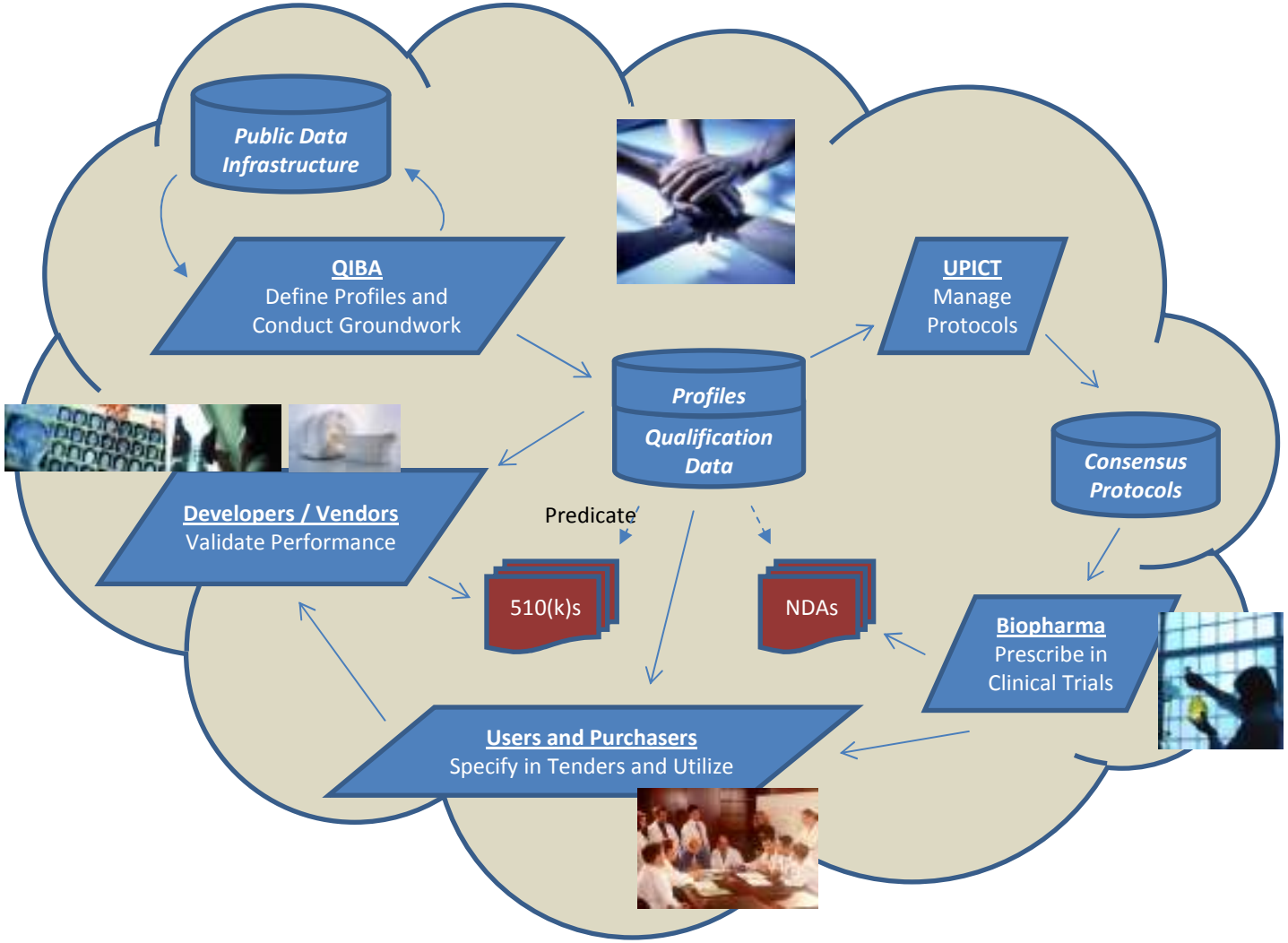
Context: How does Profiling fit in to what we are trying to accomplish overall?



Step 1: What is the clinical question being addressed, i.e., what is the biomarker?

Cancer staging
The extent of lung cancer dissemination is defined at the time of initial diagnosis of a patient in a process called staging. The schema (*TNM Classification of Malignant Tumors*?) for staging lung cancer has been updated recently so that it more accurately clusters patients who benefit from particular therapeutic interventions with predictable outcomes {Goldstraw, 2007 #7}.

A table of how staging relates to lung cancer drug therapy approaches, the imaging approaches used in those stages and issues relative to the image requirements is summarized in Table 1.

Table 1: Summary of Image Processing Issues Relative to Stage of Lung Cancer

Stage	% of Cases	5-year Survival	Imaging Focus/ Therapy Focus	Imaging Tool	Issues	Thoracic Segmentation	Hi-Res
I	16	49	Primary tumor/ Neo and adjuvant RX	sCT	Small cancers surrounded by air	Can be straightforward	Needed
II/III	35	15.2	Primary, hilar and mediastinal lymph nodes/ Combined modality	sCT, PET	Larger tumors and nodes abut other structures	Often challenging	Optional
IV	41	3	Primary/regional nodes and metastatic sites/ Chemotherapy	sCT, PET, Bone, Brain scans	Tumor response often determined outside of the chest	Often challenging	Optional

For this discussion, Stage I is considered separately as it is typically treated with surgery and has the highest potential for curability. Because Stage II is relatively uncommon, Stage II and III are clustered together as their clinical management can be similar involving combinations of radiation therapy and chemotherapy with or without surgery. Stage IV is the most common form of lung cancer; its treatment typical involves only the use of drug therapy approaches. There is a number of trial types listed in the fourth column.

