FRIENDS of CANCER RESEARCH

Leveraging AI-Enabled Tumor Assessment Tools on Radiological Images to Evaluate Treatment Effect and Support Clinical Trial Endpoints in Solid Tumors

White Paper | 2026

Executive Summary

Accurate and consistent tumor measurement is fundamental to evaluating treatment response in oncology clinical trials. For more than 25 years, Response Evaluation Criteria in Solid Tumors (RECIST) has provided a practical and widely adopted framework for these assessments, however, there are limitations. Rapid advances in artificial intelligence (AI) offer an opportunity to modernize tumor response assessment. AI-enabled tumor assessment tools can detect, segment, track, and quantify tumors across the entire body with potentially greater consistency and improved biological detail.

To realize this potential in drug development, the field must converge on clear standards for how AI-enabled tumor assessment tools are used, validated, and interpreted. Key questions include what is being measured, how results should be compared across tools, and how novel endpoints should be defined and qualified for regulatory use, particularly within pathways such as the U.S. Food and Drug Administration's Accelerated Approval program, where reliable early indicators of benefit are essential.

To drive alignment, Friends of Cancer Research (*Friends*) convened a multi-stakeholder working group to evaluate the current landscape of AI-enabled tumor assessment tools and propose a stepwise roadmap for integration into clinical trials. This white paper outlines:

- Limitations of current imaging-based endpoints and opportunities to improve upon RECIST
- Emerging AI-driven approaches, including enhanced RECIST, radiomics, volumetric analysis, and growth kinetics
- Necessary components of a framework to validate a novel AI-imaging based endpoint as an early endpoint, including outstanding questions for defining the endpoint and considerations for a meta-analysis

By advancing and aligning on methodological standards, the oncology community can enable AI-enabled tumor assessments to supplement, and in some cases surpass, traditional measurement approaches in clinical trials. This shift has the potential to improve endpoint precision, reduce trial timelines and costs, and ultimately accelerate patient access to safe and effective therapies.

Authors

Hillary Andrews, Friends of Cancer Research

Trishan Arul, Picture Health

Felix Baldauf-Lenschen, Altis Labs

Surabhi Bajpai, AstraZeneca

Vipul Baxi, Amgen Inc.

Ariel Bourla, Johnson & Johnson

Lauren K. Brady, Genmab

Nathaniel Braman, Picture Health

Ana Centelles Coll, Quibim

Stephane Champiat, MD Anderson

Andrea Corsi, radiomics.bio

Chong Duan, Pfizer, Inc.

Martin J. Edelman, Fox Chase Cancer Center

Andrea Falkoff, Medidata Solutions

Paul Galette, Pfizer, Inc.

Gregory Goldmacher, Merck and Co., Inc.

Joel Greshock, Johnson & Johnson

Diederik J. Grootendorst, Bristol Myers Squibb

Paul Hérent, Raidium

Erich Huang, The National Cancer Institute

Ohad Ilovich, Amgen Inc.

Ana Jimenez-Pastor, Quibim

Alexander Lazar, The University of MD Anderson Cancer Center

David Leung, Daiichi Sankyo Inc.

Sophie Loizillon, Guerbet

Michael T. Lotze, University of Pittsburgh

Jen-Tang Lu, Vysioneer

Léo Machado, Raidium

Chris Martersteck, *Friends* Advisory Advocate

Ross McCall, Onc.ai

Jon McDunn, Project Data Sphere

Mike Montalto, Amgen Inc.

Andrea Myers, Novartis

Jayant Narang, Takeda Pharmaceuticals Inc.

Sarthak Pati, LivAi Inc

Nicolas Pecuchet, 3DS Dassault Systèmes

Ted Pellas, GSK

Kai Qu, Amgen Inc.

Bharath Ramakrishna, Voiant

Fabian Rengier, Novartis

Lia Ridout, *Friends* Advisory Advocate

Marc-Michel Rohe, Guerbet

Lawrence Schwartz, Memorial Sloan Kettering Cancer Center

Alain Silk, Tempus AI

Chiranjiv Singh, Tempus AI

Pallavi Tiwari, LivAi Inc.

Pedro A. Torres-Saavedra, The National Cancer Institute

Sarah Vizel Genco, Quibim

Wim Vos, radiomics.bio

Yu-Shan Yang, Vysioneer

Ozlem Yarbidi, Formerly Takeda Pharmaceuticals Inc.

Seok Yeong Kim, MERIT CRO

Shinji Yue, MERIT CRO

Binsheng Zhao, Memorial Sloan Kettering Cancer Center

Table of Contents

Introduction

Novel artificial intelligence (AI)-enabled approaches for assessing tumor lesions have the potential to transform treatment response evaluation, especially in clinical trials, where accurate and consistent tumor measurement is essential. For over 25 years, Response Evaluation Criteria in Solid Tumors (RECIST) has served as the standard framework for solid tumor measurements. While widely adopted and useful, RECIST has well-recognized limitations, including its reliance on manual, inconsistent, one-dimensional assessments of a limited number of lesions—an approach that can overlook biological complexity and introduce variability. Novel AI-enabled tumor assessment tools* analyze tumors in more nuanced and multidimensional ways, enabling a new level of comprehensiveness, quantitation, and consistency that may supplement and surpass human capabilities.

