Imaging Biomarkers and Surrogate Endpoints in Oncology Clinical Trials

2

Richard G. Abramson and Thomas E. Yankeelov

Contents

R.G. Abramson, MD (✉)
Institute of Imaging Science,
Vanderbilt University, Nashville, TN, USA

Department of Radiology and Radiological Sciences,
Vanderbilt University Medical Center,
1161 Medical Center Drive,
Medical Center North, Suite CCC-1121,
Nashville, TN 37232-2675, USA

Vanderbilt-Ingram Cancer Center,
Vanderbilt University, Nashville, TN, USA
e-mail: richard.abramson@vanderbilt.edu

T.E. Yankeelov
Department of Radiology and Radiological Sciences,
Department of Biomedical Engineering,
Department of Physics and Department of Cancer Biology, Vanderbilt University, Institute of Imaging Science, 1161 21st Ave S, AA-1105 Medical Center North, Nashville, TN 37232-2310, USA

Vanderbilt-Ingram Cancer Center,
Vanderbilt University, Nashville, TN, USA
e-mail: thomas.yankeelov@vanderbilt.edu

Abbreviations

ACRIN	American College of Radiology Imaging Network
CI	Confidence interval
CR	Complete response
DCE-MRI	Dynamic contrast-enhanced magnetic resonance imaging
DFS	Disease-free survival
DW-MRI	Diffusion-weighted magnetic resonance imaging
ECOG	Eastern Cooperative Oncology Group

A. Luna et al. (eds.), *Functional Imaging in Oncology*,
DOI 10.1007/978-3-642-40412-2_2,

EGFR	Epidermal growth factor receptor
EMA	European Medicines Agency
FDA	U.S. Food and Drug Administration
FDG-PET	Fluorodeoxyglucose positron emission tomography
FLT-PET	Fluorodeoxythymidine positron emission tomography
HER2	Human epidermal growth factor receptor-2
IOM	Institute of Medicine
IWG	International Working Group
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
MRI	Magnetic resonance imaging
NCI	National Cancer Institute
NIH	National Institutes of Health
ORR	Objective response rate
PD	Progressive disease
PERCIST	PET Response Criteria in Solid Tumors
PET	Positron emission tomography
PFS	Progression-free survival
PR	Partial response
QIBA	Quantitative Imaging Biomarker Alliance
QIN	Quantitative Imaging Network
RECIST	Response Evaluation Criteria In Solid Tumors
SD	Stable disease
TTP	Time to progression

2.1 Introduction

Interest in oncology biomarkers has surged over the past decade, fueled by scientific progress toward precision medicine and by the practical search for new efficiencies in the expensive and lengthy drug development process [1]. Biomarker use has the potential to help tailor care to patient subgroups, streamline the selection of candidate drug agents, and reduce the cost and duration of clinical trials. The statistical and methodological requirements for biomarker use are still evolving, however, and the scientific, industrial, and regulatory communities continue to struggle with issues of how to properly evaluate and utilize these tools.

This chapter discusses imaging biomarkers in oncology clinical trials. After reviewing definitions for biomarkers and related terms, we discuss the motivations underlying biomarker integration into oncology drug development and the use of imaging biomarkers across the oncology drug development continuum. We then briefly review current and emerging imaging biomarkers for cancer clinical trials. We conclude by providing a brief discussion of the evaluation of imaging biomarkers as surrogate endpoints, including the perspective from regulatory agencies.

2.2 Definitions

The surge in biomarker research over the past decade has brought some confusion with regard to competing definitions for biomarkers and related terms (Table 2.1). This chapter uses the Institute of Medicine (IOM)'s modification of the National Institutes of Health (NIH) Biomarkers Consensus Group definition of a biomarker as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention" [2]. Oncology biomarkers are often divided conceptually into *prognostic biomarkers*, intended to forecast a likely disease course in the absence of treatment, and *predictive biomarkers*, intended to forecast a likely disease course in the presence of a specific treatment. Some authorities include a third category of *early response biomarkers* to draw particular attention to biomarkers that may reveal treatment response or failure earlier than conventional methods [3].

Any discussion of biomarkers in the context of oncology clinical trials must also include definitions of clinical and surrogate endpoints. The NIH Biomarkers Working Group defines a *clinical endpoint* as "a characteristic or variable that reflects how a patient feels, functions, or survives" [4]. Within a clinical trial, a clinical endpoint is a distinct measurement or analysis of

Table 2.1 Definitions

Term	Definition
Biomarker	A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to an intervention
Prognostic biomarker	Intended to forecast a likely disease course in the absence of treatment
Predictive biomarker	Intended to forecast a likely disease course in the presence of a specific treatment
Early response biomarker	Intended to reveal treatment response or failure earlier than conventional methods
Clinical endpoint	A characteristic or variable that reflects how a patient feels, functions, or survives
Surrogate endpoint	A biomarker intended to substitute for a clinical endpoint
Biomarker evaluation	The overall process for confirming a biomarker's suitability as a surrogate endpoint
Analytical validation	The first component of biomarker evaluation in the IOM framework, describing objective demonstration that the biomarker can be accurately measured
Qualification	The second component of biomarker evaluation in the IOM framework, describing objective demonstration that the biomarker is associated with the clinical endpoint of concern
Utilization	The third component of biomarker evaluation in the IOM framework, describing subjective assessment of biomarker performance in the specific context of its proposed use

disease characteristics that reflects the effect of a therapeutic intervention. In general, survival has traditionally been the clinical endpoint of greatest interest in oncology trials [5].

