Point–Counterpoint Review

Point

RECIST: No Longer the Sharpest Tool in the Oncology

Clinical Trials Toolbox

See Counterpoint and Reply by Fojo and Noonan, p. 5151 and p. 5150

Manish R. Sharma1,3, Michael L. Maitland1,2,3, and Mark J. Ratain1,2,3

Abstract

Although "response" has been an attractive term for oncologists and patients, oncologists really want to know which therapy to start for a given patient and when to discontinue that therapy in favor of an alternative. In efficacy trials, cancer therapeutics have conventionally been assessed by endpoints that are based on the categorical Response Evaluation Criteria In Solid Tumors (RECIST) system. In this article, we make the case for a new paradigm in which therapeutics are assessed on a continuous scale by evidence of efficacy, using a variety of quantitative tools that take advantage of technologic innovations and increasing understanding of cancer biology. The new paradigm relies on randomized comparisons between investigational arms and control arms, as historical controls are unavailable or unreliable for these quantitative measures. We discuss multiple limitations of RECIST, including its overemphasis on tumor regression, concerns about the accuracy of tumor measurements and the validity of comparisons with historical controls, and its inadequacy in disease settings in which tumor measurements on cross-sectional imaging are difficult or uninformative. We discuss how the new paradigm overcomes these limitations and provides a framework for answering the key questions of the oncologist and improving patient outcomes. Cancer Res; 72(20); 5145–9. ©2012 AACR.
From a practical point of view, the oncologist wants to know which therapy to start for a given patient and when to discontinue that therapy in favor of an alternative. Rather than argue about where the lines should be drawn to define response and progression, we contend that there should be no lines at all, but that efficacy should be viewed on a continuous (rather than categorical) scale. Cancer therapeutics should be evaluated by evidence of efficacy, which is best established by randomized comparisons between investigational arms and control arms.

The new paradigm

The volume of potential therapies/combinations and emerging data on interindividual and intrindividual differences in treatment effects present challenges that cannot be addressed by a categorical, response-based system. To address these challenges, we propose a new paradigm of using evidence of efficacy to evaluate new therapies. An intentionally vague term, evidence of efficacy would be established by one or more quantitative tools that take advantage of technologic innovations and increasing understanding of cancer biology. In this section, we highlight a number of reasons why this new paradigm would increase the efficiency of drug development.

In the past 2 decades, we have witnessed an explosion in the availability of diagnostics that divide cancers into molecular subtypes (6). Some of these diagnostics have led to predictive biomarkers for therapeutics, such as vemurafenib for melanoma harboring the V600E BRAF mutation (7). Molecular diagnostics have shown that clinically relevant differences exist in histologically identical tumors between patients, between tumors in the same patient, in the same tumor over time, and even in distinct clonal populations from the same tumor at the same time (8). When developing cancer therapeutics, it is reasonable to expect that different measures of efficacy may be needed for different molecular subtypes, due to differences in tumor biology and in the prognoses with available therapies. The new paradigm would allow for the flexibility needed to study subpopulations of patients with different efficacy measures.

Better resolution on cross-sectional imaging has improved our capacity to measure tumors precisely, making it possible to use change in tumor size as a measure of efficacy in patients with measurable disease. Lavin first pointed out that the use of continuous tumor size data to compare randomized groups of patients is more statistically efficient than the use of categorical assessments in nonrandomized patients (9). Karrison and colleagues took this concept one step further by showing how interindividual and intraindividual differences in tumor biology and in the prognoses with available therapies. The new paradigm would allow for the flexibility needed to study subpopulations of patients with different efficacy measures.

