Imaging Endpoints in Therapeutic Clinical Trials

Richard L. Wahl, M.D.
Division of Nuclear Medicine
Johns Hopkins Medical Institutes
Baltimore, MD
What is an Endpoint?

In clinical trials, an event or outcome that can be measured objectively to determine whether the intervention being studied is beneficial. The endpoints of a clinical trial are usually included in the study objectives. Some examples of endpoints are survival, improvements in quality of life, relief of symptoms, and disappearance of the tumor.
Types of Trials

• Diagnostic Accuracy/performance: Single and multi center
• Change in management (hypothetical)
• Change in management (actual)
• Integrated with Therapy
Figure. Paired Representative Florabetapir-PET Scans and β-Amyloid Antibody 4G8 Immunohistochemistry Photo Micrographs

Clark, C. M. et al. JAMA 2011;305:275-283
Table 2. Key Correlations for the Primary Analysis Cohort (n = 29)a.

<table>
<thead>
<tr>
<th>Cortex Region</th>
<th>Florbetapir-PET Measure</th>
<th>Pathology Reference Standard</th>
<th>Bonferroni ρ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole brain</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.78 (0.58-0.89)</td>
</tr>
<tr>
<td>Whole brain</td>
<td>SUVr</td>
<td>β-Amyloid area</td>
<td>0.75 (0.53-0.88)</td>
</tr>
<tr>
<td>Whole brain</td>
<td>Visual</td>
<td>NPS</td>
<td>0.71 (0.47-0.86)</td>
</tr>
<tr>
<td>Whole brain</td>
<td>SUVr</td>
<td>NPS</td>
<td>0.74 (0.51-0.87)</td>
</tr>
<tr>
<td>Whole brain</td>
<td>SUVr vs visual</td>
<td>NA</td>
<td>0.82 (0.64-0.91)</td>
</tr>
<tr>
<td>Whole brain</td>
<td>NA</td>
<td>β-Amyloid area vs NPS</td>
<td>0.88 (0.76-0.94)</td>
</tr>
<tr>
<td>Precuneus</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.75 (0.54-0.88)</td>
</tr>
<tr>
<td>Parietal</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.77 (0.56-0.89)</td>
</tr>
<tr>
<td>Frontal</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.69 (0.44-0.85)</td>
</tr>
<tr>
<td>Temporal</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.68 (0.42-0.84)</td>
</tr>
<tr>
<td>Posterior cingulate</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.70 (0.44-0.85)</td>
</tr>
<tr>
<td>Anterior cingulate</td>
<td>Visual</td>
<td>β-Amyloid area</td>
<td>0.74 (0.51-0.87)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; NA, data not available; NPS, neuritic plaque score; PET, positron emission tomographic; SUVr, semiautomated quantitative analysis of the ratio of cortical to cerebellar PET signal.

Correlations were assessed between key florbetapir-PET imaging measures and key pathological measures using the Spearman rank correlation coefficient.

Clark, C. M. et al. JAMA 2011;305:275-283
What Types of Therapeutic Clinical Trials?

- Phase I: Safety. Dose, PK and toxicity information. (10-20 pts)

- Phase II: Evidence of therapeutic activity in a specific situation (16-100 pts)

- Phase III: Evidence of therapeutic activity, or side effect profile, which is greater than an accepted comparator Rx.
Operational Problems

• Can we do this trial and imaging procedure?
• Can the patients tolerate the longer study?
• Can the patients tolerate an additional study?
• Is the radiation dose acceptable?
• Can I make or secure the imaging agent reliably at the needed time of day?
Can I find the Patients I need?

- Alliance with clinician seeing patients
- Shared interest and rewards
- Prioritization of protocol in Cancer Center protocol lists (if cancer protocol)
- Competition with clinical trials
- e.g. rising PSA imaging trials vs Rx trials (Rx wins) with concerns from sponsors re investigational agent use in Dx trials.
Can I measure what I need to on the Images?

