Challenges in clinical studies with multiple imaging probes ☆

Kenneth A. Krohn*, Finbarr O'Sullivan, John Crowley, Janet F. Eary, Hannah M. Linden,
Jeanne M. Link, David A. Mankoff, Mark Muzi, Joseph G. Rajendran,
Alexander M. Spence, Kristin R. Swanson

Radiology Department, Molecular Imaging Center, University of Washington, Seattle, WA 98195-6004, USA

Abstract

This article addresses two related issues: (a) When a new imaging agent is proposed, how does the imager integrate it with other biomarkers, either sampled or imaged? (b) When we have multiple imaging agents, is the information additive or duplicative and how is this objectively determined? Molecular biology is leading to new treatment options with reduced normal tissue toxicity, and imaging should have a role in objectively evaluating new treatments. There are two roles for molecular characterization of disease. Molecular imaging measurements before therapy help predict the aggressiveness of disease and identify therapeutic targets and, therefore, help choose the optimal therapy for an individual. Measurements of specific biochemical processes made during or after therapy should be sensitive measures of tumor response. The rules of evidence are not fully developed for the prognostic role of imaging biomarkers, but the potential of molecular imaging provides compelling motivation to push forward with convincing validation studies. New imaging procedures need to be characterized for their effectiveness under realistic clinical conditions to improve the management of patients and achieve a better outcome. The purpose of this article is to promote a critical discussion within the molecular imaging community because our future value to the overall biomedical community will be in supporting better treatment outcomes rather than in detection.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Multiple imaging; Biomarkers; Molecular biology

1. Introduction

Our title suggests a broad topic with multiple components, each presenting challenging statistical and analysis issues:

- When a new imaging agent is proposed, how does the imager integrate it with other biomarkers, either sampled or imaged?
- When we have multiple imaging agents, are the measurements duplicative or additive and how is this objectively determined?
- When multiple tracers are used in a single imaging session, how can the distribution pharmacokinetics of several tracers be analyzed to take advantage of common features of each tracer and most accurately

reflect the differences between tracers? This is an important question and a larger topic than can be discussed here.

When a new therapy is proposed, how does imaging help the clinician select those patients most likely to benefit from the new treatment strategy? It is no longer sufficient to show that a diagnostic procedure depicts function with some level of specificity and sensitivity. New imaging procedures need to be characterized for their effectiveness under realistic clinical conditions to improve the management of patients and achieve a better outcome [1]. Publications evaluating new imaging procedures need to address not only their diagnostic impact but also their impact on therapy and patient management and describe their contribution toward improving patient outcomes. We are obligated to show that patients are better off as a result of any new imaging procedure.

Molecular pathology provides detailed genomic and proteomic information from single biopsy specimens and other tissue samples. These approaches are leading to a better understanding of biomarkers that are being incorporated into

clinical treatment decisions and outcome measures [2]. Rigorous procedures are needed to identify biomarkers that correlate with a disease and its extent and severity and then to verify that a given biomarker changes in response to an effective treatment. Changes should be quantified to the highest degree possible. When considered together with imaging biomarkers, quantitative measures of protein expression and function should be more predictive than genomic assays.

Our understanding of genetic instability and how altered gene expression and regulation lead to a disease phenotype is incomplete. Nevertheless, advances in molecular biology are already leading to proposed new molecular treatment options for patients, with greatly reduced normal tissue toxicity [3]. Molecular imaging should be a valuable partner in objectively evaluating these new treatments. It is also becoming an important tool for selecting the treatment strategy for an individual because it characterizes the phenotypic expression resulting from all of the molecular alterations in a disease. Thus, molecular imaging must keep pace with new knowledge in molecular biology by developing and validating methods that are much more specific than FDG positron emission tomography (PET).

Imaging can help the clinician by quantifying specific regional molecular changes of disease. Both PET and molecular pathology provide important information about an individual's disease. Molecular imaging has the capacity to interrogate the whole body, but a different imaging agent is required for each unique aspect of the tumor phenotype. Thus, it is evident that molecular imaging and molecular pathology are complementary tools.

