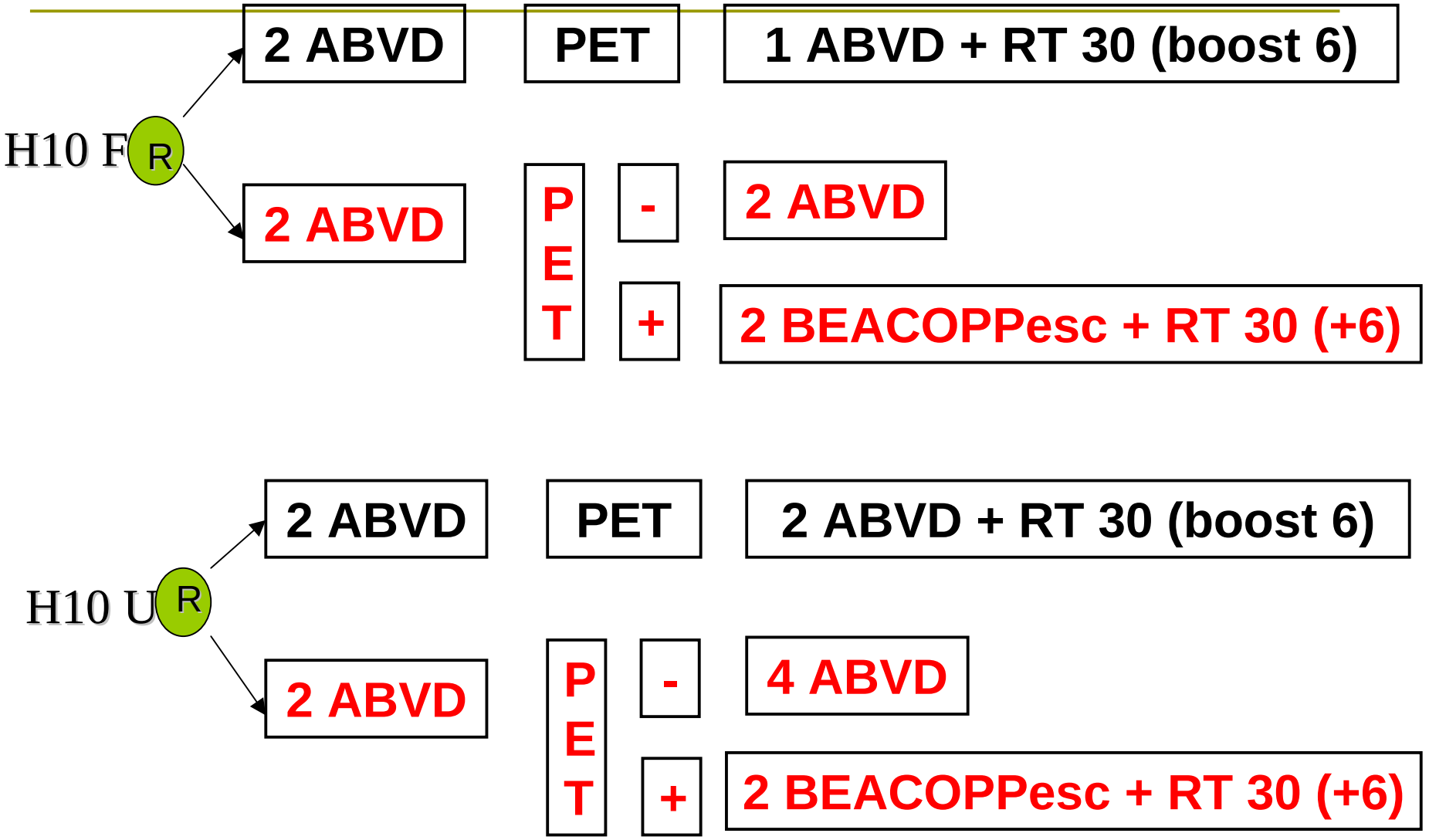
Limits of the IHP Criteria

Stéphane Bardet, Caen, France
Malik Juweid, Iowa City, USA

Menton, April 8th 2010

H10 STUDY


2 ABVD
PET
1 ABVD + RT 30 (boost 6)
H10 F
R
2 ABVD
P
E
T
-
2 ABVD
+
2 BEACOPPesc + RT 30 (+6)
2 ABVD
PET
2 ABVD + RT 30 (boost 6)
H10 U
R
2 ABVD
P
E
T
-
4 ABVD
+
2 BEACOPPesc + RT 30 (+6)

Interim PET Scan

- Performed at least 7-10 days after day 15 of the second cycle of ABVD.

- A baseline PET scan is strongly recommended, but not mandatory.