- Methods must be specified
- Some training may be needed for newer methods or the situation under study.
FDG: A Tracer of Early Steps of Glucose Metabolism

Extracellular  Intracellular  Intracellular
Overview of Lecture

• RECIST 1.1
• Alternative Anatomic applications
• PET is a powerful tool for diagnosis but mainly Qualitative interpretations
• Importance and need for Quantifying treatment response
• Introduction to PERCIST 1.0
• Validation and next steps
• Response Adapted Therapies
• Global Initiatives to standardize PET Quantitation
RECIST 1.1 vs 1.0

- RECIST 1.0, 2000
- RECIST 1.1, 2009

Major differences:
- A reduction in the number of lesions to be assessed for response from a maximum of 10 to five, and from five to a maximum of two per organ.
• New guidance on making robust measurements of lymph node involvement.
• Confirmation of response is required for trials with objective response as a primary endpoint, but is no longer required for randomized studies, since the control arm of these studies provides appropriate means for interpreting results of the experimental arm.
Difference between WHO and RECIST guidelines for lesion measurement.

Chalian H et al. Radiographics 2011;31:2093-2105

©2011 by Radiological Society of North America
Axial CT images obtained in a 42-year-old woman with breast cancer show multiple metastases in the liver (Li1, Li2) and right kidney (K) (a) and in both lungs (Lu1, Lu2) (b, c).

Chalian H et al. Radiographics 2011;31:2093-2105
Axial CT images obtained in a 42-year-old woman with breast cancer show multiple metastases in the liver (Li1, Li2) and right kidney (K) (a) and in both lungs (Lu1, Lu2) (b, c).
Axial CT images obtained in an 82-year-old woman with colon cancer metastases to the liver show incorrect (a) and correct (b) measurements of the longest diameters of two adjacent target lesions that are nearly coalescent.
Normal, nontarget, and target lymph nodes in a 64-year-old man with a history of lymphoma.
—Comparison of sum of longest diameters of target lesions on baseline CT measurements performed in clinical trial (measurement 1) and by observer in present study (measurement 2).

\[ y = 0.9169x + 0.5623 \]
\[ r^2 = 0.9740 \]

Nishino M et al. AJR 2010;195:W221-W228
—Time to progression by Response Evaluation Criteria in Solid Tumors (RECIST) 1.1 versus RECIST 1.0.

Nishino M et al. AJR 2010;195:W221-W228
<table>
<thead>
<tr>
<th>Response</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response</td>
<td>Requires all of the following: complete disappearance of all enhancing measurable and nonmeasurable disease sustained for at least 4 weeks; no new lesions; no corticosteroids; and stable or improved clinically</td>
</tr>
<tr>
<td>Partial response</td>
<td>Requires all of the following: $\geq 50%$ decrease compared with baseline in the sum of products of perpendicular diameters of all measurable enhancing lesions sustained for at least 4 weeks; no new lesions; stable or reduced corticosteroid dose; and stable or improved clinically</td>
</tr>
<tr>
<td>Stable disease</td>
<td>Requires all of the following: does not qualify for complete response, partial response, or progression; and stable clinically</td>
</tr>
<tr>
<td>Progression</td>
<td>Defined by any of the following: $\geq 25%$ increase in sum of the products of perpendicular diameters of enhancing lesions; any new lesion; or clinical deterioration</td>
</tr>
</tbody>
</table>
## RANO Criteria

### Table 4: Summary of the Proposed RANO Response Criteria

<table>
<thead>
<tr>
<th>Criterion</th>
<th>CR</th>
<th>PR</th>
<th>SD</th>
<th>PD</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1 gadolinium enhancing disease</td>
<td>None</td>
<td>≥ 50% ↓</td>
<td>&lt; 50% ↓ but &lt; 25% ↑</td>
<td>≥ 25% ↑*</td>
</tr>
<tr>
<td>T2/FLAIR</td>
<td>Stable or ↓</td>
<td>Stable or ↓</td>
<td>Stable or ↓</td>
<td>↑*</td>
</tr>
<tr>
<td>New lesion</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>Present*</td>
</tr>
<tr>
<td>Corticosteroids</td>
<td>None</td>
<td>Stable or ↓</td>
<td>Stable or ↓</td>
<td>NA†</td>
</tr>
<tr>
<td>Clinical status</td>
<td>Stable or ↑</td>
<td>Stable or ↑</td>
<td>Stable or ↓</td>
<td>↓*</td>
</tr>
<tr>
<td>Requirement for response</td>
<td>All</td>
<td>All</td>
<td>All</td>
<td>Any*</td>
</tr>
</tbody>
</table>