A *surrogate endpoint* is defined as a biomarker that is intended to substitute for a clinical endpoint [4]. Surrogate endpoints are a subset of biomarkers. Using a biomarker as a surrogate endpoint implies that the biomarker has been through a rigorous, formal process to confirm its suitability as a substitute for a clinical endpoint. This confirmatory process has been described using a variety of different terms, including evaluation, qualification, and validation. In keeping with the IOM framework, we use the general term *biomarker evaluation* to refer to the overall process for confirming a biomarker's suitability as a surrogate endpoint. The IOM framework identifies three key components in the biomarker evaluation process: analytical validation, qualification, and utilization [2]. These three steps are discussed in more detail in Sect. 2.7.

2.3 Motivations Underlying Biomarker Integration into Oncology Drug Development

Despite its relatively recent emergence onto the research agenda, the biomarker concept is now invoked at all stages of oncology drug development, from early preclinical studies to late-stage clinical trials. The search for meaningful oncology biomarkers reflects both the growing importance of precision medicine and the increasing emphasis on finding ways to streamline the drug development process [6].

In the context of oncology drug development, precision (or personalized) medicine captures the notion of cancer as a heterogeneous group of diseases characterized by a diverse array of gene expression and activity patterns [7]. Tumor analysis at the molecular level offers the potential for identifying targets that may be variably expressed among different patients or even at different tumor sites in a single patient. Linking different cancer subtypes with the presence or absence of certain biomarkers may facilitate identification of tumors in which a targeted drug agent has a higher likelihood of success. Biomarker integration may thus promote optimization of therapeutic regimens for individual patients, with specific agents being selected only in tumor subtypes associated with a particular biomarker status, e.g., trastuzumab for

HER2-overexpressing breast cancers or cetuximab for EGFR-expressing colorectal cancers lacking a concomitant KRAS mutation.

Meanwhile, integration of biomarkers into preclinical studies and clinical trials offers the potential of reducing the length and expense of the drug development process, which by some estimates can require up to 10 years and one billion dollars [8]. In preclinical and early-stage clinical development, biomarker analysis may promote improved selection of patients for clinical trials by identifying patients who are more or less likely to benefit from therapy. Biomarkers may provide valuable early information on the presence or absence of drug efficacy that can be used to terminate less promising projects before they enter into more expensive later-stage testing [6]. Performing such studies in the preclinical setting also affords the opportunity to select appropriate imaging biomarkers for application in subsequent clinical trials employing the therapy under consideration.

For investigational agents that do proceed into late-stage clinical testing, biomarkers used as surrogate endpoints offer the prospect of smaller, less expensive, and more efficient clinical trials [9]. In particular, if a biomarker is deemed to be an acceptable substitute for survival, then trials can be designed and powered to demonstrate a significant change in the biomarker rather than the clinical endpoint, typically resulting in smaller patient accrual requirements and a dramatically shortened evaluation timeframe. Furthermore, traditional survival trials in many cancer types are becoming more difficult to analyze and interpret due to increasing patient life expectancies and the proliferation of therapeutic options, e.g., in metastatic breast cancer where patients may undergo multiple different lines of treatment extending over several years. It has been argued that in certain circumstances, biomarker-driven trials may offer "cleaner" assessments of drug efficacy than survival trials with fewer problems due to patient loss to follow-up, patients undergoing additional treatment after the investigational therapy, and patients experiencing intercurrent illness and death from other causes [9, 10].

2.4 Use of Imaging Biomarkers Across the Oncology Drug Development Continuum

An important concept to emphasize is that different imaging biomarkers might be appropriately deployed at different stages of the oncology drug development process. As the goals change from preclinical studies into early- and late-stage clinical trials, the requirements for biomarker evaluation and utilization also change.

2.4.1 Preclinical Studies

In the preclinical stages of the drug development process, key priorities include target identification and validation, identification of promising drug compound leads, and demonstration of target engagement as proof of concept. This process typically occurs first in in vitro studies of appropriate cell lines and then in preclinical in vivo animal studies and may initially involve high-throughput screening of thousands of candidate compounds to identify a handful of promising leads. The most sought-after imaging biomarkers during the preclinical stage will provide information on activity at the cellular and molecular levels. A comprehensive review of imaging biomarkers for preclinical drug development is beyond the scope of this chapter, which is focused on imaging biomarkers for human clinical trials.

2.4.2 Early- and Late-Stage Clinical Trials

Investigational agents undergo increasingly rigorous clinical testing as they progress along the path toward regulatory approval. Phase 1 and 2 studies are typically smaller trials designed to establish initial safety and dosing data and to demonstrate efficacy in small study populations. Phase 3 studies are typically much larger and more costly multisite trials aimed at collecting the necessary safety and efficacy data to support a marketing application to regulatory authorities.

Given the expense involved in proceeding from phase 2 into phase 3 testing, a key objective in early clinical trials is gathering sufficient preliminary efficacy data to inform a "go or no-go" decision on continuing into phase 3 studies.

The most relevant imaging biomarkers during clinical trials will be those that provide evidence of drug efficacy as well as those that select for patient subgroups in whom drug efficacy may be higher [11, 12]. In phase 1 and phase 2 trials, investigators may choose to incorporate biomarker endpoints from a constantly expanding menu of advanced imaging techniques, presumably selecting approaches that report on functional or compositional variables correlated with the mechanism of drug action. It is in this context that the advanced techniques described in this text are currently most relevant for demonstrating the efficacy of novel molecularly targeted agents. In phase 3 trials, the most important biomarkers will be those that are reproducible across large, multisite trials and those that have been rigorously confirmed as acceptable surrogate endpoints for survival. The requirements for imaging biomarkers may be more stringent for late-stage than for early-stage clinical trials because biomarker results from early-stage clinical trials are used primarily by the trial sponsor for internal decision making, while results from late-stage clinical trials will be scrutinized by outside regulatory authorities in the drug approval process.