- Visual interpretation according to IHP (Juweid et al. JCO 2007)

Central review

6 experts in nuclear medicine

- Hôpital Henri Mondor, Créteil: M Meignan, E Itti
- Institut Gustave Roussy, Villejuif: J Lumbroso
- Centre René Hughenin, St Cloud: V Edeline
- CHU Nancy: P Olivier
- Mont-Godinne, Belgique: T Vander Borght
- Centre François Baclesse, Caen: S Bardet

Cornerstone: Positoscope

Pre-treatment PET/CT Post-treatment PET/CT Commands

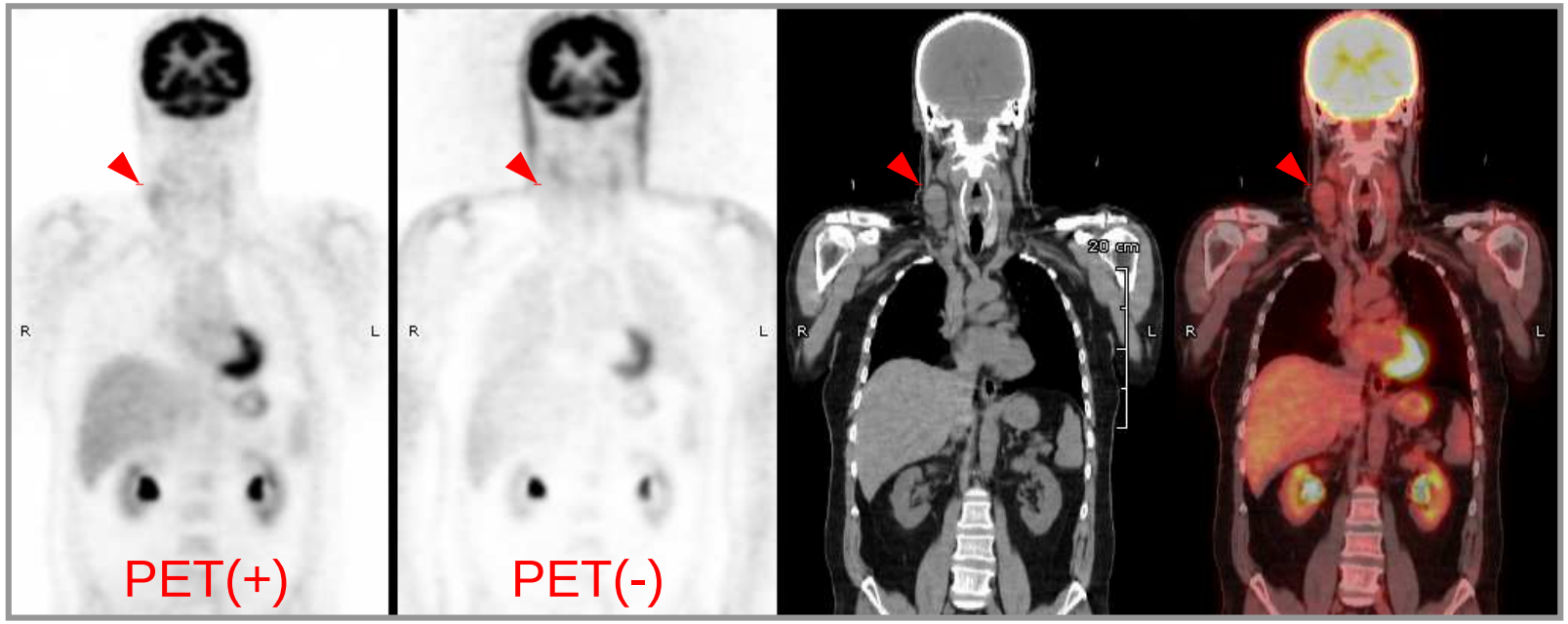
Multimodality dual screen workstation linked to the DICOM network
Side to side display of pre and post-treatment PET/CT
Complete processing: Multi slices display, MIP, triangulation, ROI, SUV

Criteria for interim PET assessment

Modified Juweid criteria

Consensus after analysis of the first 114 patients of the H10 and after having found a Kappa at 0.45

- PET (+) if SUV_{tumor} **> 25% $SUV_{reference}$**
  (mediastinum or neighboring bkg dep. on residual mass ∅)
- Necessity to interpret interim PET / baseline PET
- PET (+) if also present on non att.-corr. image



Fleiss' Kappa

	Centre Baclesse	Centre René Huguenin	IGR	CHU Henri Mondor	CHU de Nancy	Mont Godinne Yvoir
Centre Baclesse						
Centre René Huguenin	0,49					
IGR	0,59	0,51				
CHU Henri Mondor	0,54	0,54	0,49			
CHU de Nancy	0,51	0,56	0,46	0,60		
Mont Godinne Yvoir	0,64	0,66	0,64	0,70	0,63	

- The interobserver variability assessed by unweighted kappa statistics was $\kappa = 0.57$ which is a moderate agreement.