**Abbreviations:**
- RANO, Response Assessment in Neuro-Oncology
- CR, complete response
- PR, partial response
- SD, stable disease
- PD, progressive disease
- FLAIR, fluid-attenuated inversion recovery
- NA, not applicable

*Progression occurs when this criterion is present.
†Increase in corticosteroids alone will not be taken into account in determining progression in the absence of persistent clinical deterioration.
## Table 1. Comparison between WHO criteria and the irRC

<table>
<thead>
<tr>
<th></th>
<th>WHO</th>
<th>irRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>New, measurable lesions (i.e., $\geq 5 \times 5 \text{ mm}$)</td>
<td>Always represent PD</td>
<td>Incorporated into tumor burden</td>
</tr>
<tr>
<td>New, nonmeasurable lesions (i.e., $&lt;5 \times 5 \text{ mm}$)</td>
<td>Always represent PD</td>
<td>Do not define progression (but preclude irCR)</td>
</tr>
<tr>
<td>Non-index lesions</td>
<td>Changes contribute to defining BOR of CR, PR, SD, and PD</td>
<td>Contribute to defining irCR (complete disappearance required)</td>
</tr>
<tr>
<td>CR</td>
<td>Disappearance of all lesions in two consecutive observations not less than 4 wk apart</td>
<td>Disappearance of all lesions in two consecutive observations not less than 4 wk apart</td>
</tr>
<tr>
<td>PR</td>
<td>$\geq 50%$ decrease in SPD of all index lesions compared with baseline in two observations at least 4 wk apart, in absence of new lesions or unequivocal progression of non-index lesions</td>
<td>$\geq 50%$ decrease in tumor burden compared with baseline in two observations at least 4 wk apart</td>
</tr>
<tr>
<td>SD</td>
<td>50% decrease in SPD compared with baseline cannot be established nor 25% increase compared with nadir, in absence of new lesions or unequivocal progression of non-index lesions</td>
<td>50% decrease in tumor burden compared with baseline cannot be established nor 25% increase compared with nadir</td>
</tr>
<tr>
<td>PD</td>
<td>At least 25% increase in SPD compared with nadir and/or unequivocal progression of non-index lesions and/or appearance of new lesions (at any single time point)</td>
<td>At least 25% increase in tumor burden compared with nadir (at any single time point) in two consecutive observations at least 4 wk apart</td>
</tr>
</tbody>
</table>
Changes

• * The definition of disease progression has been refined so that it not only includes a 20% increase in the size of the lesion, but also a 5 mm absolute increase as well, so that changes of just a few mms in very small tumors, which may be within the range of measurement error, are not unnecessarily described as disease progression.

• * Guidance on imaging, including its use in the detection of new lesions and the interpretation of FDG-PET scan assessment.
Figure 1a

NOPR Objectives

• **Primary:** To assess the effect of FDG-PET on referring physicians’ plans of intended patient management across the spectrum of the expanded cancer indications for FDG-PET

• **Secondary:** To assess the effect of FDG-PET on referring physicians’ plans of intended patient management in relation to:
  - Specific type of cancer
  - Specific indication for FDG-PET
  - Patient performance status
  - Physician’s role as provider of cancer treatment
  - Type of FDG-PET study (PET/CT vs. conventional PET)
PET Changed Intended Management in 36.5% of Cases

<table>
<thead>
<tr>
<th>Pre-Pet Plan</th>
<th>Post-PET Plan</th>
<th>Dx n=5,616</th>
<th>Staging n=6,464</th>
<th>Restaging n=5,607</th>
<th>Recurrence n=5,388</th>
<th>All n=22,975</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treat</td>
<td>Same</td>
<td>16.0</td>
<td>46.5</td>
<td>15.8</td>
<td>20.4</td>
<td>25.5</td>
</tr>
<tr>
<td>Non-Treat</td>
<td>Same</td>
<td>52.9</td>
<td>14.0</td>
<td>48.0</td>
<td>40.7</td>
<td>37.9</td>
</tr>
</tbody>
</table>

| Non-Treat | Treat | 23.2 | 31.6 | 28.6 | 29.2 | 28.3 |
| Treat     | Non-Treat | 7.9  | 7.9  | 7.5  | 9.7  | 8.2  |

Patients with change post-PET (%)  

31.1  
39.5  
36.1  
39.0  
36.5

Hillner et al., J Clin Oncol 2008
What is PET?