Before broadly incorporating AI-enabled tumor assessment tools into clinical trials or using their outputs for clinical decision-making, identifying the setting, harmonizing the approach, and defining how to interpret the tool outputs are necessary to promote consistency across trials, build credibility, and enhance transparency. Clinical trial designs that support regulatory decision-making require analysis of accuracy and reproducibility to realize the tools' promise and ensure acceptability. For novel tool outputs, the field must first agree on key definitions including, but not limited to, what is measured, how AI-enabled tools perform assessments, what the output represents in respect to clinical implications, and the circumstances in which its use is clinically appropriate.

Friends of Cancer Research (*Friends*) established a working group of experts to provide a landscape assessment of AI-enabled tumor assessment tools, including their current and emerging applications within oncology clinical trials. The outputs from these tools hold potential for decision-making in a variety of clinical trial contexts including go/no-go decisions, dose-finding, and ultimately, as early endpoints for the U.S. Food and Drug Administration's (FDA) Accelerated Approval Pathway. For our discussions, we consider the "tool" to be the product that measures the radiographic image including any analysis (i.e., the assay) and the "tool output" as the resulting information (i.e., the biomarker). The tool output is used to understand how the biomarker changes in response to treatment (i.e., the endpoint). We consider how tool outputs could be incorporated as *early* endpoints in clinical trials, as this use case requires the most evidence. Other *Friends*-led projects have demonstrated the importance of consistency in assessment approaches, shared reference datasets, and collaborative validation when developing and evaluating new tools and endpoints, setting the stage for the current work.[1,2]

We explore how novel technologies may enable more quantitative and consistent tumor assessments, emphasizing the need for collaboration and rigor before broadly integrating AI tools into clinical trials. Our hope is that these early efforts will enable consideration of a standard repertoire of tumor assessment tools and their corresponding outputs that could be used to implement and evaluate other exploratory imaging biomarkers in the future. Ultimately, we intend to provide a framework for stepwise adoption and eventual standardization of imaging-based endpoints across studies and sponsors, enabling objective insights from

*The term AI-enabled tumor assessment tools was selected by the working group to describe the tools that assess radiologic images and provide measurements that support an understanding of tumor dynamics. The term AI-enabled tools is also used. Principles incorporated throughout can also be considered for quantitative approaches for measuring tumors that may not incorporate AI.

early-phase trials through regulatory decision-making, including late-phase approval pathways such as Accelerated Approval.

The History of RECIST and Current Challenges

As clinical cancer research increased in the last century, growing interest in comparing outcomes across trials underscored the need for a standardized framework to consistently measure and report tumor changes. During the 1980s, the World Health Organization (WHO) published a series of guidelines describing bidimensional assessment of tumor measurements. Groups, including the nascent U.S. National Cancer Institute's (NCI) cooperative groups, modified the WHO criteria, which led to a lack of standardization and made it difficult to compare results across studies or interpret tumor shrinkage as a reliable indicator of clinical benefit. In response, the European Organisation for Research and Treatment of Cancer (EORTC), the NCI, and the National Cancer Institute of Canada (NCIC) collaborated to develop RECIST as a unified approach to simplify and standardize tumor measurements.[3] RECIST was rapidly adopted by academic researchers, NCI's cooperative groups, and regulatory agencies, and later refined to address emerging needs in cancer research as RECIST v1.1.[4] More recently, Immune-related RECIST (iRECIST) was developed to evaluate atypical response patterns often observed with immunotherapies;[5] however, it has not been used as a formal endpoint in registrational trials.

While RECIST is a standardized basis for defining endpoints such as objective response rate (ORR) and progression-free survival (PFS), limitations exist. RECIST v1.1 focuses on up to five target lesions representative of all involved organs per patient and uses unidimensional measurements. These decisions were originally made for pragmatic reasons, such as ease of implementation and reduced variability among human readers. However, measuring only a sample of lesions, rather than all lesions, assumes that overall survival (OS) is determined by the behavior of the average lesion. This assumption is flawed because patient survival is limited by their most aggressive and/or treatment-resistant lesion, not the average. Another challenge is distinguishing treatment-related changes (i.e., pseudo-progression) from tumor growth. In some indications, unidimensional measurements per RECIST do not capture treatment effect as well as volume measurements.[6] Moreover, the reliance on one-dimensional measurements, combined with the subjectivity of manual lesion selection and measurement,[7] can limit the ability to fully and consistently capture tumor dynamics over time.

While RECIST has been invaluable in enabling a standardized assessment approach for comparing results between studies of individual therapies, associated limitations have prompted growing interest in complementary or alternative approaches, such as outputs from emerging AI-enabled tumor assessment tools. These innovations offer the potential not only for improving efficiency and consistency of RECIST but also to support novel imaging-derived endpoints that may more accurately reflect overall disease burden, patient outcomes while reducing time investment of radiologists. The promise of these approaches must be systemically explored and validated prior to their use. As quantitative imaging advances, such methods may facilitate more comprehensive and objective assessments while improving reproducibility compared to RECIST.

