Published in final edited form as:

Acad Radiol. 2017 August ; 24(8): 1027–1035. doi:10.1016/j.acra.2016.11.024.

Beyond Correlations, Sensitivities, and Specificities: Case Examples of the Evaluation of Advanced Imaging in Oncology Clinical Trials

Frank I. Lin, MD¹, Erich P. Huang, PhD², and Lalitha K. Shankar, MD, PhD¹

¹Cancer Imaging Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, NIH. 9609 Medical Center Drive, Bethesda, MD, 20892

²Biometric Research Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, NIH. 9609 Medical Center Drive, Bethesda, MD, 20892

Abstract

Although advanced imaging is an important component of oncology clinical trials, there has not been a lot of success in advancing its use from a research perspective. One likely reason is the lack of consensus on the methodology used to study advanced imaging in trials, which results in a disconcerted research effort and produces data that are difficult to collate for use in validating the imaging components being studied. Imaging is used in cancer clinical trials for various indications, and the study design needed to evaluate the imaging in a particular indication will vary. Through case examples, this paper will discuss how advanced imaging is currently being investigated in oncology clinical trials, categorized by the potential clinical indication for the imaging tool and offer suggestions on how development should proceed to further evaluate imaging in the given indication. Available NCI resources which can assist in this process will also be discussed.

Keywords

Clinical trial; Cancer imaging; Study design

1 Introduction

In recent years, researchers have shown significant interest in using advanced imaging to improve the efficiency and success rates of clinical trials in oncology. In a clinical trial setting, imaging can be used to serve a number of clinical indications, which when described in chronological order relative to a patient's disease process, include diagnosis and staging, prognosis, as a predictive biomarker assay, as a pharmacokinetic or pharmacodynamics

Corresponding author: Frank I. Lin, MD, 9609 Medical Center Drive, MSC 9735, Bethesda, MD 20892-9735, Phone: (240) 276-6151, Fax: (240) 276-7888, (frank.lin2@nih.gov).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

marker, early response assessment, and as the basis of a clinical trial endpoint. Definitions and examples of each of these indications are given in Table 1.

Currently, advanced imaging is most often studied as part of a secondary or correlative science objective within an oncology clinical trial investigating a therapeutic efficacy question. For example, in the trial Radiation Therapy Oncology Group (RTOG) 1106, a F18-Fluoromisonidazole (FMISO) Positron Emission Tomography (PET) scan is conducted at baseline to identify the presence of hypoxia and its role as a prognostic biomarker in non-small cell lung cancer (NSCLC) patients undergoing external beam radiation therapy. (ClinicalTrials.gov Identifier: NCT01507428). On the other hand, advanced imaging may sometimes be studied as the primary objective of a clinical trial without additional evaluation of an investigational therapeutic regimen. An example of this is American College of Radiology Imaging Network (ACRIN) 6678, a study which evaluated the role of F18-Fluorodeoxyglucose (FDG) PET as an early response marker in NSCLC (NCT00424138).

However, despite the rich variety of functions imaging can serve in a clinical trial, most current oncology clinical trials use imaging in very limited roles, most commonly as the basis of a trial endpoint via validated response evaluation criteria such as the Response Evaluation Criteria in Solid Tumors (RECIST), currently at version 1.1¹. Part of the reason that imaging is not more heavily utilized for its other potential roles is the lack of validation of the imaging modality in those roles, which stems from reasons such as a lack of knowledge and consensus on appropriate validating methodology, as well as a general lack of data from prospective clinical trials which can be used to provide such validations. The lack of consensus contributes to the current status of imaging in clinical trials today, which is characterized by a fragmented research effort with investigations that try to establish the technical (e.g. repeatability and reproducibility) and clinical (e.g. correlations with clinical outcomes) validity of the imaging study but do not produce results that can be easily collated into a unified analysis to support further validation and development such as obtaining regulatory approval.

There are several issues which contribute to this fragmentation of effort. For instance, data on novel molecular imaging agents or functional methods for a particular tumor histology may not be generalizable to other tumor types. Furthermore, data used to support the use of an imaging test in one clinical scenario such as response evaluation may not be relevant when that same imaging test is being used in a different clinical role, for example, as a predictive biomarker assay. Similarly, the evaluation of an imaging biomarker to be used for disease characterization requires a different study design compared to an evaluation in a response assessment setting, and it is important for the imaging research community to recognize these differences. To make matters more complicated, in addition to a lack of standardization on technical issues such as acquisition protocols and post-processing algorithms, there is a lack of consensus on basic issues such as imaging biomarker terminology. In order to combine results from different studies seeking to evaluate an imaging agent as an assay for a predictive biomarker, for instance, the definition of a “predictive biomarker” and how it is to be studied and validated should be standardized and understood by the community at large. In this paper, the authors will examine how advanced

imaging has been evaluated in oncology clinical trials categorized by the imaging's clinical indication. Illustrative examples will be provided to demonstrate how the imaging has been studied and suggestions will be provided on potential future studies which can be performed to further clinical evaluation of the imaging tool for that clinical indication.

2 Diagnosis and Staging

The process of differentiating benign from malignant disease can be broadly described as disease characterization and is a process central to clinical roles such as diagnosis (where disease characterization is being performed on the primary lesion) and staging (where distant lesions are being characterized). A variety of different imaging modalities can be used to characterize disease, whether it be an anatomic criterion such as lesion size, or a functional one such as uptake of the glucose derivative FDG on a Positron Emission Tomography (PET) scan. While the decision to call a lesion benign versus malignant is often based on qualitative parameters in clinical practice, there is a growing trend in clinical trials to use more quantitative measures, especially if they can be repeated reliably. Different thresholds for positivity are associated with different sensitivity and specificity for disease detection, and sometimes examining the positive predictive value (PPV) and negative predictive value (NPV) are more relevant. For validation, while histopathological correlation of the imaged lesion to a tissue biopsy is usually considered the reference gold standard, it may be possible to validate a newer imaging agent or modality against an existing accepted one if there is enough evidence to support it.

