IRF established date, presumably because investigators have lower sensitivity to subtle FLAIR changes than radiologists. Therefore, multiple dates of progression may be generated. Given that the scans are typically 6 weeks apart, this can lead to a significant impact on the determination of time to progression. Because the time to progression benefit of bevacizumab, eg, is on the order of 4–5 months, this becomes an area of great concern to the Food and Drug Administration. Therefore, improved methods for reducing adjudication rate are critical. We feel that our proposed method for marking suspicious areas of possible nonenhancing tumor progression and then retrospectively confirming them as areas of disease progression will help reduce the adjudication rate, but this remains to be determined.

Additional Suggestions for Improvement

One way to achieve greater consistency in communicating guidelines for response assessment in glioma might be to use nomenclature that is consistent with Response Evaluation Criteria In Solid Tumors (RECIST). Thus, lesions could be defined as 1) target enhancing or nonenhancing lesions (up to 5 measurable baseline lesions, at least 10 × 10 mm at baseline, amenable to repeated measurements, representative of a subject's disease), 2) nontarget enhancing lesions (all other baseline enhancing lesions, including nonmeasurable lesions and measurable enhancing lesions not chosen as target lesions, and 3) new lesions. A table for deriving overall response could address all 3 of these "domains." Confirmation requirements would be best left out of the time point response (TPR) definitions and addressed separately, because confirmation applies to best overall response (across time points) not to a given time point.

When evaluating progressive disease for small measurable lesions, we would advocate requiring both a relative increase (eg, ≥25%) and an absolute increase (eg, at least X mm increase in the sum of the longest diameters or Y mm^2 increase in the sum of the products of the perpendicular diameters), similar to RECIST 1.1. In general, the use of "after the initiation of therapy" in the TPR criteria is not compatible with the intention-to-treat principle. Perhaps a better statement would be "after the initiation of therapy or randomization" or "after baseline" to be more widely applicable. When considering borderline progressive disease, the recommendation to follow such subjects at close intervals (eg, every 4 weeks) is contrary to the statistical preference for fixed, consistent intervals applied to all subjects within a trial. We would suggest instead that such subjects be followed further on the original preplanned assessment schedule.

A few other semantic issues: for nonenhancing lesions the reference MR imaging for establishing TPR of progressive disease ("baseline or best response") needs to be specified. Is it the "best response" for nonenhancing lesions or the "best response" for measurable enhancing lesions? We advocate this should be based on "best response" for the nonenhancing lesion in question. With regard to nomenclature, it might be better to use the term "nadir" rather than "best response" to avoid confusion with the concept of "best overall response." Similarly, the progressive disease designation does not have clear guidelines for nonmeasurable enhancing disease. The T2/FLAIR component of the TPR stable disease definition ("stable compared with baseline") is inconsistent with the T2/FLAIR component of the TPR progressive disease definition ("significant increase compared with baseline or best response"). Thus, we suggest that the T2/FLAIR component of the TPR stable disease definition should be compared with either baseline or best response, to maintain consistency.

Conclusions

The RANO criteria and guidelines are a needed advance over the formerly widely adopted Macdonald criteria. This is particularly evident in antiangiogenic therapy for glioblastoma. Some ambiguities in the response criteria pose challenges to applying them consistently and rigorously, particularly in multicenter trials with both investigator and independent review facility assessment of response. Given the rapid changes in treatment strategies, the guidelines are clearly a work in progress, probably requiring more frequent updates in the future. In this editorial, we sought to highlight some challenging areas of response assessment and suggest some added details and modifications that could be incorporated into future guidelines with the hope of improving the standardization of their application.

DISCLOSURES: Whitney Pope. *Consultant:* Genentech. *Details:* Have provided consulting services to Genentech in the past regarding imaging of Avastin-treated glioblastoma; Colin Hessel. *Ownership Interest:* Exelixis, Inc. *Details:* As an employee of Exelixis, Inc. I own stock and have stock options for the company. *Other Financial Interests:* Exelixis, Inc. *Details:* I am an employee of Exelixis, Inc. a biopharmaceutical company.

References

1. 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;28:1963–72
2. Macdonald DR, Cascino TL, Schold SC, et al. **Response criteria for phase II studies of supratentorial malignant glioma.** *J Clin Oncol* 1990;8:1277–80
3. Norden AD, Young GS, Setayesh K, et al. **Bevacizumab for recurrent malignant gliomas: efficacy, toxicity, and patterns of recurrence.** *Neurology* 2008;70:779–87
4. Roth P, Wick W, Weller M. **Steroids in neurooncology: actions, indications, side-effects.** *Curr Opin Neurol* 2010;23:597–602

W.B. Pope
Department of Radiological Sciences
David Geffen School of Medicine at UCLA
Los Angeles, California
C. Hessel
Exelixis Inc.
South San Francisco, California

DOI 10.3174/ajnr.A2582