This evolution is particularly relevant in the context of the FDA's Accelerated Approval Pathway,[8] where early, reliable indicators of treatment benefit are critical for regulatory decision-making and timely patient access to promising therapies. If novel AI-enabled endpoints derived from routine medical imaging show stronger association with OS or can be assessed sooner than RECIST-based endpoints, trials may reach the same

conclusions with fewer participants or shorter timelines. Ultimately, the clinical trial process may become more efficient, bringing safe and effective drugs to patients faster, more rapidly eliminating ineffective therapies, and making better use of the existing approach to non-invasive routinely collected imaging data. Importantly, data and evidence are necessary to support the use of these tools in this context.

Oversight Considerations

For this review, we focus on tumor assessment in clinical trials for drug approval by FDA, rather than routine clinical practice. In clinical trials focused on drug safety and efficacy, tumor assessments support retrospective evaluation of treatment effect at the study level, whereas in clinical practice, they guide an understanding of individual patient responses and inform treatment decisions. Unlike clinical practice, trial tumor data may be analyzed after study completion to assess a drug's overall effect. As a result, the risk of using a tool output to define an endpoint and understand a drug's efficacy is much lower than the risk of using that same tool output to inform treatment decisions since the patients are not directly impacted.

Generally, tools used to evaluate endpoints in clinical trials do not require FDA's Center for Devices and Radiological Health (CDRH) clearance or approval (i.e., as a device). However, if the tool were independently validated and FDA-authorized, and its use in a clinical trial was within the authorized indications, drug regulators may have additional assurance when evaluating the tool's performance and validity for endpoint measurement in clinical trials. Approaches to endpoint measurement should be discussed with the FDA as appropriate, prior to their use in clinical trials.

Novel endpoints must be qualified within a specific context of use (COU) before they can be used as an early endpoint to support regulatory decisions for Accelerated Approval.[9] There are three main pathways for qualification: 1) the Biomarker Qualification Program (BQP), part of FDA's Drug Development Tool (DDT) Qualification Program; 2) collaborative group interactions with FDA's Center for Drug Evaluation and Research (CDER); and 3) qualification within a specific drug development program. Tools should be adequately verified and validated to ensure they can reliably measure and support interpretation of the endpoint for various tumor types and treatment modalities. While our focus is on regulatory oversight in the U.S. by FDA, lessons learned can translate to other global regulatory authorities.

To ensure their appropriate use in endpoint measurement, AI-enabled tumor assessment tool validation requires large, relevant, and reproducible datasets, along with a clear understanding of the tool's role within upstream and downstream clinical workflows. The availability of large multi-modal datasets, such as the UK Biobank, The Cancer Imaging Archive, and the Radiological Society of North America (RSNA) AI Challenge Collection, make it possible for tool developers to build AI models for computational use cases. There may be potential to use these databases for tool validation, however, having clear rules for training versus validation is critical. Ongoing monitoring may be necessary to ensure continued safety and efficacy as patient populations evolve, and as the tool's machine learning (ML) algorithms adapt or change over time. FDA's guidance on AI/ML-based software as a medical device (SaMD) provides standards for evaluating these tools' safety, effectiveness, and clinical impact and take ML into consideration.[10]

Approaches to Assessing Tumors

Multiple considerations are important before incorporating AI-enabled tumor assessment tool output into clinical trials for regulatory decision-making. Broadly, we need alignment on definitions, assessment approaches, data requirements, validation expectations, and appropriate clinical use. Additionally, education

for patients, providers, and regulatory agencies is important to improve comfort with AI-enabled tools. To support this, recent FDA approvals of AI-enabled devices and increased use of AI by healthcare firms over time suggest increased adoption of AI.[11,12]

Here, we propose a stepwise approach for introducing AI tools into clinical trials as technologies improve and stakeholders become more comfortable with their use. While the use of these tools in clinical trials does not necessarily require CDRH review and approval or clearance, we draw on concepts from CDRH regarding the development and testing of AI-enabled tumor assessment tools. This approach provides a useful structure for describing a stepwise incorporation of AI-enabled analysis tools into clinical trials, focusing on the level of risk as informed by a tool's use in a clinical trial and its technological characteristics.[13] These concepts help differentiate lower-risk applications, such as AI tools that support or confirm human measurements, from high-risk applications, such as tools that independently perform assessments used to determine trial endpoints.

AI-Enhanced RECIST

Introducing AI-enabled tools to perform RECIST measurements is a meaningful incremental step, maintaining a well-known paradigm and supplementing the measurements with potentially greater accuracy and reproducibility. Currently in registrational trials, a blinded independent central review (BICR) performs RECIST assessments to ensure objectivity and consistency across readers. Generally, a 2+1 paradigm is employed, in which two radiologists interpret and complete RECIST assessments independently. A third independent human reader resolves any discordant categorization through an adjudication process. Incorporating AI-enabled tumor assessment tools into individual points of this workflow may improve efficiency by automating lesion identification and measurement, reducing variability, and streamlining the adjudication process or eliminating it altogether.

Objectivity is critical, as local readings can introduce bias and categorize patients on treatment arms as responders to prolong therapy or patients on control arms non-responders to enable crossover. One lower-risk approach to incorporating AI-enabled tools is to provide AI-generated results to local radiologists as a reference, allowing them to review a tool's results after reading the scan while retaining final decision-making authority (i.e., human-in-the-loop). In this context, such tools should not be used for clinical decision-making by the local radiologist as this might impact treatment decisions.

