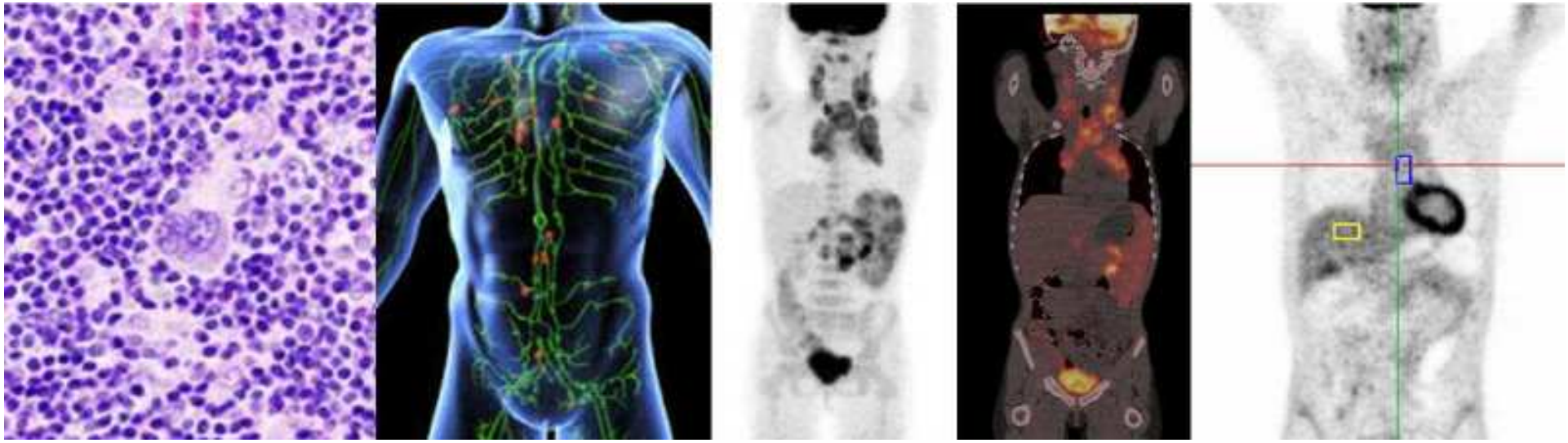
Interim FDG-PET

Visual interpretation vs. qPET

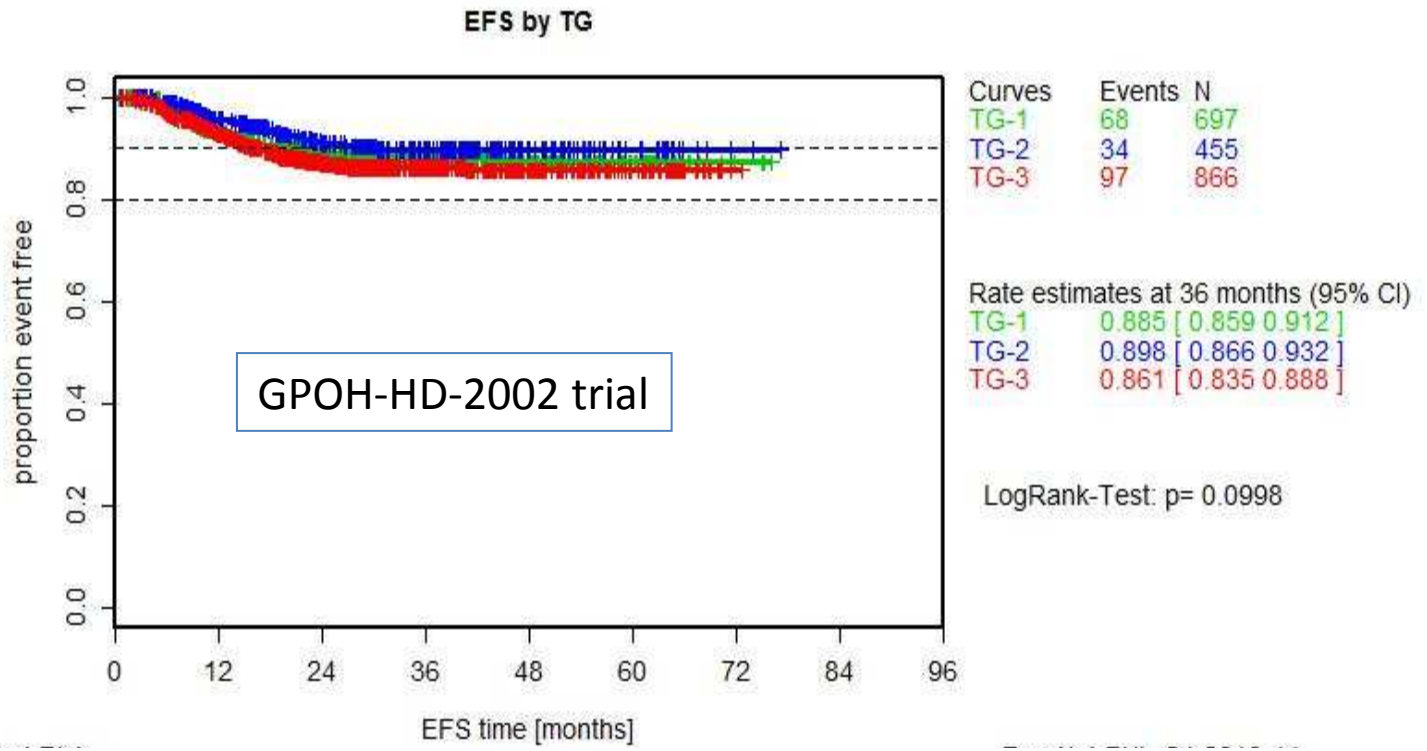


R. Kluge, D. Hasenclever, L. Kurch,
L. Chavdarova, M. Hoffmann, C. Kobe, B. Malkowski, F. Montravers,
C. Mauz-Körholz, T. Georgi, D. Körholz

Deutsche Krebshilfe. Helfen. Forschen. Informieren.
HERMES
mitteldeutsche kinderkrebsforschung
Paediatric Hodgkin Network
Executive Agency for Health and Consumers

