Standardization of Response Assessment: Methods, Analysis and Reporting: From RECIST to "PERCIST 1.0"

Richard L. Wahl, M.D.
Division of Nuclear Medicine
Johns Hopkins Medical Institutes
Baltimore, MD

PET/CT is a Qualitative and a Quantitative Method

- Most applications to date have been Qualitative
- In treatment response assessment, especially if looking for small treatment induced changes, Quantitation will be needed
- Quantitation requires greater attention to technical details than qualitative imaging
- **Standardization** of methods—including **Analytical** Methods is required

No standardized and validated quantitative metabolic response criteria exist

EORTC 1999

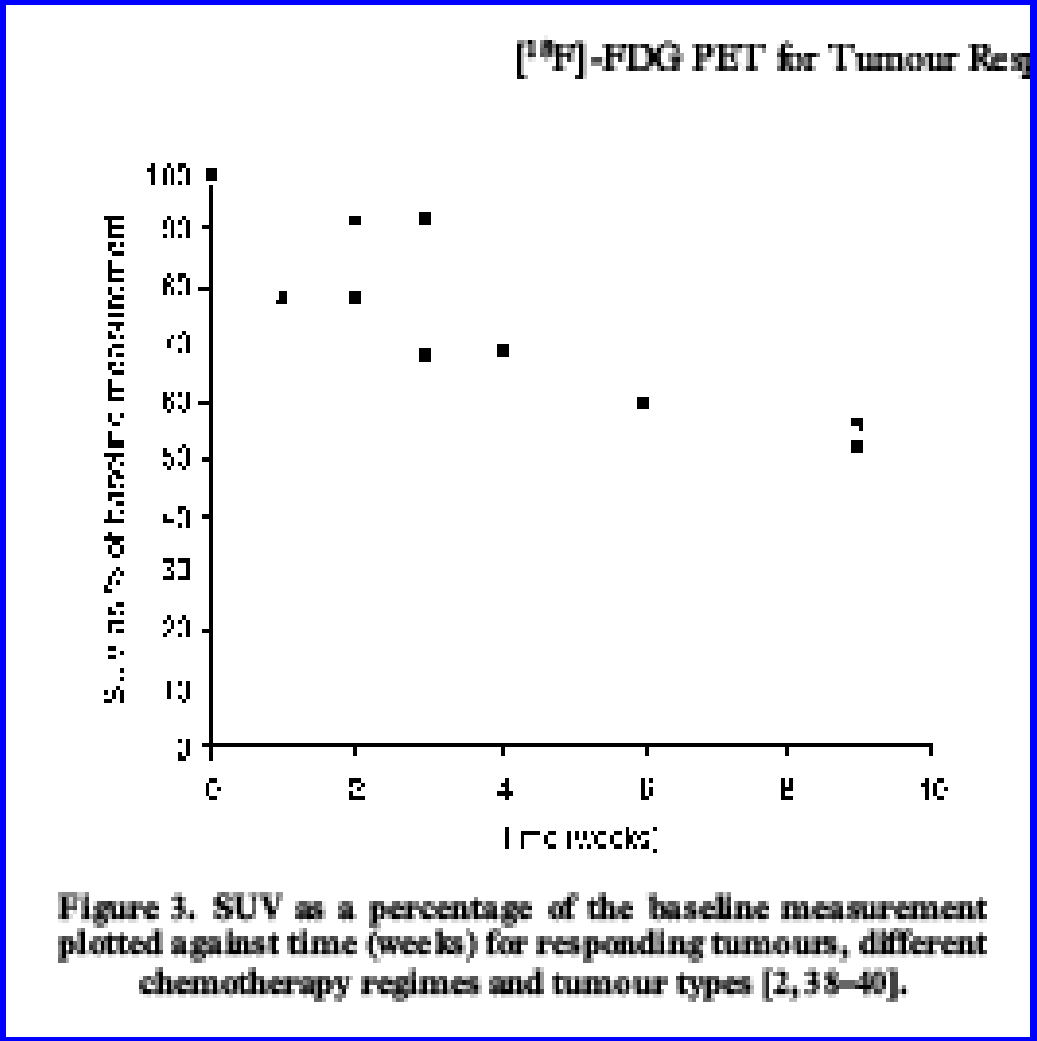


Figure 3. SUV as a percentage of the baseline measurement plotted against time (weeks) for responding tumours, different chemotherapy regimes and tumour types [2,38–40].

EORTC Response Criteria

CR	Complete Disappearance of all Metabolically Active Tumor (i.e. decreased to background levels)
PR	>15% decline in SUV after 1 cycle, >25% decline after 2 or more cycles. Reduction in extent (size) of FDG uptake is not required
SD	Increase in FDG SUV of <25% or decrease of <15% in SUV and no increase in extent of uptake (<20% in longest dimension)
PMD	Increase in SUV of over 25%, Increase in extent of FDG uptake by >20%, New FDG positive metastases

EORTC Criteria Limitations

Individual tumor changes in SUV

Retrospective data

Limited

- patient number
- tumor and treatment types

Questions Remaining?

What Size ROI

What SUV value

Best cut-off values for SUV

- percent change vs. absolute floor
- different diseases and treatments

Number of lesions

How to calculate percent change

Discrepancies on PET & anatomic imaging

If there were RECIST Criteria for PET, they Would be Defined as…

“PERCIST”

- Positron
- Emission
- Response
- Criteria in
- Solid
- Tumors

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

Richard L. Wahl[1,2], Heather Jacene[1], Yvette Kasamon[2], and Martin A. Lodge[1]

J Nucl Med 2009 50: 122S-150S

Standardized PET Techniques

National Cancer Institute

Shankar et al. *J Nucl Med* 2006;47:1059-66

Netherlands protocol

Boellaard et al. *Eur J Nucl Med Mol Imaging* 2008;35:2320-33

Key Elements of PERCIST

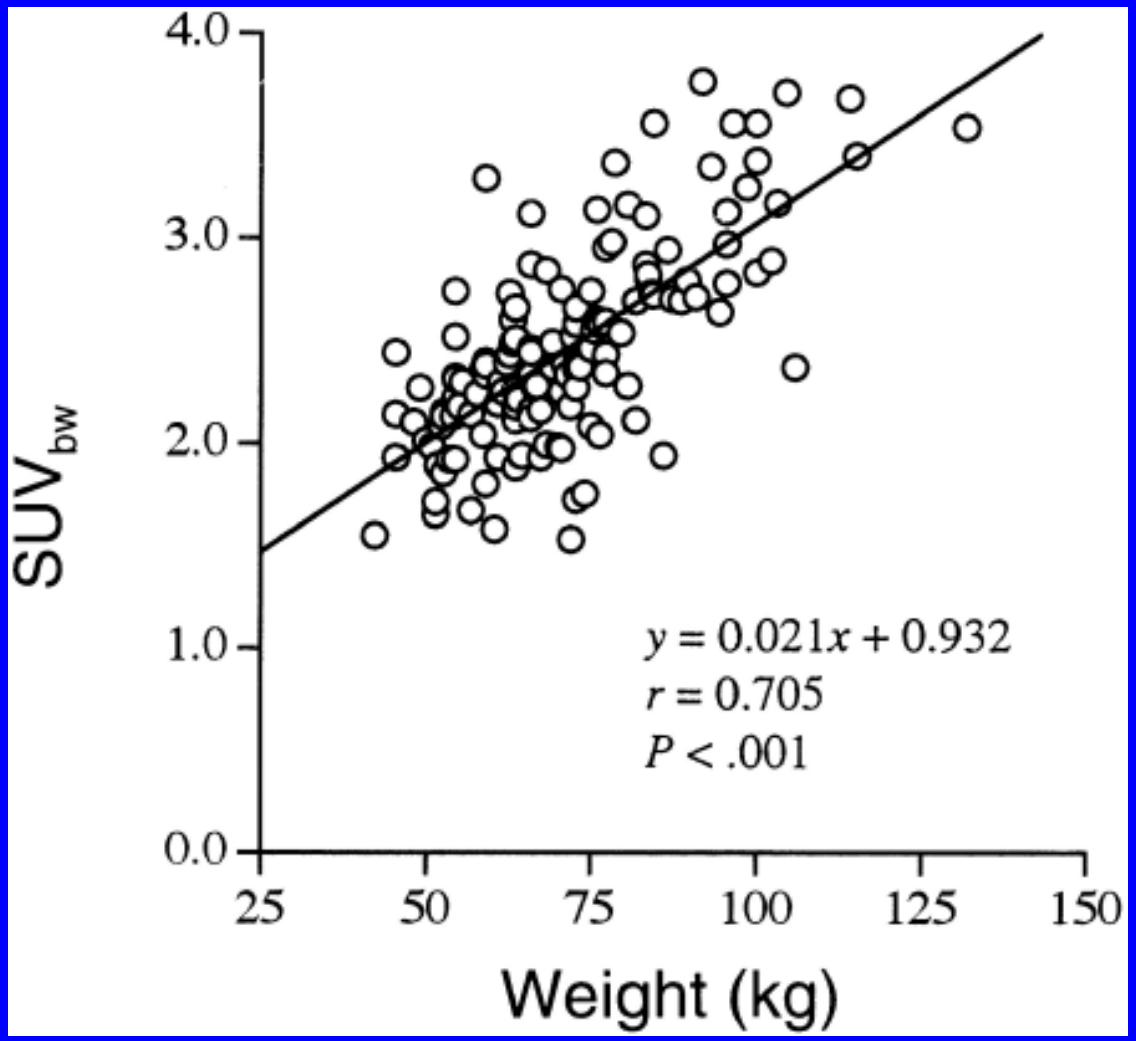
- Suggested Standardization of:

- Tumor to be assessed (hottest, statistical considerations applied vs background)
- Size of ROI applied, Type of SUV (SUV lean)
- Timing of scan (explicit allowances)
- Quality of Data: including normal tissue reference region
- Reporting: Continous, Timing, Normal tissue

What is Measured?