The AI-enabled tools may provide outputs with improved consistency and could theoretically support the BICR process by replacing one or more of the human readers. Initially, the tools could include a human-in-the-loop approach to support lesion identification or tracking, which would streamline the process by providing AI measurements subsequently checked by the radiologist. As performance of these tools and confidence in use of their results improves, an alternative approach would be to have AI-enabled tumor assessments replace one of the two primary readers and to engage the adjudicator only if the human and AI-enabled readings disagree. Defining clear criteria for when the AI can "graduate" to an independent reader to maximize efficiency will be critical, which can at least partially be accomplished by incorporating performance measures.

While enhancing RECIST with AI has potential, using AI tools to improve tumor assessments warrants validation and alignment. One of the biggest software challenges is identifying, selecting, and segmenting tumors to align with RECIST. This process requires nuanced and subjective interpretation informed by human expertise, especially in abdominal/pelvic regions where the contrast between tissues can be limited.

How similarly AI-enabled tumor assessment tools measure RECIST is unknown. *Friends* leads the ai.RECIST Project, which aims to determine the level of agreement among AI-enhanced tools that use RECIST to measure the same dataset of images.

Other Uses of AI-Enabled Tumor Assessment Tools

Beyond supporting current standard tumor measurement, AI-enabled tumor assessment tools offer the potential to generate novel insights. By analyzing complex spatial and temporal tumor characteristics that may be imperceptible by humans, AI can uncover subtle biological signals, such as intra-tumoral heterogeneity, architecture, and growth dynamics. These types of features may provide earlier and more nuanced indicators of treatment response, as well as potential evidence of response. AI-enabled tumor assessments have the potential to transform imaging and provide biologically meaningful evidence to support drug development and regulatory decision-making. As assessments move away from data interpretable by humans towards features measurable only by AI tools, the amount of evidence required to demonstrate reliability and consistency increases.

Tools should be appropriately validated to support the use of the information they provide, especially in case of "black-box" models lacking clear interpretable biological correlations.[14] Initiatives like FNIH's Advanced metrics and modeling with Volumetric CT for Precision Analysis of Clinical Trial results (Vol-PACT), RSNA's Quantitative Imaging Committee (QUIC), and the Quantitative Medical Imaging Coalition (QMIC) are working to scale, validate, and standardize novel methods and move toward broader adoption in clinical trial design and endpoint development.

Radiomic Features

AI-enabled radiomics enables quantitative assessment of macroscopic tissue architecture and dynamic tumor behavior through advanced feature extraction.[15] The ability to perform radiomic analysis at large scale can only be accomplished with AI measuring techniques due to the time needed to segment each lesion, and/or its surrounding tissue microenvironment. Texture analysis captures variations in voxel intensity that characterize intratumoral and peritumoral heterogeneity, potentially indicating differences in cellular density, necrosis, or fibrosis. Radiomics can detect gradients in perfusion, oxygenation, or metabolism that drive progression and therapeutic resistance by analyzing spatial relationships within and around the tumor.

When applied to longitudinal imaging, these features can track subtle structural and textural changes over time, offering early and quantitative insights into treatment response. Collectively, these multidimensional measures translate imaging data into a deeper understanding of tumor biology and the associated host response, which may enable the development of biomarkers that reflect dynamic changes in tissue architecture and ultimately, therapeutic effect.[16] Some approaches include:

- Characterizing tumor morphology (i.e. surface- and shape-related features) as well as local heterogeneity via textural representations
- Defining tissue architecture (e.g., tumor vascularity using the Quantitative Vessel Tortuosity (QVT) Score)[17]
- Predicting tumor infiltrating lymphocytes presence and location (e.g., inflamed, desert, or immune excluded)[18]
- Distinguishing between viable and non-viable tissue (e.g., Fractional Tumor Burden (FTB) maps)[19]

- Assessing the impact of lesion location, particularly adjacent to large vessels[20]

Volumetric Assessments

Many AI-enabled tools can provide volumetric assessments as a potentially reliable indicator of total tumor burden. Unlike unidimensional (e.g., RECIST) or bidimensional measurements (e.g., WHO criteria), volumetric assessments may provide more sensitive and biologically relevant information.[21] There are two key mathematical reasons for this: 1) AI-enabled assessments can be rapidly produced for nearly all lesions in the body, thus providing a more accurate assessment of total tumor burden, and 2) AI can perform volumetric assessments efficiently, whereas manual volumetric assessments are extremely time-consuming, tedious, and potentially variable. By enabling volumetric assessment of all lesions (i.e., total tumor burden), rather than selecting a set of target lesions, AI-enabled tools may better capture assessments of the most aggressive and treatment-resistant lesions that lead to worse outcomes (i.e., death). Some AI tools have already shown how volumetric measures can help reduce inter-observer variability and reduce the average measurement time.[22] Moreover, volumetric measures may be more sensitive to changes in the lesion burden and may detect responders and non-responders at earlier time-points. Morphological biomarkers such as surface irregularity may be considered as a prognostic or predictive factor.[23,24]