2.5 Current Imaging Biomarkers for Oncology Clinical Trials

Imaging biomarkers for oncology clinical trials have evolved over the past 50 years, driven by the need for objective standards with which to perform "apples-to-apples" comparisons of treatment response both between patients within a clinical trial and between different clinical trials. Until recently, the majority of imaging biomarkers for oncology have centered on tumor size measurement and size measurement changes. This section briefly reviews the salient features and drawbacks of size-based imaging biomarkers, with an emphasis on the Response Evaluation Criteria for Solid Tumors (RECIST), the most commonly recognized and utilized standard for assessing response in solid malignancies. We also review the important criticisms of RECIST, and we describe incremental modifications of RECIST that have been deployed in various tumor types.

2.5.1 RECIST

Imaging-based tumor size measurement assessment guidelines for solid malignancies were first codified in the 1980 World Health Organization criteria [13] and were revised as the Response Evaluation Criteria in Solid Tumors (RECIST) in 2000 [14] and RECIST 1.1 in 2009 [15]. The Macdonald criteria for supratentorial malignant glioma were proposed in 1990 [16]. The International Working Group (IWG) or Cheson criteria for hematologic malignancies were first issued in 1999 [17] and were revised in 2007 [18]. These response assessment tools have enjoyed widespread utilization in the scientific, industrial, and regulatory communities, and new drug approval applications routinely include results using these imaging biomarkers to support efficacy claims.

RECIST was explicitly designed for use in phase 2 clinical trials (although it is used at other stages in drug development and even clinically) and is essentially a guideline for assessing tumor response and progression based on changes in anatomical tumor burden over time. RECIST specifies criteria for categorizing lesions on baseline (i.e., pretreatment) imaging as either "target" or "nontarget" lesions, with the former to be followed with successive quantitative size measurements and the latter to be followed qualitatively. At follow-up imaging (i.e., during treatment), lesion burden is reassessed in standardized fashion, with target lesions reevaluated as the sum of their unidimensional size measurements and nontarget lesions reevaluated subjectively according to changes perceived by the reviewer. At each imaging timepoint (typically a predetermined follow-up interval specified in the study protocol), patients are assigned one of four response

categories: complete response (CR), partial response (PR), stable disease (SD), or progressive disease (PD). These patient-level response categories can then be used to construct summary trial endpoints including objective response rate (ORR, i.e., the percentage of patients achieving either PR or CR), time to progression (TTP, i.e., average time until PD), and progression-free survival (PFS, i.e., average time until PD or death).

As present time, most clinical trial protocols for solid malignancies specify one or more of these RECIST-derived imaging biomarkers as endpoints for assessing efficacy of the investigational agent. Although ORR is still widely utilized, PFS and other "time-to-event" biomarkers are increasingly incorporated as primary efficacy endpoints, especially in late-stage randomized clinical trials [19]. Recent National Cancer Institute (NCI) task force recommendations have specifically encouraged the use of PFS as a primary efficacy endpoint in phase 2 clinical trials [20]. The ascendancy of PFS has paralleled the introduction of cytostatic agents into the therapeutic armamentarium; these agents, in contrast to traditional cytotoxic agents, result in cell cycle arrest rather than cell death and may be less likely to produce gross tumor shrinkage, although patients may benefit from a delay in tumor progression which would be captured in PFS but not in ORR.

2.5.2 Problems with Size-Based Biomarkers

How well do tumor size-based biomarkers perform for response assessment? In the phase 2 clinical trial setting, ORR is generally accepted as a valid indicator of antitumor efficacy because objective responses are infrequent in the absence of efficacious treatment [21]. There also exists a small but important literature linking objective tumor response to clinical survival benefit [22, 23], to success in later-stage clinical testing [24], and to future regulatory approval [25]. In general, however, tumor shrinkage is considered an unreliable surrogate for survival, one that may either overestimate or underestimate a drug's effect on the relevant clinical endpoint [15, 26] and one that may have a different correlation with survival in different tumor types or with different drug agents [25, 27]. PFS, meanwhile, has been correlated with survival only in certain tumor types (particular advanced colorectal and ovarian cancers) [28], and biostatisticians have cautioned against extrapolating survival associations even in these tumor types to novel anticancer therapies [27]. Although tumor size-based endpoints have become an important basis for oncology drug regulatory approval [21, 29], there have been several high-profile examples of authorities having granted accelerated approval to new anticancer drugs based on tumor size measurement data, only to rescind or narrow the approval as postmarketing data failed to show a survival benefit; notable examples include U.S. Food and Drug Administration (FDA) approvals of gefitinib (Iressa) for non-small cell lung cancer and, more recently, bevacizumab (Avastin) for metastatic breast cancer.

The shortcomings of RECIST have been well publicized [30, 31] and fall into two general categories: practical problems with its implementation and more fundamental objections to a size-based approach to response assessment. Practical problems with implementing RECIST include difficulties choosing target lesions representative of total tumor burden, the need to assess many lesions on a qualitative and subjective basis, and high intraobserver and interobserver variability for heterogeneous lesions or lesions with irregular borders [32]. Single-axis measurements as dictated by RECIST may not adequately capture size changes in nonspherical lesions or lesions with asymmetric growth. Lesions along curved surfaces, abutting other organs, or adjacent to other pathology may be difficult to measure by RECIST guidelines. Finally, the use of categorical rather than continuous response variables may sacrifice statistical power.