- **Quantitative** Functional Nuclear Imaging test
- Patients injected with Short lived radiotracer
- Sensitive method for detecting small lesions
- Images rich in functional information
- Images much less rich in anatomic information
Do we really need to quantitate PET images?
They take too long to read already…
Typical Cancer Treatment

- Histological diagnosis of cancer
- Treat with best therapy (on average) for this group of patients
- Give the complete therapy
- Assess response at conclusion of Therapy
- Alternative: Assess response anatomically after 2 months of therapy
- Challenges: If treatment does not work, weeks to months of expense, toxicity, unnecessarily
Patients are Individuals

- Varying pharmacokinetics in whole body, organs and tumors
- Varying receptor status in tumors
- Varying proliferative/apoptotic rates in tumors
- Varying in many characteristics
- Current treatments are substantially based on the “Average Patient” not the individual patient.
- Goal is to individualize treatments to optimize responses and minimize toxicity and COSTS
- Imaging shows the phenotype in the physiological milieu—for chemo/biological therapy and can be repeated
When is Response Assessed?

• End of a Treatment Course: perhaps most relevant to have a complete response at this time

Mid treatment course

Shortly after treatment is initiated
<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Revised (includes Hodgkin’s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CR</strong></td>
<td>Normal size</td>
<td>FDG-avid tumor: mass of any size, as long as PET “negative”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variably FDG-avid or unknown: normal size</td>
</tr>
<tr>
<td><strong>CRu</strong></td>
<td>If nodes &gt; 1.5 cm, &gt; 75% decrease</td>
<td>n/a</td>
</tr>
<tr>
<td><strong>PR</strong></td>
<td>&gt; 50% decrease</td>
<td>FDG-avid tumor: &gt; 50% decrease, but at least one PET positive focus</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variably FDG-avid or unknown: regression in size</td>
</tr>
</tbody>
</table>

PET Monitoring of Chemotherapy

63 days post chemo start, 48% decline in FDG uptake in responders

No significant decline in FDG uptake in non responders, n=11

Wahl et al. JCO 11:2101-2111, 1993
Why might mid-treatment, or early in treatment, PET be superior to post-treatment?

Early PET result implies a certain rate of tumor kill
First-order kinetics

Usual size at diagnosis

With 6 cycles, need at least 1.5 logs of cell kill per cycle
Usual size at diagnosis

Lower detection limit of PET

PET likely can only measure the first 2-3 logs of cell kill (so negative PET does not mean absence of tumor)
A true negative PET after 2 cycles implies an adequate rate of tumor kill.

Usual size at diagnosis

Lower detection limit of PET

A true negative PET at end of therapy might be less predictive.
Usual size at diagnosis

Lower detection limit of PET

Logs of lymphoma cells

Cycles of chemotherapy

Cure
A true positive PET after 2 cycles suggests cure is unlikely.

Logs of lymphoma cells

Cycles of chemotherapy
Qualitative vs Quantitative or Both?

- Visual assessments can be used to assess response
- Strengths: No special instrumentation, integration of all data.
- Weaknesses: Tendency to be binary, not reproducible in several series

- Quantitative: Ignores qualitative data, can be erroneous due to technical factors, “Standardized” is not so standard after all
No standardized and validated Quantitative metabolic response criteria exist
## EORTC Response Criteria

<table>
<thead>
<tr>
<th>CR</th>
<th>Complete Disappearance of all Metabolically Active Tumor (i.e. decreased to background levels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR</td>
<td>&gt;15% decline in SUV after 1 cycle, &gt;25% decline after 2 or more cycles. Reduction in extent (size) of FDG uptake is not required</td>
</tr>
<tr>
<td>SD</td>
<td>Increase in FDG SUV of &lt;25% or decrease of &lt;15% in SUV and no increase in extent of uptake (&lt;20% in longest dimension)</td>
</tr>
<tr>
<td>PMD</td>
<td>Increase in SUV of over 25%, Increase in extent of FDG uptake by &gt;20%, New FDG positive metastases</td>
</tr>
</tbody>
</table>
Limitations of EORTC Response Criteria

- Does not specify how ROI are drawn
- Does not specify how to handle multiple lesions
- Response criteria of 15% for early PR may be within variability of the test and may be FP due to chance
- Methods for determining “increase in extent of FDG uptake of 20%” left open
- Excellent beginning but needed update
If there were RECIST Criteria for PET, they Would be Defined as…

Task of reviewing and proposing PET response criteria for JNM supplement and for a CME lecture
“PERCIST”

- Positron
- Emission
- Response
- Criteria in
- Solid
- Tumors
From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors

Richard L. Wahl¹,², Heather Jacene¹, Yvette Kasamon², and Martin A. Lodge¹
PERCIST has received considerable attention in the literature:

- Google scholar search: 238 ms have cited PERCIST since it was published in 2009
Most-Read Articles during April 2011 -- updated monthly

Most-read rankings are recalculated at the beginning of the month and are based on full-text and pdf views.