To provide a clearer understanding of spatial heterogeneity, tumor volumes are sometimes analyzed in combination with radiomic features. While volumetric and radiomic assessments may provide more detail, whether this additional detail supports stronger associations with outcomes is yet to be determined. Furthermore, the location of lesions with respect to major neurovascular structures could also bring additional information into consideration, as is currently conducted routinely in the setting of primary pancreatic cancers.[25,26] Despite these advances, there is a lack of standardization in the approach for measuring tumor volume, location, and total tumor burden. There are relatively nascent proposals for volumetric response criteria, for example, the RSNA Quantitative Imaging Biomarkers Alliance's (QIBA) CT Tumor Volume Change for Advanced Disease (CTV-AD).[27,28]

Tumor Growth Kinetics

While imaging offers static snapshots of radiomic features or volume, growth kinetics reveal the dynamic processes underlying those changes, which can help distinguish meaningful biological effects from measurement noise or natural variability. Multiple kinetic models can translate changes in imaging-based assessments into biological growth signals like growth rate or doubling time.[29] Kinetic approaches can support an understanding of whether a tumor is increasing, stabilizing, or decreasing. The potential to include quantitative kinetic endpoints into clinical trials could complement or refine categorical endpoints. As with the other approaches outlined here, agreement on the key characteristics of the approach, which need to be harmonized, and how to incorporate these measures into clinical trials for regulatory decision-making all need to be considered.

Additional Approaches Beyond Radiologic Assessment

While this white paper focuses on radiological assessment of contrast enhanced CT/MRI images, it is important to acknowledge other approaches currently under consideration for assessing tumor response. Functional and molecular imaging modalities such as FDG-PET, PSMA-PET, spectral CT/dual-energy CT, and DW-MRI may enable earlier and more nuanced detection of tumor response by revealing metabolic and

physiological changes, which may even precede lesion detection or visible size reduction. In parallel, the emergence of liquid biopsy through analysis of circulating tumor DNA (ctDNA) and circulating tumor cells (CTCs), may offer dynamic insight into resistance mechanisms and disease progression that could be further enhanced if combined with radiologic assessments. The rise of neoadjuvant therapies has elevated pathological response assessment through measures such as pathologic complete response (pCR) and major pathologic response (MPR) as more direct indicators of residual disease and treatment efficacy. Together, these advances mark a shift from traditional, size-based RECIST assessments toward a more holistic, multimodal, and biologically informed framework that integrates imaging, pathology, and computational tools to better capture treatment response and guide personalized cancer care.

Uses for Tumor Assessments in Clinical Trials

The goal of early-phase clinical trials is to establish findings that support an understanding of a novel therapy's safety, dosage, pharmacology, and preliminary efficacy. However, RECIST measurements are relatively coarse and may not capture subtle differences in tumor biology, meaning larger sample sizes are typically needed to detect meaningful changes. Because early studies are intentionally small by design, reliance on RECIST-based endpoints limits statistical power and complicates interpretation. Moreover, the association between RECIST-based endpoints and OS is weak, such that high stakes decisions are made based on poor inputs that lead to failures later in development.

AI-enabled tumor assessment tools may allow for a deeper biological and morphological characterization of tumor response, enabling a more precise understanding of drug effects earlier in development, and ultimately supporting earlier decisions about the benefit/risk profile. AI-enhanced tumor assessment tools that measure full disease burden or use advanced anatomic assessments may provide more nuanced and biologically relevant outputs that more accurately delineate between patients who have a response. Additionally, baseline imaging could inform risk modeling by identifying patients at higher risk of early treatment discontinuation. Earlier and more informative response assessments may reduce development costs through fewer imaging sessions, earlier response stratification, and more efficient allocation of clinical trial resources.

Late phase clinical trials are designed to confirm efficacy and safety in large patient populations and provide evidence for regulatory approval. If AI-enabled tools have greater consistency, sensitivity, and ease of use compared with manual RECIST measurements, they may ultimately help trials run more efficiently and generate richer data to inform regulatory decisions. RECIST-based endpoints (e.g., ORR and PFS) are often used as early or surrogate endpoints that are reasonably likely to predict a clinical benefit for Accelerated Approval. While PFS is associated with OS at the patient-level, trial-level associations are variable.[30] AI-enabled tools may provide outputs that have stronger associations with OS at the patient- and trial-level.

Establishing Novel AI-Enabled Tool Outputs Clinical Trial Endpoints

Using FDA's evidentiary framework for drug development tools provides a general roadmap for qualifying novel biomarkers that can be considered for AI-enabled tumor assessments as clinical trial endpoints.[9] It is critical to first define the unmet medical need, then the COU, a concise description of the specific circumstances under which the tool is proposed to be used in drug development and regulatory review, which will drive the level and type of evidence required to support its qualification. Although the tool could potentially be applied in multiple settings, it is essential to focus on a development setting where there is a

clear unmet need and a plausible impact on decision-making. Once a COU is established, the evidentiary framework further emphasizes understanding the data necessary to support qualification, including both analytical validation (i.e., demonstrating reliable and reproducible measurement performance) and clinical validation (i.e., demonstrating a robust relationship between the tool output and relevant clinical outcomes for the COU).