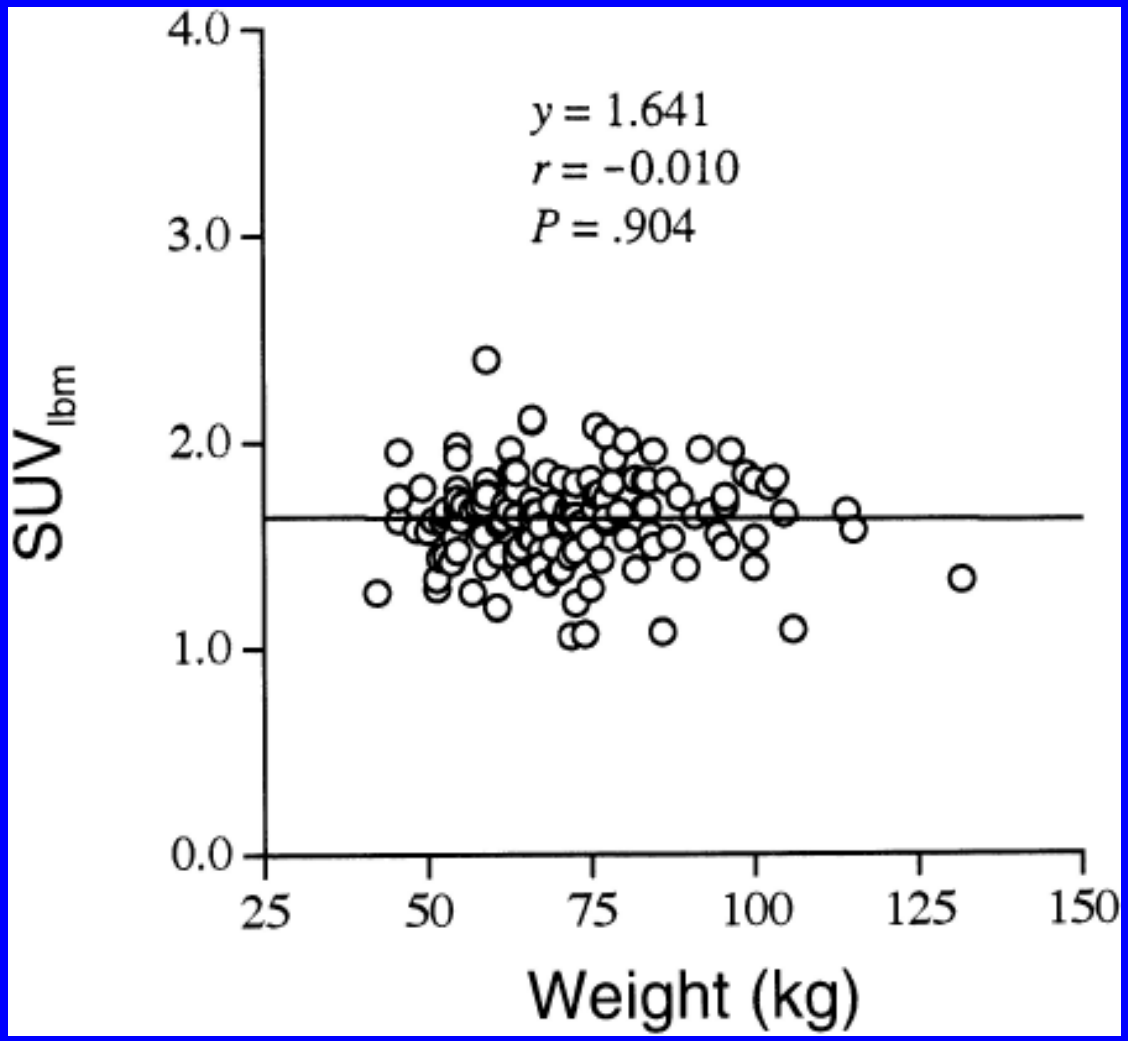
- **Lean Body Mass Corrected SUV**
- **SUL Peak**
- **In Hottest Tumor Focus**

Figure 1a. Graphs depict the relationships between patient body weight and blood SUVs: (a) SUVbw, (b) SUVibw, (c) SUVlbm, or (d) SUVbsa



Sugawara, Y. et al. Radiology 1999;213:521-525

Figure 1c. Graphs depict the relationships between patient body weight and blood SUVs: (a) SUVbw, (b) SUVibw, (c) SUVlbm, or (d) SUVbsa

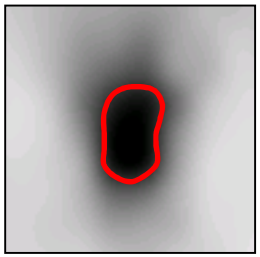


Sugawara, Y. et al. Radiology 1999;213:521-525

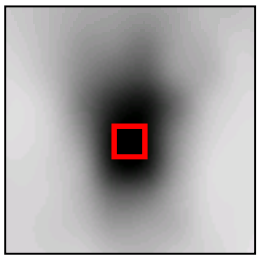
Radiology

Introduction

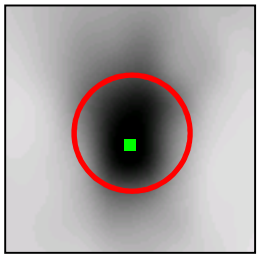
- A number of different ROI definitions have been employed including:
  - Mean within an irregular ROI defined by isocontours.
  - Mean within a fixed size ROI centered on the most metabolically active region.
  - Maximum pixel within a large ROI encompassing the entire tumor.
- SUV_{max} has been widely used although single pixel measurements of this sort may be compromised when images have high levels of noise.



Isocontour



Fixed size



Maximum pixel

Quantifying Metabolic Tumor Response to Therapy: The Influence of Image Noise on Maximum and Mean SUV.

MA Lodge, J P Leal, RL Wahl

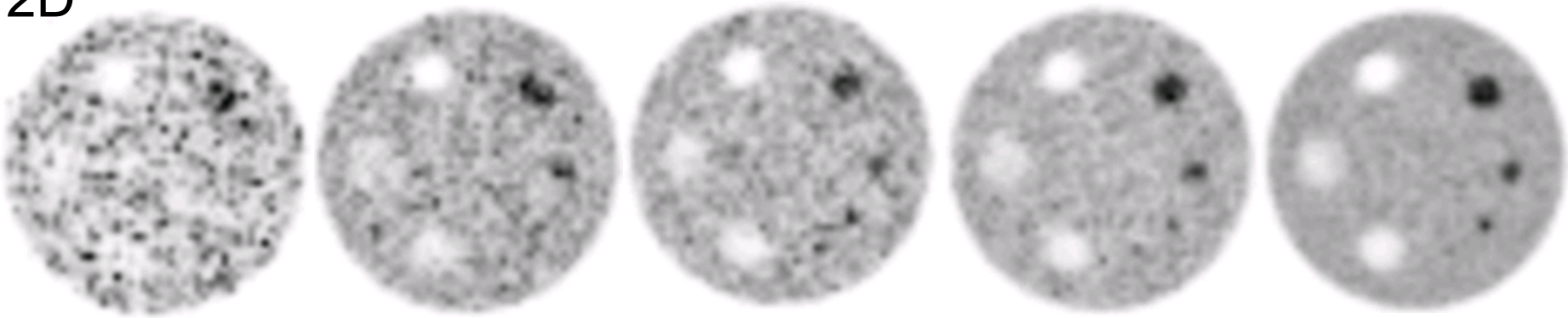
Russell H. Morgan Department of Radiology and Radiological Sciences