1. Personalizing Cancer Therapy with FDG PET: From RECIST to PERCIST: Richard L. Wahl, Heather Jacene, Yvette Kasamon, and Martin A. Lodge

   From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors

   (10.2967/jnumed.108.057307).

   Abstract  Full Text  Full Text (PDF)  Figures Only
Key Aspects of PERCIST

- Defined Acceptable Uptake Time for Study 1 and Study 2
- Must have same scanner and software
- Validation with normal tissue background (liver or blood)
- Minimum acceptable tumor lesion metabolic activity and statistical basis
- Specified # of lesions
- Use of SUV lean (SUL)
- Continuous scale
Standardized PET Techniques

National Cancer Institute

Netherlands protocol

UPICT updated protocol
Factors that Affect SUV

Uptake time
Blood glucose level
Body weight
Injection technique
Camera calibration
Partial volume
Region of interest (ROI)
Reconstruction method
Matrix size

Sugawara et al. Radiology 1999; 213:521
Torizuka T et al. Radiology 1997; 203:169
Figure 1a. Graphs depict the relationships between patient body weight and blood SUVs: (a) SUVbw, (b) SUVibw, (c) SUVlbm, or (d) SUVbsa

\[
y = 0.021x + 0.932 \\
r = 0.705 \\
P < .001
\]
Figure 1c. Graphs depict the relationships between patient body weight and blood SUVs: (a) SUVbw, (b) SUVibw, (c) SUVlbm, or (d) SUVbsa.
Introduction

- A number of different ROI definitions have been employed including:
  - Mean within an irregular ROI defined by isocontours.
  - Mean within a fixed size ROI centered on the most metabolically active region.
  - Maximum pixel within a large ROI encompassing the entire tumor.

- $SUV_{\text{max}}$ has been widely used although single pixel measurements of this sort may be compromised when images have high levels of noise.
Quantifying Metabolic Tumor Response to Therapy: The Influence of Image Noise on Maximum and Mean SUV.

MA Lodge, J P Leal, RL Wahl

Russell H. Morgan Department of Radiology and Radiological Sciences
Johns Hopkins University School of Medicine
Baltimore, MD
Results: ROI_{max}

- Insert has an SUV of 2.5 (2.5:1 insert-to-background ratio).

- 0.5 minute acquisition
  \( \text{SUV}_{\text{max}} = 5.28 \)

- 20 minute acquisition
  \( \text{SUV}_{\text{max}} = 2.57 \)
Conclusions

$\text{ROI}_{\text{max}}$
- Maximum pixel within lesion.
- Increasing positive bias as noise increased.

$\text{ROI}_{42\%}$
- Mean of all pixels within an irregular ROI based on an isocontour at 42% of the maximum pixel.
- Increasing positive bias as noise increased.

$\text{ROI}_{9\times9}$
- Mean of all pixels within a small 9 mm x 9 mm region.
- No bias found as noise increased.
SUL max Limitations

Size of ROI variable
- scanner
- matrix size
- slice thickness
- scanner diameter

Precision depends on ROI size
Single-pixel more variable due to noise*

SUL peak

1.2 cm diameter (1 cm³ volume sphere)

Centered around hottest area in tumor

Standardizes ROI size

Maybe less variance than SUL max
“Measurable Lesions”

FDG uptake and not tumor size

Minimum uptake

- $1.5 \times$ liver SUL mean $\pm$ 2 SD
- $2 \times$ blood pool SUL mean $\pm$ 2 SD
- $1.35 \times$ hepatic uptake (NOT)

Sufficient fall in SUL post-treatment
Other methods

Threshold

Varies w/variability of single pixel max

Tumor lesion glycolysis

• Less practical
• Based on threshold method
• Exploratory
Normal Background
Scan to Scan Difference