Two examples of accepted early endpoints are pCR in breast cancer and minimal residual disease (MRD) in multiple myeloma. In both cases, structured collaboration between clinical researchers, cooperative groups, and the FDA helped generate the evidence needed for regulatory consideration. For pCR, the FDA convened a working group to systematically evaluate patient-level trial data using consistent and clearly defined pCR criteria, ultimately establishing its prognostic relevance in early-stage breast cancer.[31] For MRD, the International Myeloma Working Group (IMWG) first developed a harmonized definition and standardized assessment framework,[32] which enabled its prospective incorporation into clinical trials.[33] Large-scale meta-analyses, including the EVIDENCE and I^{2}TEAMM datasets, subsequently demonstrated a strong association between MRD response and long-term outcomes.[34–36] In both breast cancer and myeloma, meta-analyses did not support qualification as a *surrogate* endpoint, which requires demonstration of both patient- and trial-level associations between the endpoint and long-term outcomes like OS.[31,37] Despite this, public discussion at Oncologic Drug Advisory Committee (ODAC) meetings reinforced the strength of the evidence, ultimately supporting the use of these biomarkers as *early endpoints* to inform regulatory decision-making.

These lessons from prior experience provide more practical considerations for developing endpoints using novel AI-enabled tools. Once the COU is established, it would be helpful to reach consensus on which output should be pursued as an endpoint. Once determined and then defined, the endpoint should ideally be incorporated prospectively into clinical trials to minimize missing data. For AI-enabled imaging tools, there may be opportunities to leverage existing imaging data from previously completed trials, since the sponsors prospectively define the timing and frequency of tumor assessment for RECIST assessments and that timing likely aligns with the requirements for a new AI-based endpoint.

The sections below outline key questions to guide alignment on COU, endpoint definitions, and the associated meta-analyses.

Determining COU

COU should consider the BEST biomarker category and specific use in drug development.[9] In the case of clinical trial decision-making, the AI-enabled tools would be used as a response biomarker. For the specific use case, consider:

- Which areas have the highest unmet need/ would benefit from this novel endpoint?
    - Cancer type
    - Treatment modality
    - Cancer stage
- Are AI-enabled tools trained in one setting vs. another?

Defining the Endpoint

1. Align the optimal approach for measuring tumor response to treatment.

    - Is the approach focused on lesion-, organ-, or patient-level assessments?

2. Validate the chosen approach.
    - Why is this approach the most appropriate approach for accurately assessing tumors in the specific COU?
    - How does the approach compare with alternative methods, including RECIST?
    - Is the approach reproducible within a single tool and across different tools?
    - Can tumor assessments be consistently applied across individual clinical settings?
    - How can consistency and comparability of tumor assessments across various tools be ensured?
    - Are there any other limitations to the approach that might impact its use in different contexts?
3. Define the endpoint.
    - How is the endpoint defined and derived? Including, what is physically measured, how the tool evaluates the assessment, and how this information is reported.
    - What is the timing of images collection and how consistent does this need to be?
    - How critical is the quality and the anatomical coverage of the images acquired?
    - Is it necessary to have a cut point to determine responder vs. non-responder?
        - Traditionally, criteria such as RECIST rely on set cut points (e.g., a percentage change in tumor assessments) to classify patients as responders or non-responders. With new assessment techniques, especially those producing continuous scores, are strict cut points necessary?
        - If yes, what best practices guide determination of this cut point?
        - If not, how are continuous scores used to determine response?
    - How are “progressors” defined? Is this definition necessary if the endpoint is continuous?
        - How do we avoid excluding patients with 'pseudo-progression’ from trials due to apparent disease progression?
        - How do we account for variable times between the baseline (i.e., pre-therapy study), therapy initiation, and the first treatment assessment scan?
        - Can non-pharmacological interventions (e.g., radiotherapy, surgical excision) be considered instead of defining patients as "non-evaluable"?

Meta-Analysis Considerations

1. Identify dataset clinical relevance and characteristics.
    - What patient-level data from the clinical trial are necessary? Would defining minimum-viable data characteristics be useful for exploratory analysis versus regulatory submission?
        - How is quality data defined and assessed?
        - How can we ensure the data obtained meets these quality standards?

- What imaging protocol and scanner metadata are necessary to support reliable assessments? (Consider practical limitations on collecting extensive data in both clinical trials and routine practice.)

2. Determine clinical trial characteristics.
   - How many trials and patients are needed for robust meta-analysis? Randomized control trials and/or single arm trials?
     - Prioritize trials with survival outcomes
     - Include clinical trials with negative results to reduce bias
   - Could the data be stored in a repository for future validation? If yes, how does this affect data needs and how can we structure the dataset to support both this meta-analysis and future analyses?
3. Identify and select trials for inclusion in the meta-analysis.
   - What are potential risks for sponsors sharing data?
     - Consider concerns about re-adjudicating results - data could be anonymized, shared by drug class, or partially included.
     - How can we ensure patient consent allows for these analyses?
     - How can we update consent forms for future trials to allow for these analyses?
4. Develop and execute an aligned statistical analysis plan.

Conclusions

AI-enabled imaging tools offer the potential to transform tumor burden and treatment response assessments in oncology clinical trials. By generating more comprehensive, quantitative, and reproducible assessments than traditional manual approaches, these technologies may provide earlier and more sensitive indicators of therapeutic activity, supporting more efficient drug development and regulatory decision-making. However, realizing this promise requires deliberate coordination to progress toward harmonized standards.