A more fundamental objection to RECIST is that an exclusive focus on tumor size may exclude other potentially meaningful features, including morphologic, compositional, and functional parameters that may provide a more comprehensive assessment of tumor status.

Tumor size change may lag weeks to months behind a tumor biological response or may never occur at all. Size measurement criteria may therefore underestimate or fail to capture antitumor efficacy, especially of newer targeted agents that produce a cytostatic rather than cytotoxic effect.

These considerations are motivating ongoing efforts within the oncologic imaging community to improve upon current tumor size measurement biomarkers. These efforts fall into two general categories: incremental modifications to size-based biomarkers and novel imaging techniques reporting on parameters other than tumor size.

2.5.3 Incremental Modifications to Tumor Size Measurement Techniques

Other important modifications of tumor size measurement techniques under current investigation include three-dimensional volumetric measurement approaches [33] and customized response assessment guidelines tailored to specific tumor types. Examples of the latter include the Choi criteria for gastrointestinal stromal tumors [34], the modified RECIST (mRECIST) criteria for hepatocellular carcinoma [35], and the immune-related response criteria (irRC) for melanoma immune modulator therapies [36].

In addition, the RECIST and IWG criteria themselves are dynamic systems that continue to evolve and incorporate new techniques. The most recent IWG criteria incorporate fluorodeoxyglucose positron emission tomography (FDG-PET) on a variable basis depending on the FDG avidity of the lymphoma subtype [18]. RECIST 1.1 also incorporates FDG-PET, albeit on a limited basis as an indicator of disease progression [15]. In general, despite much enthusiasm for the potential of metabolic imaging, response assessment guidelines have been slow to incorporate FDG-PET, especially for solid malignancies. This is due, at least in part, to the great challenge in optimizing and standardizing techniques to enable comparable results to be obtained by different vendors and institutions.

2.6 Emerging Techniques

The cancer imaging community is actively engaged in developing new oncologic imaging biomarkers based on advanced methods of tumor characterization. Development of new biomarker tools has proceeded across several different modalities, but in general all of the newer methods aim to interrogate for functional, molecular, or compositional changes that may report on tumor response earlier and/or with greater specificity than conventional methods. Emerging techniques providing candidate biomarkers include perfusion imaging (including dynamic contrast enhancement-MRI (DCE-MRI), perfusion CT, and microbubble contrast ultrasound techniques), diffusion imaging (including newer whole-body MR diffusion approaches), advanced imaging tools for molecular compositional analysis (including MR spectroscopy, magnetization transfer, and chemical exchange saturation transfer techniques), new elastography approaches (both MR and ultrasound based), and hybrid techniques facilitating registration of functional and anatomical information (including PET-CT and PET-MR). A brief introduction is presented here, but the reader is referred to other chapters in this volume for more detailed information on these techniques.

2.6.1 Imaging Methods Reporting on Vascular Status

When a malignant tumor reaches approximately 1–2 mm^3 in volume, it can no longer rely on the passive diffusion of metabolites from host tissue blood vessels in order to continue to proliferate, so new vasculature must develop in order for the tumor to continue to thrive [37, 38]. This process of neovascularization or angiogenesis is a signature of neoplasms and one of the principal potential targets for quantitative imaging [39]. In contrast to mature blood vessels that are the result of normal physiologic processes, tumor vessels produced by angiogenesis are characteristically leaky, fragile, and incompletely formed. It is believed that virtually all solid tumors are

dependent upon angiogenesis for survival [40] and many anti-angiogenic drugs are currently in clinical trials [41]. Thus, methods for imaging and quantitatively assessing this phenomenon are quite promising as biomarkers for application in preclinical and clinical studies.

Currently, one of the most widely employed methods for characterizing tumor neovasculature is DCE-MRI. (Other important methods for interrogating tumor vascularity include contrast-enhanced CT [42] and microbubble enhanced sonography [43].) Changes in the parameters obtained from DCE-MRI can be used to assess vascular changes within a tumor and, in particular, how a tumor is responding to treatment. The method is based on measurements and pharmacokinetic models of how a (typically) gadolinium based contrast agent perfuses through such vessels. Healthy vessels in normal tissues may be characterized by a range of parameters measuring blood flow, vessel permeability, and tissue volume fractions (i.e., fractions of a given sample of tissue that can be attributed to intravascular or extravascular space). These parameters are known to be different in vessels associated with tumors. Furthermore, as tumor blood vessels are known to change in response to anti-angiogenic drugs, the method provides a way of quantifying those changes. It is thus a plausible hypothesis that parameters measuring treatment induced changes in pathologic vessels will be predictive of response at an earlier time than changes in longest dimension. Indeed, many studies across a range of tumor types have shown just that.

2.6.2 Imaging Methods Reporting on Cell Density

Perhaps the most basic definition of cancer is that it is a set of diseases characterized by unregulated cell growth and proliferation. Furthermore, since many anticancer drugs have as their ultimate goal destruction of tumor cells, imaging methods sensitive to changes to tissue cellularity are of great importance. The typical application of histology and molecular biology in *living* systems is somewhat limited as it cannot provide a noninvasive deep tissue visualization of cells and molecules of interest. In particular, the visualization of cellular and molecular activity in animals normally requires the sacrifice and destruction of the organism to allow for analysis by histology and molecular biology. The development of cellular and molecular imaging techniques has begun to bridge this gap.