2.1 Current investigations

Currently, anatomic imaging is most commonly used to determine lesion positivity in a clinical trial, using parameters such as size, morphology, and contrast enhancement. While histopathology will always be the reference standard for determining lesion positivity, tumor characterization by imaging is invaluable in situations where the lesions are too numerous or too risky to biopsy.

When characterizing a lesion using qualitative measures, consistency in interpretation becomes an important issue because reader judgement plays a significant role in qualitative evaluations. Because of this, there are attempts to improve the consistency of existing imaging methods across different institutions and readers by standardizing diagnostic interpretation. One such example when imaging with F18-FDG PET is the Deauville Criteria for lymphoma, which categorizes the characterization of each lesion to a 5-point scale based on how the lesions' Standardized Uptake Value (SUV) compares to its surrounding and to reference regions such as the liver and the mediastinum. Studies have shown that good interobserver agreement is achieved when using the 5-point scale in Hodgkin's Lymphoma^{2,3}, Diffuse Large B-Cell Lymphoma⁴, and in Follicular Lymphoma⁵ with kappa values ranging from approximately 0.75 to 0.85.

When disease characterization is applied to the evaluation of regional or distant lesions, staging is being performed. Staging disease has always been an important role for imaging, which has the advantage of being non-invasive and can provide a more holistic view by looking at large areas of the body simultaneously. In addition to having prognostic value,

accurate staging information can often times impact the treatment course and options available to a patient significantly. An excellent example of how a patient's management can be affected by improved staging information from imaging can be seen in the currently open trial ACRIN 6685 (NCT00983697). In ACRIN 6685, the primary objective is to determine whether a negative F18-FDG PET of the neck (i.e. no abnormal nodal uptake of F18-FDG) can accurately indicate the absence of regional nodal involvement in head and neck cancer, in patients who are being planned for nodal dissection clinically. Since nodal pathology will be available, the negative predictive value of the F18-FDG PET can be determined. If the F18-FDG PET is found to have sufficiently high negative predictive value for nodal involvement, going forward, patients may potentially be spared a highly morbid nodal dissection of the neck.

2.2 Future Directions

One can assume intuitively that decreasing the inter-observer variability in image interpretation would be relevant, especially in the setting of a multi-center clinical trial, although it might still be useful to conduct studies which will directly evaluate the clinical utility of categorized interpretations such as that done using the Deauville Criteria.

Although the results from ACRIN 6685 may very well suggest that F18-FDG PET has a high enough NPV to indicate which patients do not require a neck dissection, further studies will need to be conducted before this can be put into clinical practice. If this study is positive, a subsequent study will need to be done to test the hypothesis that having the staging information obtained from the F18-FDG PET can positively affect clinical outcomes. One possible study design to test this hypothesis would be a randomized-controlled trial where patients are randomized into either a control group, who would receive standard-of-care treatment, or an experimental group who would have treatment decisions made based on information from the F18-FDG PET. The clinical outcomes of these two groups would then be compared. Superior clinical outcomes in the experimental arm would provide evidence of the clinical utility of F18-FDG PET in head and neck cancer staging. Huang et al [reference other paper] provide details on the designs and statistical analyses of such studies.

3 Imaging-Based Prognostic Markers

In oncology, prognosis is the expected outcome on standard therapy for a given stage. Because outcomes on standard therapy can vary widely even among patients of the same cancer stage, prior to the initiation of treatment, it would be useful to identify low-risk patients whose expected outcomes are good enough that they may perhaps forego more intensive and toxic treatments that may only provide marginal improvement in clinical outcome, as opposed to high-risk patients with poor expected outcomes that may benefit from more aggressive alternative treatments⁶. A prognostic marker is a baseline characteristic that can be measured and can indicate expected outcomes on standard treatment. In a clinical trial setting, prognostic markers can be useful not only in disease management, but also in determining clinical trial eligibility as well, since patients with poor

prognosis may be more reasonably be assigned to experimental therapies with higher potential toxicities than a patient with good baseline prognosis.

3.1 Current Investigations

Imaging is currently used for prognosis in clinical trials for some disease types. Tumor size by anatomic imaging is a very intuitive and basic imaging-based prognostic marker that is widely used and accepted. Because tumor size has prognostic value independent of other variables, it is one of the key determinants of a patient's TNM Classification of Malignant Tumors (TNM) staging for many malignancies. For functional imaging, while it is not extensively validated, there is evidence to support the hypothesis that the metabolic activity of a primary non-small cell lung cancer tumor, as measured by the Standardized Uptake Value (SUV) on an F18-FDG PET scan, is prognostic for survival⁷. Furthermore, a 2008 meta-analysis performed by the International Association for the Study of Lung Cancer demonstrated that primary lung tumors with high SUV is a poor prognostic factor for survival⁷. This association between higher SUV of the primary tumor and worse prognosis is also seen in head and neck squamous cell carcinoma^{8,9}. Similar findings have also been reported in gastro-esophageal cancer^{10,11}. In the esophageal cancer example, a meta-analysis of SUV measurements by F18-FDG PET showed that patients classified as having “high SUV” in the primary tumor have worse survival compared to those without¹¹, although the usefulness of these results are limited due to differences in how the threshold SUV for “high” vs. “low” uptake is defined (range SUV 4.5 to 15).