Step 1: What is the clinical question being addressed, i.e., what is the biomarker?

Courtesy Jim Mulshine

Cancer staging

The extent of lung cancer dissemination is defined at the time of initial diagnosis of a patient in a process called staging. The schema (*TNM Classification of Malignant Tumors*?) for staging lung cancer has been updated recently so that it more accurately clusters patients who benefit from particular therapeutic interventions with predictable outcomes {Goldstraw, 2007 #7}.

A table of how staging relates to lung cancer drug therapy approaches, the imaging approaches used in those stages and issues relative to the image requirements is summarized in Table 1.

Table 1: Summary of Image Processing Issues Relative to Stage of Lung Cancer

Stage	% of Cases	5-year Survival	Imaging Focus/ Therapy Focus	Imagin			Hi-Res
I	16	49	Primary tumor/ Neo and adjuvant RX	sC			Needed
II/III	35	15.2	Primary, hilar and mediastinal lymph nodes/ Combined modality	sCT, PET	er tumors and nodes		ptional
IV	41	3	Primary/regional nodes and metastatic sites/ Chemotherapy	sCT, PET, Bone, Brai scans	e chest	Often	Optional

Primary/regional nodes and metastatic sites

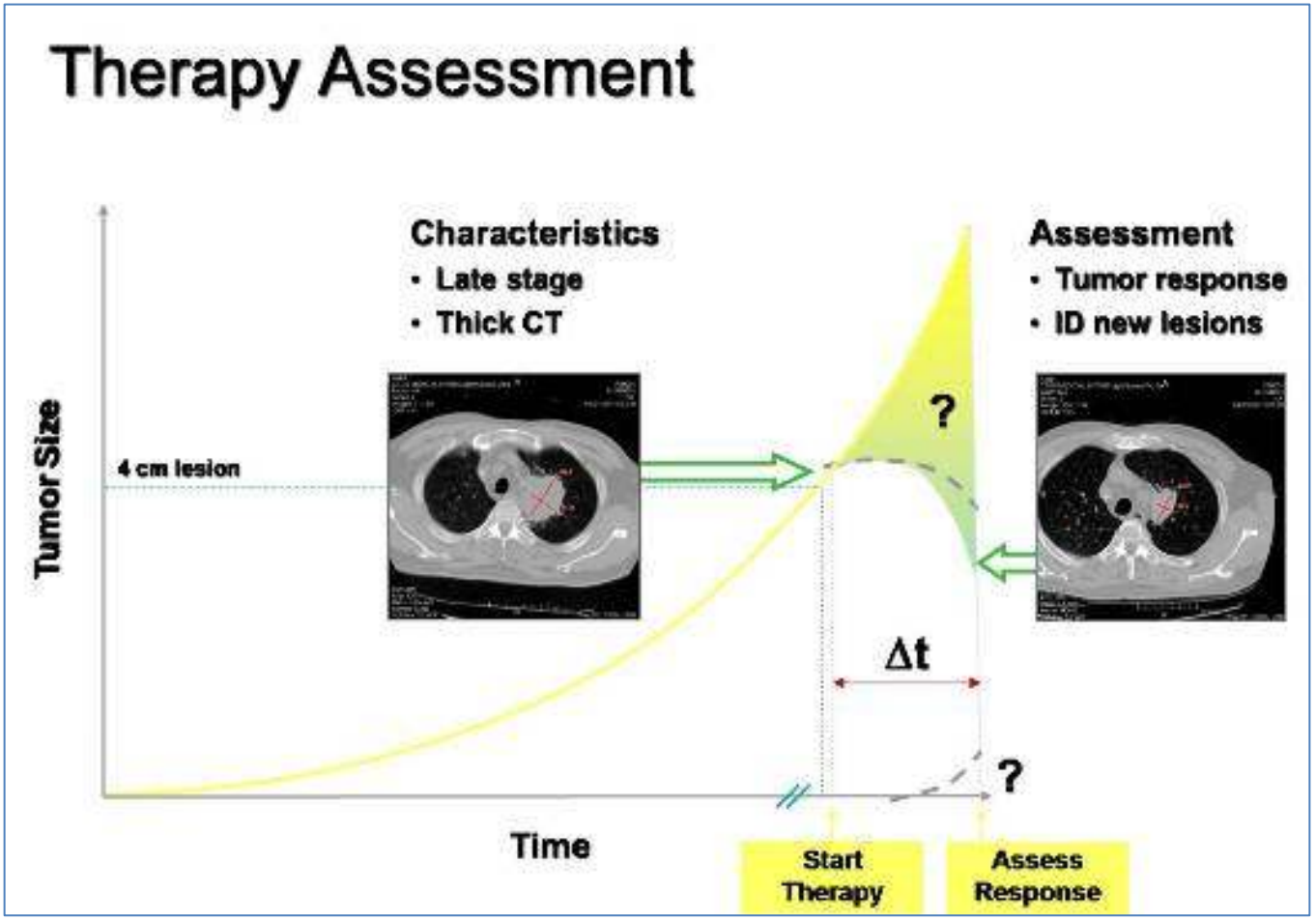
Primary, hilar and mediastinal lymph nodes

Primary tumor/ Neo and adjuvant RX

For this discussion, Stage I is considered separately as it is typically treated wi
Because Stage II is relatively uncommon, Stage II and III are clustered togethe
combinations of radiation therapy and chemotherapy with or without surgery. Stage IV is the most common form of lung cancer; its treatment typical involves only the use of drug therapy approaches. There is a number of trial types listed in the fourth column.