Results

- **564 patients** read with Imotep Network from **June 2007 to January 2010** by the **GELA's team**.
- 440 patients **(78%)** are **negative** after 2 cycles.
- 124 patients **(22%)** are **positive** after 2 cycles.
- 421 patients **(75%)** were interpreted **with baseline PET**.

Interpretation with the 5-points scale for patients scored positive using the IHP criteria

Patients

- Patients with positive interim PET using the IHP criteria until January 2010 among patients reviewed in the Imotep Network

- Patients with baseline PET/CT available (PET alone excluded, unvailable baseline PET excluded)

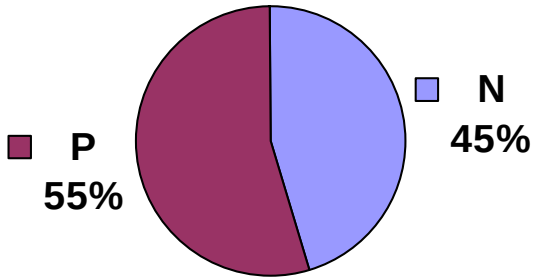
→ **n=64**

Methods

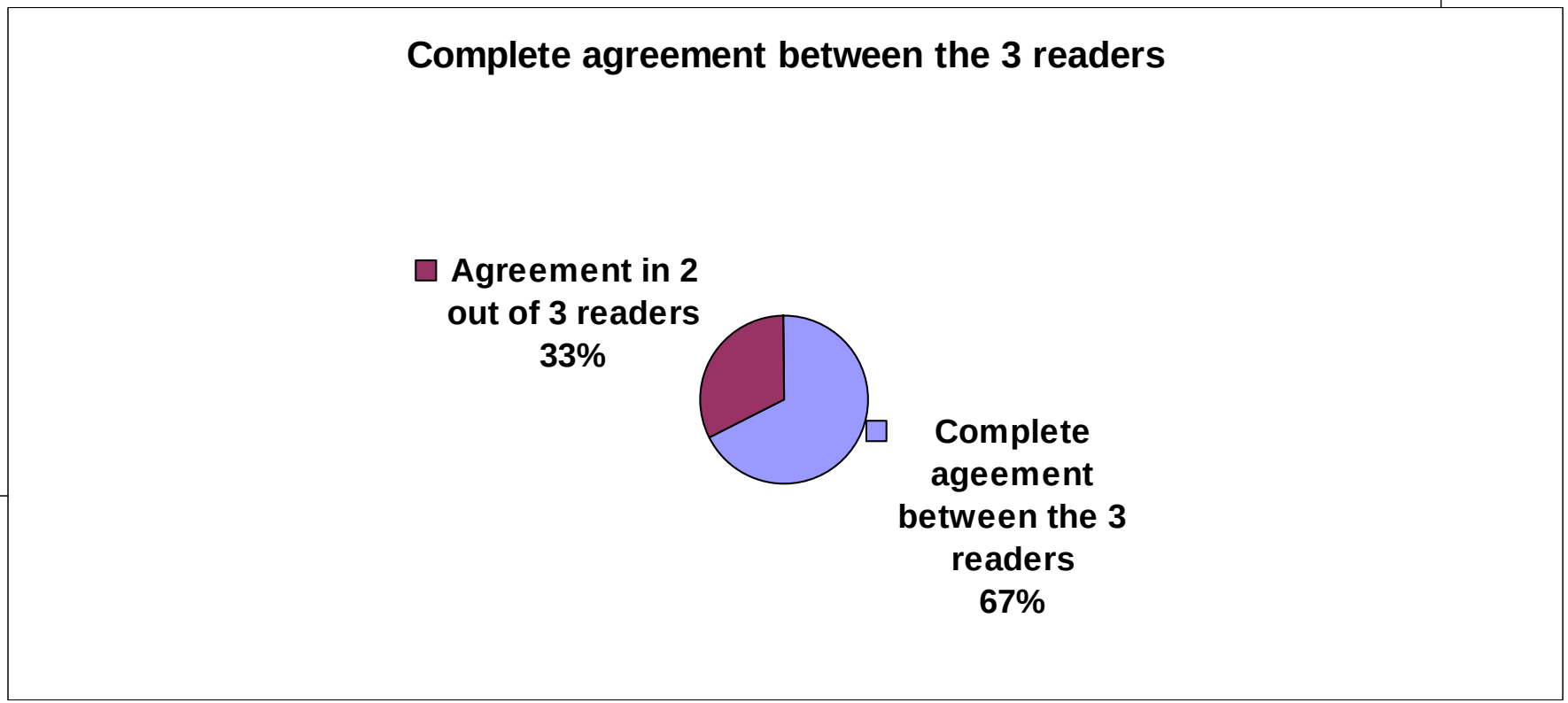
- Second interpretation with the 5-points scale: Positive if ≥ 4 (liver as a reference), negative otherwise