The recent explosion in knowledge about molecular biology has not yielded a comparable improvement in the treatment outcome for patients. There are a few notable exceptions, such as imatinib in CML or gastrointestinal stromal tumors expressing c-kit and trastuzumab and lapatinib in HER2/neu-positive breast cancer, but the successes have been sparse [4]. Clearly, the discovery of molecular pathways is growing faster than our understanding of their regulation and interactions or their potential as therapeutic targets. The dynamic balance of cell regulation in the complex tumor microenvironment is poorly observed in culture by blotting techniques or RT-PCR. Furthermore, it is becoming increasingly clear that animal models of human disease are often inadequate for predicting the human study. Thus, clinicians need the best imaging possible to characterize the in vivo tumor phenotype in patients in the context of a given tumor's microenvironment during the whole course of treatment.

There are two promising roles for molecular characterization of disease: (a) more rational selection of therapy and (b) early assessment of therapeutic response. Molecular imaging measurements before therapy are expected to predict the aggressiveness of disease and identify therapeutic targets [3]. This improves choosing the optimal therapy for an individual. Measurements of specific biochemical processes made during or after therapy should be more sensitive measures of tumor response than the conventional approach of imaging structural changes by RECIST criteria [5]. When a particular therapy is effective, as determined by a change in a critical tumor function, it can be continued with confidence. Perhaps more importantly, when the therapy fails to impair a critical aspect of tumor function, it is unlikely to be effective but may still be damaging normal tissues. Molecular imaging should be able to identify reliably when to stop one treatment and implement an alternative. Imaging changes will not be subtle if the imaging probe is sufficiently specific for the appropriate function. Useful molecular imaging procedures must have more than a small incremental impact on outcomes.

Molecular pathways associated with disease are often complex, with cross paths providing many ways to express a phenotype. The pathways get even more complicated with each month's journals. Cancer cells, by definition, have dysregulated growth and/or death pathways and thus provide a challenge for developing molecularly targeted therapeutics. The success of imatinib is related to the fact that a single mutation in c-kit sustains cancer growth [6]. Other molecularly targeted drugs, for example, the EGFR inhibitors, have been much less successful as single agents, likely because other pathways with other distinct targets provide redundant growth signaling to sustain the cancer phenotype [3]. Recent drug development is focusing on "dirty" tyrosine kinase inhibitors, ones that cross-react with many structurally related receptor tyrosine kinases [7]. This may suggest that chemists developing new molecular imaging agents should focus on molecules that image a general aspect of the tumor phenotype, something more specific than FDG but less specific than a radioligand for mutant vIII of the epidermal growth factor receptor. For example, Hanahan and Weinberg [8] have distilled the complexity of cancer signaling pathways down to six attributes that they called hallmarks of cancer:

1. Self-sufficiency in growth signals
2. Insensitivity to antigrowth signals
3. Tissue invasion and metastasis
4. Limitless replicative potential
5. Sustained angiogenesis
6. Evading apoptosis

Each of these could be an important imaging goal. The challenge for the imaging community is to develop radiopharmaceuticals with a useful level of specificity without making the imaging agent so specialized that it would never be studied with enough patients to be validated, say, nothing of becoming commercially viable.

2. Prognosis and prediction: more than diagnostic accuracy

Molecular imaging is expected to play a role that is more demanding than diagnostic screening, staging or using

imaging evidence to make probability statements about the etiology or characteristics of a disease. In general, our patients will have known disease that is evaluated for specific characteristics that help select the most appropriate therapy from multiple treatment options. Imaging scientists must provide compelling supporting evidence that understanding important molecular differences and how they change during treatment will ultimately lead to better monitoring of treatments and, hence, better outcomes [3].

Our vision has always been to integrate PET imaging with treatment planning [9]. This has several implications for an overall imaging research strategy. Instead of a general PET imaging agent to detect disease, for example, FDG to detect cancer, the molecular imaging community is developing more specific imaging agents for biological characterization of disease. Models for the value of a new imaging procedure must include the imaging measurement, along with other prognostic factors, and a prognosis-specific action [10–12]. This approach, when optimized and validated, will be better than treating all patients within a particular disease histology and stage in the same statistically derived way. If we chose an appropriate imaging agent that reflects the biological mechanism that was targeted by the treatment being evaluated, then we should detect major changes.