The new paradigm would increase the efficiency of drug development. First, the durability of drug effect is an essential principle required for successful implementation of the new paradigm. Regardless of the endpoint(s) selected, a few guiding principles are required for successful implementation of the new paradigm. First, the durability of drug effect is an essential component of the evidence of efficacy. A drug that has a small effect with long duration might be a better initial therapy than one that has a larger effect with shorter duration. For example,
everolimus for advanced pancreatic neuroendocrine tumors does not cause substantial tumor regression in many patients (RECIST response rate of 5% vs. 2% with placebo) but inhibits tumor growth for an extended period of time in many patients (PFS rate of 34% at 18 months vs. 9% with placebo; ref. 16). On the other hand, chemotherapy (5-fluorouracil, cisplatin, and streptozocin) for advanced neuroendocrine tumors causes substantial tumor regression in many patients (RECIST response rate of 33%), but these effects are of relatively short duration (median, 36 weeks; ref. 17). Second, because these endpoints rely on technologic innovations that make historical data unavailable or unreliable, randomization and blinding between an investigational arm and a control arm are essential for declaring that a novel therapy has evidence of efficacy compared with alternatives. Furthermore, randomized allocation of treatment and objective, quantitative measurement by individuals blinded to the assigned treatment can overcome biases introduced by recognition of drug-specific toxicities. Figure 1 shows how the new paradigm offers multiple approaches for answering the key questions of the treating oncologist. In a randomized phase II trial with a control arm, comparisons based on one or more of the proposed endpoints would help sponsors decide whether or not to conduct a phase III trial. Investigators would prospectively define a clinically meaningful benefit based on these endpoints, considering the disease and patient population.

RECIST falls short of the new paradigm

In recent years, most novel therapies have been developed based on activity against a molecular target or pathway that is a putative driver of the malignant phenotype. As Ratain and Eckhardt have pointed out, such drugs may be active without meeting criteria for a RECIST response (18). As of December 2011, we are aware of at least 4 drugs in 5 indications with evidence of efficacy as monotherapy leading to U.S. Food and Drug Administration approval despite RECIST response rates of 10% or less in phase III trials: sorafenib and everolimus in renal cancer, panitumumab in colorectal cancer, and sunitinib and everolimus in pancreatic neuroendocrine cancer (16, 19–22). Given these examples and the multitude of drugs in development that might also cause growth inhibition or slight tumor regression, the wisdom of continuing to use RECIST response rate as an endpoint for clinical trials is questionable. The increasing use of “waterfall plots” to depict the percentage change in tumor size for individual patients reflects the understanding that RECIST response rates do not tell the whole story, but it remains impossible to make valid conclusions from these data in the absence of a comparator arm. In contrast to RECIST, the new paradigm treats tumor size as a continuous variable, so a drug that causes a mean 0% change in tumor size at a certain time point would have evidence of efficacy relative to a control arm with a mean 10% increase in tumor size.

Concerns have been raised about the accuracy of tumor measurements and the validity of conclusions in trials using RECIST-based endpoints. Oxnard and colleagues conducted a study in patients with NSCLC showing high intermeasurement variability in the size of lung target lesions on CT scans conducted within 15 minutes of each other and concluded that increases or decreases of less than 10% might be the result of this variability alone (23). Given that sharp contrasts between lung tumors and the surrounding air-filled parenchyma allow for easier identification of tumor boundaries, this variability might be even higher for tumors in other visceral organs. The use of categorical response criteria magnifies the impact of measurement variability, whereas comparisons of continuous data minimize this impact. Tang and colleagues showed that, in single-arm phase II trials with RECIST response rate as the primary endpoint, small errors in historical response rate assumptions can result in a 2- to 4-fold increase in the rate of phase II trials that are falsely interpreted as positive (24). Randomization could minimize erroneous conclusions based on RECIST-based endpoints, but the power of such trials would lag behind the power of trials that use comparisons of continuous data.

---

**Figure 1.** Key questions for the oncologist and reasonable approaches for obtaining answers.

---

**Table:**

<table>
<thead>
<tr>
<th>Questions</th>
<th>Approaches</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Which therapy should be started for a given patient?</td>
<td>Randomized, blinded comparative studies in a carefully selected population (by clinical or molecular subtype) with any of the following endpoints or combinations thereof:</td>
<td>1. The therapy that has the greatest evidence of efficacy</td>
</tr>
<tr>
<td>2. When should an ongoing therapy be discontinued in favor of an alternative?</td>
<td>1. Log ratio of tumor size compared with baseline (with volumetric measurements)</td>
<td>2. When there is no longer evidence to support using that therapy</td>
</tr>
</tbody>
</table>