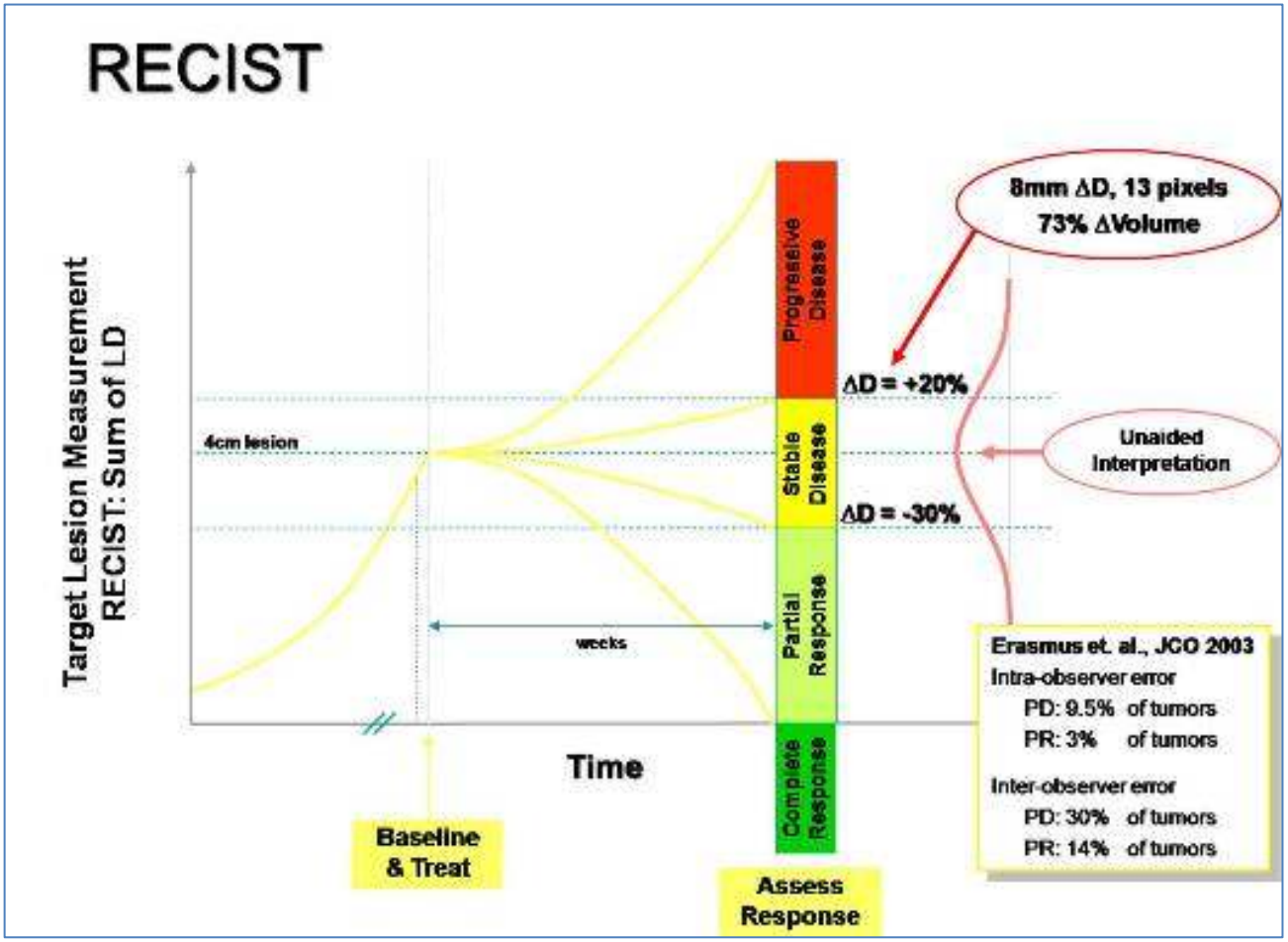
Step 2: How does the field understand the end point involved?