Paediatric Hodgkin's lymphoma

EFS of early, intermediate and advanced stages

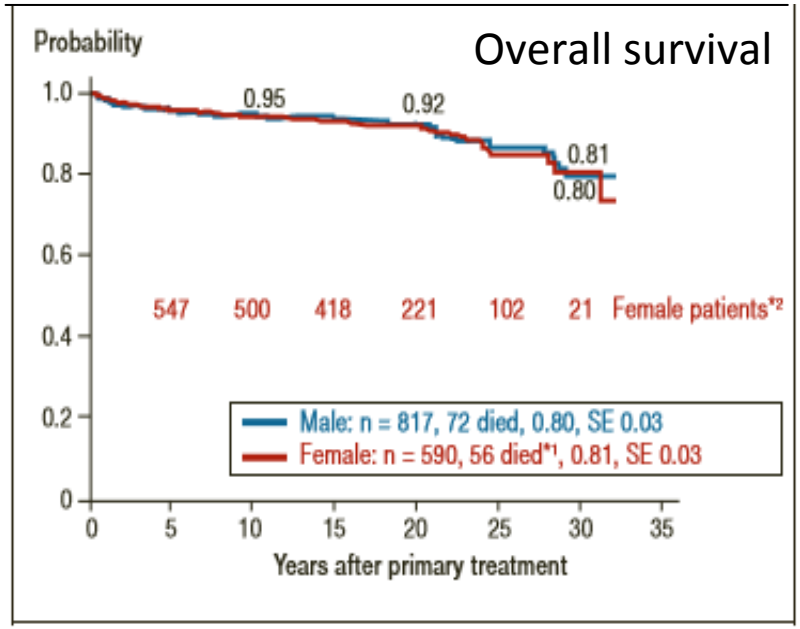


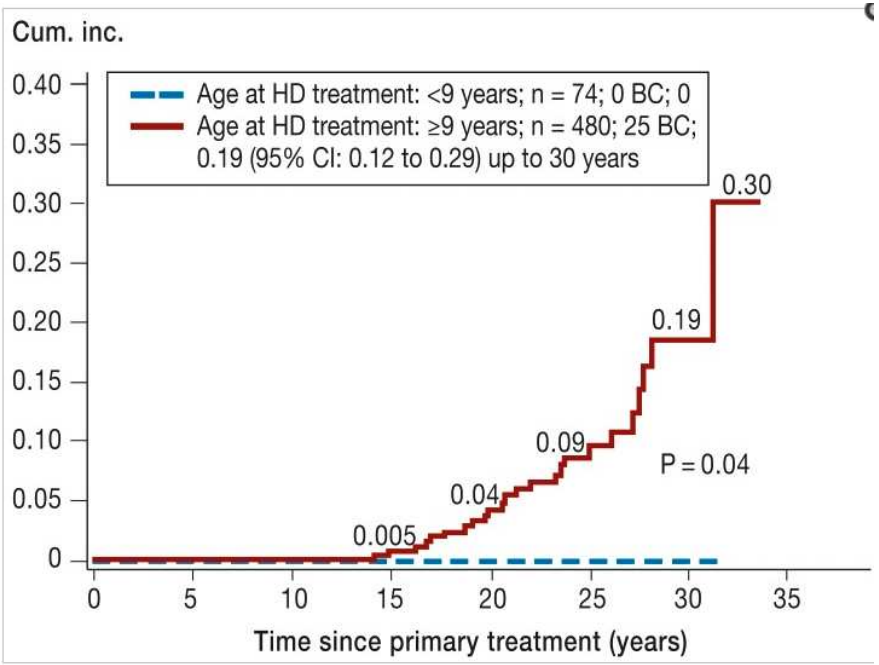
All patients have a good chance to be cured if enough treatment is applied.