Statistical methods for evaluating factors that predict the course of disease for groups of patients defined by several prognostic factors and for ranking the relative importance of these factors are not widely practiced or well developed. This is in contrast with therapeutic trials or epidemiology, where statistical principles and methods are well developed, generally accepted and rigorously enforced by journal editors. The only way to learn who should be treated and how, the holy grail of molecular medicine in support of tailored therapy, is to generate a clinical trial that includes both the imaging test and a prognostic model that specifies the treatment. In many cases, both the test and the treatment will be experimental and may be developed in parallel. This implies a detailed protocol and validation analysis to provide evidence that imaging in a prognostic model improves the outcome of patients by altered medical practice. This represents the long-range goal of molecular imaging. It is clearly beyond the general practice of evaluating a new diagnostic test with respect to reproducibility and positive and negative predictive value.

The process of validating an imaging study's predictive capability starts with observational studies in established treatment protocols for testing imaging variables and how they predict response over the course of specific treatments. Useful radiopharmaceuticals should add substantially to conventional imaging. Showing that the imaging procedure under study is predictive of patient outcomes should not be only a simple univariate test comparing the imaging result to clinical response. Analysis must also include a proper multivariate analysis considering simultaneously the influence of various potential prognostic factors on an appropriate outcome variable [10–13]. These observational studies need

to be carried out in a way that mimics the careful design standards used in clinical trials. The primary goal is to identify imaging procedures where a clinically important improvement in predictive accuracy, well beyond that attained with existing clinical and diagnostic methods, can be achieved. The improved prediction will have clear therapeutic implications, which can be generalized in larger cooperative trials.

Early observational trials are particularly challenging to design and analyze because there is no opportunity for randomization, which is an important mechanism to protect the investigator against bias. In addition, they may involve retrospective analysis. Thus, these pilot trials need to be interpreted cautiously, but they are critical to designing a rigorous prospective trial with clearly stated primary and secondary hypotheses. However, it is important to appreciate that any randomization will be at the level of intent to treat and that all subjects should be imaged at this level without regard to the treatment arm into which they fall.

2.1. Some examples

These general principles are best appreciated by considering several examples taken from our studies in cancer imaging. The first examples show situations where imaging studies with two different radiopharmaceuticals provided much more information than one study alone. The latter examples tested the added predictive power of FMISO and FDG in head and neck cancer and of FES plus HER2/neu overexpression in breast cancer. In the last case, an imaging study was combined with a sampled biomarker assay.

The first example led to a mechanistic conclusion. The regional metabolic rate of glucose is most commonly studied with $[^{18}F]$-FDG, but it can also be imaged with 1-$[^{11}C]$-glucose. In subjects where both studies were done in one imaging session, the ratio of the metabolic rate measured with ^{18}F versus the ^{11}C rate was analyzed as an imaging measure of the regional lumped constant [14]. The results of the ratio image showed that this cancer involved grossly altered enzymology. The ratio image was uniform in normal brain but greatly increased in a glioblastoma. This is an example where a statistical test was not needed to provide convincing evidence of the value of the combined study for revealing a fundamental fact about the biology of a specific tumor.

A second set of examples shows that imaging protocols that compare local delivery with regional metabolism are more useful than a single study of either parameter. In heart imaging, this concept has been used to assess myocardial viability by comparing flow images, often with $[^{13}N]$-ammonia, and metabolism images with FDG [15]. Here, we present a similar example from Mankoff's group of a study of blood flow measured with $[^{15}O]$-water and a single-compartment model paired with dynamic FDG PET in 35 patients with locally advanced breast cancer. In these studies, an imbalance between pretherapy tumor glucose metabolism and blood flow, measured by dual-tracer PET studies, predicted poorer disease-free survival (DFS) [16]. Blood

flow and the FDG transport parameter, K_1, were correlated before chemotherapy ($P<.001$), but neither the phosphorylation rate, k_3, nor the net FDG flux was well correlated with blood flow, $P=.52$ and $P=.05$, respectively. Blood flow and FDG flux were more closely matched after chemotherapy, $P<.001$, mostly because of changes in k_3. The ratio of FDG flux to transport after chemotherapy was predictive of ultimate response, with low ratios associated with a favorable outcome. This example illustrates how dual-tracer PET studies can identify changes in pathway kinetics with therapy that indicate changes in tumor enzymology and are predictive of patient outcome.