Johns Hopkins University School of Medicine

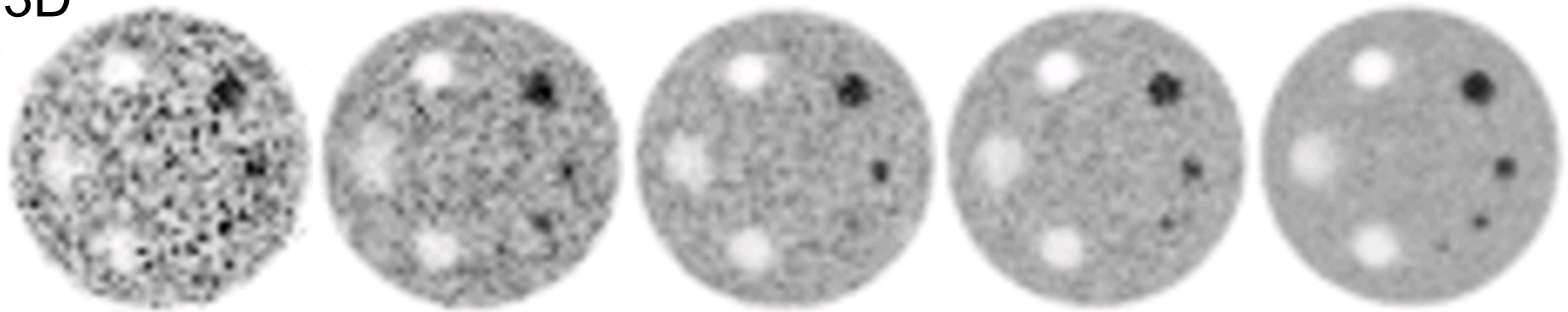
Baltimore, MD

Results: Statistical Quality

2D



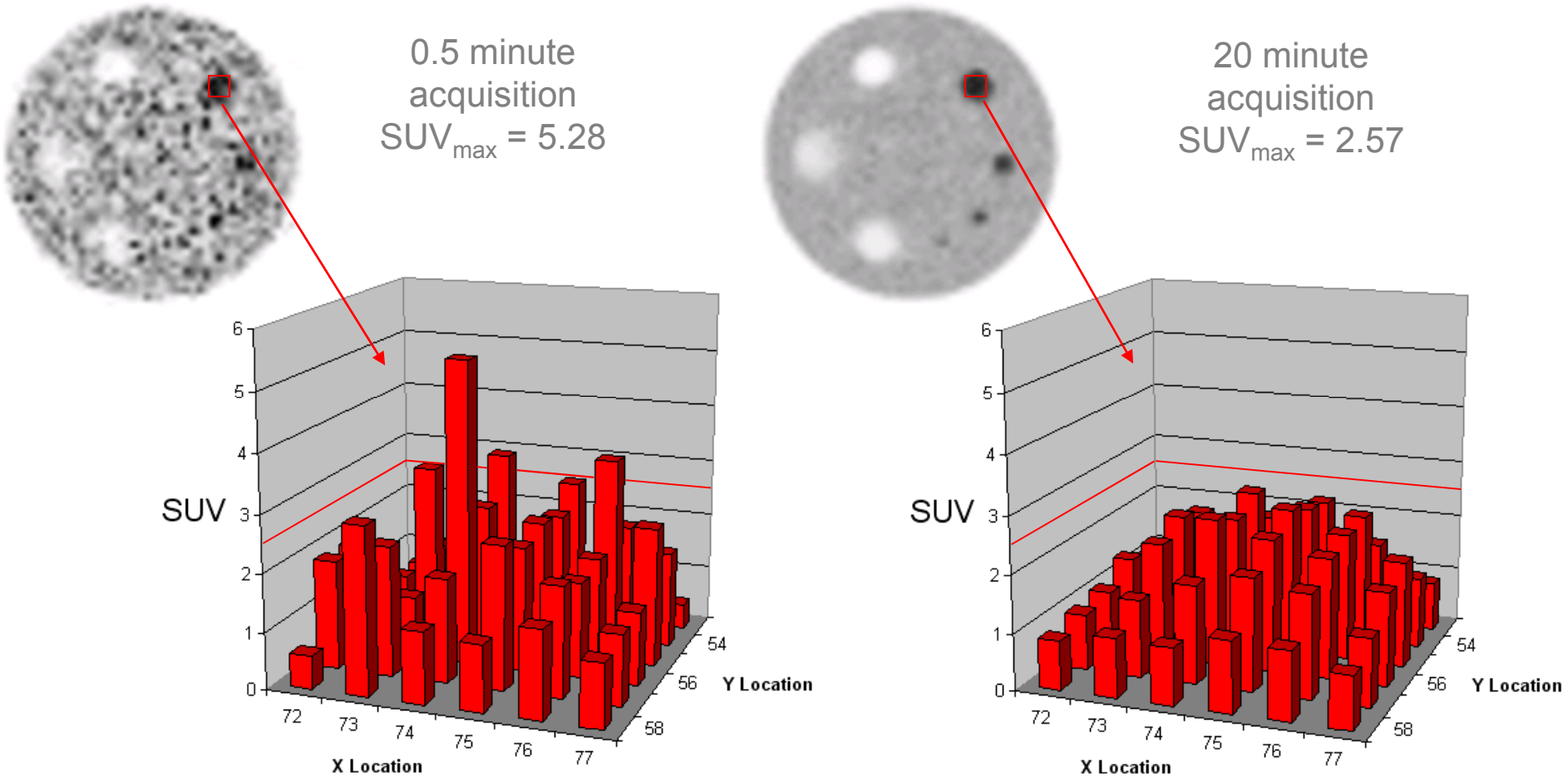
3D



0.5 min 2 min 3.5 min 5 min 20 min

- Statistical quality of the images deteriorates with decreasing scan duration for both 2D and 3D.
  - Decay of the isotope also contributes to increasing noise.

Results: ROI_{max}



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

Conclusions

ROI_{max}

- Maximum pixel within lesion.

- Increasing positive bias as noise increased.

$ROI_{42\%}$

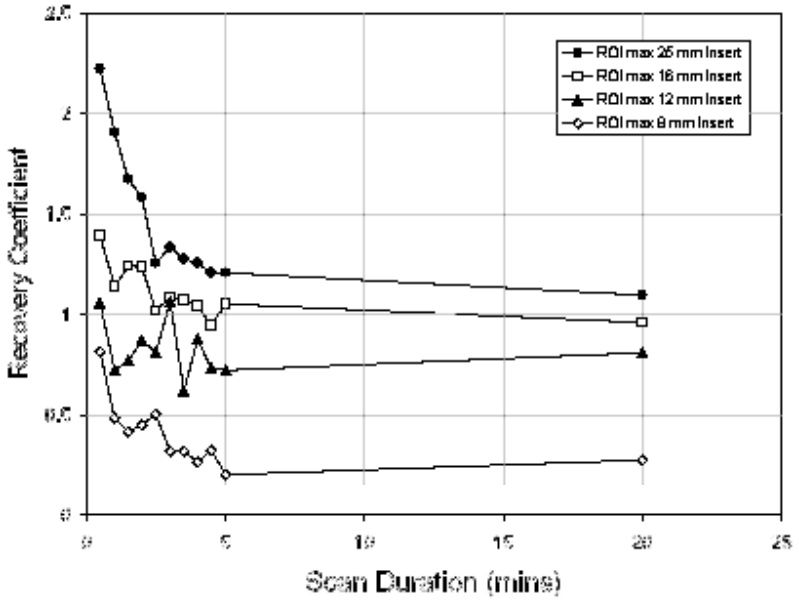
- Mean of all pixels within an irregular ROI based on an isocontour at 42% of the maximum pixel.

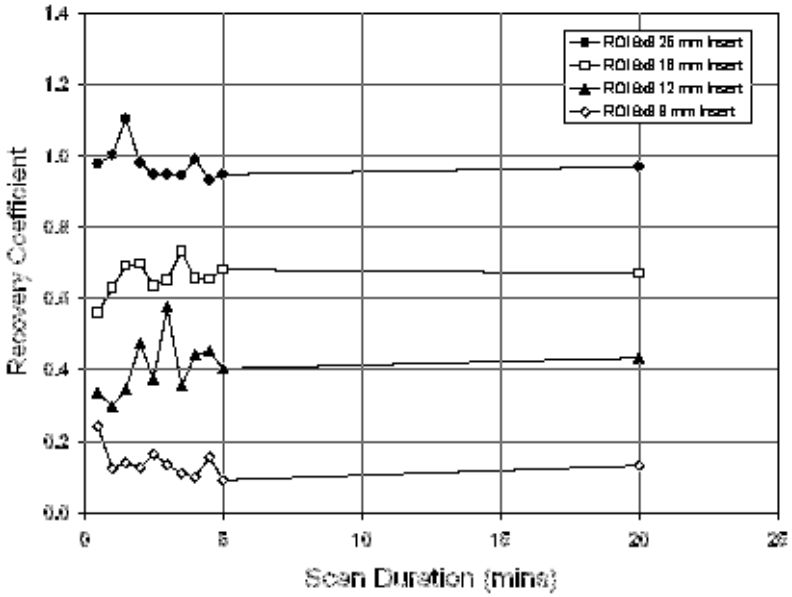
- Increasing positive bias as noise increased.

ROI_{9x9}

- 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*

Nahmias, Wahl LM. *J Nucl Med* 2008;49:1804
Boellard et al. *J Nucl Med* 2004;45:1519

SUL peak

1.2 cm diameter (1 cm^3 volume sphere)

Centered around hottest area in tumor

Standardizes ROI size

Maybe less variance than SUL max

Intra-subject Variability of SUV

Same tumor measured multiple times

Large tumors with high metabolism

Best ~ 6-10%

Worst ~43%

Larger fixed ROIs more reproducible

Current scanners smaller voxel size

Minn et al. *Radiology* 1995;196:167
Weber et al. *J Nucl Med* 1999;40:1771
Nakamoto et al. *Mol Imaging Biol* 2002;4:171

Other methods

Threshold

Varies w/variability of single pixel max

Tumor lesion glycolysis

- Less practical
- Based on threshold method
- Exploratory
- Still promising

Partial-volume Effect

Not included in PERCIST

Size measurements

- possible with PET/CT
- errors – particularly small tumors

No standardized “accepted” method

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
Hamberg et al . *J Nucl Med* 1994; 35:1308
Weber WA et al. *J Nucl Med* 1999; 40:1771
Torizuka T et al. *Radiology* 1997; 203:169
Jaskowiak CJ et al . *J Nucl Med* 2005; 46:424
Schoder H, et al. *J Nucl Med* 2004; 45:559

Consistency in SUV lean

- Same Scanner or Model
- Calibrated
- Same software
- Similar technical parameters

Sugawara et al. *Radiology* 1999; 213:521
Hamberg et al . *J Nucl Med* 1994; 35:1308
Weber WA et al. *J Nucl Med* 1999; 40:1771
Torizuka T et al. *Radiology* 1997; 203:169
Jaskowiak CJ et al . *J Nucl Med* 2005; 46:424
Schoder H, et al. *J Nucl Med* 2004; 45:559

What is Measured?