- 3 independant readers (T Van der Borght, M Meignan, S Bardet) on similar positoscope workstations

Proportion of positive and negative patients with the 5-point scale



Proportion of positive and negative patients with the 5-point scale



Proportion of positive and negative patients with the 5-point scale

Complete agreement between the 3 readers

	Créteil	Mont Godinne	Caen
Créteil			
Montgodinne	0,59821429		
Caen	0,57663479	0,51666667	

Kappa 0,56

Conclusions

- In patients with early stage HL, the use of IHP criteria probably leads to an excess of positive interim PET (20-25%).

- The use of the 5-PS with liver as a reference reduces the proportion of positive interim PET (10-15%), closer to that expected.

- The use of the 5-PS with liver as a reference makes the visual interpretation easier with a better interobserver variability.

- The prognostic impact of both sets of criteria should be assessed *a posteriori* in the classical arm of the H10 study.

Interpretation of end-of-therapy scans with the 5-points scale using the IHP and liver-based criteria

Patients

- Patients with HL or aggressive NHL treated with 4-8 cycles of chemtx, PET/CT done within 3-12 wks post-tx with adequate F/U for ≥ 12 mo or until progression/evidence of persistent disease (*median F/U 44 mo; 48 mo in pts without progression*)

-No baseline PET/CT used for interpretation, although available on some patients

→ **n=50**

Methods

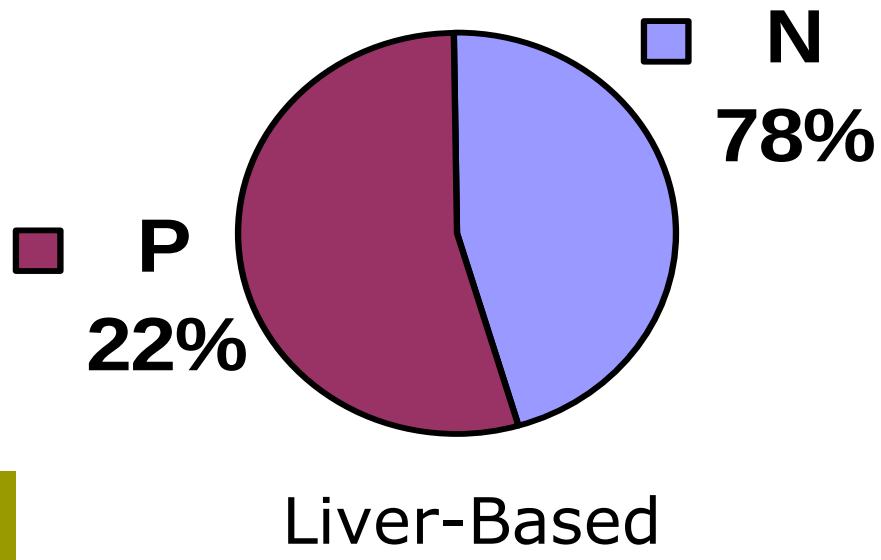
- One interpretation with the 5-points scale using mediastinal blood pool structures (MBPS) as reference (IHP) and another using liver as reference; in both schemes positive if score ≥ 4, otherwise negative

- 3 independant readers: A (M Juweid), B (D Bushnell) and C (M Graham) on similar workstations

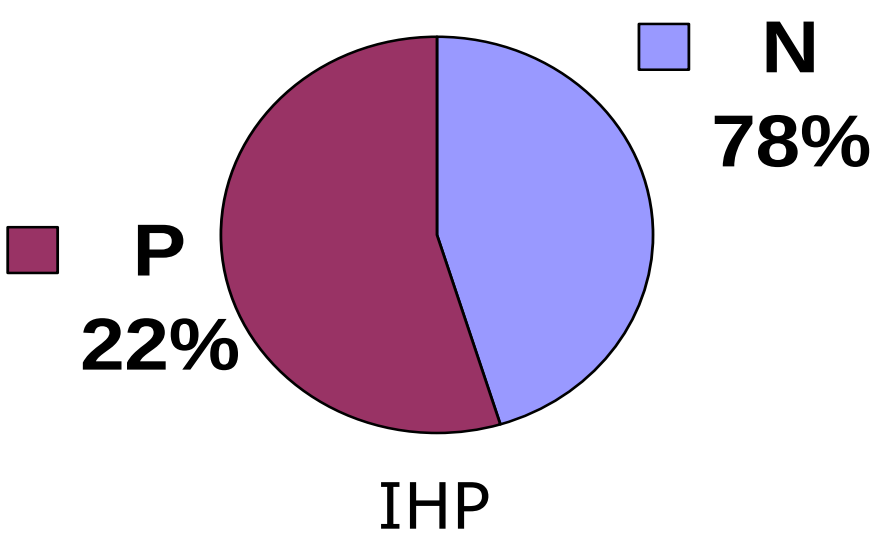
Agreement between the Three Readers

- **IHP-based:**
  - **90% complete agreement**
  - **98% between A and B, 92% A and C and 90% B and C**
- **Liver-based:**
  - **86% complete agreement**
  - **94% between A and B, 90% A and C and 88% B and C**