There have been many agents and modalities which have shown promise in a prognostic context, but investigations of them have most commonly consisted of assessing associations of the imaging variables with clinical outcomes. One such example is the RTOG 1106 trial, which investigate F18-FMISO PET as a prognostic marker in NSCLC patients undergoing standard radiation therapy (NCT01507428). Based upon the pre-clinical finding that oxygen is required for the cancer cell killing effects of ionizing radiation, it was postulated and later confirmed in a pre-clinical setting that hypoxia confers resistance to ionization radiation in tumors. In RTOG 1106, an optional baseline F18-FMISO PET scan will be obtained at select institutions, after which investigators will assess correlations between F18-FMISO PET parameters and overall survival. The role of imaged hypoxia as a prognostic marker for clinical outcomes was also demonstrated in the recently published trial ACRIN 6684¹², which showed that the SUV-peak on a F18-FMISO PET have independent prognostic value in patients newly diagnosed with glioblastoma. This study was also provided data which supports the use of the vascular permeability parameter k-trans, as determined by dynamic susceptibility contrast (DSC) and dynamic contrast-enhanced (DCE) MRI, as a prognostic imaging biomarker.

3.2 Future Studies: Evaluating Additional Prognostic Value and Clinical Utility of the Imaging

Statistically significant correlations do not necessarily translate into utility in patient disease management or clinical trial design. In addition to showing an association with clinical outcome, it is important to demonstrate that the imaging has additional prognostic value on top of that of standard prognostic variables such as age and performance status. This

assessment involves showing that prognostic markers involving the imaging, with or without standard variables, forecast the expected outcome significantly more accurately than prognostic markers involving standard variables alone.

The evaluation of F18-FMISO PET in the RTOG 1106 study is a good model for the development of investigational imaging because it is based on a known observed phenomenon that has been reproduced in the pre-clinical model. Although the therapeutic portion of RTOG 1106 is a phase 2 efficacy trial, the F18-FMISO PET imaging component should be considered a pilot evaluation of the tracer for this indication. If a correlation between F18-FMISO uptake and radiation therapy resistance is found, this would provide the necessary preliminary information to then design a phase 2 study where patients can be stratified according to their F18-FMISO PET results. The phase 2 study would be needed at this point because simple correlations seen in the pilot study do not necessarily translate into evidence which can be used to change patient management or to base medical decisions off of. Prior to embarking on such a phase 2 study, however, important information such as an appropriate cut-off threshold value for what defines a poor prognosis versus a better prognosis will first need to be established, including an estimation of the variance in measurement obtained through a test-retest study design. Furthermore, standardized procedures for image acquisition, analysis, and interpretation will need to have been established. After these basic frameworks for image interpretation have been laid out, one can then further establish the true value of an imaging test as a prognostic biomarker by demonstrating that the imaging study still have independent prognostic value after controlling for standard prognostic variables such as age and disease stage.

4 Imaging-Based Assays of Predictive Biomarkers

A predictive biomarker in cancer is a baseline tumor characteristic that indicates whether improved clinical outcome can be expected on an investigational treatment relative to standard therapy. An example of a predictive biomarker in oncology would be the presence of estrogen receptors (ER+) to predict response to estrogen hormonal therapy. Similarly, for non-small cell lung cancer, the presence of the EGFR and ALK mutations should be predictive of response to agents which target the EGFR and ALK gene products, respectively. In imaging, the best example of a predictive biomarker is the presence of radioiodine uptake on Iodine(I)-131 and I-123 gamma camera scans, or a I-124 PET scan, in well-differentiated thyroid carcinoma, in the prediction of successful therapy with radioactive I-131. Conversely, the absence of radioactive iodine uptake on the imaging is predictive for poorer response to radioiodine therapy with I-131 compared to patients with iodine-avid disease.

It should be noted that whether an imaging-based assay is considered predictive depends significantly on the treatment with which the imaging is being associated. For example, F18-FDG PET scans, when used in conjunction with a treatment agent such as an insuling-like growth factor-1 receptor (IGFR-1) inhibitor which affects a target's ability to utilize glucose, might very well be considered predictive for the efficacy of that particular agent because there is a direct connection between the biological mechanisms that are being measured.

However, if the F18-FDG PET is being used with an agent which have no immediate metabolic consequences, this predictive role could be less easy to establish.

4.1 Current Investigations

To date, investigations of imaging-based predictive biomarker assays have consisted of assessing the association between the imaging measurement and an endpoint such as response in patients uniformly treated with the investigational treatment. As an example, several single center studies have shown that in ER-positive breast cancer patients undergoing treatment with aromatase inhibitor or fulvestrant, the SUV obtained from a baseline F18-fluoroestradiol (FES) PET is significantly higher in those that ultimately respond to treatment than in those that do not¹³⁻¹⁵. Building upon these promising single center results for F18-FES PET imaging, ECOG-ACRIN is currently in the process of initiating a multi-center phase 2 evaluation of this promising PET agent tool to be used with endocrine therapy in patients newly diagnosed with metastatic breast cancer (Trial EAI142, NCT02398773). In this study, the hypothesis being tested is that a negative F18-FES PET, as defined by having no lesions with SUV-max $\geq$ 1.5, is predictive of poorer clinical outcome in patients being treated with endocrine therapy, since patients with lesions that have low F18-FES uptake presumably have only low levels of ER expression. The primary outcome measure is the negative predictive value of a F18-FES PET scan, which will be correlated with patients who have progressive disease.