- **PERCIST 1.0**1. Measurable target lesion is hottest **single** tumor lesion SUL of “maximal 1.2 cm diameter volume ROI in tumor” (SUL peak). SUL peak is at least 1.5 fold greater than liver SUL mean+ 2SD(in 3 cm spherical ROI in normal right lobe of liver).

- If liver is abnormal, primary tumor should have uptake >2.0 times SUL mean of blood pool in a 1 cm diameter ROI in the descending thoracic aorta extended over 2 cm Z axis.

- The tumor with the maximum SUL peak is assessed post -treatment. While typically this is in the same region of the tumor with the highest SUL peak at baseline , it need not be.

What is Measured?

- Uptake measurements should be made for the peak and maximum single voxel tumor SUL.
- Other SUV metrics including SUL mean at 50 or 70% of SUV peak can be collected as exploratory data, TLG can be collected ideally based on voxels more intense than 2SD above liver mean SUL
- These parameters on up to 5 measurable target lesions can be recorded as exploratory data, typically the 5 hottest lesions, which are typically the largest, not over 2 /organ. Tumor size of these lesions can be determined per RECIST 1.1.

“Measurable Lesions”

FDG uptake and not tumor size

Minimum uptake

- 1.5 x liver SUL mean + 2 SD

To allow for sufficient fall in SUL post-treatment

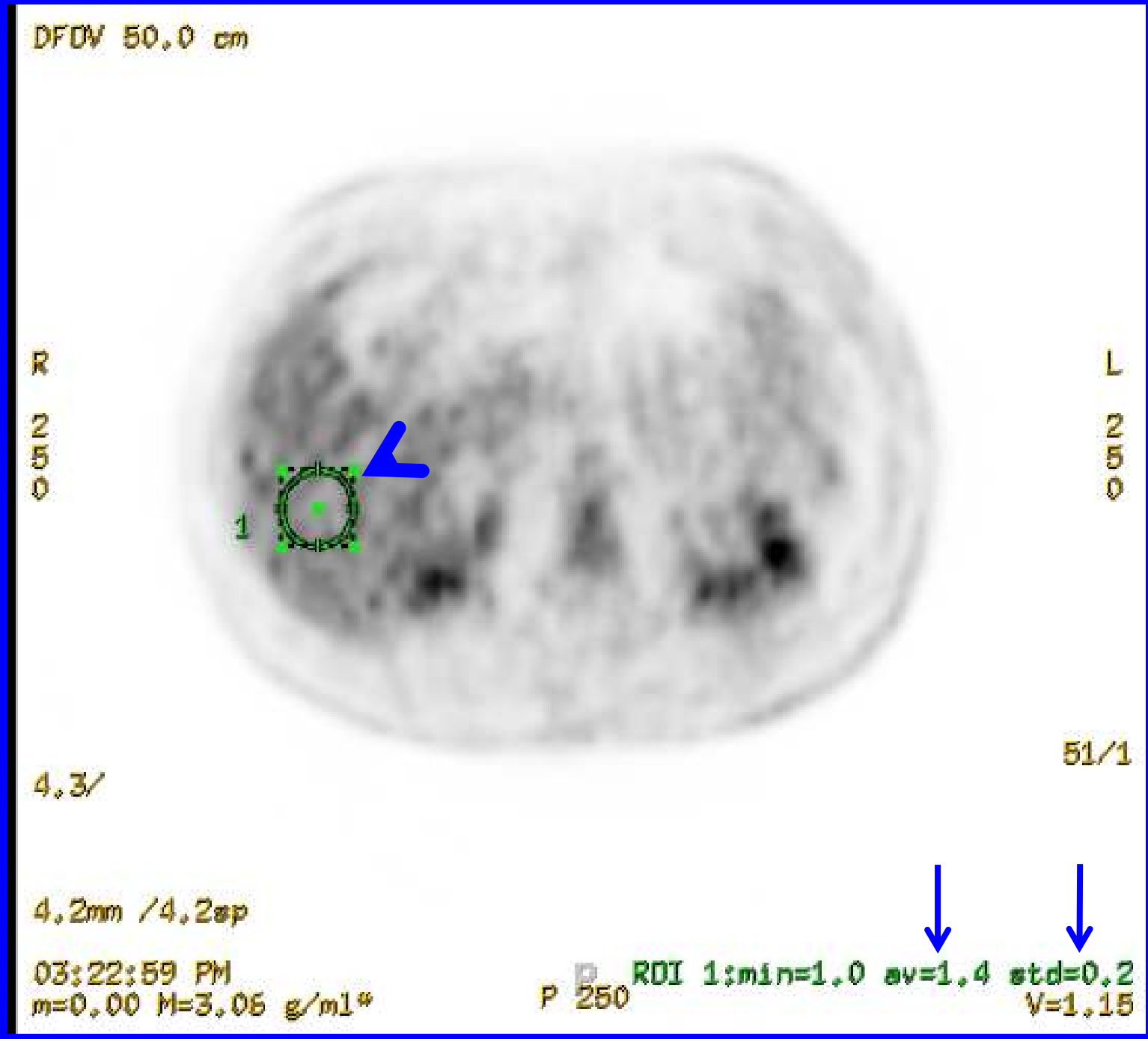
Alternatives:

- 2 x blood pool SUL mean + 2 SD
- 1.35 x hepatic uptake (prob too low as exploratory alternative)

Normalization for Quality Control

- -Normal liver SUL must be within 20% (and <0.3 SUL mean units) for baseline and follow up study to be assessable.-If liver is abnormal, blood pool SUL must be within 20% (and <0.3 SUL mean units) for baseline and follow up study to be assessable.-

- Uptake time of baseline study and follow up study 2 must be within 15 minutes of one another to be assessable. Typically, these are at a mean of 60 minutes post injection, but not less than 50 min post injection. –


DFOV 50.0 cm
R
250
L
250
1
51/1
4.3/
4.2mm /4.2sp
03:22:59 PM
m=0.00 M=3.06 g/ml*
P 250
ROI 1:min=1.0 av=1.4 std=0.2
V=1.15

Normal Background Scan to Scan Difference

Within

± 20% and

0.3 SUL units

Comparable uptake times!