Within

± 20% and
0.3 SUL units

Comparable uptake times!
Reproducibility of Liver and Mediastinal Blood pool F-18 Activity as Normal Reference Tissues

David Chien, Martin Lodge, Richard L Wahl
Division of Nuclear Medicine
Russell H. Morgan Department of Radiology
The Johns Hopkins University, Baltimore, MD
Results
MD & CV of Baseline vs Follow-up Study

Liver SUL
Mean Diff = 0.002
Coeff of Var = 10.8%
Percist 1.0 preferred ROI

Left Atrium SUL
Mean Diff = -0.052
Coeff of Var = 22.3%

Descending Aorta SUL
Mean Diff = -0.008
Coeff of Var = 17.7%
Percist 1.0 alternative ROI

Left Ventricle SUL
Mean Diff = 0.069
Coeff of Var = 24.6%
Conclusion

• Right hepatic lobe 3 cm ROI is a reproducible measure of background activity
  • More reliable than multiple mediastinal blood pool structure

• Descending aorta 1 cm diameter x 2 cm z-axis ROI can be used as an alternative background
  • More reproducible than LA and LV

• Subgroup analysis suggests different scanner type did not decrease reproducibility of SUL

• Findings support the normal tissue ROI recommendations outline in PERCIST 1.0
FDG PET Quantitative Metabolic Tumor Response Assessment: Is the Number of Target Lesions Evaluated Important?

H.A. Jacene and R.L. Wahl
Russell H. Morgan Department of Radiology and Radiological Science
Johns Hopkins University
Baltimore, MD
Number of Lesions

- PERCIST 1.0 only evaluates the SUL peak of the hottest tumor. This is a possible limitation of the approach, but lesions and their responses are highly correlated in general.

- Additional data are required to determine how many lesions should be assessed over 1.

- A suggested option is to include the 5 hottest lesions, or the 5 observed on RECIST 1.1 which are the most measurable. % change in SUL can be reported for the single lesion with the largest increase in uptake or the smallest decline in uptake. Additional studies will be needed to define how many lesions are optimal for assessment.
All are considered
Target and non-target lesions
Objective

To determine if there is a difference in quantitative metabolic tumor response classification on FDG PET/CT based on 1, 3, 5 or 6 target lesions
Methods

30 pts receiving RIT for lymphoma

6 target lesions with highest SUV (SULmax)

SUVs of these 6 target lesions summed
  • Pre-RIT summed SUV
  • Post-RIT summed SUV

Percent change in summed SUV determined

Repeated w/1, 3, & 5 lesions w/highest SUVs

## Correlation % Change in $\sum SUL$

<table>
<thead>
<tr>
<th>No. Lesions</th>
<th>5</th>
<th>3</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.99</td>
<td>0.98</td>
<td>0.80</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.97</td>
<td>0.81</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>0.85</td>
</tr>
</tbody>
</table>

$p<0.01$
Biologically Relevant?

1º response predicts outcomes in metastases
Same lesions before and after treatment
Worst responding lesion
• lesion with least change
• highest uptake before and after treatment
PERCIST: Primary Response Analysis

Single hottest lesions: NOT always the same lesion at scan 1 and scan 2

Percent change in SUL peak
Complete Metabolic Response

- **Complete metabolic response (CMR)** complete resolution of [18F]-FDG uptake within the measurable target lesion so that it is less than mean liver activity and indistinguishable from surrounding background blood pool levels.

- Disappearance of all other lesions to background blood pool levels. % decline in SUL should be recorded from measurable region as well as (ideally) time in weeks after treatment was begun (i.e. CMR -90, 4).

- No new FDG avid lesions in a pattern typical of cancer. If progression by RECIST must verify with follow up
Continued declines out to 24 weeks
## Partial Response

<table>
<thead>
<tr>
<th>Target</th>
<th>All others</th>
</tr>
</thead>
<tbody>
<tr>
<td>↓30% $\text{SUL}_{\text{peak}}$</td>
<td>No ↑30% SUL or size</td>
</tr>
<tr>
<td>↓0.8 SUL units</td>
<td>No new FDG avid lesions</td>
</tr>
<tr>
<td></td>
<td>Anatomic PD – verify</td>
</tr>
</tbody>
</table>
Stable Metabolic Disease

- Stable metabolic disease (SMD) Not CMR, PMR nor PMD. Note, the SUL peak in metabolic target lesion should be recorded as well as (ideally) time from start of most recent therapy in weeks (i.e. SMD - 15,7). No new lesions
Progressive metabolic disease (PMD)

- >30% increase in FDG SUL peak, with >0.8 SUL unit increase in tumor SUV peak from the baseline scan in pattern typical of tumor and not of infection/treatment effect.