Challenge: Cure patients but avoid late effects

Treatment-related effects

-Cardiopulmonary events
-Secondary cancer

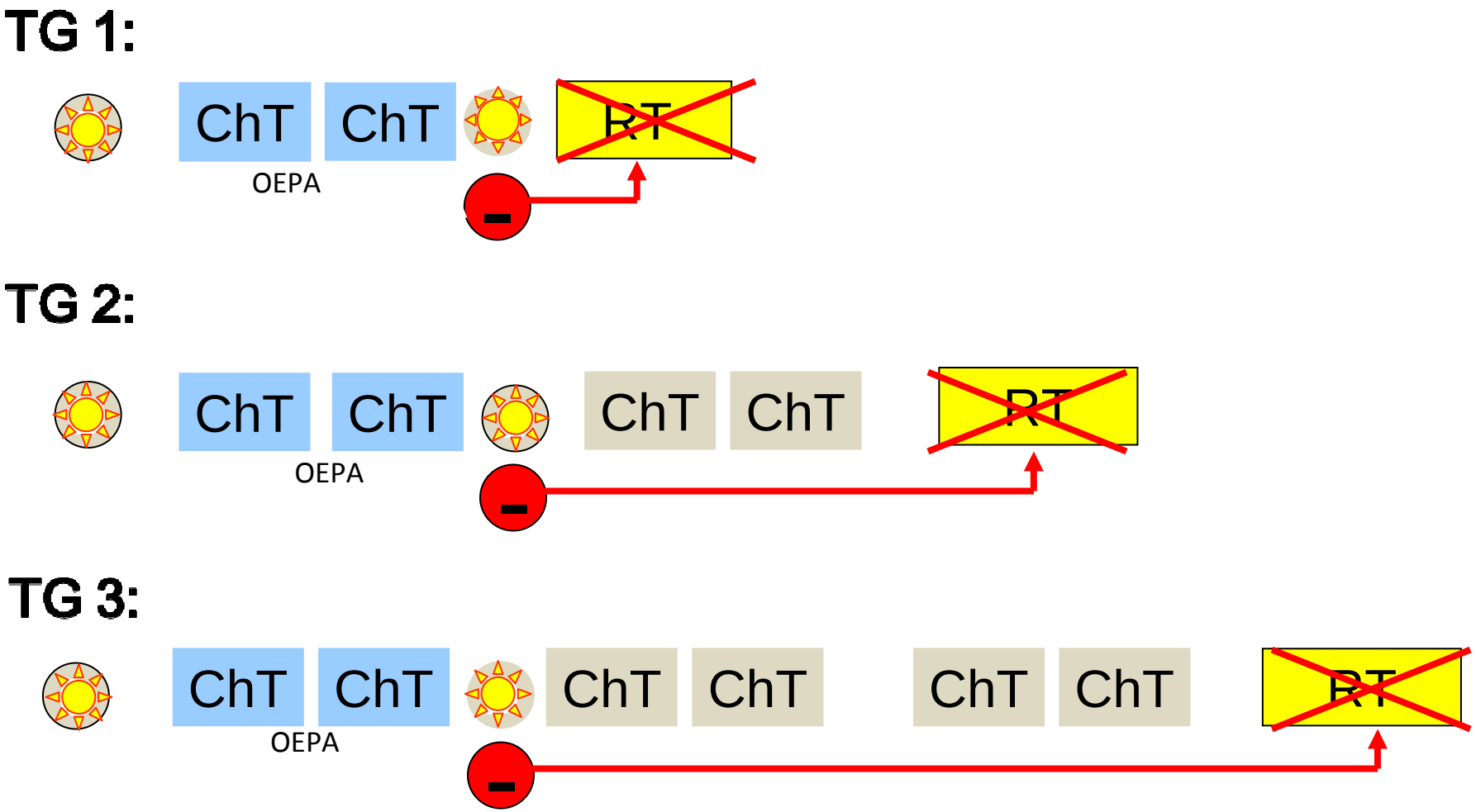




Schellong et al. Dtsch Arztebl Int. Jan 2014; 111(1-2): 3–9.

EuroNet-PHL-C1

Treatment of low, intermediate and high risk patients



EuroNet-PHL study group

EuroNet-PHL-C1

- 2131 registered patients
- Central review of all imaging
- Paediatric Hodgkin Network
- 47% PET-negative - no radiotherapy
- EFS: slight, non-significant reduction after 36 months

IHP Criteria (2007) were used in C1-study

0 – 0	completely negative PET in all initially involved regions
0 – 1	slightly diffuse enhanced uptake < mediastinal blood pool (if residuum > 2 cm)

1 – 2	uptake > mediastinal blood pool in residual area > 2 cm or any enhanced uptake in an involved area < 2 cm
1 – 3	strongly enhanced uptake

Deauville Criteria (2010)

Score	Residuals in Interim-PET/CT
1	No Uptake over background
2	Uptake ≤ Mediastinum
3	Uptake > Mediastinum but ≤ Liver
4	Uptake moderately > Liver
5	Uptake strongly > Liver

1-3 = complete metabolic remission?

Sensitive cut for treatment reduction studies?

But: Borderline cases and differences in interpretation

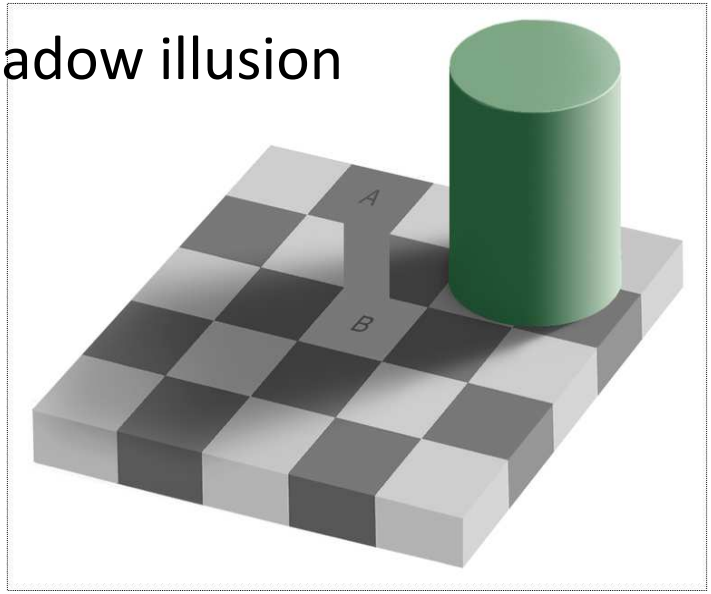
Problems:

- Identify the most intensive residual
- Compare correctly
  - Inhomogeneity of reference levels
  - What is the “hottest” part of the residual?

Optical illusions

Visual contrast illusion

Checker shadow illusion



Estimation of gray levels is influenced by the pattern

Inter-reader study
Design

- N=100 consecutive cases
  - Presented to
- 5 readers (R1 – R5)
- Readers were asked to score up to three involved sites with highest uptake of each case with
  - EuronetScore used in EuroNet-PHL-C1
  - DeauvilleScore

Frequency of Deauville scores by reader

DV	R1	R2	R3	R4	R5
1	49	52	10	49	14
2	12	10	13	18	49
3	19	24	52	16	18
4	15	13	21	14	14
5	5	1	4	3	5

61 62 23 67 63

Minimum

Maximum

Readers differ in frequency of using specific Deauville scores.

Method

- We estimate the probability that two random readers concord on the score in a random case (Uebersax-J 1983).
  - Overall
  - Given that one reader has assigned category k

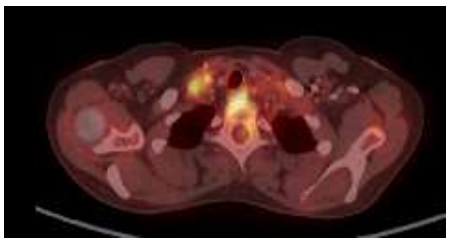
- See http://www.john-uebersax.com/stat/raw.htm#genera