4.2 Future Studies: Assessing Outcome Differences from Treatments Among Benefitters and Non-Benefitters

If the results of EAI142 are positive, i.e. it is found that patients who had low SUVs on the F18-FES PET had worse clinical outcomes as compared to those who did not, it might be enticing to draw conclusions from these findings about the predictive value of F18-FES PET for hormonal therapy. However, these types of evidence would be insufficient by themselves to validate F18-FES PET as a predictive biomarker, due to the fact that EAI142 is a non-randomized study. The observed differences in such a non-randomized study may be a result of confounding factors such as the prognostic value of the imaging, and similar results may also be observed among patients undergoing standard therapy.

In the in vitro biomarker literature, one study design used in practice to disentangle these confounding factors from the predictive value of the underlying biomarker is the Marker by Treatment Interaction Design. This involves demonstrating clinical benefit on the experimental treatment (e.g. longer survival) among biomarker-positive patients and no such benefit among biomarker-negative patients through a randomized controlled trial where patients undergo baseline imaging and are randomized to either the investigational treatment or standard therapy. However, this type of study would often be difficult to conduct in practice for imaging in many cases. Not only are in vitro predictive biomarker assays often widely available (e.g. immunohistochemistry for ER expression), making randomization to standard therapy difficult to justify, but use of the class of treatments associated with the predictive biomarker may be so ubiquitous that investigators are likely to be unwilling to randomizing patients to have that treatment withheld. Methodology for evaluating imaging-based predictive biomarker assays in light of these considerations is sparse in the literature,

but Huang et al propose some study designs that may be useful here (reference companion paper).

5 Imaging-Based Pharmacodynamic and Pharmacokinetic Markers

In this scenario, the imaging is being performed to obtain a non-invasive confirmation of whether a drug has reached their intended target (pharmacokinetics) or to confirm that the drug is having the desired effect on the body (pharmacodynamics). Imaging allows for a whole body evaluation of both the pharmacokinetic and pharmacodynamic properties of a drug that is non-invasive, repeatable, and usually gives this information in real time. This capability is of considerable utility in the early evaluation of targeted agents for which there is a clear intended drug target with predictable distribution patterns that can be confirmed against theoretic or modeled patterns. Furthermore, having the ability to evaluate a drug via imaging allows for the very early evaluation of drug efficacy at the cellular or molecular level, as well as potentially help in determining optimal dose levels or schedules. This in turn allows for the go versus no-go decision regarding continued development to be made earlier in the development process, potentially saving much time and resources that can be diverted to other agents of interest.

5.1 Current Investigations

In clinical trials, perfusion imaging has been used to evaluate the pharmacodynamic effects on tumors of new anti-angiogenesis drugs which are designed to alter perfusion patterns at the target tumor lesion. Changes in perfusion parameters such as the k-trans used in the analysis of perfusion imaging can be seen as early as days to hours after administration of an anti-angiogenesis drug. These findings have been seen in a variety of different tumor types using different imaging modalities such as perfusion CTs and MRIs performed with Dynamic Contrast Enhanced (DCE) and Dynamic Susceptibility Contrast (DSC) protocols. Despite promising early results, it has been difficult to provide conclusive evidence that these observed early changes in the perfusion imaging parameters are associated with clinical outcomes. For instance, although changes in the k-trans parameter can be seen in renal cancer just hours after treatment with drugs such as bevacizumab and cetuximab, these changes have not been found to be associated with later clinical response evaluation metrics such as progression-free survival (PFS) or overall survival (OS)^{16,17}. Nevertheless, for the purposes of determining PK and PD, it is not necessary for these imaging studies to be associated with clinical outcomes.

In addition to perfusion imaging, there has been interest recently in using PET imaging to evaluate the pharmacodynamics effects of a drug. Such an example can be seen in a clinical trial being proposed by the NCI Cooperative Group ECOG-ACRIN, in the study EA2131 (NCT02194829). EA2131 is a combined phase I/II clinical trial evaluating the drug combination nab-paclitaxel/gemcitabine + an experimental agent AZD1775 in pancreatic cancer. In this study, there is inclusion of a secondary study objective which uses F18-Fluorothymidine (FLT) PET to evaluate whether the drug has reached its intended target. In a phenomenon known as “FLT flare,” it has been observed that treatment with the DNA-damaging drug gemcitabine induces a S-phase arrest in tumor cells which leads to a build-up

of the radio-labelled thymidine analogue F18-FLT in cells. This phenomenon can be observed in targets as an increase in the lesion's SUV (i.e. “flare”), which can then be abrogated with successful treatment of agents which releases cells from mitotic arrest such as the WEE-1 inhibitor AZD1775¹⁸. This is an exciting new application of imaging because abrogation of the F18-FLT flare provides mechanistic evidence that the therapy is having the intended effect on its target tumor cells. Moreover, the fact that this phenomenon can be seen at 24 hours means that F18-FLT PET imaging can potentially be utilized as a very early biomarker of efficacy in drugs which works via the CHK-1 pathway. Not only is there a direct link between the PET agent's and therapeutic agent's mechanisms of actions, the PET imaging is being used to directly assess the therapeutic agent's efficacy at the cellular level.

5.2 Future Work

As management of malignancies are based more on their genetic subtypes, it will be increasing important to be able to evaluate tumors using biomarkers that reflect their underlying genetic subtype or expressed phenotype such as receptor status. Furthermore, as therapy becomes more specific and mechanistic, evaluations of the therapeutic effects of these highly-targeted drugs also need to be more tied to the underlying mechanism of action. Although imaging studies such as perfusion CT or MRI and F18-FLT PET provide evidence for the PK and PD of a drug, they can only provide this information indirectly. A more direct method would be if the drug itself can be imaged, and there have been attempts by investigators to label therapeutic drugs with agents which will allow the drug to be imaged, such as the labelling of panitumumab with Zr89 and erlotinib with F18, both positron emitters which allow the tagged drug to be imaged using PET. While imaging using this tagging methodology may give results that closely resemble the true PK and PD of an agent, the assumption still has to be made that the addition of the imaging label of Zr89 and F18 does not significantly change the PK and PD of the drug. An imaging modality which does not alter the drug chemically, such as hyperpolarized MRI, may offer the best solution in terms of accurately and faithfully imaging the PK and PD of a drug, but is much more difficult to carry out technically.