Courtesy Rick Avila



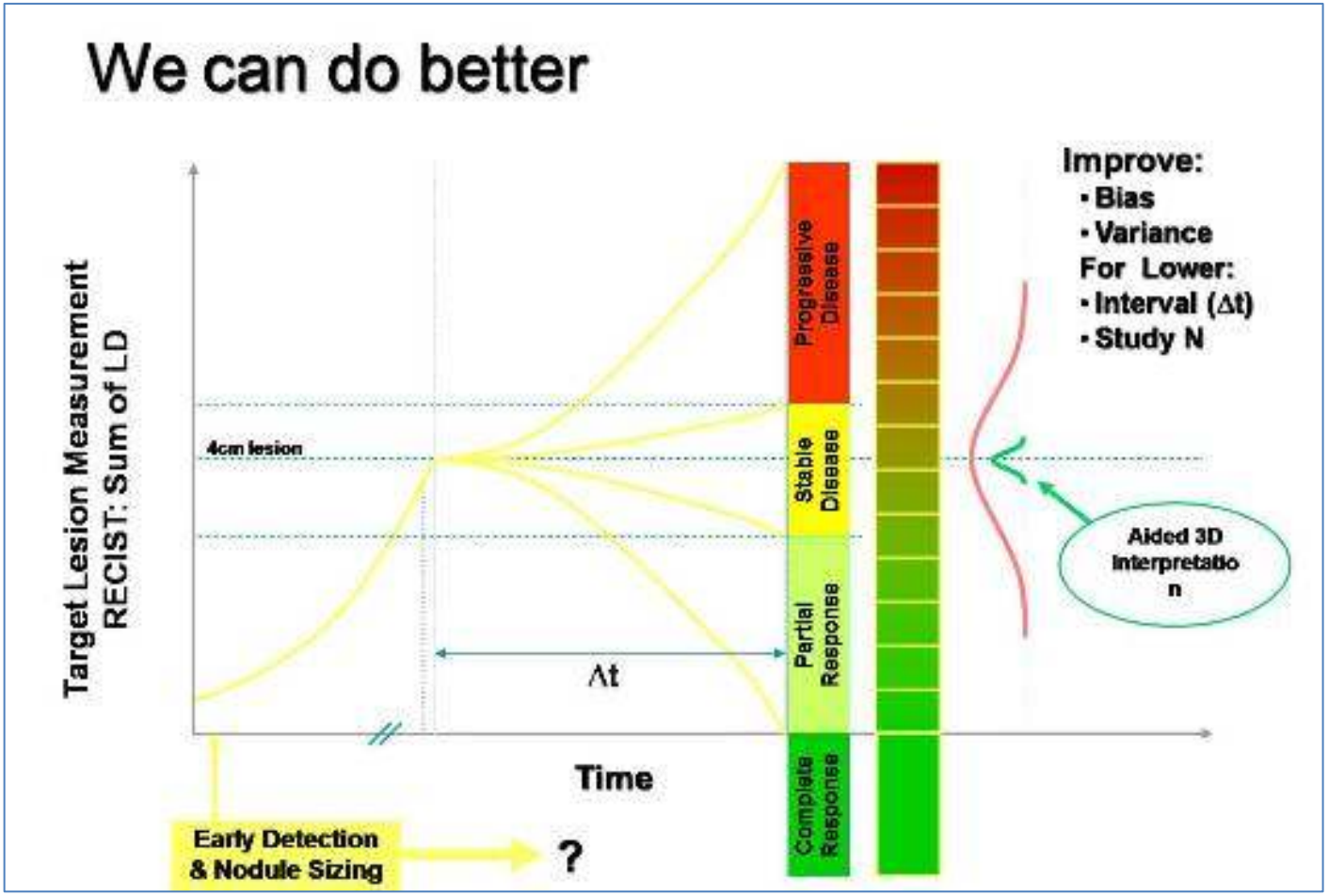
Step 3: What is the performance of the currently accepted methodology?