Key areas for consensus include defining what is being measured, how tumor assessments are generated, how endpoints should be calculated and interpreted, and under what conditions AI-enabled tumor assessment tools can supplement or replace existing frameworks such as RECIST. Lessons learned from FDA guidance and prior qualification efforts demonstrate that robust validation depends on shared datasets, collaborative analyses, cross-tool evaluation, and clear methodological transparency. Importantly, previously completed clinical trials may provide opportunities to retrospectively analyze existing imaging data, accelerate development, and reduce barriers to early validation.

By advancing this work collaboratively across stakeholders, including sponsors, technology developers, imaging experts, regulators, and the clinical research community, we can move toward a future where AI-driven imaging biomarkers and tumor assessment tools are deployed consistently, transparently, and at scale. Such progress has the potential not only to modernize oncology trial endpoints but also to expand the evidence base for Accelerated Approval and other pathways, ultimately advancing timely access to promising cancer therapies for patients.

References

1. Friends of Cancer Research Digital PATH Project. Available from: https://friendsofcancerresearch.org/digital-pathology/.
2. Friends of Cancer Research ctMoniTR Project. Available from: https://friendsofcancerresearch.org/ctdna/.
3. Therasse P, Arbuck SG, Eisenhauer EA et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205–216. DOI: 10.1093/jnci/92.3.205.
4. Eisenhauer EA, Therasse P, Bogaerts J et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). Eur J Cancer 2009;45:228–247. DOI: 10.1016/j.ejca.2008.10.026.
5. Seymour L, Bogaerts J, Perrone A et al. iRECIST: Guidelines for Response Criteria for Use in Trials Testing Immunotherapeutics. Lancet Oncol 2017;18:e143. DOI: 10.1016/S1470-2045(17)30074-8.
6. Hayes SA, Pietanza MC, O'Driscoll D et al. Comparison of CT Volumetric Measurement with RECIST Response in Patients with Lung Cancer. Eur J Radiol 2016;85:524–533. DOI: 10.1016/j.ejrad.2015.12.019.
7. Suzuki C, Torkzad MR, Jacobsson H et al. Interobserver and Intraobserver Variability in the Response Evaluation of Cancer Therapy According to RECIST and WHO-Criteria. Acta Oncol 2010;49:509–514. DOI: 10.3109/02841861003705794.
8. FDA. Accelerated Approval – Expedited Program for Serious Conditions (Draft Guidance, December 2024). Available at http://www.fda.gov/regulatory-information/search-fda-guidance-documents/accelerated-approval-expedited-program-serious-conditions.
9. FDA. Biomarker Qualification: Evidentiary Framework (Draft Guidance, December 2018). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/biomarker-qualification-evidentiary-framework.
10. FDA. Artificial Intelligence-Enabled Device Software Functions: Lifecycle Management and Marketing Submission Recommendations (Draft Guidance, January 2025). Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing.
11. Dozens of New AI-Powered Devices Make FDA's List of Approvals. Available from: https://radiologybusiness.com/topics/artificial-intelligence/dozens-new-ai-powered-devices-make-fdas-list-approvals.
12. Nguyen TD, Whaley CM, Simon K et al. Adoption of Artificial Intelligence in the Health Care Sector. JAMA Health Forum 2025;6:e255029–e255029. DOI: 10.1001/jamahealthforum.2025.5029.
13. Petrick N, Chen W, Delfino JG et al. Regulatory Considerations for Medical Imaging AI/ML Devices in the United States: Concepts and Challenges. J Med Imaging (Bellingham) 2023;10. DOI: 10.1117/1.JMI.10.5.051804.
14. Huang EP, O'Connor JPB, McShane LM et al. Criteria for the Translation of Radiomics into Clinically Useful Tests. Nat Rev Clin Oncol 2023;20:69–82. DOI: 10.1038/s41571-022-00707-0.

15. Nakajima EC, Simpson A, Bogaerts J et al. Tumor Size Is Not Everything: Advancing Radiomics as a Precision Medicine Biomarker in Oncology Drug Development and Clinical Care. A Report of a Multidisciplinary Workshop Coordinated by the RECIST Working Group. JCO Precis Oncol 2024;8. DOI: 10.1200/PO.23.00687.

16. Cottrell TR, Lotze MT, Ali A et al. Society for Immunotherapy of Cancer (SITC) Consensus Statement on Essential Biomarkers for immunotherapy Clinical Protocols. J Immunother Cancer 2025;13:10928. DOI: 10.1136/jitc-2024-010928.

17. Chae YK, Velcheti V, Zhang K et al. QVT Score, A Radiomic Biomarker of Vascular Complexity, Enables Prognostication and Monitoring of NSCLC Immunotherapy. medRxiv 2025;:2025.09.08.25335020. DOI: https://doi.org/10.1101/2025.09.08.25335020

18. Park S, Ock CY, Kim H et al. Artificial Intelligence-Powered Spatial Analysis of Tumor-Infiltrating Lymphocytes as Complementary Biomarker for Immune Checkpoint Inhibition in Non-Small-Cell Lung Cancer. J Clin Oncol 2022;40:1916–1928. DOI: 10.1200/JCO.21.02010.

19. Iv M, Liu X, Lavezo J et al. Perfusion MRI-Based Fractional Tumor Burden Differentiates between Tumor and Treatment Effect in Recurrent Glioblastomas and Informs Clinical Decision-Making. AJNR Am J Neuroradiol 2019;40:1649–1657. DOI: 10.3174/ajnr.A6211.