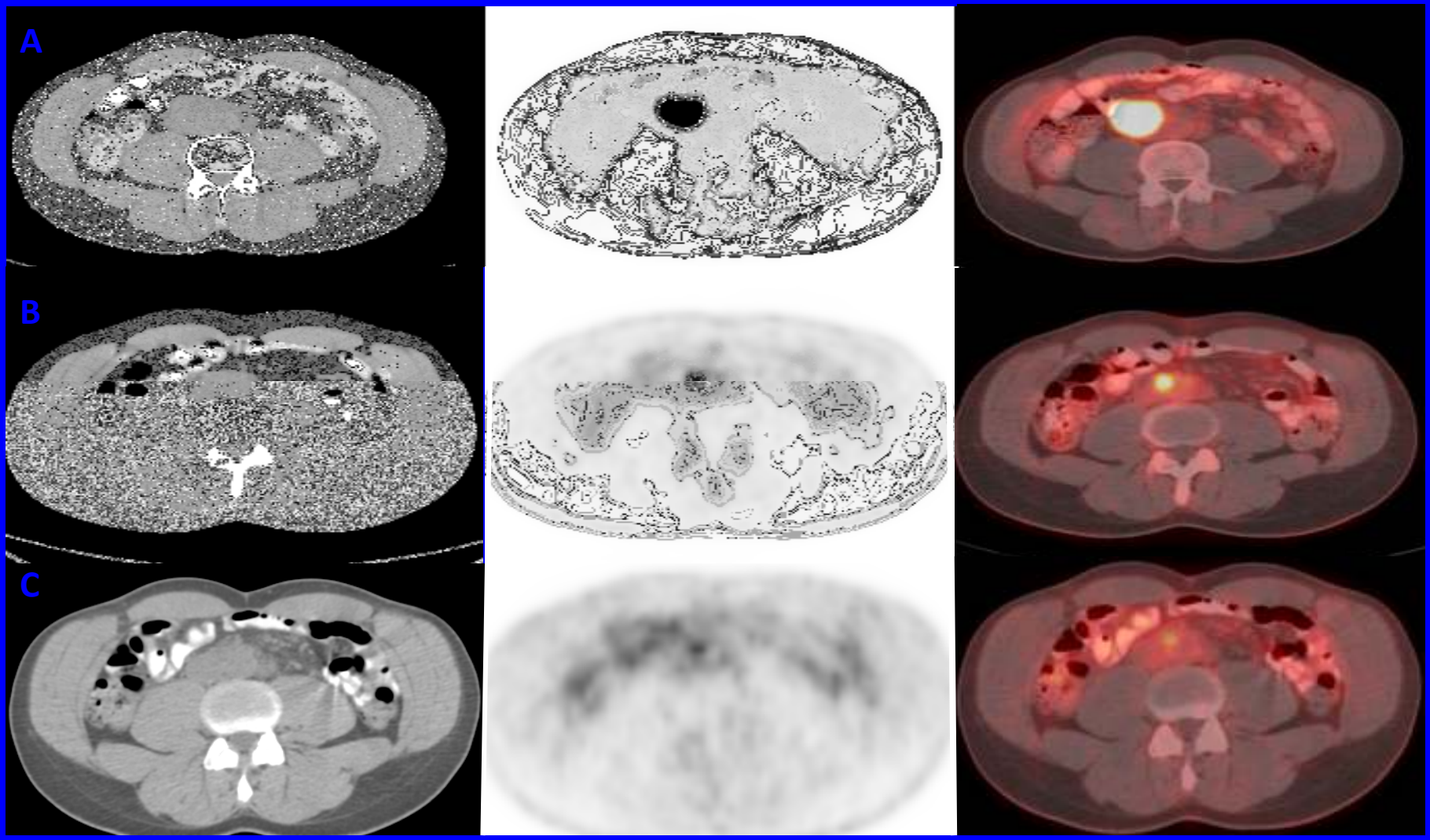
Complete Metabolic Response

- **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 Metabolic Response (PMR)

- Reduction of a minimum of 30% in target measurable tumor FDG SUL peak. Absolute drop in SUL must be at least 0.8 SUL units, as well.

- Measurement is commonly in the same lesion as the baseline, but can be another lesion if that lesion was previously present and is most active lesion after treatment.

- ROI does not have to be in precisely the same area as baseline scan, though typically it is.

- No increase, >30% in SUL or size of target or non target lesions (i.e. no PD by RECIST or IWC) (If PD anatomically, must verify with follow up).

A reduction in the extent of the tumor FDG uptake is not a requirement for partial metabolic response. % decline in SUL should be recorded as well as (ideally) time in weeks after treatment was begun (i.e. **PMR -40, 3). No new lesions.**

Partial Response

↓30% SUL peak

- EORTC: 15-25%
- 10-20% variability of SUV
- Lower thresholds, medically relevant
- 25% of a low number not much change

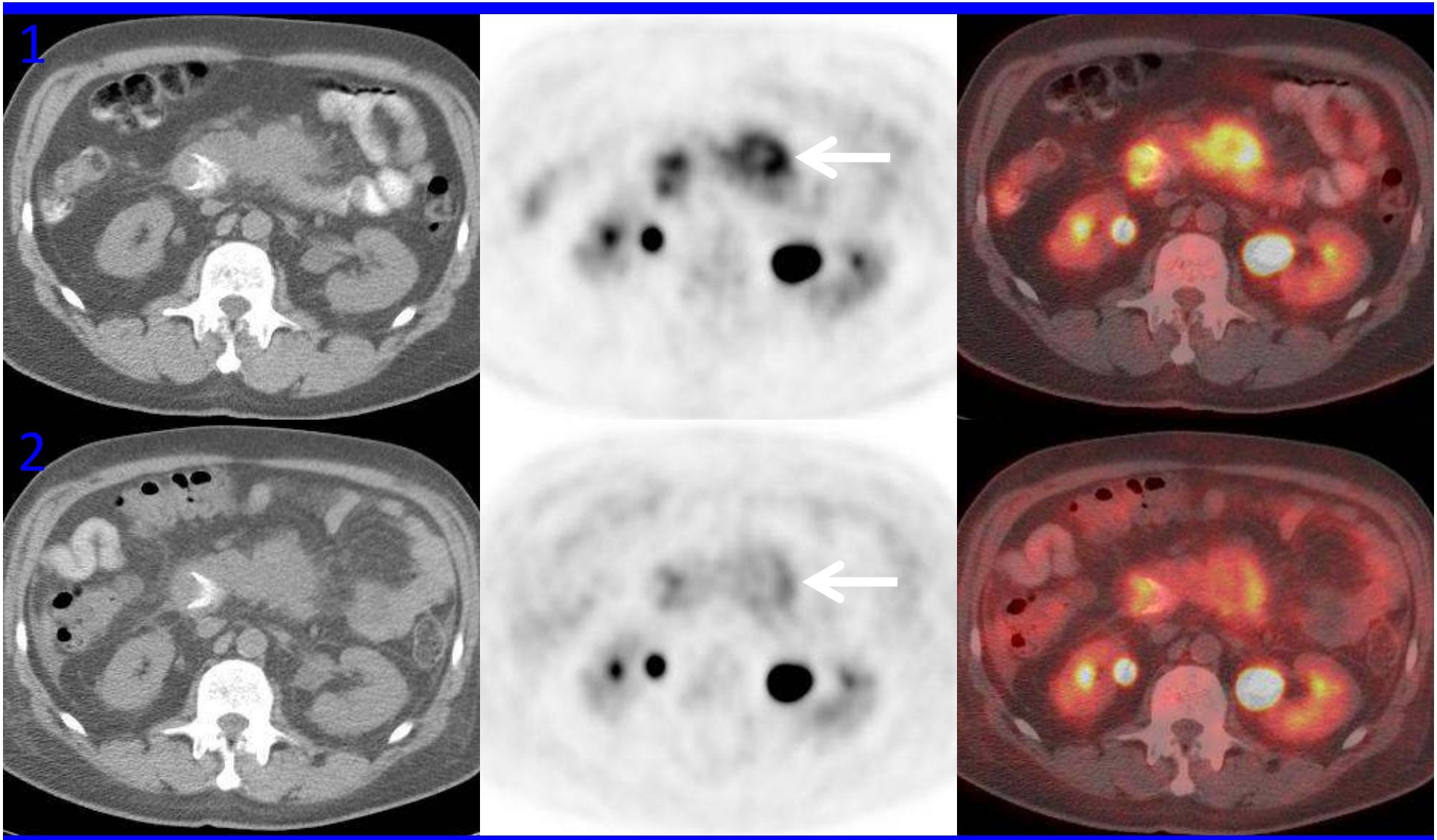
↓0.8 SUL units

- 0.9 and 0.5 SUV units previously proposed*

*Weber et al. *J Nucl Med* 1999;40:1771
*Nahmias et al. *J Nucl Med* 2008;49:1804

Partial Response

Target	All others
↓30% SUL_{peak}	No ↑30% SUL or size
↓0.8 SUL units	No new FDG avid lesions
	Anatomic PD – verify