We anticipate numerous other examples where a comparison between two images from two radiopharmaceuticals or at two different times or even between two rate parameters from a single radiopharmaceutical will provide more prognostic information that comes from a single snapshot image. For example, a comparison of a proliferation image and an apoptosis image should be more predictive of a successful therapy than either image alone because some treatments stop DNA synthesis while others greatly accelerate apoptosis. The ratio image provides an amplified signal of a therapeutic effect that halts the uncontrolled growth of tumor. If these two images are obtained before and soon after a treatment regimen, the added value should be even greater [17]. In another example, an imaging biomarker study, FDG PET, combined with a serum biomarker, thyroglobulin, identifies those thyroid cancers most likely to cause death and thereby directs more aggressive treatment [18].

2.2. Testing for added value of a new imaging procedure

Rajendran's group in our program has been doing FMISO PET imaging in patients with head and neck cancer and recently reported an evaluation of 73 patients imaged pretherapy, many of whom also underwent FDG PET imaging [19]. Significant hypoxia was seen in 58 patients, and there were 28 deaths in the follow-up period. In univariate analyses, FMISO measurements as either maximum tumor/blood ratio (T/B_{max}) or hypoxic volume (HV) and the presence of nodes were strong predictors. In a multivariate analysis including FDG SUV_{max} (limiting the analysis to 53 subjects with both imaging studies), no variable was predictive at $P<.05$. However, in the larger group without complete FDG data, nodal status and T/B_{max} were both highly predictive, $P=.02$ and $P=.006$, respectively. This pilot study provided critical data for planning further prospective studies with FMISO PET in head and neck cancer as well as cervical cancer.

The next level of analysis tests the independent predictive capability of a new procedure, for example, FMISO PET, as compared to established or putative predictive factors. In the example we cite in Section 2.3, we use FDG PET as an example of a test with predictive value for many cancers. FDG PET has particular importance to cancer physicians and PET imagers, given its widespread use in clinical practice. The FDG SUV parameter may be used to integrate prior prognostic information, including imaging, into a single variable, and the FMISO PET parameter could become a new variable to test for added predictive value. The relationship between the clinical response [time to progression or overall survival (OS)] and all of the measured prognostic factors is most conveniently evaluated using a standard statistical analysis based on the Cox proportional hazards model [20]. This analysis permits an examination of the influence on outcome of a new molecular imaging measurement while controlling for the impact of histology, prognostic group and treatment procedures, as well as other potentially relevant patient information such as age and gender. The Cox model allows us to assess the percentage change in the risk of an event associated with increasing the new molecular imaging parameter value by one unit, while keeping all other variables fixed.

When initially evaluating an imaging protocol in patients, we are ethically constrained to an observational study. The patient's clinical management must not be tainted by the new imaging measurement. Nevertheless, some meaningful results may come from the early studies. For example, Linden et al. [21] analyzed retrospectively the data gathered from several RDRC protocols that were each designed to develop basic information about the pharmacokinetics of [^{18}F]-fluoroestradiol (FES) in patients with breast cancer. One of these protocols was developing biodistribution kinetic results to calculate dosimetry, another was measuring the heterogeneity of FES uptake in subjects with advanced disease, another was comparing FES uptake with in vitro measurements of the estrogen receptor (ER) level by IHC and yet another was measuring differences in FES kinetics in patients who were undergoing endocrine therapy without cytotoxic chemotherapy. The patient's treatment was not altered by results from the FES PET study. From this very diverse group of patients, 47 could be identified with recurrent or metastatic breast cancer with ER-positive tumors and with known progesterone receptor and HER2/neu status. None of the patients had received tamoxifen within 2 months before FES, and the type of endocrine therapy was at the discretion of the treating oncologist. The authors found a response rate of 23% to salvage hormonal therapy in this heavily pretreated group, similar to other trials with aromatase inhibitors and other endocrine agents. In spite of a highly heterogeneous population in a retrospective analysis, the use of quantitative FES PET to select patients for endocrine therapy could have increased this response rate to as high as 40% without excluding any patients who had a response [21]. These findings need to be confirmed in prospective trials with predefined quantitative thresholds, but the data gleaned from this initial analysis were useful. However, this report emphasizes many of the practical limitations associated with pilot clinical trials of new imaging agents.