6 Early Response Assessment

One of the potential roles that imaging can play in the management of a cancer patient is to assess whether the likely outcome on a particular treatment is promising before the completion of therapy. This type of evaluation has been termed “early response evaluation” or “interim response evaluation” and is intended to differentiate patients who will be eventual responders from eventual non-responders. Referred to as “adaptive therapy,” cancer treatment that is adjusted prior to completion based on interim results is currently being used in an number of clinical investigations.

Over the past several decades, there have been several attempts to codify and standardize how response to therapy is to be determined by imaging. Currently, the most widely accepted method of determining response is via anatomic imaging, using changes in lesion size as the most important factor in determining response. This response by size criteria approach was first standardized by the World Health Organization (WHO)¹⁹, which was

later improved upon and made into the Response Evaluate Criteria for Solid Tumors (RECIST) criteria in 2000²⁰, and most recently revised in 2009 to RECIST 1.1¹. While anatomic imaging is useful and adequate in many malignancies, anatomic changes such as size changes are often a late response phenomenon and true response to therapy can frequently be detected much earlier using functional and molecular imaging. As a result, efforts to create new imaging response criteria based on these functional and molecular imaging have been dramatically increased in recent years. While none of these new response criteria currently have the wide impact that the RECIST criteria had on imaging in a clinical trial setting, specific response criteria tailored to a particular disease or group of diseases have proven to be successful and some have been incorporated into standard-of-care setting guidelines such as the National Comprehensive Cancer Network (NCCN) guidelines. For example, the Cheson or International Working Group (IWG) criteria for lymphoma²¹, based heavily on the use of F18-FDG PET, are currently the preferred method of determining response in the management of lymphoma. Other examples of response criteria based on functional or molecular imaging include the PET Response Criteria in Solid Tumors (PERCIST) Criteria for evaluating tumor response using F18-FDG PET²² and the Response Assessment in Neuro-Oncology (RANO) Criteria for brain tumors²³. However, before a set of response criteria can be accepted by the scientific and clinical community as being accurate and truly reflective of response, a series of studies and clinical trials need to be done to demonstrate that an imaging modality can be appropriately used to measure therapeutic response.

6.1 Current Investigations

Currently, imaging scans are being widely used for response evaluation not only in a clinical trial setting but in clinical practice. Imaging results such as changes in lesion size on anatomic imaging and changes in the Standardized Uptake Value on a F18-FDG PET scan are now part of the standard reporting of these studies and is used by clinicians in the management of the cancer patient. However, unlike in a study trial setting, imaging used in clinical practice to evaluate therapy response tends to be more qualitative than quantitative in nature and is often not reported using an evidence-based response criteria, except in the few malignancies where such imaging response criteria have been defined, such as lymphoma.

Currently, there are many imaging agents being studied as potential response evaluation biomarkers in a response evaluation role. Most commonly, imaging is done at baseline and after completion of therapy, with the interval change of a particular parameter at a particular threshold being determined to be indicative of response to therapy. For example, in RECIST 1.1, a decrease in the summed lesion diameters of more than 30% is considered to be a partial response to therapy.

For example, in RTOG 1106, metabolic response on an interim F18-FDG PET scan is being used to modify the intensity and coverage of radiation therapy used to treat non-small cell lung cancer (NSCLC) patients. In the Alliance A021302 trial (NCT02485834), lack of a response on F18-FDG PET imaging after one cycle of pre-operative chemotherapy (i.e. F18-FDG PET non-responders) is the criteria being used to place gastric cancer patients on the trial so that they can receive salvage chemotherapy in an attempt to improve clinical

outcomes. While the use of F18-FDG PET in these two studies to modify therapy may seem intuitive and logically, it must be mentioned that a series of studies need to be performed in the particular cancer histology being studied, and in the same time frame, prior to testing the imaging study in such a capacity. For instance, in order to provide enough evidence to use F18-FDG PET non-response as a selection criterion in the A021302 trial, previous studies needed to show, and in fact did show, that differing response on F18-FDG PET scan after neo-adjuvant chemotherapy does indeed forecast different clinical outcomes²⁴⁻²⁷. This is necessary not only to provide the scientific basis for conducting the study, but also to avoid any ethical considerations for using a non-proven imaging tool to make treatment decisions.

6.2 Future Work

As therapeutic paradigms shift in oncology, researchers engaged in imaging research need to be aware of these changes and adapt to the new opportunities that are created. A good example of this is the rise of immunotherapy in oncology. Unlike other therapeutic agents for cancer, which tend to stabilize or decrease lesions after drug administration, immunotherapy mechanism of action involves the recruitment of immune cells into the lesion which causes temporary enlargement of the lesion, creating a dilemma in measuring response to therapy which are most often based on size changes. It therefore becomes difficult to differentiate between responders from progressors as the lesion both increase in size on short follow-up. Imaging methods which could interrogate metabolic or other functional aspect of a treated lesion need to be tested in this setting.