© 2012 American Association for Cancer Research
The "one size fits all" philosophy of RECIST limits its applicability to trials in a number of disease settings. Cancer is a biologically heterogeneous group of diseases, and the idea that a single set of criteria can assess the response to therapy for different cancers is analogous to the suggestion that the same could be done for distinct diseases in other medical specialties (e.g., rheumatoid arthritis and systemic lupus erythematosus). The response to therapy in Hodgkin lymphoma and diffuse large B-cell lymphoma is increasingly assessed by PET/CT rather than conventional CT because of increased sensitivity at early time points (25). Similarly, gastrointestinal stromal tumors may respond to therapy without RECIST sensitivity, resulting in a new set of CT response criteria for this disease that correlate better with meaningful clinical outcomes (26, 27). In ovarian cancer, the challenge of measuring tumor deposits in the peritoneum has led to alternative approaches using CA-125 levels to assess response (28). The absence of reliable measurable disease is also true of CRPC with predominantly bony and lymph node metastases, leading to expert recommendations on biochemical (PSA) and other non-RECIST endpoints for clinical trials (29). The Response Assessment in Neuro-Oncology Working Group was recently formed to overcome the limitations of RECIST and the closely related Macdonald criteria for assessing responses in primary brain tumors and brain metastases and has published new criteria for high-grade gliomas that more precisely define measurable disease and incorporate contrast enhancement, treatment with corticosteroids, and clinical status as part of the assessment (30). Whereas RECIST-based endpoints have little or no value for clinical trials in each of these tumor types, the new paradigm would encourage measures of efficacy that reflect the biology of each disease and that can be measured accurately with modern technology. Randomization and blinding between an investigational arm and a control arm would preclude the need to validate each endpoint prospectively in each disease setting, as they would minimize the risk of making false conclusions (i.e., type I and type II errors).

Conclusions
Although RECIST served a historical purpose and our colleagues who have worked tirelessly to develop and refine it should be applauded, it falls short of the ideal as an assessment of efficacy for cancer therapeutics in development. Rather than suggest a modification to RECIST or an alternative set of criteria that can be applied broadly to solid tumors, we have argued for a new paradigm in which the tools to establish evidence of efficacy are chosen to reflect the biology of the disease in the population being studied. This new paradigm takes advantage of the state of the science in oncology, including knowledge about the biologic heterogeneity of cancer, improved precision in cross-sectional imaging, the availability of functional imaging studies, biomarkers that have emerged from technologic innovations, and advanced computational methods for modeling longitudinal data. Randomization and blinding between investigational arms and control arms are essential components of this new paradigm, allowing the use of endpoints that are not previously validated. Adoption of this new paradigm will more efficiently answer questions about which therapy to start and when to discontinue therapy and consider alternatives to maximize outcomes for our patients.

Disclosure of Potential Conflicts of Interest
M.L. Maitland has received confidential data and reimbursement for travel expenses from GlaxoSmithKline for related research. No potential conflicts of interest were disclosed by the other authors.

Authors’ Contributions
Conception and design: M.R. Sharma, M.L. Maitland, M.J. Ratain
Development of methodology: M.R. Sharma
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.R. Sharma
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M.R. Sharma
Writing, review, and/or revision of the manuscript: M.R. Sharma, M.L. Maitland, M.J. Ratain

Grant Support
The work was supported by NIH grant T32GM007019 and a Conquer Cancer Foundation Translational Research Professorship award to M.J. Ratain; National Cancer Institute Mentored Career Development Award K23CA124802 to M.L. Maitland.

Received January 6, 2012; revised April 9, 2012; accepted April 11, 2012; published OnlineFirst September 4, 2012.

References
18. Ratain MJ, Echardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. J Clin Oncol 2004;22:4442–45.
RECIST: No Longer the Sharpest Tool in the Oncology Clinical Trials Toolbox—Point

Manish R. Sharma, Michael L. Maitland and Mark J. Ratain


Updated version
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-12-0058

Cited articles
This article cites 29 articles, 17 of which you can access for free at:
http://cancerres.aacrjournals.org/content/72/20/5145.full.html#ref-list-1

Citing articles
This article has been cited by 1 HighWire-hosted articles. Access the articles at:
/content/72/20/5145.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.