Comparison of FES PET results in this study to the presence or absence of HER2 overexpression by tissue assay illustrates another point about combining imaging and tissue biomarkers. The retrospective analysis revealed that no

patients whose tumor overexpressed HER2/neu responded to endocrine treatment although there was no correlation between FES uptake and HER2/neu expression. This suggested that HER2 provided a growth signal independent of the ER, which sustained cancer growth even when the ER pathway had been interrupted by treatment. Exclusion of patients with both overexpression of HER2/neu and a positive FES PET would have increased the rate of response to salvage therapy to as high as 55% [21]. In this case, the combination of the tissue biomarker and imaging biomarker would have improved the overall predictive capability. Future FES PET imaging trials will need to control for HER2/neu and use uniform selection criteria and treatment regimens to fully evaluate the predictive value of this imaging protocol.

2.3. The challenge of sample size calculations

One of the more challenging problems in developing a research trial is to determine how many observations or measurements will be needed to arrive at a conclusion. When experimental imaging is being evaluated as an observational study added to a therapeutic protocol, several factors contribute to variability and thus to the required sample size [20]. Among these are the prevalence of the cancer subtypes being imaged, the number of prognostic factors that will be evaluated and the statistical algorithm for evaluating the added value of a new prognostic variable. For these reasons, conventional power calculations may not be as helpful in designing imaging trials as they have been for testing hypotheses in other studies with matched cases and controls. The assessment of biomarkers by conventional epidemiologic methods is costly and time consuming. We need to use mathematical simulations to predict the effect of new biomarker imaging technologies on overall outcomes [2,22,23], as can be demonstrated from our prospective trial for FMISO PET in cervical cancer.

Rajendran is currently evaluating the predictive value of FMISO PET for imaging hypoxia. This study tests the hypothesis that there is a relationship between the risk of disease progression at time t and the pretherapy FMISO measure of the maximum tumor-to-blood ratio, T/B_{max}. Based on previous experience with several cancers, pretherapy FDG (maximum tumor SUV) is used as a control for confounding by all other prognostic variables, including TNM staging. Letting X represent the FMISO measurement and Z be the clinical FDG SUV, the Cox proportional hazards model specifies that the logarithm of the hazard for progression at time t ($\lambda(t|Z,X)$) relative to the baseline hazard ($\lambda_0(t)$) is a linear combination of effects explained by the FMISO and FDG results; that is,

$$\log\left[\frac{\lambda(t|Z,X)}{\lambda_0(t)}\right] = \beta X_{FMISO} + \gamma Z_{FDG}.$$

This approach will estimate the joint effects of FMISO (β) and FDG (γ), and then a z test can be applied to the estimated

FMISO coefficient for assessment of the null hypothesis, $\beta=0$. Note that if $\beta=0$, then, after adjustment for FDG, pre-FMISO is unrelated to outcome. A one-sided test with a .05 level is used because we anticipate $\beta>0$, reflecting our understanding of how the PET FMISO measurement (X) relates to outcome; more hypoxia can only be worse for the patient. For this sample calculation, we have taken values of β and γ from analysis of FMISO and FDG imaging in a pilot series of 18 patients with cervical cancer. With X and Z standardized, we found β to be in the 0.40–0.60 range for a simulated study enrolling 60 patients, similar to what we have reported in a large series of patients with head and neck cancer [19]. A value of $\beta=0.50$ implies that median survival for patients whose FMISO measurement is one standard deviation above the mean FMISO value will be 65% shorter than that for patients whose FMISO measure is at the mean level. Smaller effects have been found with FDG in our cervical cancer series; however, our data show substantial variability. We used $\gamma=0.40$, matching the current result from our patients with head and neck cancer. We assume uniform accrual of study patients and an underlying exponential failure pattern. The median DFS for patients in our study population, mostly consisting of Stage II and III disease, is expected to be 3.0 years, and the median time to death is 6 years. We project analysis after 2 and 7 years of patient follow-up. Power calculations obtained by simulation for 60 subjects have been obtained with and without adjustment for confounding (i.e., inclusion of FDG information). We also assessed the effect of 15% measurement error in PET parameters. Table 1 presents results for two different outcome parameters: DFS, in the top two rows, and OS, in the bottom two rows.