Another area of need is improving the imaging tools that can be used to assess the determination of response in bone lesions. While most non-measurable disease by RECIST criteria can be safely excluded from studies without major losses to accrual or resulting clinical utility, this is not the case in bone lesions. Several malignancies can present with only bony lesions, such as prostate cancer, and the lack of a validated response evaluation imaging tool for these patients remains one of the key stumbling blocks to conducting clinical trials in these patient populations. Not having a validated approach to assess patients with bone-only disease is also costing these patients the opportunity to enroll on therapeutic clinical trials, many of which, due to the need to have measurable endpoints, exclude these cases. There have been some single center investigations examining the use of F18-FDG PET in malignancies such as breast cancer as a response evaluation tool in metastatic bone disease²⁸⁻³⁰. Despite these early promising results, however, more prospective evaluations in a multi-center setting are needed.

7 Imaging-based Trial Endpoints

Imaging-based trial endpoints are alternative measures of the efficacy of a therapy regimen. Examples include complete metabolic response (CMR) rate according to F18-FDG PET, which has been used in Phase II trials of triapine in cervical cancer (NCT01835171), and overall response rate (ORR) according to anatomic imaging, which has been used in numerous Phase II trials across many disease types. These are often ascertainable earlier than more definitive endpoints such as OS, particularly in malignancies with long expected

survival such as prostate or breast cancer, thus expediting drug development and saving valuable time and resources.

Imaging-based Phase II trial endpoints are often measures of the anti-tumor activity of a treatment, such as a decrease in tumor size or in metabolic activity. These Phase II trial endpoints do not need to be direct measures of patient benefit such as OS as the purpose of these studies is to determine whether to proceed to the Phase III trial rather than to gather evidence for regulatory approval. Meanwhile, imaging-based Phase III trial endpoints do need to be surrogates for the direct endpoints (i.e. treatment effects on a more definitive endpoint such as OS need to be captured by treatment effects on the imaging-based endpoint). Thus, evaluation of a Phase III trial endpoint proceeds differently from that of a Phase II endpoint.

7.1 Current Investigations

Currently, many imaging-based Phase II endpoints used in practice are based on changes in lesion size as measured through anatomic imaging using CT or MRI. The RECIST 1.1 criteria is most often used for determining response for solid tumors based on anatomic imaging. Researchers have also shown interest in using other tumor characteristics for this purpose. In some cases, changes in lesion size may not be a particularly accurate reflection of the mechanism of the drug or tumor response. Other measures such as changes in F18-FDG uptake or metabolic response may be superior metrics of anti-tumor activity, and standardized methodologies to evaluate trial response by F18-FDG PET have been published by the European Organisation for Research and Cancer Therapy (EORTC) in 1999³¹ and by Wahl et. al in the PERCIST criteria²². Such endpoints are often used without formal evaluation of their performance in this context, which requires showing that most therapy regimens that move forward to the Phase III trial based on the imaging endpoint are truly efficacious. As mentioned in Huang et al (reference other paper), the positive predictive value (PPV) of the Phase II trial results based on the imaging in predicting the result of the associated Phase III trial should be shown to be reasonably high. Ratain performs this analysis for overall response rate (ORR). However, similar studies have not been done for other imaging modalities to date.

7.2 Future Work

While RECIST is validated for solid tumors in the body, primary CNS malignancies in the brain are excluded. Hence, response to therapy seen on brain imaging has traditionally been done using the McDonald criteria³², and more recently by the RANO criteria which takes into account enhancing regions on T2 weighted MRIs²³. Currently, work is being done to provide further refinements and to provide validated data sets in the RANO 2.0 endeavors. In other disease processes and treatments such as the new class of immunotherapy agents, successful and responsive treatment typically involves an increase in size due to immune cell infiltrations rather than a decrease, a phenomenon which would make the incorrect evaluation of a tumor as having progressed when in reality it is responding quite well to therapy. To account for this phenomenon, new guidelines for the evaluation of immune therapy activity in solid tumors have been developed in a modified version of RECIST called the immune-related response criteria (iRECIST)³³. Having a way to evaluate the response of

immunotherapy agents, whether it be by iRECIST or another something else, will be critical for the development of these agents and for the field.

While the use of F18-FDG PET as part of an imaging clinical trial endpoint is becoming more accepted in some cancers such as lymphoma, there is interest in the field to apply functional imaging such as F18-FDG PET more broadly into other malignancies. To this end, the RECIST working group is currently in the process of constructing a data warehouse with collections of relevant imaging data from existing clinical trials with the goal of evaluating F18-FDG PET for incorporation into the next update to RECIST.

8 NCI Resources

The National Cancer Institute (NCI), one of the member Institutes of the National Institutes of Health (NIH), has a long history of funding and supporting clinical trials in oncology, many of which have significant advanced imaging components in either a primary or in a secondary/correlative role. Currently, the NCI views and evaluates imaging studies both as a standard of care tool as well as a biomarker, depending on the context of use. This classification of imaging studies as biomarkers is reflected in the fact that funding for imaging studies is covered by the biomarker funding mechanism Biomarker, Imaging, Quality of Life Studies Funding Program (BIQSFP) (<http://www.cancer.gov/about-nci/organization/ccct/other-programs/biqsfp>), as well as the fact that imaging studies are eligible for review through the NCI's Biomarker Review Committee (BRC). Borrowing the terminology defined in the BQSFP program announcement, biomarker studies are defined as either "integral" or "integrated" as they relate to a study's primary objective. Integral studies are defined as tests that must be performed in order for the trial to proceed and to complete its primary objective. On the other hand, integrated studies are those that are clearly part of the clinical trial design but their inclusion is meant only to identify or validate the biomarker for planned use in future trials. Currently, the majority of advanced imaging seen in clinical trials can best be described as integrated biomarkers. However, more studies using imaging as integral biomarkers are needed as they are considered to have more fundamental impacts on disease management and therefore have higher priority for development.