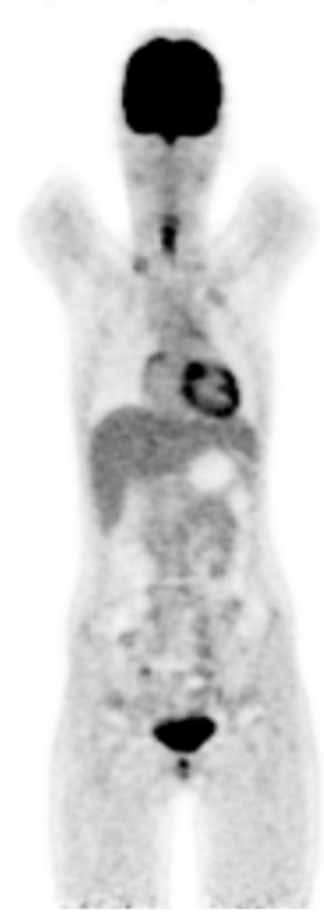
Probability of concordance

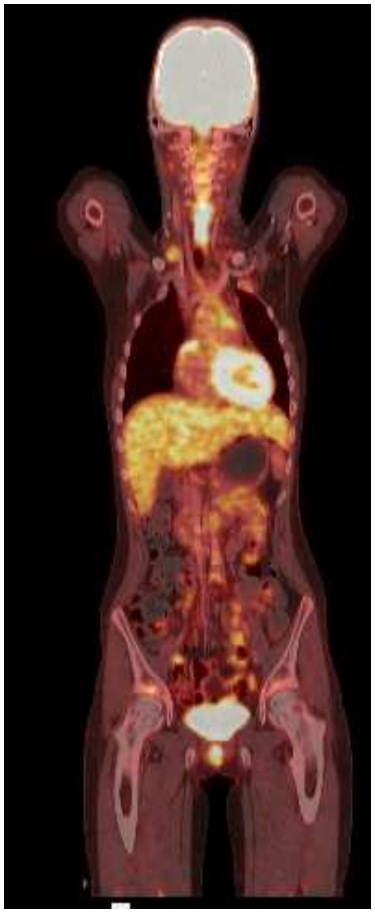
Overall and category-specific

```
Five categories
P0    |    Ps1    Ps2    Ps3    Ps4    Ps5
0.422     0.520 0.255 0.360 0.494 0.556

Three Categories  1-2 versus 3 versus 4-5
P0    |    Ps12   Ps3    Ps45
0.604     0.705  0.360 0.642

Two Categories 1-2 versus 3-5
P0    |    Ps12   Ps345
0.674     0.705  0.636

Two Categories 1-3 versus 4-5

P0    |    Ps123    Ps45
0.864     0.916    0.642
```

Case 2545 – Neck/supra/infraclavicular



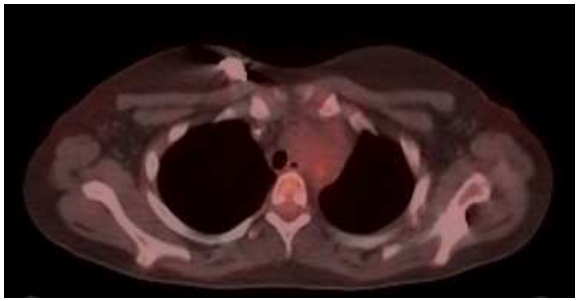




Readers:

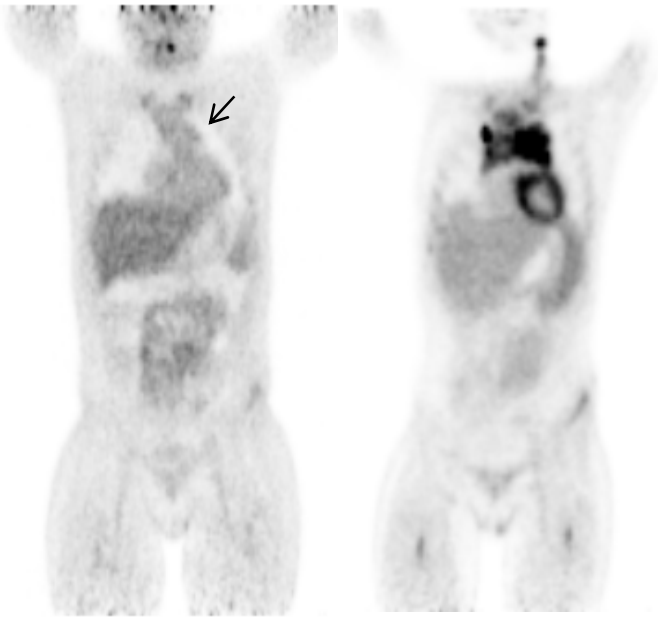
4-4-3-3-3
in lower neck,
supra- or
infraclavicular

Case 2848 - Mediastinum



Readers:
3-3-4-4-4
in upper or middle mediastinum

Case 2670 - Mediastinum



Readers:
1-2-2-3-3
in upper or
middle
mediastinum

Summary visual reading

- The probability that two random reader concord on the exact DV score of a random case is less than 50%.
- Concordance is particularly low in cases considered for DV 2 or 3.
- The binary decision DV 1-3 versus DV 4-5 is more reliable: Concordance is 86%
- BUT: This is mainly due to clearly negative cases.
- In cases considered for positivity at all: only 64%
- Summary:

  Visual Deauville scoring shows only limited - moderate reproducibility in

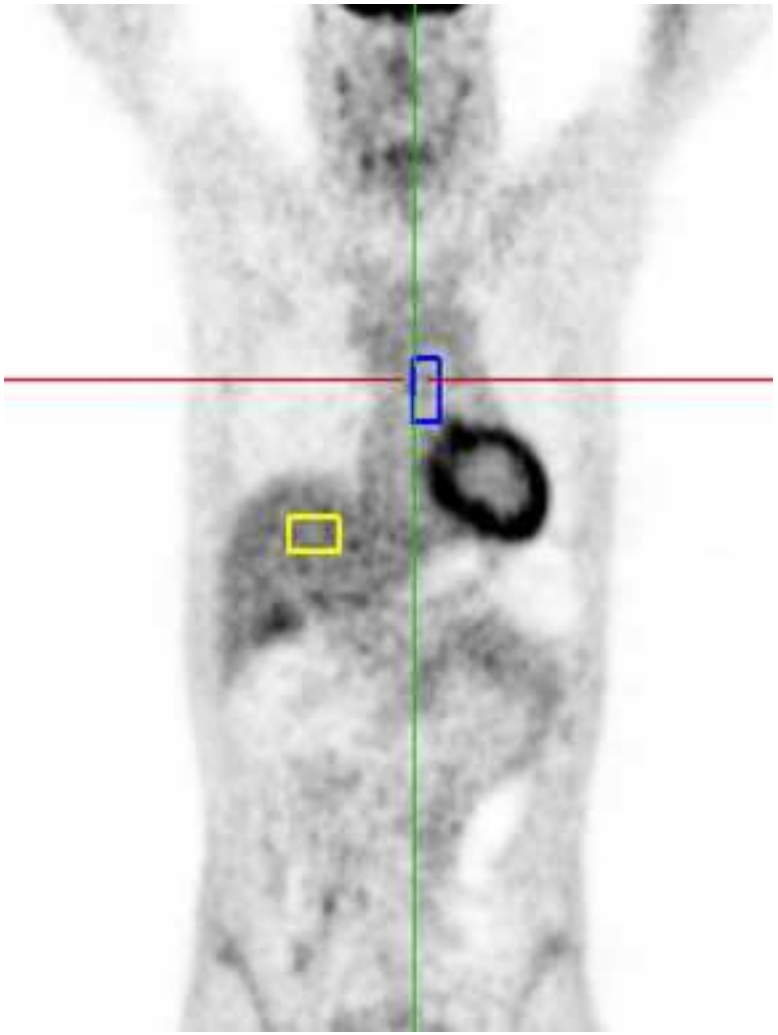
  our setting.