Two of the most promising methods for noninvasively probing tissue cellularity are diffusion-weighted magnetic resonance imaging (DW-MRI) to probe tissue cellularity and fluorodeoxythymidine-PET (FLT-PET) to image cell proliferation. DW-MRI exploits the microscopic thermally induced behavior of molecules moving in a random pattern, i.e., self-diffusion or Brownian motion. The rate of diffusion in cellular tissues is described by means of an apparent diffusion coefficient (ADC) that largely depends on the number and separation of barriers that a diffusing water molecule encounters. DW-MRI methods have been developed to map the ADC, and in well-controlled situations the variations in ADC have been shown to correlate inversely with tissue cellularity [44]. Many studies, both preclinical and clinical, have shown that exposure of tumors to chemotherapy leads to measurable increases in water diffusion in cases of favorable treatment response [45, 46].

FLT-PET exploits the increased uptake of thymidine in malignant tumors. Thymidine is a native nucleoside taken up by cells via surface nucleoside transporters and phosphorylated inside the cell by thymidine kinase 1 (TK1) into thymidine monophosphate. TK1 activity is upregulated during active DNA synthesis. FLT works in an analogous fashion: FLT is phosphorylated by TK1 into FLT monophosphate which is subsequently modified into FLT diphosphonate and triphosphonate by thymidylate kinase and diphosphate kinase, respectively. As FLT triphosphonate cannot be incorporated into the growing DNA chain, there will be an accumulation of FLT mono-, di-, and trisphosphate in such cells, thus

forming the basis of FLT-PET imaging of cell proliferation [47]. FLT-PET is thus a promising biomarker for cell proliferation in both preclinical and clinical studies.

2.6.3 Imaging Methods Reporting on Metabolic Events

The PET radiotracer most frequently used in clinical practice is fluorodeoxyglucose (FDG). As a glucose analogue, FDG is taken up by tumor cells via the GLUT1 and GLUT3 transporters and phosphorylated by hexokinase to FDG-6-phosphate. However, unlike glucose-6-phosphate, FDG-6-phosphate is not metabolized further in the glycolytic pathway and therefore remains trapped intracellularly because tumor cells do not have a significant amount of glucose-6-phosphatase to reverse this reaction. As the rate of glucose metabolism can differ significantly between healthy and malignant tissues, FDG can selectively accumulate in tumors, and quantification of this accumulation is a biomarker of tumor glucose metabolism. As discussed above, FDG-PET has been incorporated into the most recent response assessment guidelines for both solid and hematologic malignancies, but further incorporation into RECIST has been deferred pending development of further image acquisition and analysis standards. The PET community has taken bold steps in proposing a preliminary independent metabolic response standard, the PET Response Criteria in Solid Tumors (PERCIST) [48].

Magnetic resonance spectroscopy (MRS) can noninvasively provide data on the presence and relative concentrations of different metabolites in tumors [49]. MRS imaging (MRSI) extends this approach to provide spatial resolution of metabolite concentrations at the cost (typically) of increased scan time and reduced signal-to-noise ratios. These techniques have been employed for several decades to detect the altered metabolic signatures of cancer cells in both the diagnostic and prognostic settings. In particular, many malignancies demonstrate elevated levels of choline (due to increased membrane turnover in proliferating tumor cells) and lactate (due to increased anaerobic glycolysis).

2.7 Evaluation of Imaging Biomarkers as Surrogate Endpoints for Clinical Trials

As new imaging biomarkers advance to the point of possible adoption into clinical trials, investigators must contend with achieving the proper level of evaluation to assure that the biomarker is providing valuable information to its users. Just as different biomarkers might be appropriately deployed at different stages in the drug development process, there may be different requirements for biomarker evaluation depending on the intended use of the biomarker itself. In early-stage clinical studies, trial sponsors may determine for themselves whether an imaging biomarker is biologically relevant and has been sufficiently evaluated so as to be useful for internal decision making [6]. Conversely, if imaging biomarker data is to be collected for submission to regulatory authorities as part of a drug approval application, that biomarker data will be accepted as evidence of drug efficacy only if the biomarker has been previously evaluated and confirmed to be a valid or at least reasonably likely surrogate endpoint [21]. This section discusses the biomarker evaluation framework proposed by the IOM as it pertains to imaging biomarkers and surrogate endpoints for cancer clinical trials. We also briefly discuss the regulatory perspective.

2.7.1 The IOM Biomarker Evaluation Framework

The IOM framework establishes three discrete steps in the biomarker evaluation process: analytical validation, qualification, and utilization. *Analytical validation* involves demonstration that the biomarker can be reliably measured. For a biomarker to be analytically valid, its detection

and/or quantitative measurement must be accurate, reproducible across multiple clinical settings, and feasible over time [2]. Analytical validation of a biomarker includes generation of data on limits of detection, limits of quantification, and reference normal values.

A particular component of analytical validation that is underexplored, and critical to the acceptance of quantitative imaging biomarkers, is the repeatability and reproducibility of individual measures. Repeatability concerns variability in successive measurements (made by the same operator) and defines the difference between two scans that can be attributed to protocol and noise as opposed to true physiological changes. Reproducibility is defined as the degree of agreement between measurements made by different operators and is specified in several parameters including the 95 % confidence interval (CI) of the mean, which denotes the inter-user variability of the group mean parameter value. Measuring institutional repeatability and elaborating rigorous imaging protocols to ensure reproducibility are both important to inform comparisons between imaging sessions separated in time and are both crucial to enable calculation of the magnitude of observed effect required to conclude that a true biological change has occurred (e.g., in assessing the response of tumor to therapeutic intervention). Unfortunately, there is a fairly limited literature on the repeatability and reproducibility of quantitative imaging metrics. For example efforts in MRI or PET, the interested reader is referred to [50–54] and [55–57], respectively.