Within the NCI, the Cancer Imaging Program (CIP) is tasked with fostering advances in *in vivo* medical imaging sciences through funding as well as strategic and regulatory support of imaging research. As an example, CIP runs and funds the funding opportunity PAR-14-166: Early Phase Clinical Trials in Imaging and Image-Guided Interventions (<http://grants.nih.gov/grants/guide/pa-files/PAR-14-166.html>), which is an R01 grant intended to support clinical trials that conduct preliminary evaluation of the safety and efficacy of imaging agents, among other priorities. To facilitate the conduct of advanced imaging trials in NCI-supported clinical trials, CIP initiated a program called the Centers of Quantitative Imaging Excellence (CQIE, <https://www.acrin.org/corelabs/ncicqiequalificationprogram/sitequalificationmaterials.aspx>), which qualified NCI Cancer Centers to conduct advanced imaging studies in PET, CT, and MRI. Furthermore, within the NCI's two newly consolidated clinical trial networks, the Experimental Therapeutics Clinical Trials Network (ETCTN) and the National Clinical Trials Network (NCTN), a centralized imaging resource

called the Imaging and Radiation Oncology Core (IROC, <https://www.irocqa.org/>) was created. IROC is designed to assist the development and incorporation of advanced imaging into clinical trials supported by the NCI. Other imaging resources funded and supported by CIP include the Quantitative Imaging Network (QIN, <http://imaging.cancer.gov/programsandresources/specializedinitiatives/qin>), which is designed to promote research and development of quantitative imaging methods for the measurement of tumor response to therapies in clinical trial settings. CIP also works frequently with other entities from professional societies such as the Quantitative Imaging Biomarkers Alliance (QIBA) of the Radiological Society of North America (RSNA) to help promote imaging research. These and other NCI resources are available to assist in the development of advanced imaging as biomarkers in clinical trials.

9 Conclusion

Molecular and functional imaging have the potential to be vital tools in the conduct of cancer-related clinical trials. However, there has been a lack of consensus on the evaluation methodology, which severely hinders the ability to collate resultant data into evidence that can be used for validation of the various imaging agents and modalities in their various roles in clinical trials. Trials are needed to demonstrate the efficacy of new imaging agents and modalities in various cancer histologies as well as in different clinical scenarios. We hope this and the companion paper address the issue by providing a framework and context of the current status of imaging in some of the major clinical indications as well as some suggested future directions.

References

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European journal of cancer* (Oxford, England : 1990). 2009; 45(2):228–247.
2. Barrington SF, Qian W, Somer EJ, et al. Concordance between four European centres of PET reporting criteria designed for use in multicentre trials in Hodgkin lymphoma. *European journal of nuclear medicine and molecular imaging*. 2010; 37(10):1824–1833. [PubMed: 20505930]
3. Furth C, Amthauer H, Hautzel H, et al. Evaluation of interim PET response criteria in paediatric Hodgkin's lymphoma—results for dedicated assessment criteria in a blinded dual-centre read. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2011; 22(5): 1198–1203.
4. Itti E, Meignan M, Berriolo-Riedinger A, et al. An international confirmatory study of the prognostic value of early PET/CT in diffuse large B-cell lymphoma: comparison between Deauville criteria and DeltaSUVmax. *European journal of nuclear medicine and molecular imaging*. 2013; 40(9):1312–1320. [PubMed: 23649463]
5. Dupuis J, Berriolo-Riedinger A, Julian A, et al. Impact of [(18)F]fluorodeoxyglucose positron emission tomography response evaluation in patients with high-tumor burden follicular lymphoma treated with immunochemotherapy: a prospective study from the Groupe d'Etudes des Lymphomes de l'Adulte and GOELAMS. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012; 30(35):4317–4322. [PubMed: 23109699]
6. Simon R. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Personalized medicine*. 2010; 7(1):33–47. [PubMed: 20383292]
7. Berghmans T, Dusart M, Paesmans M, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by

the European Lung Cancer Working Party for the IASLC Lung Cancer Staging Project. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2008; 3(1):6–12.

8. Higgins KA, Hoang JK, Roach MC, et al. Analysis of pretreatment FDG-PET SUV parameters in head-and-neck cancer: tumor SUVmean has superior prognostic value. *International journal of radiation oncology, biology, physics*. 2012; 82(2):548–553.
9. Zhang B, Li X, Lu X. Standardized uptake value is of prognostic value for outcome in head and neck squamous cell carcinoma. *Acta oto-laryngologica*. 2010; 130(7):756–762. [PubMed: 20085442]
10. Zhu W, Xing L, Yue J, et al. Prognostic significance of SUV on PET/CT in patients with localised oesophagogastric junction cancer receiving neoadjuvant chemotherapy/chemoradiation:a systematic review and meta-analysis. *The British journal of radiology*. 2012; 85(1017):e694–701. [PubMed: 22337686]
11. Pan L, Gu P, Huang G, Xue H, Wu S. Prognostic significance of SUV on PET/CT in patients with esophageal cancer: a systematic review and meta-analysis. *European journal of gastroenterology & hepatology*. 2009; 21(9):1008–1015. [PubMed: 19352191]
12. Gerstner ER, Zhang Z, Fink JR, et al. ACRIN 6684: Assessment of Tumor Hypoxia in Newly Diagnosed Glioblastoma Using 18F-FMISO PET and MRI. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2016; 22(20):5079–5086. [PubMed: 27185374]
13. Peterson LM, Mankoff DA, Lawton T, et al. Quantitative imaging of estrogen receptor expression in breast cancer with PET and 18F-fluoroestradiol. *J Nucl Med*. 2008; 49(3):367–374. [PubMed: 18287268]
14. Linden HM, Kurland BF, Peterson LM, et al. Fluoroestradiol positron emission tomography reveals differences in pharmacodynamics of aromatase inhibitors, tamoxifen, and fulvestrant in patients with metastatic breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2011; 17(14):4799–4805. [PubMed: 21750198]
15. Kurland BF, Peterson LM, Lee JH, et al. Between-patient and within-patient (site-to-site) variability in estrogen receptor binding, measured in vivo by 18F-fluoroestradiol PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2011; 52(10):1541–1549.
16. Flaherty KT, Rosen MA, Heitjan DF, et al. Pilot study of DCE-MRI to predict progression-free survival with sorafenib therapy in renal cell carcinoma. *Cancer biology & therapy*. 2008; 7(4):496–501. [PubMed: 18219225]
17. Hahn OM, Yang C, Medved M, et al. Dynamic contrast-enhanced magnetic resonance imaging pharmacodynamic biomarker study of sorafenib in metastatic renal carcinoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2008; 26(28):4572–4578. [PubMed: 18824708]
18. McArthur GA, Raleigh J, Blasina A, et al. Imaging with FLT-PET demonstrates that PF-477736, an inhibitor of CHK1 kinase, overcomes a cell cycle checkpoint induced by gemcitabine in PC-3 xenografts. *ASCO Meeting Abstracts*. 2006; 24(18_suppl):3045.
19. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981; 47(1):207–214. [PubMed: 7459811]
20. 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. *Journal of the National Cancer Institute*. 2000; 92(3):205–216.
21. Cheson BD, Horning SJ, Coiffier B, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1999; 17(4):1244. [PubMed: 10561185]
22. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2009; 50(Suppl 1):122S–150S.

23. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010; 28(11):1963–1972. [PubMed: 20231676]
24. Weber WA, Ott K, Becker K, et al. Prediction of response to preoperative chemotherapy in adenocarcinomas of the esophagogastric junction by metabolic imaging. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2001; 19(12):3058–3065. [PubMed: 11408502]
25. Ott K, Weber WA, Lordick F, et al. Metabolic imaging predicts response, survival, and recurrence in adenocarcinomas of the esophagogastric junction. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2006; 24(29):4692–4698. [PubMed: 16966684]
26. Krause BJ, Herrmann K, Wieder H, zum Buschenfelde CM. 18F-FDG PET and 18F-FDG PET/CT for assessing response to therapy in esophageal cancer. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2009; 50(Suppl 1):89S–96S.
27. zum Buschenfelde CM, Herrmann K, Schuster T, et al. (18)F-FDG PET-guided salvage neoadjuvant radiochemotherapy of adenocarcinoma of the esophagogastric junction: the MUNICON II trial. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*. 2011; 52(8):1189–1196.
28. Al-Muqbel KM, Yaghan RJ. Effectiveness of 18F-FDG-PET/CT vs Bone Scintigraphy in Treatment Response Assessment of Bone Metastases in Breast Cancer. *Medicine*. 2016; 95(21):e3753. [PubMed: 27227942]
29. Specht JM, Tam SL, Kurland BF, et al. Serial 2-[18F] fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) to monitor treatment of bone-dominant metastatic breast cancer predicts time to progression (TTP). *Breast cancer research and treatment*. 2007; 105(1):87–94. [PubMed: 17268819]
30. Stafford SE, Gralow JR, Schubert EK, et al. Use of serial FDG PET to measure the response of bone-dominant breast cancer to therapy. *Academic radiology*. 2002; 9(8):913–921. [PubMed: 12186440]
31. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *European journal of cancer (Oxford, England : 1990)*. 1999; 35(13):1773–1782.
32. Macdonald DR, Cascino TL, Schold SC Jr, Cairncross JG. Response criteria for phase II studies of supratentorial malignant glioma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 1990; 8(7):1277–1280. [PubMed: 2358840]
33. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2009; 15(23):7412–7420. [PubMed: 19934295]

Table 1
Clinical indications for which imaging can be performed in a clinical trial setting

Role	Definition	Examples
Diagnosis and Staging	To determine whether a lesion is positive or negative for malignancy	F18-FDG PET in Lymphoma Nodal staging using F18-FDG PET in head and neck cancers (ACRIN 6685)
Prognostic marker	To determine the expected outcome under standard therapy for the patient's disease stage	Lesion size on anatomic imaging such as CT or MRI "High" vs. "Low" SUV on F18-FDG PET in head and neck SCC, NSCLC, and gastro-esophageal cancers.
Predictive biomarker assay	To differentiate between patients expected to benefit clinically on one treatment relative to another from those not expected to experience such a benefit	I-123 scan predictive for I-131 therapy in thyroid cancer F18-FES PET predictive for hormonal therapy in breast cancer (EAI142)
Pharmacokinetics marker	To confirm that the drug has reached the intended target	F18-FLT PET "flare" in pancreatic cancer (EA2131)
Pharmacodynamic marker	To measure the effects of the drug on the body	Perfusion CT and DCE/DSC MRI in anti-angiogenesis targeted therapy
Early response indicator	To determine the expected ultimate outcome on a particular therapy regimen from changes in a tumor characteristic following a few cycles of treatment	F18-FDG PET response in gastric cancer after neoadjuvant chemotherapy (A021302) During-treatment F18-FDG PET evaluation of external beam radiation in NSCLC (RTOG 1106)
Basis of a Phase II trial endpoint	A pre- to post-treatment change measurement used to determine whether to proceed to the subsequent Phase III investigation.	Complete metabolic response according to F18-FDG PET in cervical cancer
Basis of a Phase III trial endpoint	A pre- to post-treatment change that serves as a surrogate for a definitive clinical endpoint.	PFS based on anatomic imaging