Proportion of positive and negative patients with the 5-point scale using the IHP and Liver-Based criteria

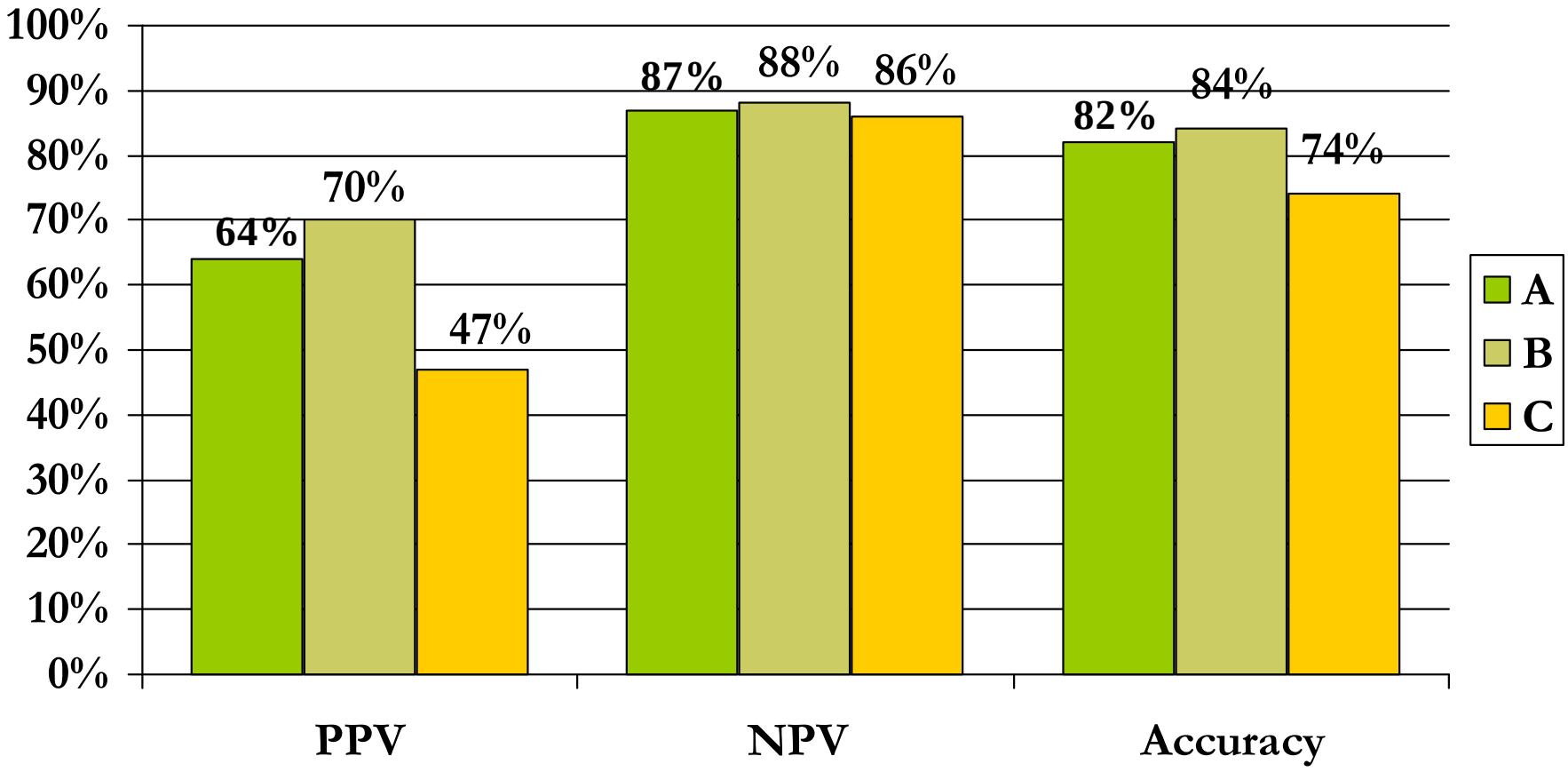


Liver-Based

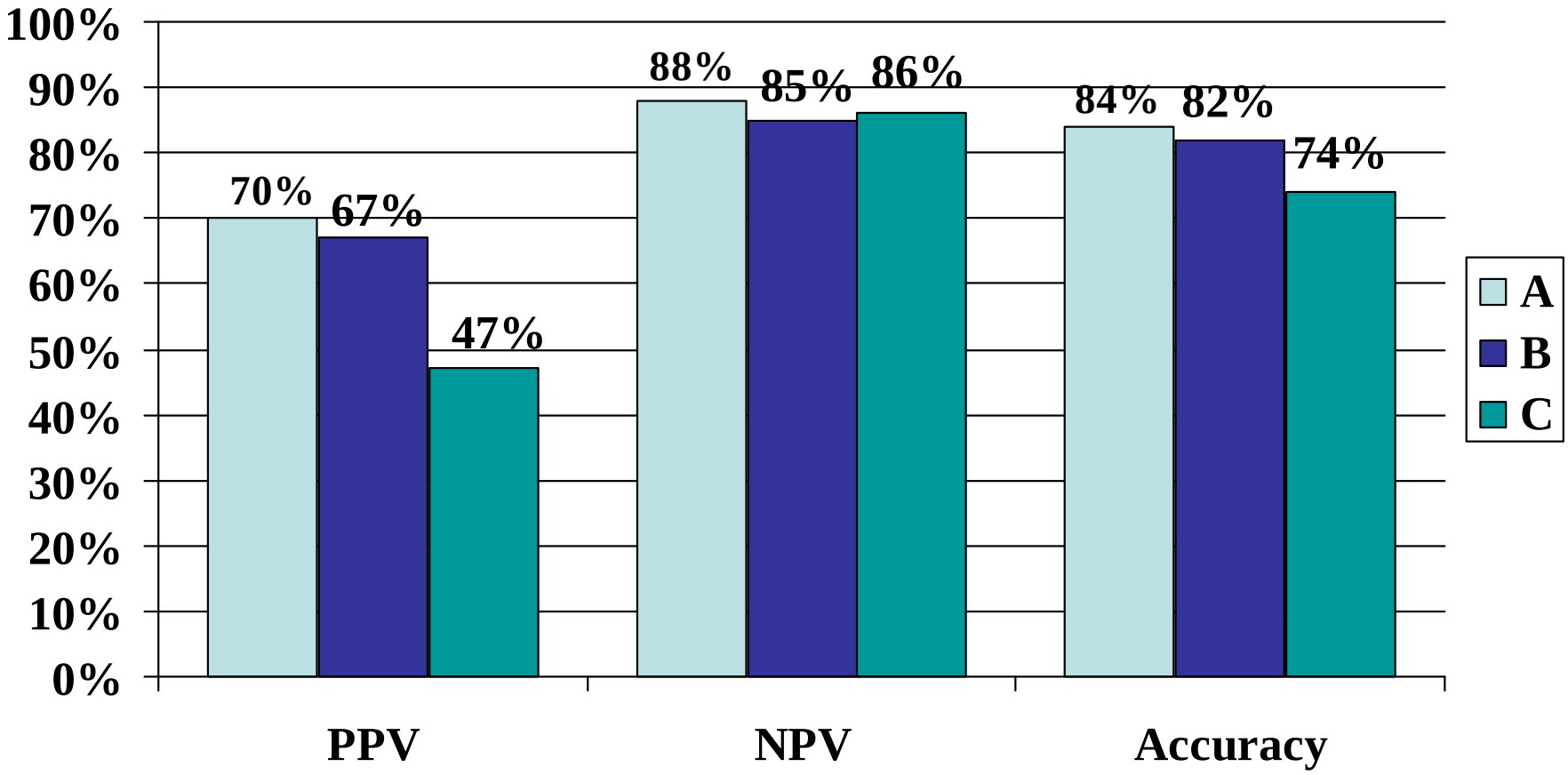


IHP

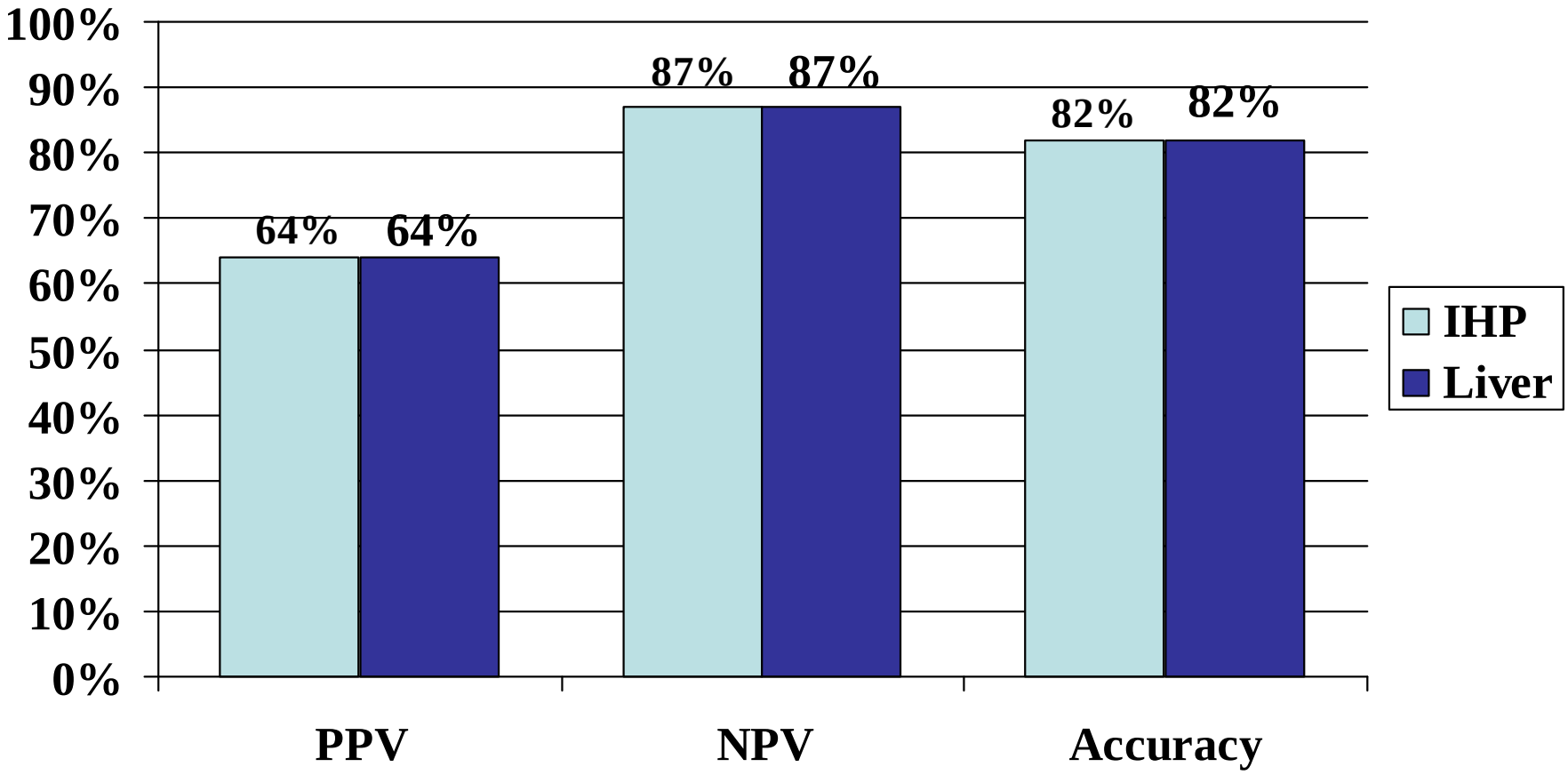