Power increases with follow-up. For the DFS outcome, the study has 90% power with 2 years of follow-up and this rises to 95% with 5 additional years of follow-up. For OS, the power for the study is 80% and increases to 91% with more extensive follow-up. Even with the assumption of 15% errors in PET measurements (which is excessive) and provided that we adjust for confounding using FDG, 86% power is achieved with adequate follow-up. This calculation underlines the importance of the FDG PET study, without which there is a loss of power.

Table 1
Power analysis for the relationship between pretherapy PET (FMISO and FDG) and patient outcome

n	b	$\exp(b)-1$	FU	P	P_0	P_M
60	0.5	0.65	2 years	0.90	0.87	0.85
			7 years	0.95	0.92	0.91
60	0.5	0.65	2 years	0.80	0.76	0.73
			7 years	0.91	0.88	0.86

FU, length of follow-up; P, power with adjustment for PET confounder; P_0, power obtained if confounder is not incorporated; P_M, effect of 15% measurement error on covariates. Median DFS (first two rows) is 3 years and median OS (last two rows) is assumed to be 6 years.

3. Assessing response in individual patients: beyond RECIST with the help of molecular imaging and mathematical modeling

Swanson's group in our program is developing mathematical models of brain tumor growth over time and calibrating these models with either [^{11}C]-thymidine or [^{18}F]-FLT PET and MRI in longitudinal studies [24]. One of the tools revealed by this integrative modeling approach has been the development of model-defined "virtual controls" for each patient against which treatment effect can be measured in individual patients. Specifically, Swanson has found that when accounting for hallmarks of biological aggressiveness, such as the net proliferation of glioma cells and their net invasion of the normal appearing surrounding parenchyma, individual gliomas behave surprisingly predictably but are widely variable across histological grade. Specifically, the mathematical model suggests that, if untreated, the mean radius of the imageable portion of the tumor (which misses the diffusely invasive component of the tumor peripheral to the imaging abnormality) tends to grow linearly in time (with a constant velocity). She has validated this model-predicted behavior in hundreds of gliomas followed serially prior to any treatment [24]. This study revealed wide ranges of biological aggressiveness, quantified by these mathematical model terms, net rates of proliferation and migration, even within a single histological grade.

Clearly, as molecularly targeted therapies become progressively individualized, our means for assessment of individuals must accommodate phenotypic heterogeneity so as to differentiate the signal (the treatment effect) from the noise (the heterogeneity among patients). RECIST, the standard approach for assessing responsiveness in clinical studies [3,5], ignores the natural history and future growth curve of each patient's disease. Consider the example of two histologically similar patients with imaged lesions of the same approximate size and location. On FLT PET imaging and/or serial MRI imaging, it is found that one of the lesions is "fast" growing and highly proliferative and the other is much slower. Using the RECIST criteria as a response measure would provide a neutral measure of "stable disease" for the slow growing tumor even if there was no significant treatment effect that altered the disease from its slow growth curve. However, for the fast tumor, specifically if the follow-up anatomical imaging is delayed relative to the treatment effect, the RECIST criteria could provide an unfavorable classification of nonresponder even if growth of the fast tumor was halted, albeit temporarily, by the treatment under consideration. Hence, the timing of the imaging used for treatment assessment and the analysis technique for characterizing the response are critically important. This is one of the natural places for molecular imaging to shine — by providing a real-time assessment of activity of the molecular target during a course of therapy. Individually targeted therapies require individually targeted techniques for assessing efficacy.