Qualification involves objective demonstration that the biomarker is associated with the clinical endpoint of concern. While the exact requirements for establishing this relationship have evolved over time and remain a subject of much debate and research, the current consensus is that qualification can be based on a "correlation approach": the biomarker should be prognostic for disease outcome in the absence of treatment, and the effect of intervention on the surrogate should be sufficiently correlated with the effect on the true endpoint [27]. With respect to the latter criterion, Prentice originally proposed that the biomarker must capture the full range of the treatment effect [58], but this requirement has been recognized as too strict to be practically useful and has been replaced with the criterion that the biomarker captures a substantial portion of the treatment effect, for example, more than 50 % [59]. Sargent et al. note that demonstrating correlation between a prospective biomarker and a clinical endpoint is not sufficient to qualify the biomarker, as such a correlation may be a result of prognostic factors influencing both the biomarker and clinical endpoint rather than the result of a similar treatment effect on both variables; rather, the true test for biomarker validity is whether it captures treatment effect at the trial level, as assessed by a meta-analysis of phase 3 trials in which both variables are measured [59].

A number of challenges exist for validation of imaging biomarkers using this statistical construct. First, even with a planned meta-analysis of several trials, it is difficult to obtain adequate power to show that a substantial portion of the treatment effect at the trial level is captured by the prospective biomarker [59]. Second, separate qualification of surrogate endpoints is required in the setting of different treatments; i.e., if a biomarker is qualified as a surrogate endpoint with respect to one treatment, it cannot be assumed that it is automatically qualified as a surrogate when evaluating a novel treatment with a different mechanism of action [27]. Third, this qualification approach assumes consensus on the appropriate clinical endpoint, which may not always be present. In particular, many observers have called attention to the difficulties in using overall survival as the primary clinical endpoint in solid tumors for which several lines of treatment may be available and have proposed instead that PFS may be a more appropriate endpoint for many phase 3 clinical trials [10].

Utilization involves assessment of biomarker performance in the specific context of its proposed use. From a pragmatic point of view, candidate biomarkers may be evaluated not solely on the basis of statistical qualification but also with respect to their biological plausibility and clinical usefulness [60]. Lassere has proposed

a formal schema for grading the surrogacy relationships between proposed biomarkers and clinical endpoints based on a weighted evaluation of biological, epidemiological, statistical, clinical trial, and risk-benefit evidence [61]. Even if an imaging biomarker is well correlated with clinical response, it may not demonstrate important drug side effects or toxicities [9]; the utilization component of the IOM framework therefore provides for a holistic assessment of a biomarker's usefulness for decision making.

2.7.2 The Regulatory Perspective

Regulatory considerations have had profound influence on the incorporation of imaging biomarkers into oncology clinical trials, especially during late-phase drug development. This section provides a brief history of how the regulatory perspective on imaging biomarkers has evolved, with a focus on the U.S. FDA. A detailed review of requirements from different agencies is beyond the scope of this chapter.

Two different routes are available for FDA approval of a new therapeutic agent: regular approval and accelerated approval. In the 1970s, before the creation of accelerated approval, the FDA commonly granted regular approval for cancer drugs based on ORR as determined either by imaging or by physical examination measurements. During the mid-1980s, however, the FDA determined that regular approval for cancer drugs should require more direct evidence of clinical benefit, particularly improved survival, improved quality of life, or improvement in an established surrogate for at least one of these [21]. Over the next decade, several endpoints were established as acceptable surrogates for clinical benefit, including improved disease-free survival (DFS) in selected adjuvant settings, durable CR in leukemia, and a high substantiated ORR in select solid tumors, provided that ORR data are considered alongside response duration, drug toxicity, and relief of tumor-related symptoms. It should be emphasized, therefore, that the requirement for established surrogacy of nonclinical endpoints did not eliminate drug approvals based on imaging biomarker data; indeed, improved ORR in conjunction with improvement in symptoms and adequate response duration has continued to support regular approval in several clinical settings [62].

The accelerated approval route was created in 1992, partially in response to public demand for quicker approval of new anticancer drugs. Accelerated approval allows for the consideration of surrogate endpoints that are "reasonably likely" to predict clinical benefit, a lower standard than that required for regular approval. A drug is approved under the accelerated approval regulations on the condition that the manufacturer conducts postmarketing studies to verify and describe the actual clinical benefit. If postmarketing studies fail to demonstrate clinical benefit, the drug may be removed from market under an expedited process. ORR has been the most commonly used nonclinical endpoint in support of accelerated approval [62].

A review of FDA cancer drug approvals between 1990 and 2002 showed that out of 71 total approvals, 57 were regular approvals and 14 were accelerated approvals. Thirty-nine out of the 57 regular approvals (68 %) were based on endpoints other than survival, mostly ORR but also DFS and TTP, occasionally but not always supplemented by evidence of relief of tumor-based symptoms. All 14 of the accelerated approvals were based on surrogate endpoints, again mostly ORR but also DFS and TTP [21].

Over the past decade, both the FDA and the European Medicines Agency (EMA) have hosted several workshops on biomarkers and surrogate endpoints in new drug development [63, 64]. The FDA has also established a public-private Biomarkers Consortium, including representation from the FDA, NIH, and pharmaceutical industry, seeking to identify and qualify new and existing biomarkers [65]. Both the FDA and EMA routinely issue guidance documents to industry on biomarker-related topics including methods of demonstrating drug efficacy [66], appropriate endpoints for cancer clinical trials [62], and considerations for adaptive design clinical trials [67].