Courtesy Rick Avila

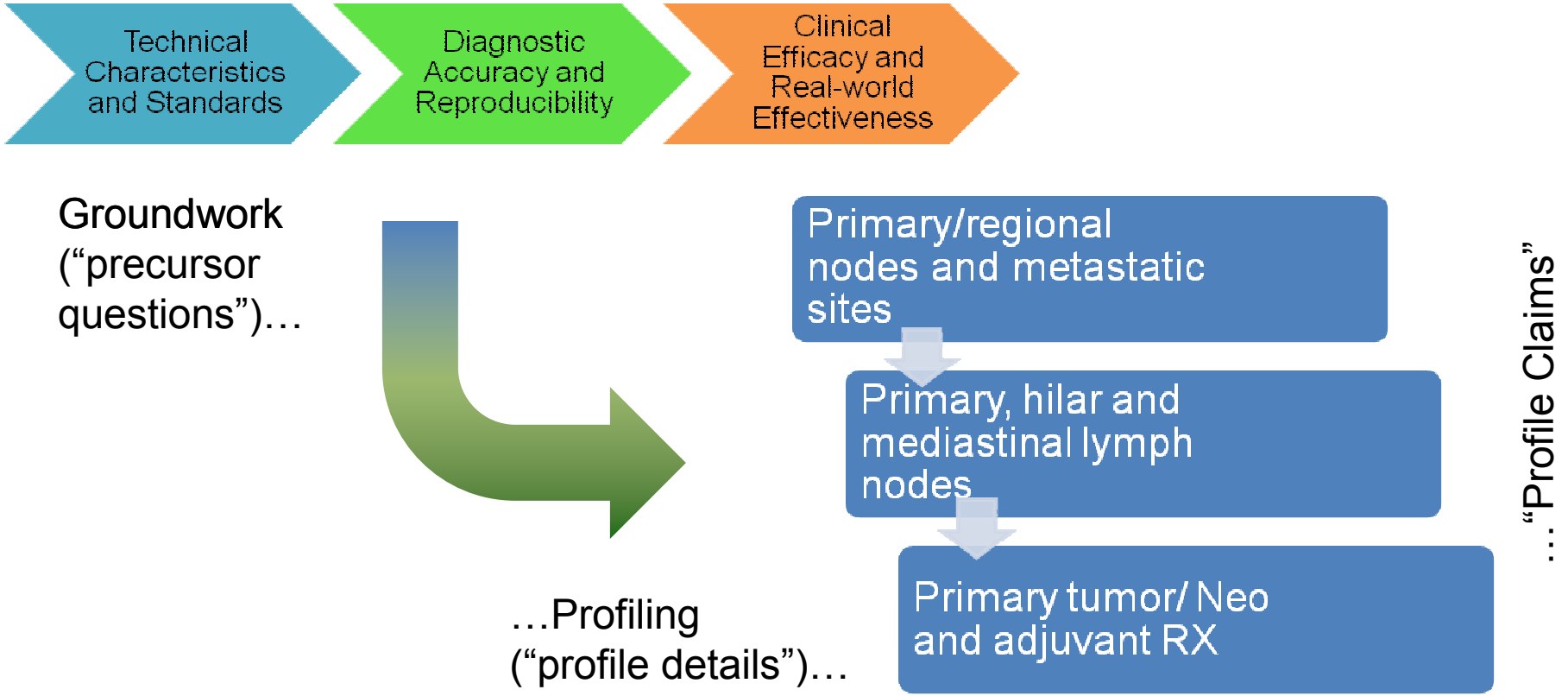


Step 4: What is the claim for the new biomarker, in statistically rigorous terms?

Courtesy Rick Avila



Step 5: What is the roadmap for groundwork, retrospective, prospective, and analytical?



Step 5a: Technology Groundwork

Technical Characteristics and Standards

- Prospective Single-Center Phantom
- Retrospective Clinical, including “No Change” Condition
- Prospective Multi-center Phantom

Step 5b: Statistical Groundwork

Diagnostic Accuracy and Reproducibility

1. Determine performance with respect to statistical power needed
2. Set acquisition standards necessary
3. Determine what type of evaluations are necessary to qualify

Step 5c: Clinical Groundwork

Clinical Efficacy and Real-world Effectiveness

1. Determine intra- and inter-reader sensitivity and specificity using new biomarker
2. Correlate performance of new method with currently accepted method

Step 6: How are the Profile's clinical Claims represented statistically?

Profile Claims (what users will be able to achieve)

Claim #1: Can create, store, retrieve images of lung tumors

Claim #2: Can create, store, retrieve linear, area and volume measurements made on lung tumor images

Claim #3: Can create, store, and retrieve mark ups of lung tumors, i.e., region of interest (ROI) boundaries

Precursor: Need Sample Implementation Chest CAD polylines or New DICOM Segmentation objects (by pixel) are likely sufficient, but should try out a sample implementation to confirm (and identify key Details to require in the Profile). Possibilities for data storage include polylines, voxels, and polygons/triangles. See also **Segmentation and Markup Formats**

Claim #4: Can measure lung tumor volume with repeatability of 18% for tumors greater than 10mm in Longest Diameter.

Rationale: For uniformly expanding cubes and solid spheres, an increase in the RECIST defined uni-dimensional Longest Diameter of a Measurable Lesion corresponds to an increase in volume of about 72%. To diagnose Progressive Disease at a change of about one half that volume, 36%, the noise needs to be less than about 18%. The claim is thus set to be "twice as sensitive as RECIST". *<What do we mean by repeatability> How should the repeatability be expressed? It's easier to meet % targets for larger tumors. Should we use mm3 instead? Or should we state % for a certain sized tumor? There is a description in Jim Mulshines work that we can copy here?* Precursor: Demonstrate this accuracy and repeatability is easily achievable Groundwork: Test-Retest measurements of FDA phantoms i.e., very-best-case-scenario, with variability one order of magnitude less than variability in "real life", i.e., algorithm returns variability of less than 1.5% *<Relevant Groundwork Link 2:>* Test-Retest measurements of small sample of NIST cases, i.e., nearly-best-case-clinical-scenario, with variability for measurement of isolated, simple lung tumors of less than 3% (up to 4 times the noise in phantoms and less than one fifth the noise expected in real life scenarios). *<Relevant Groundwork Link 3:>* Test-Retest measurements of a few well behaved masses in the MSKCC coffee break study of less than 10% between Image Set 1 and Image Set 2 of each patient studied twice in succession. This 10% threshold is somewhat capriciously based on the assumption that the precision of measurement in selected MSKCC coffee break tumors will be twice as good as that which can be achieved in most clinical trial scenarios. Precursor: Should thought be given to revising the RECIST definitions?

Claim #5: Can retrieve and/or contribute images, measurements and markups from/to caBIG. *Are we and caBIG ready to get into this now or is it OK to leave this until our next profile, e.g. volume change, when our ideas and caBIGs infrastructure are more mature/stable?*

Step 7: What Details must be specified to meet the Claim, based on the groundwork?

Profile Details (what equipment and users must do to achieve it)

The Profile defines the following roles and several **transactions** and **activities** they participate in:

Acquisition System
Measurement System
Measurer
...

Activity: Acquisition System Calibration

Detail: Site staff shall conform to the QA program defined by the device manufacturer.