Based on the validity of the time progression models of individual virtual controls calibrated in a limited number of patients and times, the simulations can be done for a population to evaluate the accuracy and, ultimately, the value of the imaging biomarker. This approach has been used to estimate the cost-effectiveness of screening for colorectal cancer, and it should become more common [25,26].

4. Conclusions and recommendations

It is an extraordinary time to be engaged in developing new imaging in support of the rapid advances in molecular biology and medicine. The popular media seem unbounded in enthusiasm for the wondrous potential of individualized molecular medicine [27–29]. Yet, the drug pipeline is not overflowing with molecular treatments, and the number of drugs that are making it through the FDA approval process is below that during the 1990s [30]. Partly, this is because our knowledge of how to exploit new genomic information and molecular targets is embryonic; systems-level interactions in complex signaling pathways are particularly daunting. However, there are some more focused limitations that the imaging community should have a role in resolving. Molecular imaging, along with molecular pathology, should have a role in defining the appropriate clinical setting and selecting the appropriate patients with the particular target for a new molecular therapy [2,3,22]. Both sampled and imaged biomarkers will have important roles in developing more effective use of molecular therapy and will likely expand the range of approved indications. Research is needed to match the best-targeted imaging agents with each therapeutic target so as to reveal the time course and regional heterogeneity of expression of the target.

The community of investigators developing sampled biomarkers, both serum and tissue, is engaged in a vigorous literature discussion of appropriate methods to validate the predictive value of new tests, either alone or in panels [22,23,29,30]. The added value of biomarkers over conventional anatomic or clinical systems for categorizing disease, such as that from the TNM classification in cancer or *DSM-III* criteria in psychiatric disorders, needs rigorous testing [22]. We are not yet aware of this discussion within the molecular imaging community, and our goal is to stimulate the discussion through this article. It has implications for how our strategies for evaluating new imaging agents are judged by NIH review groups and by journal editors and reviewers, as well as by FDA and CMS, and thus, it behooves the laboratory scientists developing new imaging methods to propose valid and rigorous standards. Our future value to the overall biomedical community will be in supporting better treatment outcomes rather than in detection. The rules of evidence are not fully developed for the prognostic role of imaging biomarkers, but the enormous potential of molecular imaging provides compelling motivation to push forward with convincing validation studies. Much standardization is

needed and consensus will be difficult to achieve. It will always be easier to test the hypothesis that a new imaging agent is better than the earlier imaging procedure for the same target by setting some arbitrary test statistic. The hard research is to show that imaging this target really results in a better patient outcome but that is the only result that the larger biomedical community wants to read. Every player in the chain of development for new imaging technology is obligated to plan tests that show imaging provides a benefit for the sick patient.

Acknowledgments

References

[1] Jarvik JG. Fundamentals of clinical research for radiologists. The research framework. AJR Am J Roentgenol 2001;176:873–8.

[2] Hartwell L, Mankoff D, Paulovich A, Ramsey S, Swisher E. Cancer biomarkers: a systems approach. Nat Biotechnol 2006;24:905–8.

[3] Kelloff GJ, Krohn KA, Larson SM, Weissleder R, Mankoff DA, et al. The progress and promise of molecular imaging probes in oncologic drug development. Clin Cancer Res 2005;11:7967–85.

[4] Dancey JE. Recent advances of molecular targeted agents. Cancer Biol Ther 2003;6:601–9.

[5] Therasse P, Eisenhauer EA, Verweij J. RECIST revisited: a review of validation studies on tumour assessment. Eur J Cancer 2006;42: 1031–9.

[6] Stroobants S, Goeminne J, Seegers M, Dimitrijevic S, Dupont P, Nuyts J, et al. [18]FDG-positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate (Glivec). Eur J Cancer 2003;39:2012–20.

[7] Motzer RJ, Hoosen S, Bello CL, Christensen JG. Sunitinib malate for the treatment of solid tumours: a review of current clinical data. Expert Opin Investig Drugs 2006;15:553–61.

[8] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100: 57–70.