Objective of qPET

- Use **semi-automatic quantification**
  - To eliminate optical illusions.
  - To avoid different interpretation of the reference levels
  - To avoid different interpretation of the maximum residual uptake

- **Additional effects**:
  - Extend the ordinal Deauville scale to a **quantitative** scale
  - Enable **novel types of mathematical analysis** helping to define what is a “normal” metabolic response.

1. step: Quantify physiological uptake in mediastinum and liver

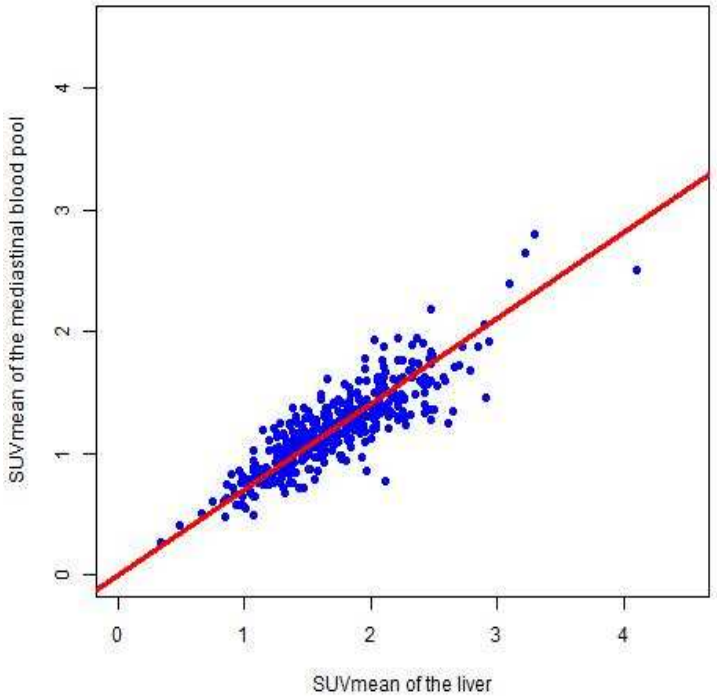


Place **standardised** VOIs to measure reference uptake.

Liver:	cuboid VOI of 30 ml
Mediastinum:	cuboid VOI of 13 ml

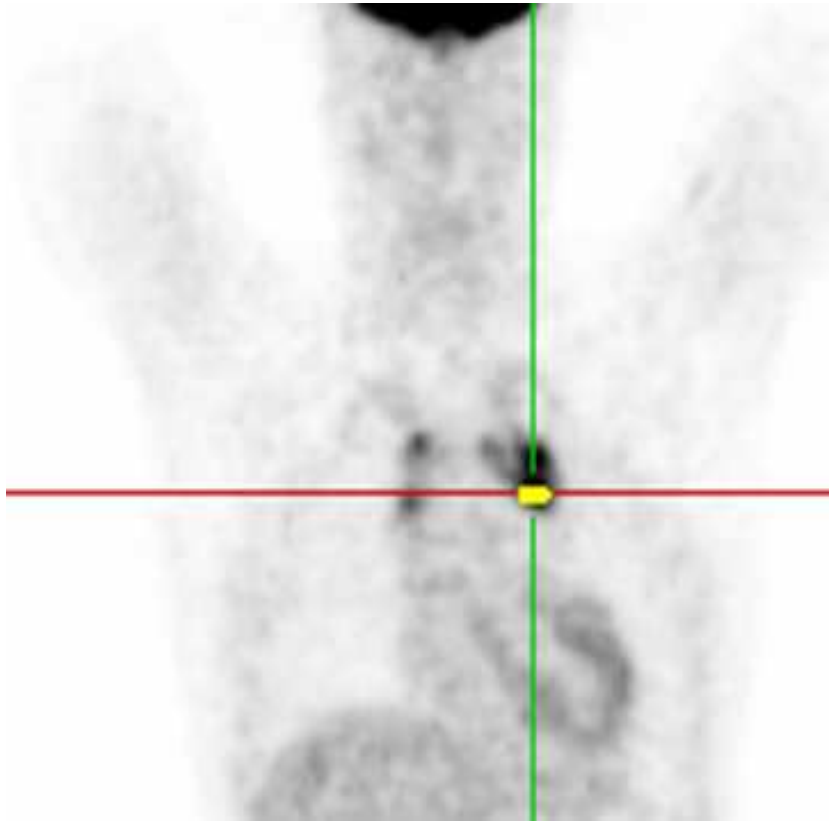
Use **average uptake**.