20. Newcomer K, Bifolco A, Reeves S et al. Macrovascular Tumor Infiltration and Circulating Tumor Cell Cluster Dynamics in Patients with Cancer Approaching the End of Life. Nat Med 2025;31. doi:10.1038/S41591-025-03966-3. DOI: 10.1038/s41591-025-03966-3.

21. Zhao B, Oxnard GR, Moskowitz CS et al. A Pilot Study of Volume Measurement as a Method of Tumor Response Evaluation to Aid Biomarker Development. Clin Cancer Res 2010;16:4647–4653. DOI: 10.1158/1078-0432.CCR-10-0125.

22. Machado L, Alberge L, Philippe H et al. A Promptable CT Foundation Model for Solid Tumor Evaluation. NPJ Precis Oncol 2025;9. DOI: 10.1038/s41698-025-00903-y.

23. Lin P, Pang JS, Lin YD et al. Tumour Surface Regularity Predicts Survival and Benefit from Gross Total Resection in IDH-wildtype Glioblastoma Patients. Insights Imaging 2025;16:42. DOI: 10.1186/s13244-025-01900-2.

24. Lu SL, Chang YC, Liang CH et al. 2012P A Post-hoc Analysis of the CROWN Study in ALK+ NSCLC Using AI to Predict Progression-Free Survival Based on Early Response. Annals of Oncology 2025;36:S1146. Available from: https://www.annalsofoncology.org/article/S0923-7534(25)03556-2/fulltext.

25. Zaky AM, Wolfgang CL, Weiss MJ et al. Tumor-Vessel Relationships in Pancreatic Ductal Adenocarcinoma at Multidetector CT: Different Classification Systems and Their Influence on Treatment Planning. Radiographics 2017;37:93–112. DOI: 10.1148/rg.2017160054.

26. Zins M, Matos C, Cassinotto C. Pancreatic Adenocarcinoma Staging in the Era of Preoperative Chemotherapy and Radiation Therapy. Radiology 2018;287:374–390. DOI: 10.1148/radiol.2018171670.

27. QIBA CT Volumetry Biomarker Committee. QIBA Profile: CT Tumor Volume Change for Advanced Disease (CTV-AD). Quantitative Imaging Biomarkers Alliance 2022. Available from: https://qibawiki.rsna.org/index.php/Profiles.

28. Zhao B, Obuchowski N, Yang H et al. Comparing Quantitative Imaging Biomarker Alliance Volumetric CT Classifications with RECIST Response Categories. Radiology advances 2025;2. DOI: 10.1093/RADADV/UMAF001.

29. Maitland ML, Wilkerson J, Karovic S et al. Enhanced Detection of Treatment Effects on Metastatic Colorectal Cancer with Volumetric CT Measurements for Tumor Burden Growth Rate Evaluation. Clin Cancer Res 2020;26:6464–6474. DOI: 10.1158/1078-0432.CCR-20-1493.

30. Shahnam A, Hitchen N, Nindra U et al. Objective Response Rate and Progression-Free Survival as Surrogates for Overall Survival Treatment Effect: A Meta-Analysis Across Diverse Tumour Groups and Contemporary Therapies. Eur J Cancer 2024;198:113503. DOI: 10.1016/j.ejca.2023.113503.

31. FDA Pathological Complete Response in Neoadjuvant Treatment of High-Risk Early-Stage Breast Cancer: Use as an Endpoint to Support Accelerated Approval (Final Guidance, July 2020). Available at https://www.fda.gov/regulatory-information/search-fda-guidance-documents/pathological-complete-response-neoadjuvant-treatment-high-risk-early-stage-breast-cancer-use.

32. Kumar S, Paiva B, Anderson KC et al. International Myeloma Working Group Consensus Criteria for Response and Minimal Residual Disease Assessment in Multiple Myeloma. Lancet Oncol 2016;17:e328–e346. DOI: 10.1016/S1470-2045(16)30206-6.

33. Landgren O, Devlin SM. Minimal Residual Disease as an Early Endpoint for Accelerated Drug Approval in Myeloma: A Roadmap. Blood Cancer Discov 2024;6:13. DOI: 10.1158/2643-3230.BCD-24-0292.

34. i^{2}TEAMM Presentation to Support MRD as Accelerated Approval Endpoint Oncologic Drugs Advisory Committee (ODAC). Available from: https://www.fda.gov/media/177738/download.

35. Paiva B, Zherniakova A, Nuñez-Córdoba JM et al. Impact of Treatment Effect on MRD and PFS: An Aggregate Data Analysis from Randomized Clinical Trials in Multiple Myeloma. Blood Adv 2023;8:219. DOI: 10.1182/bloodadvances.2023010821.

36. Landgren O, Prior TJ, Masterson T et al. EVIDENCE Meta-Analysis: Evaluating Minimal Residual Disease as an Intermediate Clinical End Point for Multiple Myeloma. Blood 2024;144:359. DOI: 10.1182/blood.2024024371.

37. Combined FDA and Applicants ODAC Briefing Document Drug Topic: Use of Minimal Residual Disease (MRD) as an Endpoint in Multiple Myeloma Clinical Trials. April 2024. Available from: https://www.fda.gov/media/177652/download.