Interpretation vs. Outcome for the Three Readers Using IHP Criteria



Interpretation vs. Outcome for the Three Readers Using Liver-Based Criteria



Interpretation vs. Outcome Based on Agreement of 2 of 3 independent readers IHP vs. Liver-Based



SUV-Based Analysis vs. IHP for end-of-tx PET/CT (Patient-Based Analysis)

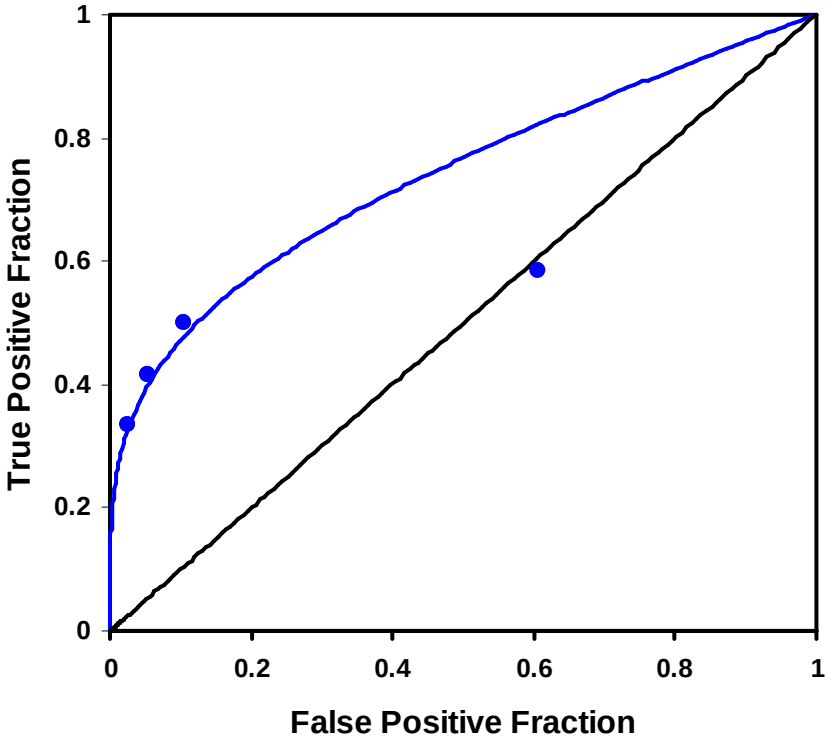
- **Improvement in PPV for all readers without compromising NPV**
- **Reasonable area under the ROC curve (0.765):**

