PET interpretation issues: experience in NHL with 5PS

Emmanuel Itti, MD PhD
H. Mondor Hospital, AP-HP
Paris-Est University, Créteil, FR

Menton April 8th, 2010
Methods

- 49 IVS patients from 4 PET centers (Créteil n=15; Dijon n=14; Cuneo n=11; Rouen n=9)
- PET/CT at baseline and 2 cycles
- Interpretation by 3 observers using the 5PS
- Transfers/readings on Positoscope workstations
- Inter-observer agreement (Kappa)
- Quantification with $\Delta$SUV (66% cut-off)
### 5-point scale weighted Kappa (Cohen)

<table>
<thead>
<tr>
<th>Observer B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>(15.3%)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>(14.3%)</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>(14.3%)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>(22.4%)</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>(32.7%)</td>
</tr>
</tbody>
</table>

Weighted Kappa: 0.744

<table>
<thead>
<tr>
<th>Observer B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>(18.4%)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>(25.5%)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>(20.5%)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>(20.4%)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(8.2%)</td>
</tr>
</tbody>
</table>

Weighted Kappa: 0.568

### Landis and Koch scale

<table>
<thead>
<tr>
<th>Interval</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0</td>
<td>no agreement</td>
</tr>
<tr>
<td>0.00 – 0.20</td>
<td>slight</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>fair</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>moderate</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>substantial</td>
</tr>
<tr>
<td>0.81 – 1.00</td>
<td>almost perfect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Observer B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>(18.4%)</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>(25.5%)</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>(20.5%)</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>(20.4%)</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(8.2%)</td>
</tr>
</tbody>
</table>

Weighted Kappa: 0.604
## 5-point scale binary (cut-off ≥3, MBP) Kappa (Cohen)

### Landis and Koch scale

<table>
<thead>
<tr>
<th>Landis and Koch scale</th>
<th>&lt; 0</th>
<th>0.00 – 0.20</th>
<th>0.21 – 0.40</th>
<th>0.41 – 0.60</th>
<th>0.61 – 0.80</th>
<th>0.81 – 1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>no agreement</td>
<td>slight</td>
<td>fair</td>
<td>moderate</td>
<td>substantial</td>
<td>almost perfect</td>
</tr>
</tbody>
</table>

### 5-point scale binary (cut-off ≥3, MBP) Kappa (Cohen)

<table>
<thead>
<tr>
<th>Observer A</th>
<th>Créteil</th>
<th>Dijon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>31</td>
</tr>
</tbody>
</table>

- **Kappa:** 0.712
- **Standard error:** 0.110
- **95% CI:** 0.496 to 0.923

<table>
<thead>
<tr>
<th>Observer A</th>
<th>Cuneo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer B</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Kappa:** 0.533
- **Standard error:** 0.124
- **95% CI:** 0.289 to 0.775

### Overall Kappa (Fleiss) (3 obs.)

<table>
<thead>
<tr>
<th>Observer A</th>
<th>Dijon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer B</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

- **Kappa:** 0.533
- **Standard error:** 0.124
- **95% CI:** 0.289 to 0.775

Overall Kappa (Fleiss) (3 obs.) \( \kappa = 0.58 \)
5-point scale binary (cut-off $\geq 4$, liver)

### Kappa (Cohen)

<table>
<thead>
<tr>
<th>Observer B</th>
<th>Observer A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Kappa 0.833
Standard error 0.073
95% CI 0.683 to 0.990

### Overall Kappa (Fleiss) (3 obs.) $\kappa = 0.61$

```
<table>
<thead>
<tr>
<th>Observer A</th>
<th>Crétai</th>
<th>Dijon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>24</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>
```

Kappa 0.555
Standard error 0.113
95% CI 0.323 to 0.787

### Landis and Koch scale

- **< 0** no agreement
- **0.00 – 0.20** slight
- **0.21 – 0.40** fair
- **0.41 – 0.60** moderate
- **0.61 – 0.80** substantial
- **0.81 – 1.00** almost perfect
Créteil:
2-y EFS: 88% vs. 63%

\[ P = .02 \quad \chi^2 = 5.05 \]

Dijon:
2-y EFS: 91% vs. 63%

\[ P = .02 \quad \chi^2 = 5.59 \]

Cuneo:
2-y EFS: 83% vs. 61%

\[ P = .03 \quad \chi^2 = 4.48 \]

→ Generates false-positives

5-point scale
(cut-off ≥3, MBP)
Event-free survival

# of events = 14
Median f-u = 25 mo
P = .001
$\chi^2 = 10.14$

Créteil:
2-y EFS: 89% vs. 54%

Dijon:
2-y EFS: 88% vs. 57%

Cuneo:
2-y EFS: 87% vs. 32%

$P = .005$
$\chi^2 = 8.04$

$P < .0001$
$\chi^2 = 25.14$

5-point scale
(cut-off $\geq 4$, liver)
Event-free survival

→ Reduction of false-positives
→ Cuneo’s interpretation ++

# of events = 14
Median f-u = 25 mo
5-point scale
(cut-off ≥5, >>liver)
Event-free survival

Créteil:
2-y EFS: 86% vs. 42%

P = .0004
χ² = 12.59

Dijon:
2-y EFS: 86% vs. 40%

P < .0001
χ² = 17.50

Cuneo:
2-y EFS: 78% vs. 0%

→ Créteil-Dijon’s interpretations ++
→ Cuneo: generates false-negatives

# of events = 14
Median f-u = 25 mo
Quantification
ΔSUV (cut-off >66%)
Event-free survival

Créteil:
2-y EFS: 81% vs. 46%
P = 0.003
χ² = 8.97

Dijon:
2-y EFS: 81% vs. 45%
P = 0.002
χ² = 9.42

Cuneo:
2-y EFS: 81% vs. 45%
P = 0.002
χ² = 9.96

→ Better agreement between observers

# of events = 14
Median f-u = 25 mo
Conclusions

- 5PS: moderate to substantial agreement ($\kappa$ 0.58-0.61)
- Ref. background must be high for interim PET/NHL
- Subjectivity $\rightarrow$ need for different opinions
  foci that were considered by Créteil/Dijon as “moderately increased above liver” (4) were considered “equal to liver” (3) by Cuneo
- Quantification may help the definition of scores 3-4
- $\Delta$SUV is not observer-dependent for EFS prediction
125% liver

150% liver

200% liver

P = 0.003
χ² = 8.68

P = 0.0001
χ² = 14.60

P < 0.0001
χ² = 18.67

92 patients from Haioun, *Blood* 2005

[¹⁸F]fluoro-2-deoxy-D-glucose positron emission tomography (FDG-PET) in aggressive lymphoma: an early prognostic tool for predicting patient outcome

Corinne Haioun, Emmanuel Itti, Alain Rahmouni, Pauline Brice, Jean-Ddier Rain, Karim Belhadj, Philippe Gaulard, Laurent Garderet, Eric Lepage, Felix Reyes, and Michel Meignan

Itti, Juweid, Haioun, et al. *SNM* 2010