1
2

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**

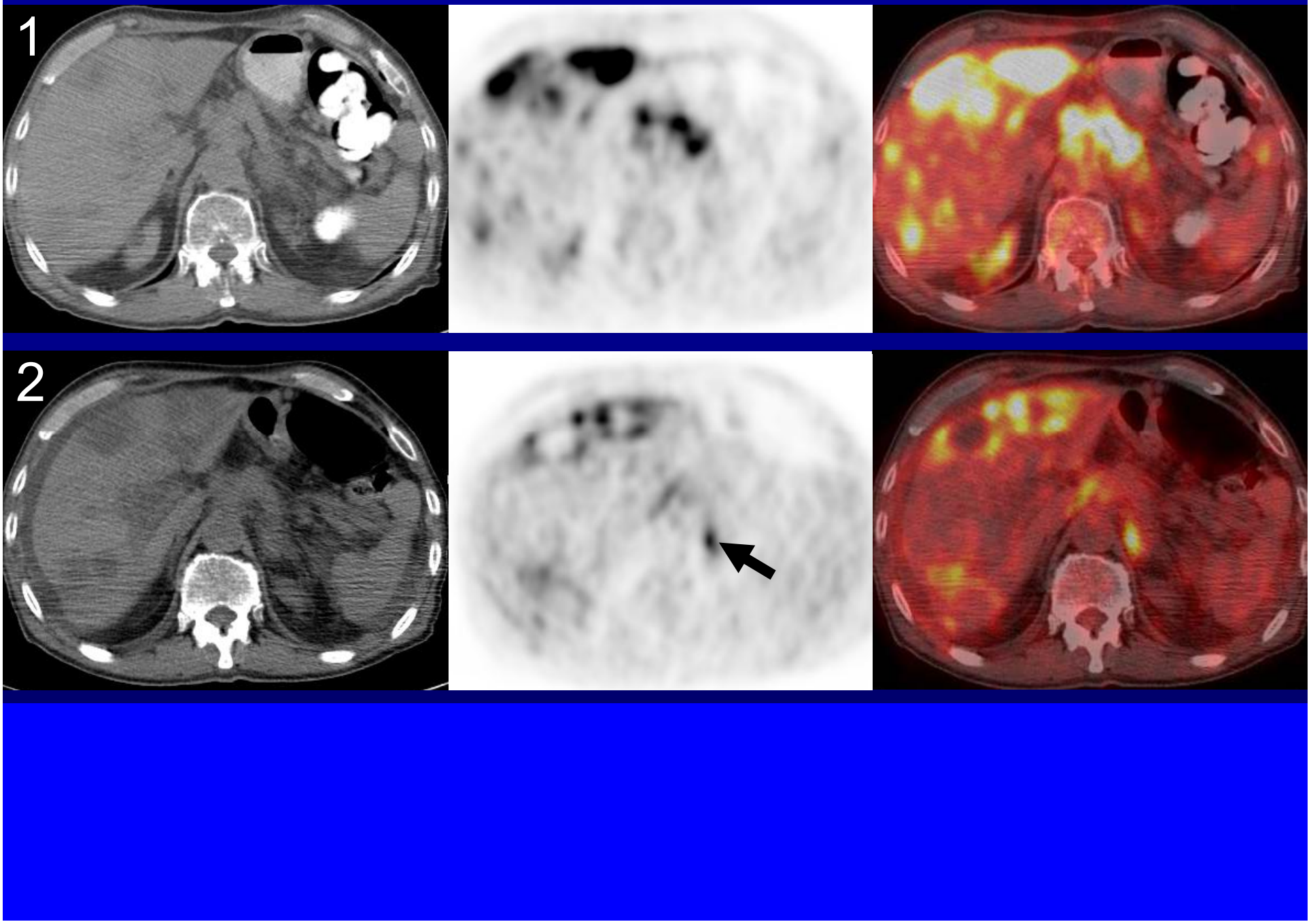
Progressive metabolic disease

- >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**).


1
2

PERCIST Continuous Response Scale

- SUL is a continuous variable
- Dividing response criteria into a limited number of somewhat arbitrary response categories loses much data.
- PERCIST preserves percent declines in the SUV peak in each reported category.
- The rapidity with which a scan normalizes is important (faster appears better), PERCIST 1.0 includes time from start of treatment as part of the reporting.
- 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.

Number of Lesions: PET

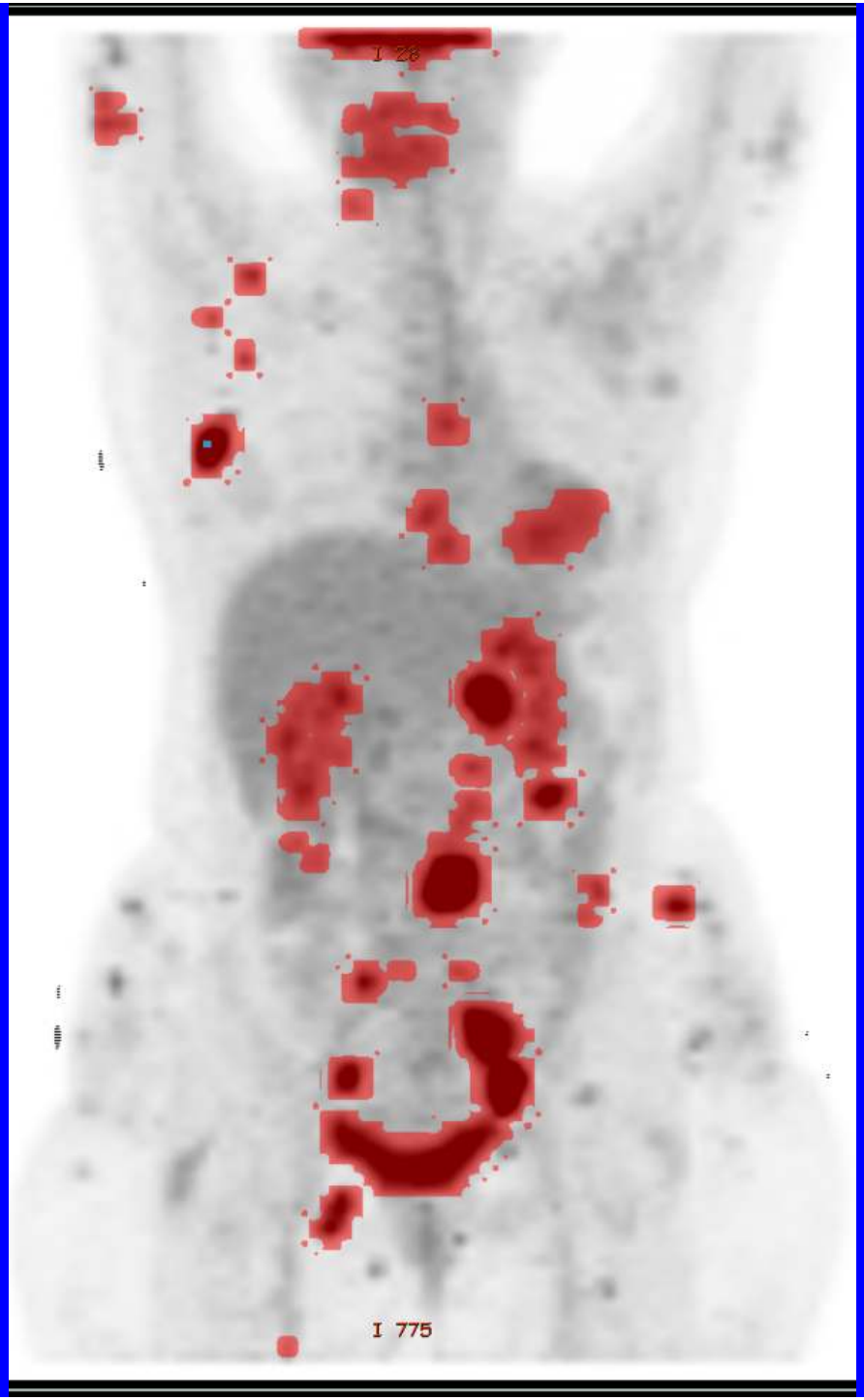
Early studies: 1 large lesion

Not specified in 1999 EORTC criteria

Various approaches

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.
- An 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 and whether summed TLG of all lesions is superior


I 28
I 775

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

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

Jacene et al. *J Nucl Med* 2009;50:8

Correlation % Change in ∑SUL

No. Lesions	5	3	1
6	0.99	0.98	0.80
5		0.97	0.81
3			0.85

$p<0.01$

Jacene et al. *J Nucl Med* 2009;50:8

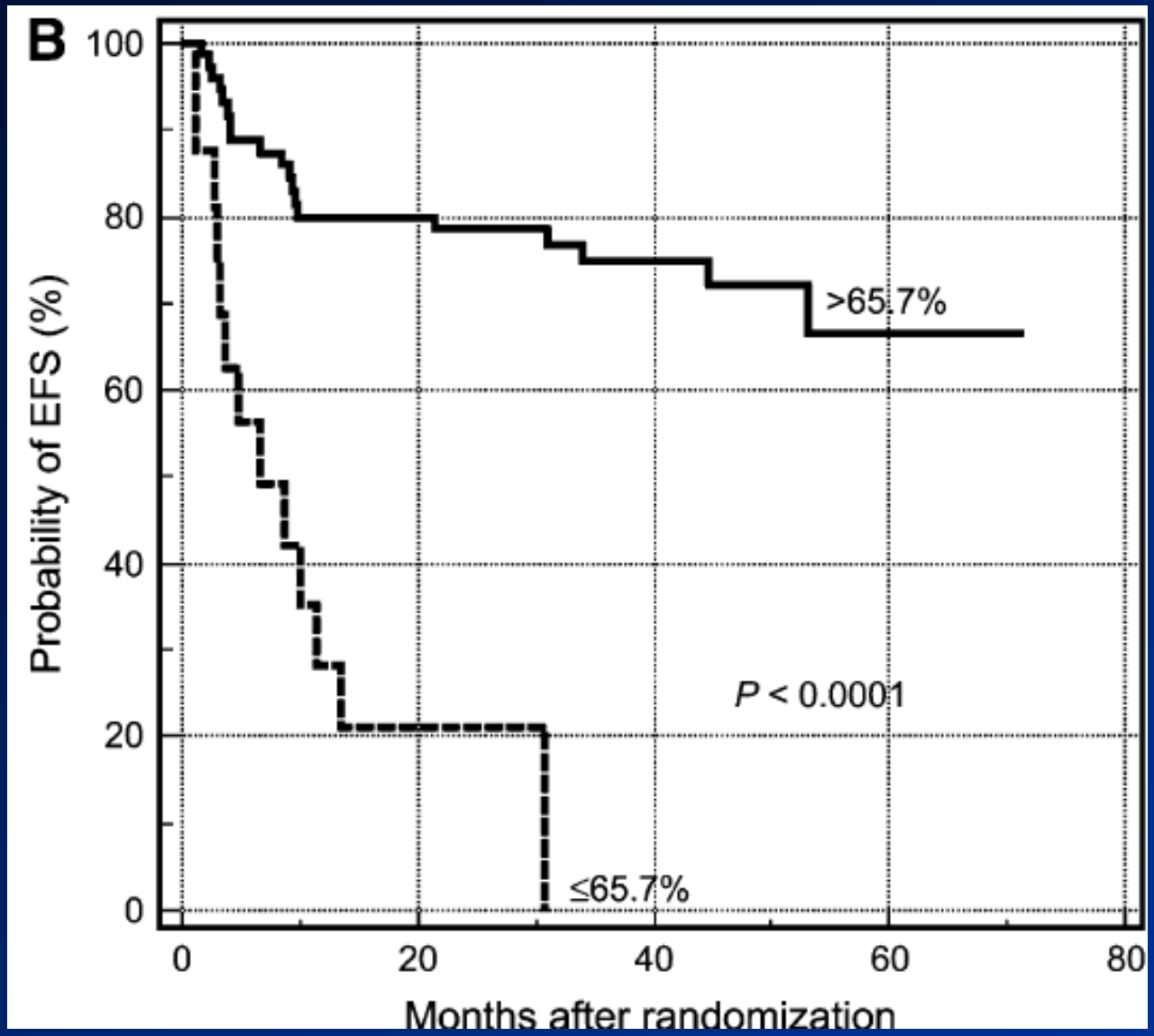
Biologically Relevant?