- OR - Visible increase in the extent of [18F]-FDG tumor uptake (75% in TLG volume with no decline in SUL).

- OR - new [18F]-FDG avid lesions which are typical of cancer and not related to treatment effect or infection. PMD other than new visceral lesions should be confirmed on follow up study within 1 month unless

- PMD also is clearly associated with progressive disease by RECIST 1.1. PMD should be reported to include % change in SUV peak, (ideally time post treatment in weeks) and whether new lesions are present/absent and their number (i.e PMD, +35, 4, New-5).
PERCIST Continuous Response Scale

- Because SUL is a continuous variable, dividing response criteria into a limited number of somewhat arbitrary response categories loses much data.

- For this reason PERCIST preserves percent declines in the SUV peak in each reported category. Because the rapidity with which a scan normalizes is important (faster appears better), PERCIST asks for the time from start of treatment as part of the reporting.

- For example, a CMR 90, 1 is probably superior to a CMR 90, 10, especially if the latter patient were SMD 20,1. More than one measurement of PET response may be needed at differing times and it may be treatment type dependent.
Elements of Reporting

• Time from injection until imaging
• SUV mean of liver
• Serum Glucose
• SUV Lean Peak of hottest lesion (and max)
• Presence and number of new lesions
• Structured reporting including key parametric indices
Exploratory Analyses

SUL peak for up to 5 lesions
Change in summed SUL
Total lesion glycolysis
Inter-Observer Variability of SUV

Same tumor data set measured multiple times by independent observers

100% agreement in SUV determination
  – Minn et al, Radiology, 1995
  – 10 tumors each measured twice by 2 independent observers
  – Semi-automated image analysis software

“Good” inter-observer agreement
  – Marom et al J Thorac Imaging 2006
  – 5 readers measured 20 primary tumors four times

Untreated primary lung cancers on average > 2 cm
Some of the Limitations

Actually getting SUL peak
Is the minimum value too high?
Lack of good data for progression
Needs VALIDATION by variety of groups
Objectives

To directly compare inter-observer reproducibility of

1) SUV & CT size measurements in malignant tumors pre- and post-therapy

2) % change in SUV & CT size measurements in response to therapy

Percent change $\text{SUV}_{\text{bw max}}$

$\text{ICC} = 0.94$

Percent change Longest CT size


ICC – 0.70
Percent change 2D CT size

ICC – 0.33

Feasibility and Potential Limitations of PERCIST Criteria 1.0 for Comparison of Sequential Studies in Routine Clinical Practice

David Chien and Richard L Wahl
Division of Nuclear Medicine
Russell H. Morgan Department of Radiology
The Johns Hopkins University, Baltimore, MD
Objective

To determine the feasibility and potential limitations of PERCIST 1.0 criteria for comparison of sequential studies in routine clinical practice
Materials and Methods

Patient Demographics

• From 4/1/10 – 7/31/10, there were 53 patients (27 M, 26 F) that had multiple studies

• Of the 53 patients, there were 111 studies
  48 patients with 2 studies
  5 patients with 3 studies

• Earliest of multiple scan (Baseline study)

• Subsequent studies were compared to the baseline study

• Total of 58 pairs of comparisons

• Referred for known or suspected malignancy
Results

Meeting PERCIST 1.0 Criteria for Comparison

- Most follow-up studies had comparable normal background tissue SUL
  - 20% and < 0.3 SUL (avg) unit variation from baseline to follow-up study

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal background tissue SUL</td>
<td>1.47 ± 0.25</td>
</tr>
<tr>
<td>Follow-up studies met criteria</td>
<td>79% (46)</td>
</tr>
</tbody>
</table>
Results

Meeting PERCIST 1.0 Criteria for Comparison

- Most follow-up studies had comparable uptake time
  - Within 15 min of each other
  - No less than 50 min after injection
  - Typically at a mean of 60 min after injection