Activity: Patient Preparation

Detail: Staff shall prepare the patient according to the local standard of care.
Precursor: Decide if we need/can be prescriptive about any of the details in efforts to "standardize human behavior" or local procedures. Some protocols call for patient to be "comfortably positioned", in "comfortable clothes" at "comfortable temperature" and with an empty bladder; presumably to minimize patient motion and stress (which might affect the imaging results) and any unnecessary patient discomfort. **Detail:** The following details shall be recorded in the <???> System, manually by the Staff if necessary.
Contrast administration: (Agent, dose, route) The standards for this are currently evolving. To be comparable (e.g. to subtract to get change values), measurements must be made under consistent contrast administration. Consistency includes contrast type, route of administration, rate of administration, interval between start of contrast and start of scan, use of power injector? Requiring no contrast (like ACRIN 6678 did) would be less of an issue than requiring contrast which has potential patient health issues.
Creatinine Clearance: (renal function). **Patient Positioning**: Target: Supine/Arms Up/HeadFirst Acceptable: any but not inconsistent with prior scan Consistency is required to avoid unnecessary variance in attenuation and gravity induced shape. Target is provided as a default to drive some consistency when details of prior scans are not available. **Breath Hold**: Target: Single Breath Hold at Full Inspiration Acceptable: *suspended respiration with high % of end inspiration = single breath hold at full inspiration, right?*
Breath hold reduces motion which degrades the image. Full inspiration inflates lungs which is necessary to separate structures and make lesion more conspicuous.
Bladder State: Target: empty bladder Acceptable: any The target here is purely for patient comfort.
Precursor: How should the details be recorded about the preparation of each actual patient? DICOM provides a way to encode most of these details in the image headers, but we may need to require the operator to enter them.

Activity: Image Acquisition

Detail: The acquisition system shall support saving and easily calling up saved acquisition protocols.
Precursor: Do we need standard naming? Could use UPICT or ACRIN proper name. Sites might prefer "site recognizable" aliases (but need to still know it is the prescribed protocol)
Detail: The acquisition system shall produce images with the following characteristics:
Precursor: Determine which characteristics of the resulting images matter? **Slice width** - Ideal: <= 1 mm Target: 1-2.5mm Acceptable: 5mm
direct component of voxel size; determines resolution along patient (z) axis
Slice interval - Ideal: contiguous or 20% overlap Target: contiguous or 20% overlap Acceptable: contiguous
gaps may "truncate" the spatial extent of the tumor
Isotropic Voxels - Ideal: yes Target: yes Acceptable:
isotropic voxels reduce the volume measurement error effect of tumor orientation (which is difficult to control) requiring isotropic voxels means requiring that the same value be selected for both slice width and voxel size.
would it ever be acceptable to allow a slight difference in the values, e.g. 1mm slice width, .8mm voxel size? **Anatomical Coverage**
E.g. Above lung apices to symphysis pubis **Field of View:Voxel Size** - Ideal: Rib-to-rib: 0.55mm - .75mm Target: Outer Thorax: 0.7mm - .8mm. Acceptable: Complete Thorax: 0.8 to 1.0mm
smaller voxels reduce partial volume effects and (likely) provide higher precision (i.e. higher spatial resolution) but larger voxels increase field of view and thus encompass more anatomy
Scan Plane - 0 azimuth **Motion Artifact** - Ideal: no artifact Target: no artifact Acceptable: "minimal??" motion artifacts may produce false targets and distort the size of existing targets
Noise Level - Ideal: "minimal?" Target: "low" Acceptable: "predictable?" greater levels of noise may degrade segmentation by humans or algorithms
<NOT sure how to deal with this one either; the noise changes with square root to slice thickness, so thinner slices will not be low noise; also recon algorithm affects noise just as significantly. can I pair these up? How about if I do pair slice thickness and noise? see below> **Slice width and Noise** - Ideal: <= 1 mm, std. deviation in 20 cm water phantom < 40 HU Target: 1-2.5mm, std. deviation in 20 cm water phantom < 40 HU Acceptable: 5mm, std. deviation in 20 cm water phantom < 40 HU
slice width determines voxel size and resolution in longitudinal (z) direction of patient; also has a significant impact on noise - thinner slices have much higher noise than thicker slices for a given mAs (or effective mAs) setting. Here noise is recommended to be constant across slice thickness; this would be accomplished by increasing mAs for thinner slices and reducing for thicker slices. **Spatial Resolution**: Ideal: 7-8 lp/cm Target: 6-8 lp/cm Acceptable: 6-8 lp/cm
Resolution is the number of resolvable line-pairs per cm in a scan of an ACR resolution phantom (or equivalent) Higher spatial resolution is necessary to distinguish borders of tumors
Spatial resolution is determined by scanner geometry (not under user control) and reconstruction algorithm (which is under user control.
Detail: The acquisition system shall support configuration of the following acquisition parameters:
Precursor: Determine which acquisition parameters matter **KVP** - Ideal: 120 Target: 110-130 Acceptable: 110-140
kVP determines contrast between tissues and also influences noise and radiation dose; should be consistent for all scans of patient **effective mAs (medium patient)** - Ideal: 80 to 120 Target: 60 to 200 Acceptable: 40 to 350 **effective mAs (large patient)** - Ideal: Target: Acceptable: effective mAs = (mA*time/pitch) higher mAs lowers noise but increases dose **Rotation Speed** - Ideal: Target: Acceptable: faster rotation reduces the breath hold requirements and reduces the likelihood of motion artifacts **Collimation width** - Ideal: 20 to 40 mm Target: 10 to 80 mm Acceptable: 5 to 160 mm wider collimation widths can increase coverage and shorten acquisition, but can introduce cone beam artifacts which may degreade image quality
# of channels - Ideal: 64 or greater Target: 16 or greater Acceptable: 1 or greater **Mod**
do we need to specify? Some protocols call for "Helical Mode", but if the detector is wide enough to span the anatomy there is no need to do helical
Some protocols call for "High Speed Mode". Do we need to state this if we already state pitch?
Table speed *do we need to specify?* E.g. 7.5mm/sec? Not needed if no helical. If it is helical, what is the purpose of specifying? Is it to imply the intended length of scan time vis-a-vis breath hold?
Recon. Kernel Characteristics: - Ideal: slightly enhnacing Target: standard to enhancing Acceptable: soft to overenhancing <the relationship between kernel characteristics and our goals/claims is likely complex. What can we say or at least identify as needing investigation>
Recon. Kernel Name – informational **Scanner Model** - informational
indicates the model has been used successfully with the described parameters
Precursor: What value ranges for each parameter constitute an acceptable "baseline"? Late Stage (IIIb and IV) Lung Cancer in World Wide Clinical Trials (pharma base case): Outer ring of quality must be RI = 5 mm; next ring RI = 3 mm. Inner ring specified by Professor Mulshine and colleagues for top-shelf clinical trials of neoadjuvant therapy in earlier stage disease at RI < 1.5 mm.
Note some earlier stage NSCLC trials done with radiofrequency ablation (RAPTURE, R. Lencioni, PI) included some stage I cancers, with all lesions < 3.5 cm} Some parameters need ranges to "normalize" results across different patient sizes
Consider existing protocols: ACRIN 6678 Quality Control Parameters for CT Scan Tumor Volumetric Measurements specified: (Slice width, Slice interval, Voxel Size, Absence of Motion Artifact) &
(KVP, mAs, Rotation Speed, Collimation width, # of channels, Scanner Model, Recon. Algorithm, Non-use of Intravenous Contrast)
Note: many parameters are specified as a range, some depending on the size of the patient NLST (National Lung Screening Trial) Acquisition Parameters specified: (Slice width, Slice interval, # of Images) & (KVP, mA, mAs, Effective mAs, Rotation Speed, Collimation width, # of channels, Detector "width", "MODE", Pitch, Table increment, Table speed, Scan time, Scanner Model, Recon. Algorithm, Dose)
Note: some of these parameters are redundant (i.e. can be calculated from other parameters), many parameters are specified as a range, some depending on the size of the patient The ACRIN protocol may be prefereable since NLST was for a screening study, not for measuring progressive disease *<Insert link to UPICT protocol specifications by Professor McNitt-Gray and colleagues>?*