1^{o} response predicts outcomes in metastases

Same lesions before and after treatment

Worst responding lesion

- lesion with least change
- highest uptake before and after treatment



18%

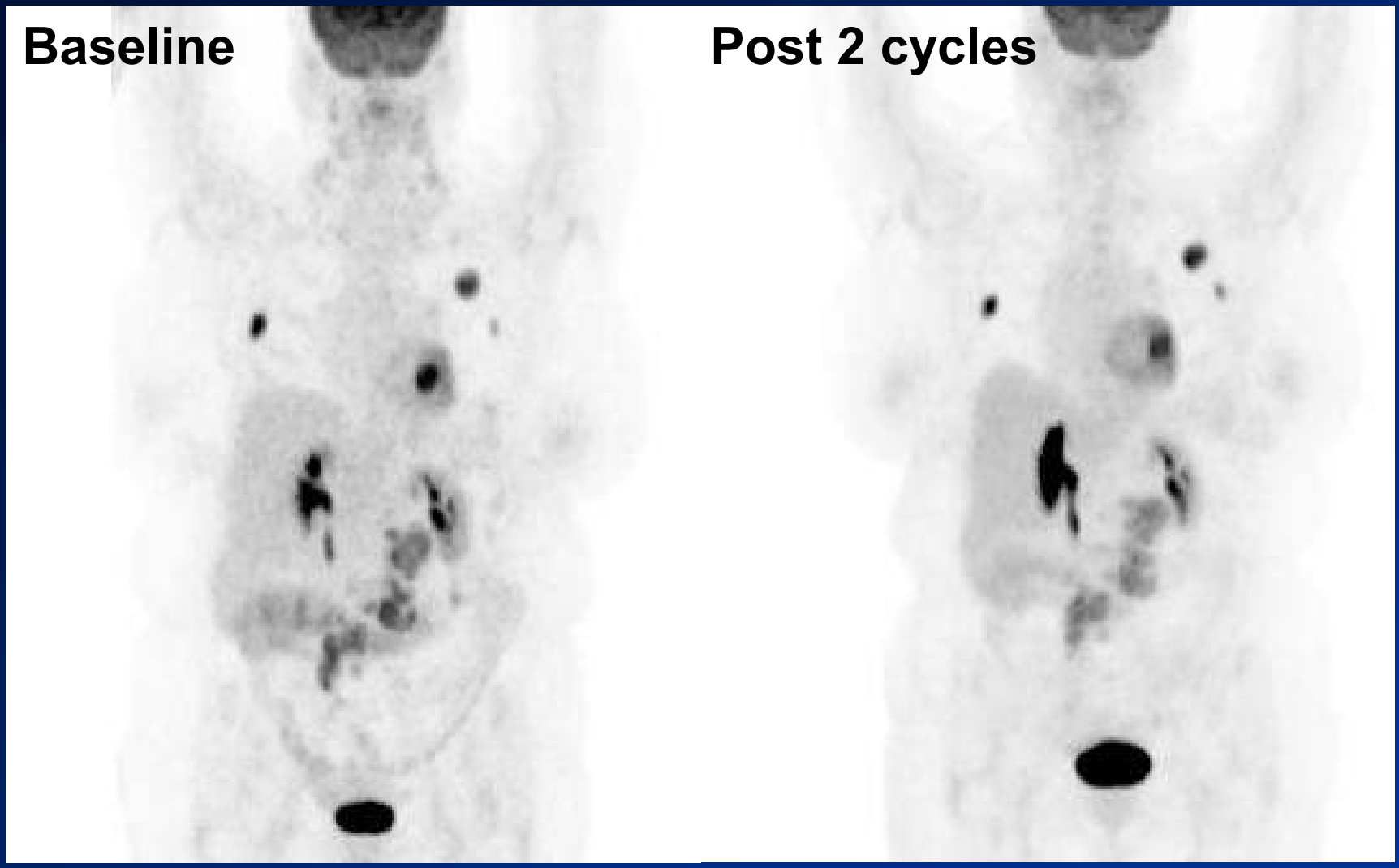
Lin et al. *J Nucl Med* 2007;48:1626

PERCIST: Primary Response Analysis

- Single hottest lesions
- Percent change in SUL peak

Elements of Reporting

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



SMD, -20, 8

Exploratory Analyses

SUL peak for up to 5 lesions

Change in summed SUL

Total lesion glycolysis

Some of the Limitations

Actually getting SUL peak

Is the minimum value too high?

Lack of good data for progression

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

Assessment of Inter-observer Reproducibility in Quantitative FDG PET and CT Measurements of Tumor Response to Therapy

HA Jacene, S Leboulleux, S Baba, D Chatzifotiadis, B Goudarzi, O Teytelbaum, Horton, I Kamel, K Macura, H Tsai, J Kowalski and RL Wahl

J Nucl Med 2009;50:1760-9.