<table>
<thead>
<tr>
<th></th>
<th>Mean ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uptake time (mins)</td>
<td>61.3 ± 6.4</td>
</tr>
<tr>
<td>Follow-up studies met criteria</td>
<td>83% (48)</td>
</tr>
</tbody>
</table>
Conclusion

• Review of 58 patients with multiple studies showed a vast majority of follow-up studies met SUL for background tissue (79%) and injection-to-scan uptake time (83%) limits and variability recommended by PERCIST 1.0
PERCIST Software

- PERCIST 1.0 Analytical software offered by at least 4 vendors
- 2 camera makers have displayed
- 2-3 independent workstation manufacturers
Performance of a PERCIST 1.0 Based CAD System in the Evaluation of Head and Neck Cancer

Jeffrey P. Leal BA, Muhammad Chaudhry, MD, and Richard L. Wahl, MD

Johns Hopkins Medical Institutions, Baltimore, MD 21287 USA
Performance of Automated Algorithm

Automated Algorithm vs Average of Two Readers, $n=85$
Intra-Class Correlation : 0.88
Successful detection at both thresholds

Original Study  PERCIST Threshold  TLG Threshold

CAD Detected Disease
The PERCIST 1.0 based CAD tool demonstrated a high level of concordance with the expert reader in untreated head and neck cancer, despite the fact that there exists a high level of physiological FDG uptake in the vicinity of these tumors in the brain, palatine tonsils, salivary glands, etc.

that the automated normal reference detection algorithm failed to measure reference metabolic activity in a statistically identical fashion to the expert reader.
Reduction in 2′-deoxy-2′-[18F]fluoro-d-glucose (FDG) uptake is associated with nonprogression.

Zander T et al. JCO 2011;29:1701-1708
Comparison of PET Response Criteria in Solid Tumors (PERCIST) and Response Evaluation Criteria in Solid Tumors (RECIST) in Neoadjuvant Chemotherapy for Esophageal Cancer:

M. Yanagawa, MD, PhD 1), M. Tatsumi 1), MD, PhD, E. Morii, MD, PhD 2), T. Watabe, MD 3), K. Isohashi, MD, PhD 3), H. Kato, MD, PhD 3), E. Shimosegawa, MD, PhD 3), N. Tomiyama, MD, PhD 1), J. Hatazawa MD, PhD 3)

From the Department of Radiology 1), the Department of Pathology 2), the Department of Nuclear Medicine 3), Osaka University Graduate School of Medicine

2011 Correlative Imaging Council Walter Wolf Young Investigator Award
Response-adapted therapy
Aggressive NHL, any stage, any IPI

(R)CHOP for 2 or 3 cycles

PET -
- complete conventional therapy

PET +
- if no disease progression
  - (R)ESHAP or (R)ICE x 2
- High dose therapy and ABMT
QIBA

- Formally began May, 2007
- Mission: Improve value and practicality of quantitative imaging biomarkers by reducing variability across devices, patients, and time.
  - Build “measuring devices” rather than “imaging devices”.
  - “Industrialize imaging biomarkers”.
Quantitation Practicalities

• **Feasible**
  - an idea or program whose end goals can likely be achieved in a specific timeframe and that has a reasonable prospect of producing the expected outcomes; ideal programs are those which could result in regulatory qualification of a biomarker in three years.

• **Practical**
  - leverages preexisting resources (e.g., intellectual capital, personnel, facilities, specimens, reagents, data) wherever possible; warrants access to RSNA resources and support.

• **Collaborative**
  - would uniquely benefit from the multi-stakeholder composition and approach of QIBA and could be feasibly executed under its policies e.g. resulting in extension or adoption in product development among hardware, software, or imaging agents. The biomarker has the support of the stakeholder community with the organizational impetus to sustain continued efforts.
Quantitation and PERCIST

- Quantitation appears important in early response assessment
- Methods must add clinical value
- Methods must be quick and robust for routine clinical deployment
- Quantitation can add consistency across readers and help less experienced readers
- Quantitative methods should allow earlier and more precise assessments of subtle physiological changes and allow earlier changes in therapy
- European adoption of UPICT protocol and evaluation of PERCIST 1.0 encouraged
Thanks to many:

- Jeff Leal
- Martin Lodge
- Heather Jacene
- Yvette Kasamon
- Dan Sullivan