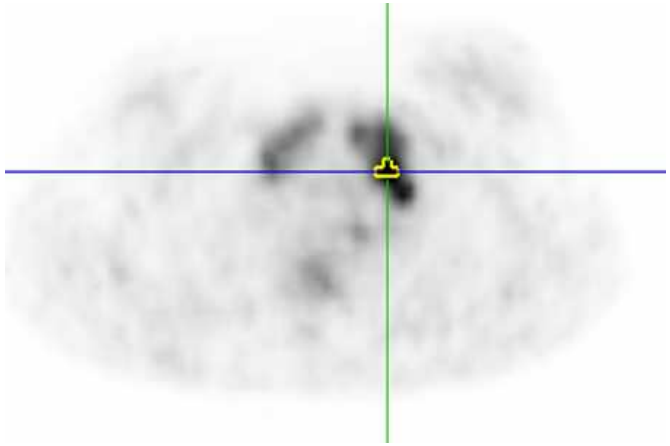
Liver preferred as reference region for qPET



- **Uptake in mediastinum and liver roughly proportional.**
  - On average the SUV_{mean} in the mediastinum is 0.714 of the SUV_{mean} in the liver.
- **VOI is easier to place in the liver**
  - Mediastinum anatomically complex and frequently involved

Measure peak uptake in tumour residuals





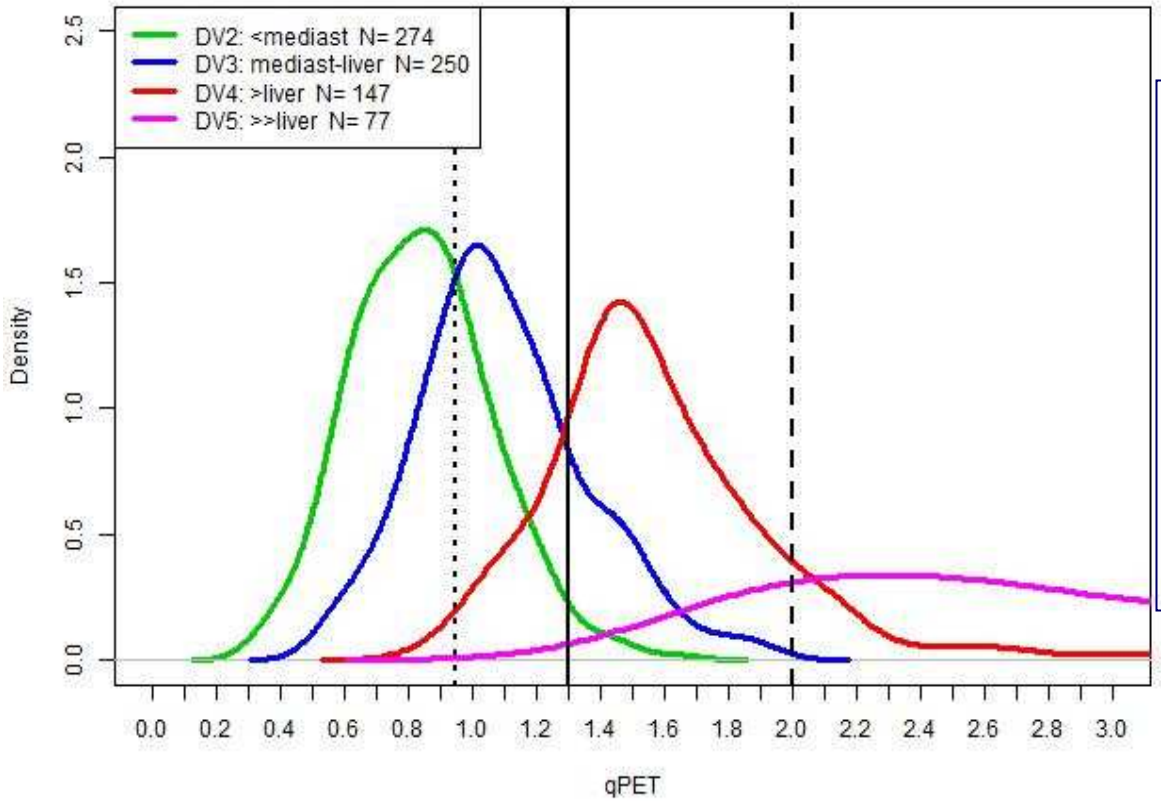
- Click on **focal residual uptake**.
- **TumourFinder™-software** determines outer contour based on user-adjustable threshold.
- Identifies **the hottest voxel** as well as **three hottest adjacent voxels**.
- **Average** over these hottest four voxels.

$$\textbf{qPET:} = \frac{\text{peak residual uptake}}{\text{average uptake in liver}}$$

Data

- **N=898** patients from EuroNet-PHL-C1 and a subsequent German registry (GPOH-HD).
- Deauville scoring
  - FDG-PET at staging was co-registered
  - Independently by two readers.
  - Consensus after discussion if discordant.
- qPET measurements after visual scoring.
  - 150 patients (16.7%) had
    - no detectable residual uptake (N=80) or
    - diffuse uptake too weak to be quantified (N=70).
  - **N=748** qPET signals.

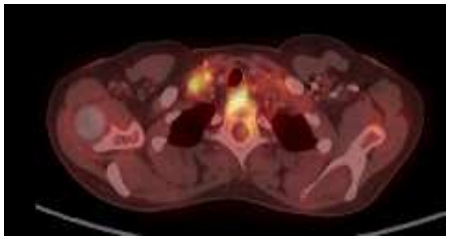
qPET values of cases visually scored as Deauville 2, 3, 4 or 5