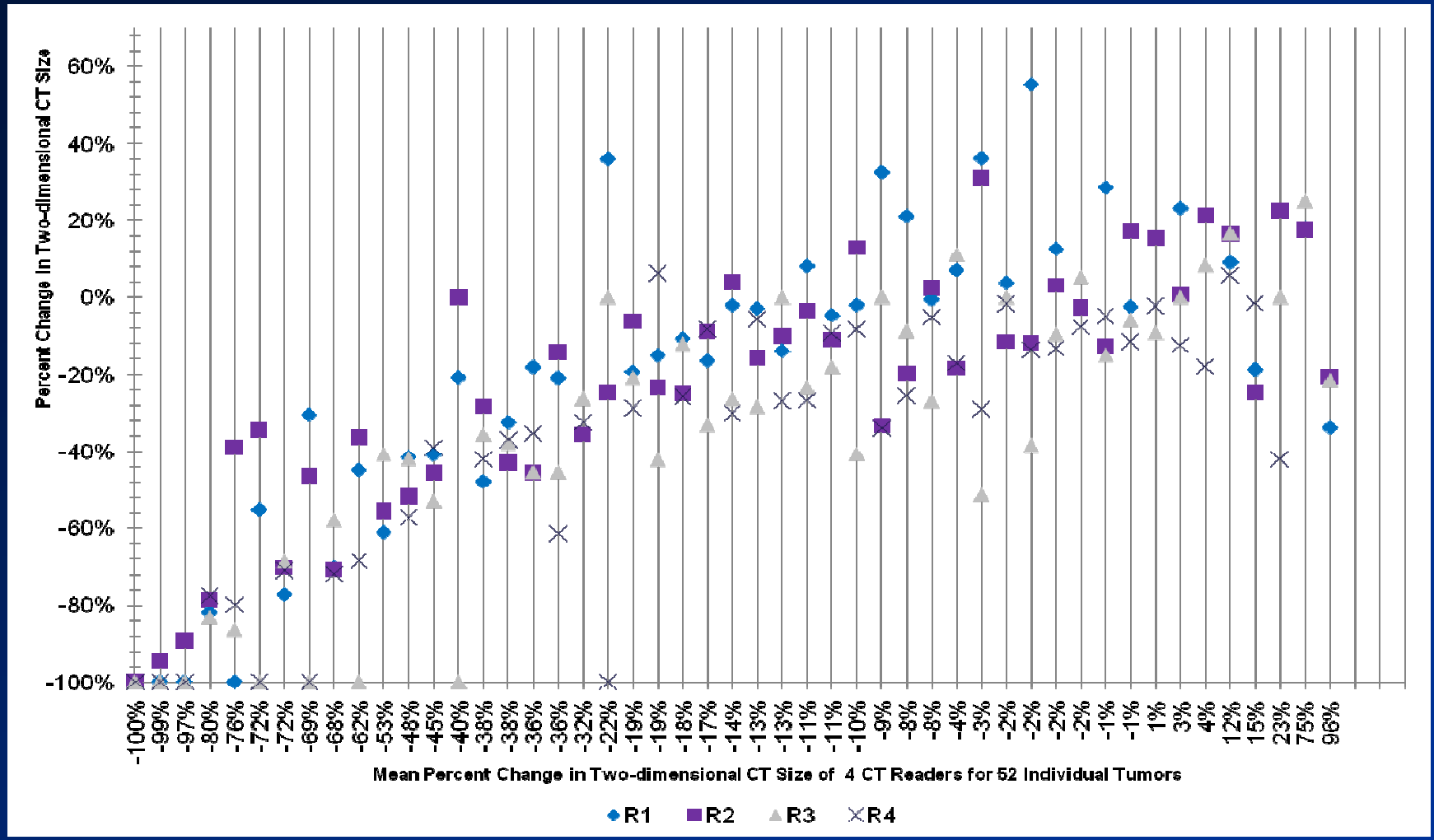
Objectives

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

Jacene et al. *J Nucl Med* 2009;50:1760-9

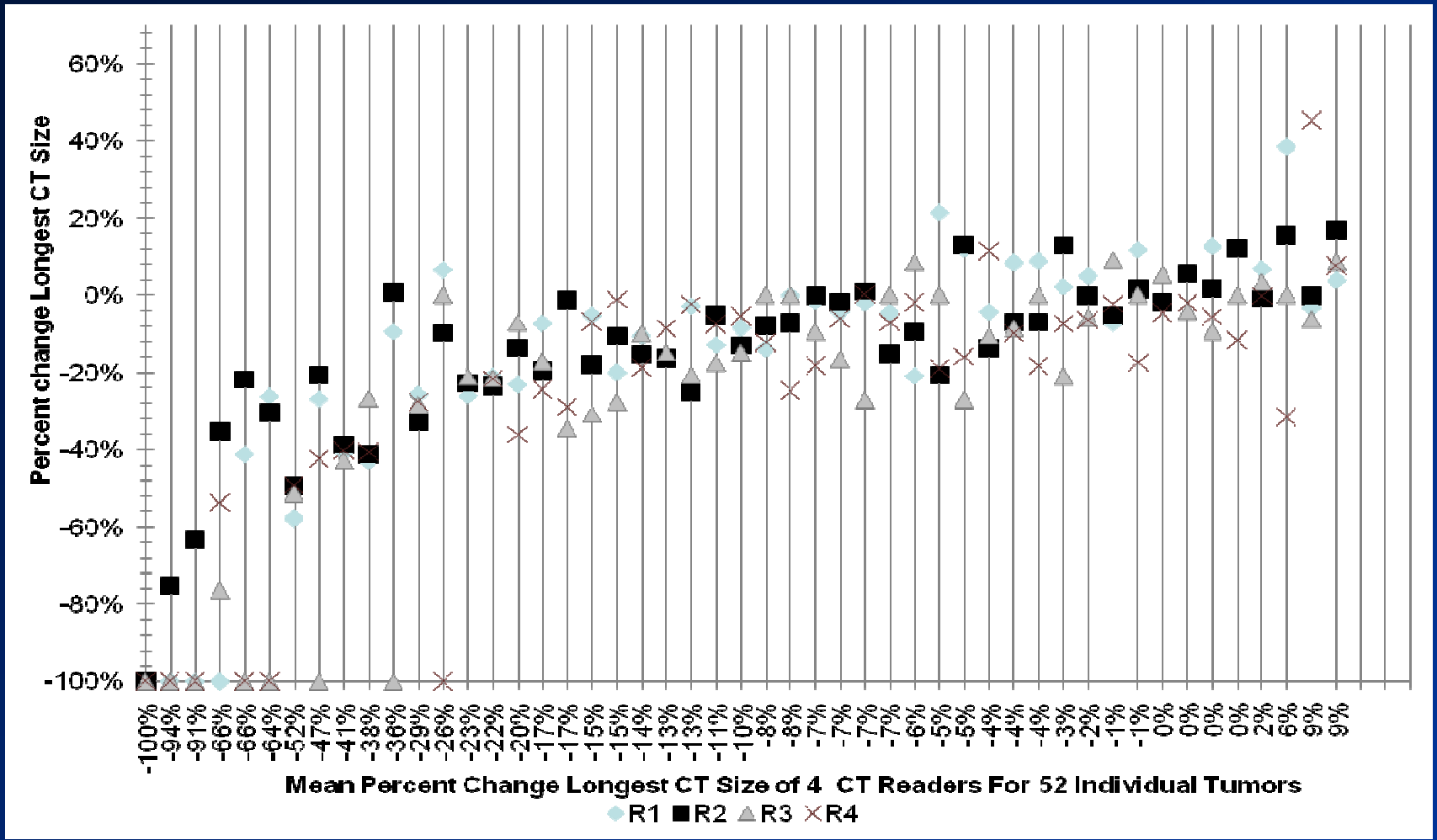
Percent change 2D CT size



ICC – 0.33

Jacene et al. *J Nucl Med* 2009;50:1760-9

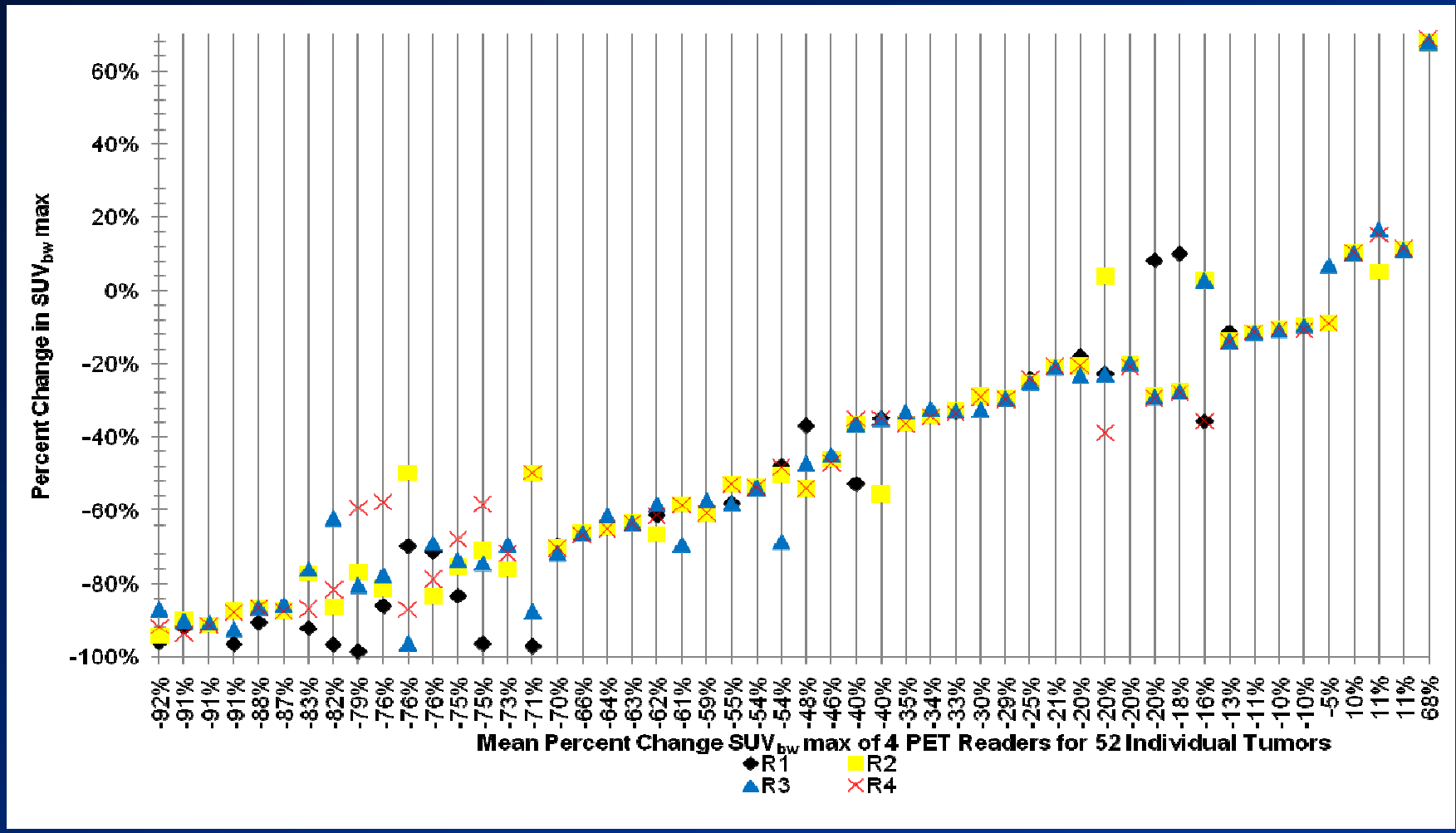
Percent change Longest CT size



ICC – 0.70

Jacene et al. *J Nucl Med* 2009;50:1760-9

Percent change SUV_{bw} max



ICC – 0.94

Jacene et al. *J Nucl Med* 2009;50:1760-9

PET Metric of SUV More Reproducible than Tumor Size for Rx Response

- Tests of change in metabolism and lesion size by experienced anatomic and metabolic readers.
- WHO 2D size change least reproducible
- RECIST 1D size change intermediate
- PERCIST-like SUV max change most reproducible

Suggestions for PERCIST 1.0 Implementation

- Collect the data in a consistent manner
- Determine prognostic value
- Determine deficiencies
- Refine
- Consider including PERCIST elements in structured reports (i.e. uptake time, normal tissue SUV lean, SUL hottest lesion).
- Suggested starting point for clinical trials and practice including PET quantitation.

Summary

- Standardization of acquisitions and analysis of PET data is essential for comparability of studies across centers
- PERCIST provides a framework for analysis and allows exploration of alternative metrics
- Comparison of reader studies of treatment response shows quantitation of PET response assessments are MUCH more reproducible than “accepted” WHO or RECIST criteria.

Acknowledgements

- Heather Jacene, M.D.
- Martin Lodge, Ph.D.
- Yvette Kasamon, M.D.