Cut-off SUV_{max} of 2.5 results in:

PPV = 75% vs. 64%, 70%, 47%
for the 3 readers

NPV = 86% vs. 87%, 88%, 86%

Accuracy= 84% vs. 82%, 84%, 74%



SUV-Based Analysis vs. IHP for end-of-tx PET/CT (Lesion-Based Analysis; 60 lesions in 30 pts)

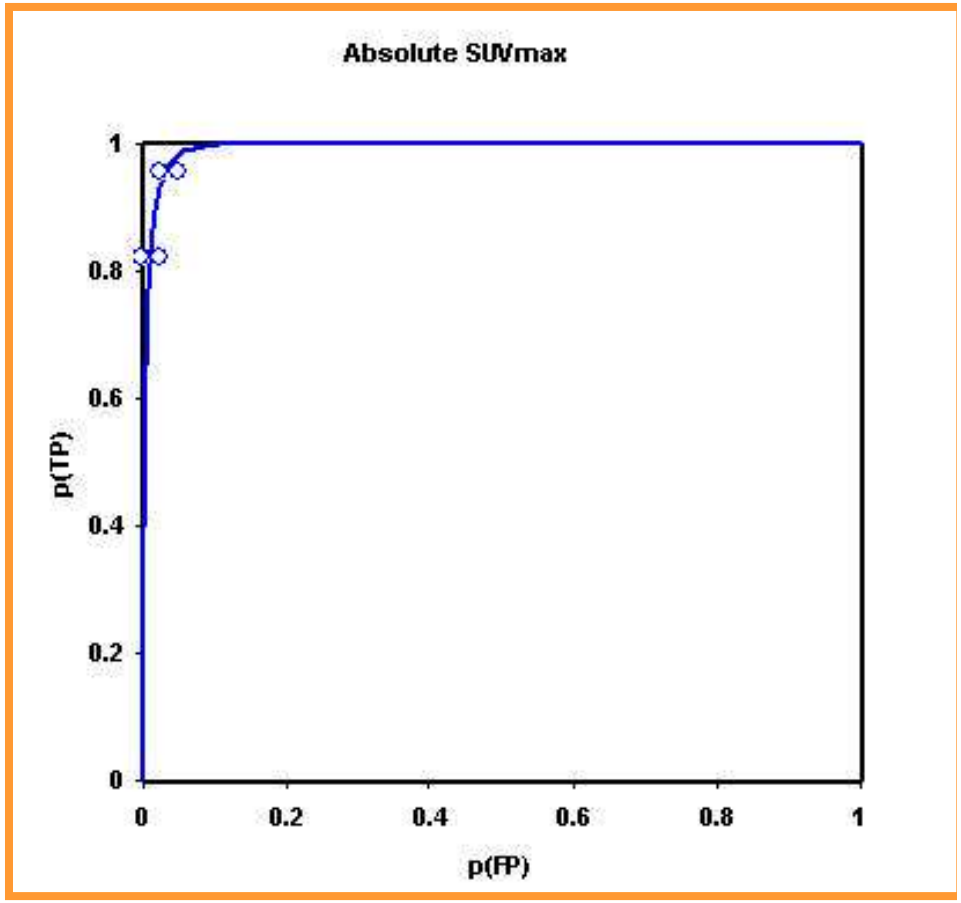
- **Substantial improvement in PPV for all readers with almost perfect area under the ROC curve (0.99):**

Cut-off SUV_{max} of 2.5 results in:

PPV = 92% vs. 81%, 79%, 67%
for the 3 readers

NPV = 100% vs. 100%, 100%, 100%

Accuracy= 97% vs. 92%, 90%, 82%



Conclusions

- There was good agreement between readers in interpretation of end-of-therapy PET/CT scans with both the IHP- and liver-based criteria using the 5-point scale. The IHP criteria tended to result in only slightly greater agreement than the liver-based criteria.

- The higher fraction of pos scans identified by one reader in contrast to the other two resulting in substantially lower PPV and accuracy emphasizes the need for training for using the 5-point scale using a training set; this should probably occur more universally through educational sessions at national and international nuclear medicine/radiology meetings

- Semiquantitative analysis appears to improve the PPV for all readers (to different extents) and accuracy for some readers