[9] Mankoff DA, Krohn KA. PET imaging of response and resistance to cancer therapy. Chapter 5. In: Teicher B, editor. Cancer discovery and development: cancer drug resistance. Totowa (NJ): Humana Press Inc.; 2006. p. 105–22.

[10] Schumacher M, Hollander N, Schwarzer G, Sauerbrei W. Prognostic factor studies. In: Crowley J, editor. Handbook of statistics in clinical oncology. New York: Marcel Dekker Inc.; 2001. p. 321–78.

[11] Hilden J, Habbema DF. Prognosis in medicine: an analysis of its meaning and roles. Theor Med 1987;8:349–65.

[12] Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer 1994;69:979–85.

[13] Windeler J. Prognosis — what does the clinician associate with this notion? Stat Med 2000;19:425–30.

[14] O'Sullivan F, Muzi M, Graham MM, Spence AM. Parametric imaging by mixture analysis in 3-D validation of dual-tracer glucose studies. In: Myers R, Cunningham VJ, Bailey DL, Jones T, editors. Quantification of brain function with PET. San Diego: Academic Press; 1996. p. 297–300.

[15] Berman DS, Hachamovitch R, Shaw LJ, Friedman JD, Hayes SW, Thompson LEJ, et al. Roles of nuclear cardiology, cardiac computed tomography, and cardiac magnetic resonance assessment in patients with suspected coronary artery disease. J Nucl Med 2006;47:74–82.

[16] Tseng J, Dunnwald LK, Schubert EK, Link JM, Minoshima S, Muzi M, et al. ^{18}F-FDG kinetics in locally advanced breast cancer: correlation with tumor blood flow and changes in response to neoadjuvant chemotherapy. J Nucl Med 2004;45: 1829–37.

[17] Yagle KJ, Eary JF, Tait JF, Grierson JR, Link JM, Lewellen B, et al. Evaluation of ^{18}F-annexin V as a PET imaging agent in an animal model of apoptosis. J Nucl Med 2005;46:658–66.

[18] Robbins RJ, Wan Q, Grewal RK, Reibke R, Gonen M, Strauss HW, et al. Real-time prognosis for metastatic thyroid carcinoma based on 2-[^{18}F]fluoro-2-deoxy-D-glucose-positron emission tomography scanning. J Clin Endocrinol Metab 2006;91:498–505.

[19] Rajendran JG, Schwartz DL, O'Sullivan J, Paterson LM, Ng P, Scharnhorst J, et al. Tumor hypoxia imaging with [F-18]fluoromisonidazole positron emission tomography in head and neck cancer. Clin Cancer Res 2006;12:5435–41.

[20] Green S, Benedetti J, Crowley J. Clinical trials in oncology. Boca Raton: Chapman & Hall/CRC; 2000.

[21] Linden HM, Stekhova SA, Link JM, Gralow JR, Livingston RB, Ellis GK, et al. Quantitative fluoroestradiol positron emission tomography imaging predicts response to endocrine treatment in breast cancer. J Clin Oncol 2006;24:2793–9.

[22] Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845–56.

[23] Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.

[24] Harpold HLP, Alvord EC, Swanson KR. The evolution of mathematical modeling of glioma proliferation and invasion. J Neuropathol Exp Neurol 2007;66:1–9.

[25] Loeve F, Brown ML, Boer R, van Ballegooijen M, van Oortmarssen GJ, Habbema DF. Endoscopic colorectal cancer screening: a cost-saving analysis. J Natl Cancer Inst 2000;92:557–63.

[26] Vijan S, Hwang EW, Hofer TP, Hayward RA. Which colon cancer screening test? A comparison of costs, effectiveness, and compliance. Am J Med 2001;111:593–601.

[27] See the special issue of Science, October 2004, dedicated to "personalized medicine". Science 2004;306:640.

[28] Zerhouni EA. Translational and clinical science — time for a new vision. N Engl J Med 2005;353:1621–3.

[29] Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. Nat Rev Cancer 2004;4:309–14.

[30] Rothenberg ML, Carbone DP, Johnson DH. Improving the evaluation of new cancer treatments: challenges and opportunities. Nat Rev Cancer 2003;3:303–9.