Conclusion

Imaging biomarkers hold great potential for streamlining drug development and optimizing clinical care, but there is much to be done in order to develop, validate, and standardize emerging methods before they can be accepted as surrogate endpoints in multicenter trials. Ideally, biomarker development would be closely coupled with development of candidate compounds [11], but add-on costs have been a significant barrier, especially in clinical trials [68]. Part of the challenge has now been taken up by multiple government-industry partnerships, including the Quantitative Imaging Biomarker Alliance (QIBA), organized by the Radiological Society of North America (RSNA); the NCI's Quantitative Imaging Network (QIN); and the FDA and EMA biomarker consortia mentioned above. In the USA, the recent merging of the Eastern Cooperative Oncology Group (ECOG) with the American College of Radiology Imaging Network (ACRIN) provides an additional promising venue for advancing imaging biomarkers in oncology.

While traditional size-based imaging biomarkers will likely remain the dominant non-survival endpoints for the foreseeable future, major efforts are now being directed toward pushing response assessment beyond anatomical and morphological imaging to more fundamental metrics at the physiological, cellular, and molecular levels. These concepts will be developed more fully in subsequent chapters.

References

1. Institute of Medicine. A national cancer clinical trials system for the 21st century: reinvigorating the NCI cooperative group program. Washington, DC: National Academies Press; 2010.
2. Institute of Medicine. Evaluation of biomarkers and surrogate endpoints in chronic disease. Washington, DC: National Academies Press; 2010.
3. Hricak H. Oncologic imaging: a guiding hand of personalized cancer care. Radiology. 2011;259(3):633–40.
4. Group BDW. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther. 2001;69(3):89–95.
5. Wolff AC, et al. Research issues affecting preoperative systemic therapy for operable breast cancer. J Clin Oncol. 2008;26(5):806–13.
6. Richter WS. Imaging biomarkers as surrogate endpoints for drug development. Eur J Nucl Med Mol Imaging. 2006;33:S6–S10.
7. Varmus H. Ten years on–the human genome and medicine. N Engl J Med. 2010;362(21):2028–9.
8. Morgan S, et al. The cost of drug development: a systematic review. Health Policy. 2011;100(1):4–17.
9. Smith JJ, et al. Biomarkers in imaging: realizing radiology's future. Radiology. 2003;227(3):633–8.
10. Sargent DJ, Hayes DF. Assessing the measure of a new drug: is survival the only thing that matters? J Clin Oncol. 2008;26(12):1922–3.
11. Park JW, et al. Rationale for biomarkers and surrogate end points in mechanism-driven oncology drug development. Clin Cancer Res. 2004;10(11):3885–96.
12. Tan DS, et al. Biomarker-driven early clinical trials in oncology: a paradigm shift in drug development. Cancer J. 2009;15(5):406–20.
13. Miller AB, et al. Reporting results of cancer treatment. Cancer. 1981;47(1):207–14.
14. Therasse P, 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(3):205–16.
15. Eisenhauer EA, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur J Cancer. 2009;45(2):228–47.
16. Macdonald DR, et al. Response criteria for phase II studies of supratentorial malignant glioma. J Clin Oncol. 1990;8(7):1277–80.
17. Cheson BD, et al. Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group. J Clin Oncol. 1999;17(4):1244.
18. Cheson BD, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25(5):579–86.
19. Kay A, et al. Randomized controlled trials in the era of molecular oncology: methodology, biomarkers, and end points. Ann Oncol. 2012;23(6):1646–51.
20. Seymour L, et al. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. Clin Cancer Res. 2010;16(6):1764–9.
21. Johnson JR, et al. End points and United States Food and Drug Administration approval of oncology drugs. J Clin Oncol. 2003;21(7):1404–11.
22. Buyse M, et al. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Meta-Analysis Group in Cancer. Lancet. 2000;356(9227):373–8.