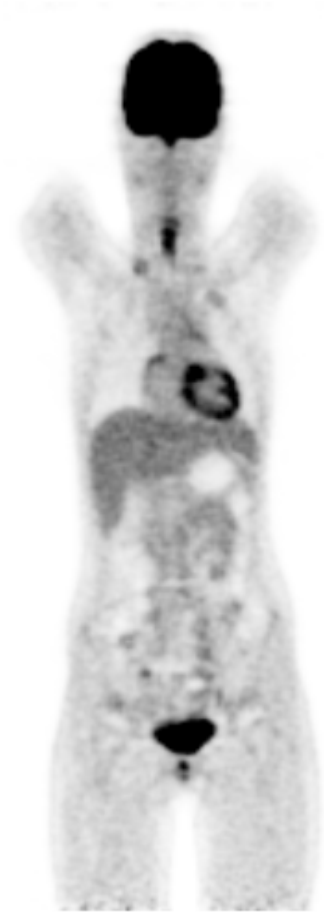


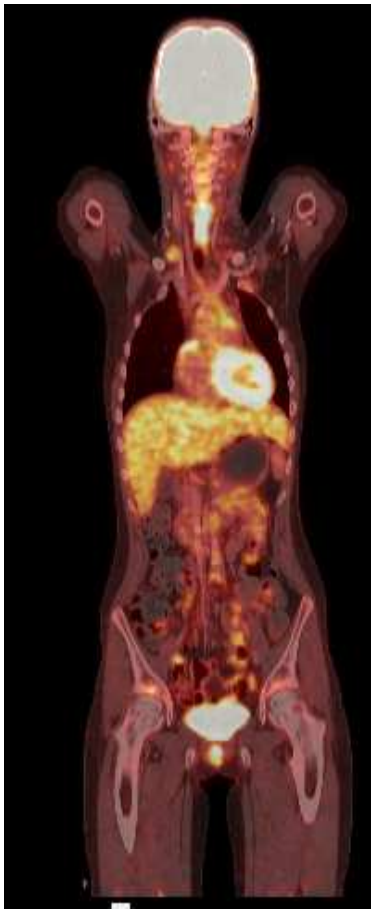
- **Thresholds**
- **qPET = 0.95 Deauville 2/3**
- **qPET = 1.3 Deauville 3/4**
- **qPET = 2.0 Deauville 4/5**

qPET is a quantitative extension of the Deauville Criteria

Case 2545 – Neck/supra/infraclavicular







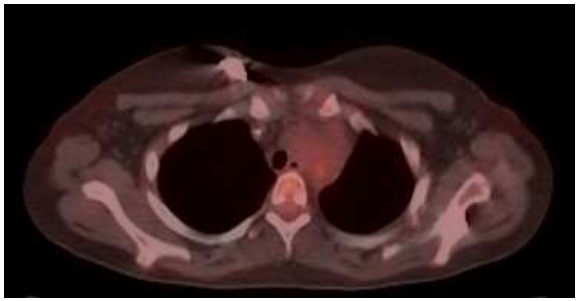
Readers:

4-4-3-3-3
in lower neck, supra- or infraclavicular

qPET=1,13

≙Deauville 3

Case 2848 - Mediastinum

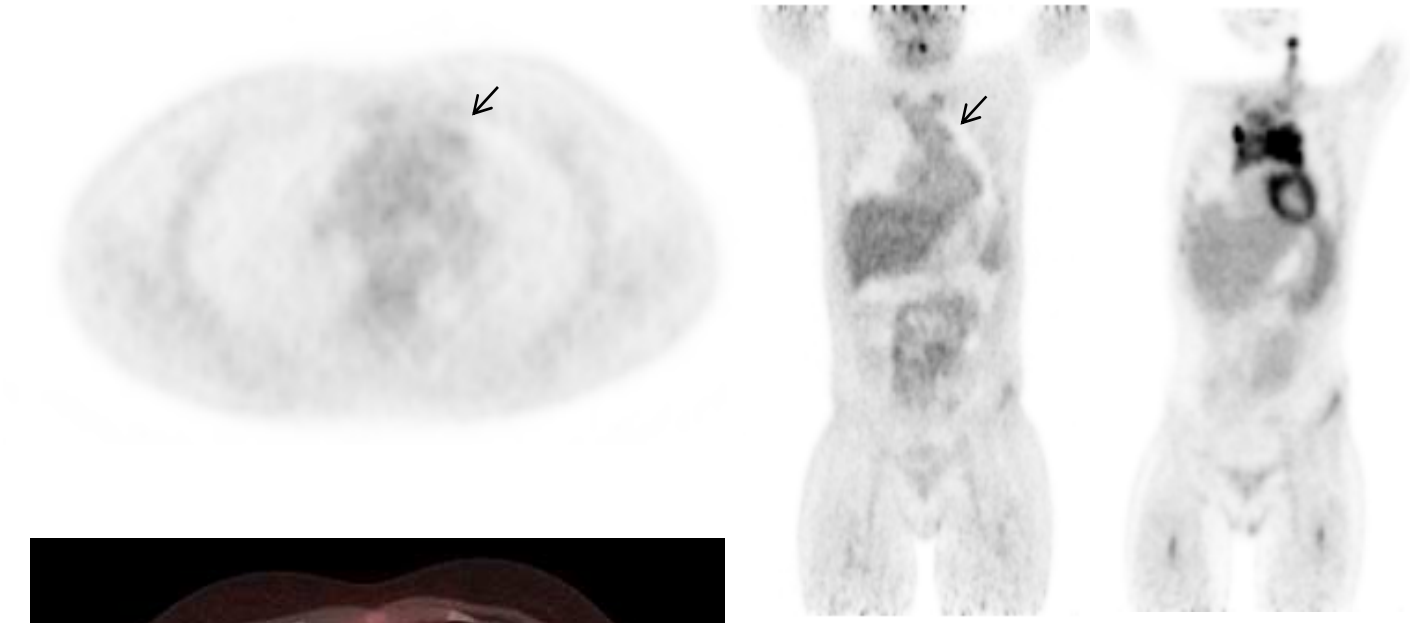


Readers:
3-3-4-4-4
in upper or middle mediastinum

qPET=1,43

≙Deauville 4

Case 2670 - Mediastinum

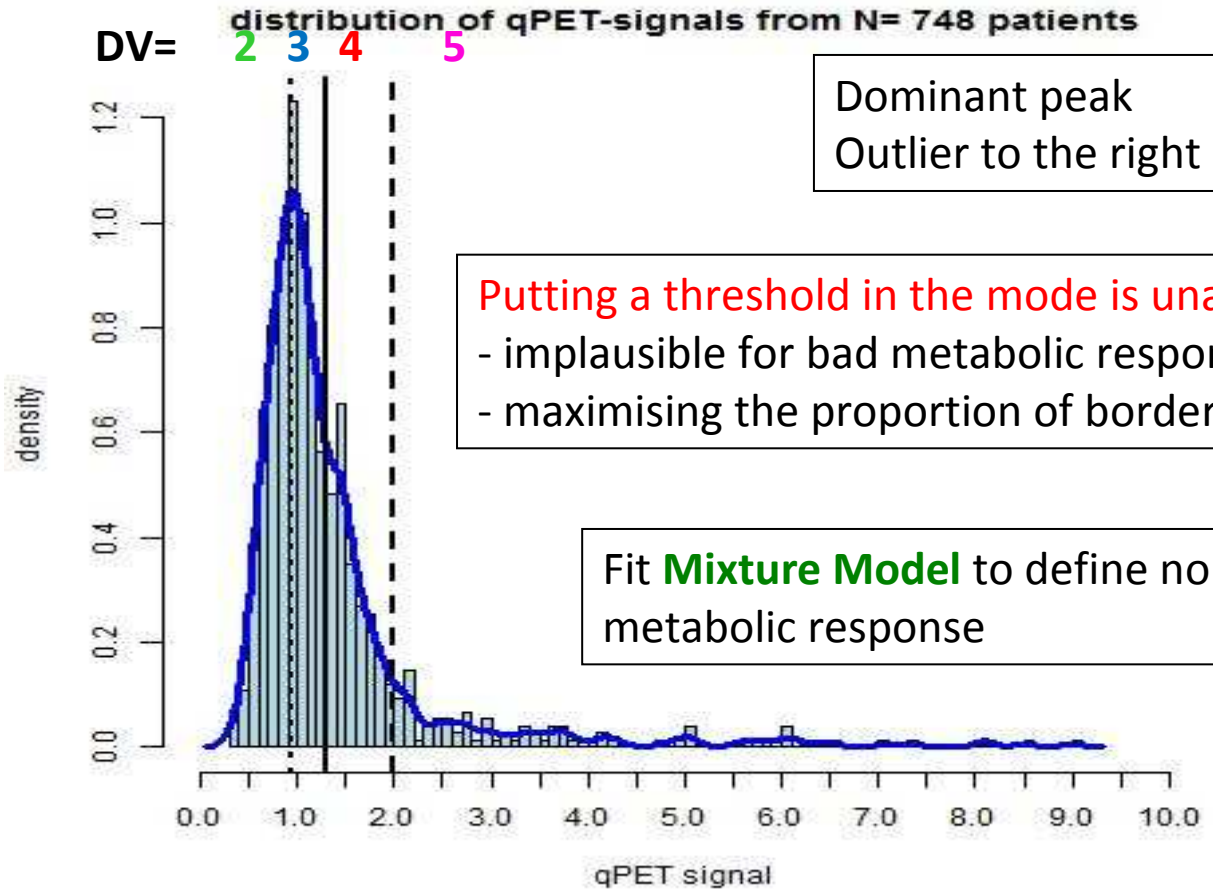


Readers:
1-2-2-3-3
in upper or
middle
mediastinum

qPET=0.80

≙Deauville 2

qPET as continuous extension of Deauville

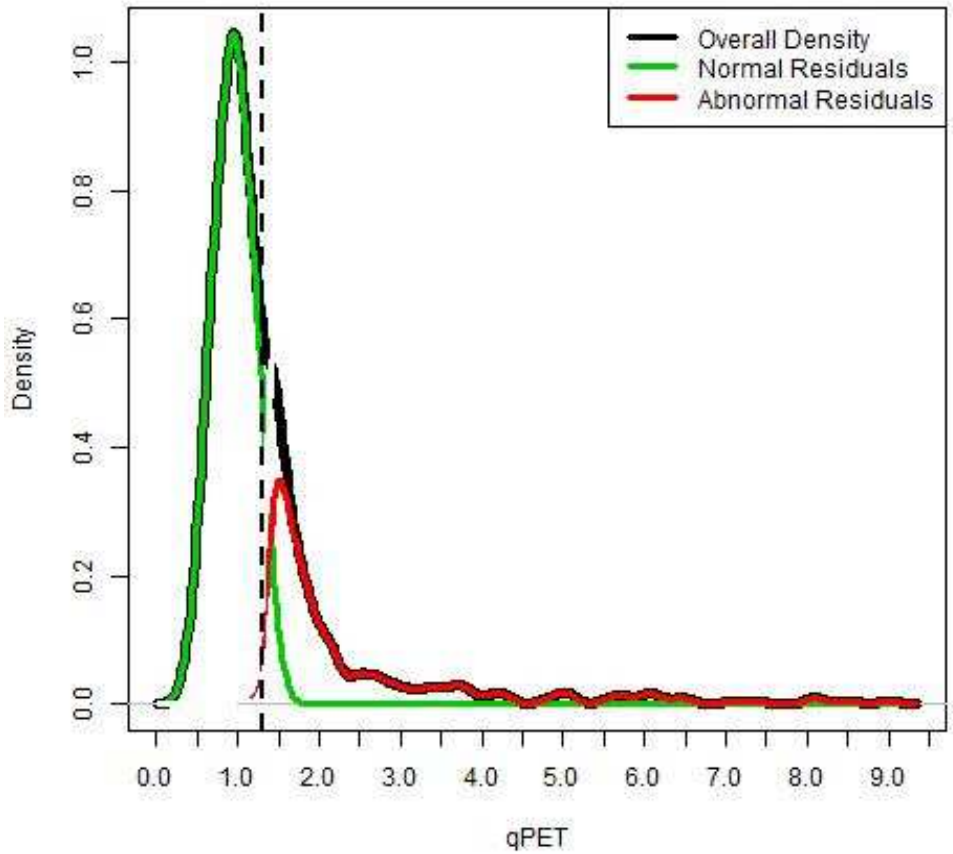


Dominant peak
Outlier to the right

Putting a threshold in the mode is unadvisable -
- implausible for bad metabolic response
- maximising the proportion of borderline cases

Fit **Mixture Model** to define normal versus abnormal metabolic response

Deviation from symmetry of peak



Density of qPET in black
is modelled as mixture of
"normal" and "abnormal" signals.

Deviation from Symmetry
at threshold at 1.3

Using another model higher threshold
at 2.0.

Conclusion

- qPET methodology provides semi-automatic quantification for interim FDG-PET response in HL.
- **qPET extends the ordinal Deauville scoring to a continuous scale.**
- Deauville categories correspond to defined qPET values. Approximate translation is possible.

Conclusion II

- The qPET thresholds corresponding to Deauville borders **should not depend on the particular clinical setting**
  - since only comparison to reference organs is involved.

- **Thresholds between normal and abnormal response can be derived from the qPET-distribution** based on a mixture model without use of follow-up data.
  - Location of the peak may depend on the clinical setting.
  - But form of qPET distribution – peak + outliers should be general

- The continuous qPET scale allows cut point optimisation for prognostication.