RAPTURE Trial Phase II trial NCT00690703 (now closed) at ClinicalTrials.Gov Perhaps Dr. Lencioni could suggest methods which would have helped him assess the results in this trial?

Precursor: What uniform language should be used for documenting image acquisition protocols in the profile. DICOM is working on a new object for storing protocols electronically (prescribed or performed) Perhaps Manufacturers should provide CDs with acquisition parameters as they did for the MRI study of brain volumes to the Alzheimer's Disease Neuroimaging Trial sponsored by the Foundation for NIH UPICT is working on common terms for the protocol parameters and possibly a standard presentation CT Acquisition Protocol Groundwork

Activity: Image Reconstruction

<Is there any reason not to fold the Image Reconstruction activity into the Image Acquisition activity and just treat them as a pair? Is there any need/value for them to be separate?>
E.g. what kernel to use? Kernel will be important. Even more so in liver than lung, and in spinal mets assessments.
Detail: The acquisition system shall be able to perform reconstruction with the following parameters:
Reconstruction interval: 5mm without any gaps Kernel: <???>
Further discussion may be necessary. Some sites complain that when compared to slices with an 8mm interval and a 5mm gap (a common clinical standard), the use of a 5mm interval with no gap slows down throughput and increases reading time, radiation exposure and storage requirements. It seems likely that 5mm interval with no gap is necessary to achieve the claims, so the related costs are unavoidable and manageable. This may also tie in to specifying the characteristics of the resulting images rather than the parameters of reconstruction for certain makes/models. Specify what to achieve, rather than how to achieve it. **Transaction: Transfer Images**
Detail: The acquisition system shall support DICOM CT Storage as SCU.
Detail: The measurement system shall support DICOM CT Storage as SCP and DICOM Q/R as SCU