23. Paesmans M, et al. Response to chemotherapy has predictive value for further survival of patients with advanced non-small cell lung cancer: 10 years experience of the European Lung Cancer Working Party. Eur J Cancer. 1997;33(14):2326–32.
24. Sekine I, et al. Relationship between objective responses in phase I trials and potential efficacy of non-specific cytotoxic investigational new drugs. Ann Oncol. 2002;13(8):1300–6.
25. Goffin J, et al. Objective responses in patients with malignant melanoma or renal cell cancer in early clinical studies do not predict regulatory approval. Clin Cancer Res. 2005;11(16):5928–34.
26. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125(7):605–13.
27. Buyse M, et al. Biomarkers and surrogate end points–the challenge of statistical validation. Nat Rev Clin Oncol. 2010;7(6):309–17.
28. Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? J Clin Oncol. 2012;30(10):1030–3.
29. Sridhara R, et al. Review of oncology and hematology drug product approvals at the US Food and Drug Administration between July 2005 and December 2007. J Natl Cancer Inst. 2010;102(4):230–43.
30. Ratain MJ, Eckhardt SG. Phase II studies of modern drugs directed against new targets: if you are fazed, too, then resist RECIST. J Clin Oncol. 2004; 22(22):4442–5.
31. Tuma RS. Sometimes size doesn't matter: reevaluating RECIST and tumor response rate endpoints. J Natl Cancer Inst. 2006;98(18):1272–4.
32. Erasmus JJ, et al. Interobserver and intraobserver variability in measurement of non-small-cell carcinoma lung lesions: implications for assessment of tumor response. J Clin Oncol. 2003;21(13):2574–82.
33. Zhao B, et al. A pilot study of volume measurement as a method of tumor response evaluation to aid biomarker development. Clin Cancer Res. 2010;16(18): 4647–53.
34. Choi H, et al. Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. J Clin Oncol. 2007;25(13):1753–9.
35. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Semin Liver Dis. 2010;30(1):52–60.
36. Wolchok JD, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. Clin Cancer Res. 2009;15(23):7412–20.
37. Folkman J. Role of angiogenesis in tumor growth and metastasis. Semin Oncol. 2002;29(6 Suppl 16):15–8.
38. Ribatti D, et al. The discovery of angiogenic factors: a historical review. Gen Pharmacol. 2000;35(5): 227–31.
39. Atri M. New technologies and directed agents for applications of cancer imaging. J Clin Oncol. 2006;24(20):3299–308.
40. Carmeliet P, Jain RK. Angiogenesis in cancer and other diseases. Nature. 2000;407(6801):249–57.
41. Verhoef C, et al. Angiogenesis inhibitors: perspectives for medical, surgical and radiation oncology. Curr Pharm Des. 2006;12(21):2623–30.
42. Koh TS, et al. Primary colorectal cancer: use of kinetic modeling of dynamic contrast-enhanced CT data to predict clinical outcome. Radiology. 2013;267(1):145–54.
43. Fleischer AC, et al. Sonographic depiction of microvessel perfusion: principles and potential. J Ultrasound Med. 2004;23(11):1499–506.
44. Anderson AW, et al. Effects of cell volume fraction changes on apparent diffusion in human cells. Mag Reson Imaging. 2000;18(6):689–95.
45. Galons JP, et al. Early increases in breast tumor xenograft water mobility in response to paclitaxel therapy detected by non-invasive diffusion magnetic resonance imaging. Neoplasia. 1999;1(2):113–7.
46. Lee KC, et al. Prospective early response imaging biomarker for neoadjuvant breast cancer chemotherapy. Clin Cancer Res. 2007;13(2 Pt 1):443–50.
47. Soloviev D, et al. [(18)F]FLT: an imaging biomarker of tumour proliferation for assessment of tumour response to treatment. Eur J Cancer. 2012;48(4): 416–24.
48. Wahl RL, et al. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med. 2009;50 Suppl 1:122S–50S.
49. Griffiths JR, Glickson JD. Monitoring pharmacokinetics of anticancer drugs: non-invasive investigation using magnetic resonance spectroscopy. Adv Drug Deliv Rev. 2000;41(1):75–89.
50. Padhani AR, et al. Reproducibility of quantitative dynamic MRI of normal human tissues. NMR Biomed. 2002;15(2):143–53.
51. Yankeelov TE, et al. Repeatability of a reference region model for analysis of murine DCE-MRI data at 7T. J Magn Reson Imaging. 2006;24(5):1140–7.
52. Barnes SL, et al. Assessing the reproducibility of dynamic contrast enhanced magnetic resonance imaging in a murine model of breast cancer. Magn Reson Med. 2013;69(6):1721–34.
53. Dula AN, et al. Amide proton transfer imaging of the breast at 3 T: establishing reproducibility and possible feasibility assessing chemotherapy response. Magn Reson Med. 2013;70(1):216–24.
54. Alonzi R, et al. Reproducibility and correlation between quantitative and semiquantitative dynamic and intrinsic susceptibility-weighted MRI parameters in the benign and malignant human prostate. J Magn Reson Imaging. 2010;32(1):155–64.
55. Whisenant JG, et al. Reproducibility of static and dynamic (18)F-FDG, (18)F-FLT, and (18)F-FMISO MicroPET studies in a murine model of HER2+ breast cancer. Mol Imaging Biol. 2013;15(1):87–96.

56. Tseng JR, et al. Reproducibility of 3′-deoxy-3′-(18) F-fluorothymidine microPET studies in tumor xenografts in mice. J Nucl Med. 2005;46(11):1851–7.
57. Dandekar M, et al. Reproducibility of 18F-FDG microPET studies in mouse tumor xenografts. J Nucl Med. 2007;48(4):602–7.
58. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Stat Med. 1989; 8(4):431–40.
59. Sargent DJ, et al. Validation of novel imaging methodologies for use as cancer clinical trial end-points. Eur J Cancer. 2009;45(2):290–9.
60. Green E, et al. Surrogate endpoint validation: statistical elegance versus clinical relevance. Stat Methods Med Res. 2008;17(5):477–86.
61. Lassere MN. The biomarker-surrogacy evaluation schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. Stat Methods Med Res. 2008; 17(3):303–40.
62. U.S. Food and Drug Administration. Guidance for industry: clinical trial endpoints for the approval of cancer drugs and biologics. Rockville, 2007.
63. Dis Markers. NIH-FDA conference: biomarkers and surrogate endpoints: advancing clinical research and applications. Abstracts. Dis Markers. 1998;14(4): 187–334.
64. European Medicines Agency. 2006 EMEA/EFPIA Workshop on Biomarkers. Available at: http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/events/2009/11/event_detail_000077.jsp&mid=WC0b01ac058004d5c3. Accessed 23 Aug 2013.
65. The biomarkers consortium. Available at: http://www.biomarkersconsortium.org/. Accessed 23 Aug 2013.
66. U.S. Food and Drug Administration. Guidance for industry: providing clinical evidence of effectiveness for human drug and biological products. Rockville, 1998.
67. U.S. Food and Drug Administration. Guidance for industry: adaptive design clinical trials for drugs and biologics (draft guidance). Rockville, 2010.
68. Woodcock J. A framework for biomarker and surrogate endpoint use in drug development. Available at: www.fda.gov. Accessed 23 Aug 2013.