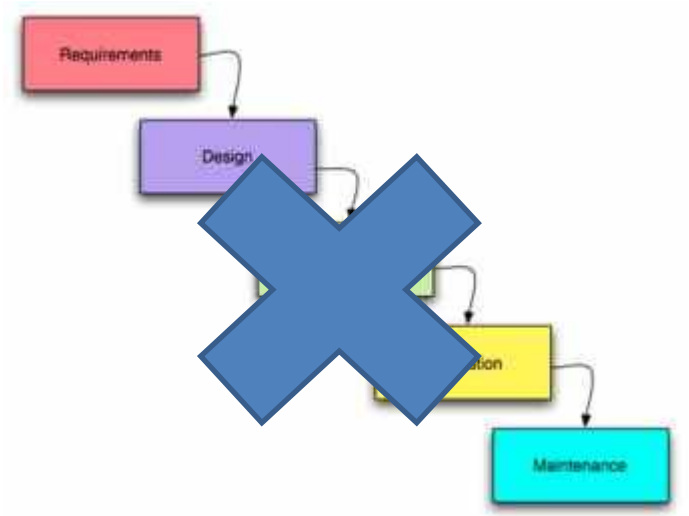
Activity: Measurement

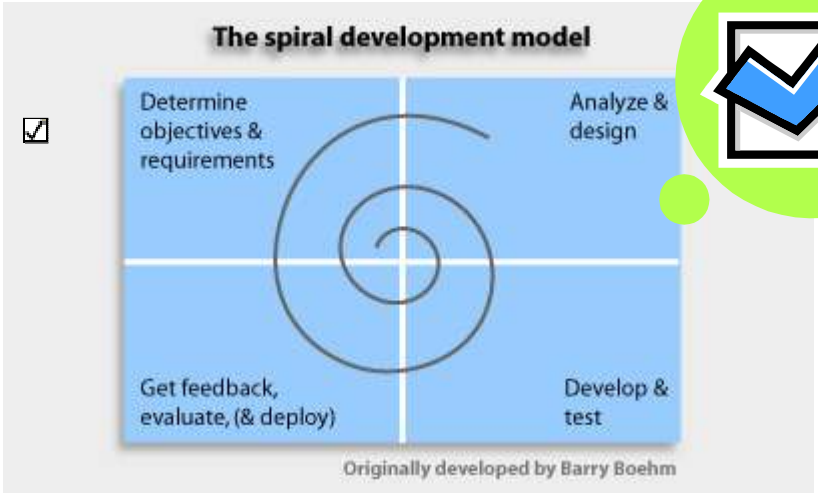
Detail: The measurement system shall support the following measurements:
Precursor: What measurements are useful for evaluating lung tumors Bitvol <because it is the typical "detailed" volume measurement>
RECIST <because it is the current gold standard and we need it to compare>
Modified RECIST (J. Natl. Cancer Inst. 2008;100:698-711) <to support wider cancer etiology than HCC>
<consider just adding a bunch of tools if they are easy to implement> Precursor: What types of cases/issues must the measurement system demonstrate being able to handle? E.g. attachment points,
Precursor: What accuracy is initially sufficient to be useful? Precision of measurement is the primary objective. Accuracy is less important to the base case for pharma. Accuracy becomes increasingly important to the inner rings of quality, reaching its maximum in screening studies of asymptomatic people with risk factors for lung cancer. Precursor: What repeatability is initially sufficient to be useful? Precursor: What accuracy/repeatability can be easily achieved? *<Insert link to relevant Groundwork> <Insert link to very preliminary image analysis in very-best-case-scenario of extremely well demarcated, simple lung tumors which suggest test-retest variability is less than 1% when RI is 5 mm>* Precursor: What is the theoretical limit for accuracy/repeatability with typical equipment *<Insert link to relevant Groundwork> <Insert links to image analysis of well demarcated tumors in the MSKCC coffee break images as the most optimistic boundary, and analysis of complex masses invading solid tissues as the most realistic boundary. First link will be to image analysis by RadPharm, Inc. Other links will be provided by software developers as the data become available.* **Detail:** The measurer shall be able to diagnose Progressive Disease at one half the change in volume associated with RECIST line-lengths.
Precursor: What do we need to specify about the measurer? Human oversight will be required. In the first stage, a trained technologist or image analysis specialist will select tumors for automatic boundary demarcation. In the next stage, the image analysis specialist will be able to manually correct portions of the boundary where either the algorithm failed or the mass becomes too complex to reliably follow over the course of treatment. In the final stage, a trained radiologist will accept or revise the mark ups. Precursor: What is the limit on accuracy/repeatability due to the measurer (reader)? The limit of inter-rater reliability will be such that thresholds for diagnosing Progressive Disease will be within one time-point assessment in a series of time-points for patients enrolled in longitudinal trials. The need for adjudication between discrepant time-point assessments will be less for volumetric image analysis than for ordinary RECIST 1.1 assessments. See 1A Reader Variability Study *<Should we add an Activity: Measurer Training to train/confirm the skill of each measurer>*

Transaction: Transfer Measurements

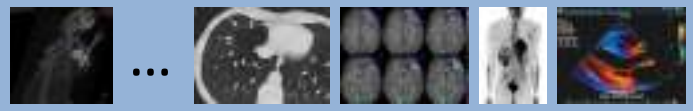
Detail: The Measuring System shall support storage of the measurements in <???> format. **Detail:** The Measuring System shall support storage of the segmentation in <???> format.
Precursor: Need to choose **Segmentation and Markup Formats** ... Consider CDISC as a way to require measurement systems to provide data in a format that is easily consumed by Clinical Trials systems/databases. Note that CDISC has done image work on CT Oncology (related to RECIST). Might not be interested in CDISC change categories, but the measurements they specify is useful (have included volume).
<Insert link to CDISC imaging work>
IHE has worked with CDISC on some general IT profiles and so there may be some IHE transactions we could borrow.

Approach: Start using it immediately, even while it is still under development





IIBE



realizing the power of

imaging in medicine