PET-CT for Staging & Early Response: Results from ‘Response Adapted Therapy in Advanced Hodgkin Lymphoma’ (RATHL) (CRUK/07/033)

Sally F Barrington¹, Amy A Kirkwood², Antonella Franceschetto³, Michael J Fulham⁴,⁵, Thomas H Roberts², Helén Alquist⁶, Eva Brun⁷, Karin Hjorthaug⁸, Zaid N Viney⁹, Lucy C Pike¹, Massimo Federico¹⁰, Stefano Luminari¹⁰, John Radford¹¹, Judith Trotman⁵,¹², Alexander Fosså¹³, Leanne Berkahn¹⁴, Daniel Molin¹⁵, Francesco D’Amore¹⁶ Donald A Sinclair¹, Paul Smith², Michael J O’Doherty¹, Lindsey Stevens², Peter W Johnson¹⁷

Running Head: PET-CT for staging and response in the RATHL trial

PET Imaging Centre, Division of Imaging Sciences and Biomedical Engineering, King’s College London, King’s Health Partners, St. Thomas’ Hospital, London, SE1 7EH, United Kingdom ¹;
Cancer Research UK and UCL Cancer Trials Centre, UK, London, UK ²; Department of Nuclear Medicine, University of Modena and Reggio Emilia, Modena, Italy ³; Department of Molecular Imaging (PET-CT), Royal Prince Alfred Hospital⁴, and Sydney Medical School, University of Sydney⁵, Australia; Department of Medical Imaging and Physiology⁶ and Department of Oncology and Radiation Physics⁷, Skane University Hospital, Lund University, Sweden; Department of Nuclear Medicine & PET Centre, Aarhus University Hospital, Aarhus, Denmark⁸; Department of Radiology, Guy’s and St Thomas’ NHS Foundation Trust, London, UK⁹; Oncology Unit, Department of Diagnostic, Clinical and Public Health Medicine, University of Modena and Reggio Emilia, Modena, Italy¹⁰; The University of Manchester and the Christie NHS Foundation Trust, Manchester, UK¹¹; Concord, Repatriation General Hospital¹² and Sydney Medical School, University of Sydney, Australia⁵; Department of Oncology, Norwegian Radium Hospital, Oslo, Norway¹³; Haematology, Auckland City
Hospital, Auckland, New Zealand 14; Department of Genetics and Pathology, Experimental and Clinical Oncology, Uppsala University, Uppsala, Sweden 15; Haematology, Aarhus University, Aarhus, Denmark 16; Cancer Research UK Centre, University of Southampton, Southampton, UK 17.

**Corresponding Author:** Dr Sally F Barrington

PET Imaging Centre, St Thomas’ Hospital, Westminster Bridge Road

London SE1 7EH UK

+ 44 207 188 4988 (phone)

+44 207 620 0790 (fax)

sally.barrington@kcl.ac.uk

Presented in part at the 12th International Conference on Malignant Lymphomas, Lugano Switzerland June 2013

**Key points:**

PET-CT is the modern standard for staging Hodgkin Lymphoma and can replace contrast enhanced CT in the vast majority of cases.

Agreement between expert and local readers is sufficient for the Deauville criteria to assess response in clinical trials and the community.
Abstract

International guidelines recommend PET-CT should replace CT in Hodgkin Lymphoma (HL).

The aims of this study were to i) compare PET-CT with CT for staging and ii) measure agreement between expert and local readers, using a five-point scale (Deauville criteria), to adapt treatment in a clinical trial ‘Response Adapted Therapy in Advanced Hodgkin Lymphoma’ (RATHL) www.cancer.gov/clinicaltrials, reference NCT00678327. Patients were staged for the trial using clinical assessment, CT and bone marrow biopsy (RATHL stage).

PET-CT was performed at baseline (PET0) and after 2 chemotherapy cycles (PET2) in a response-adapted design. PET-CT was reported centrally by experts at 5 national core labs. Local readers optionally scored PET2 scans. The RATHL and PET-CT stages were compared.

Agreement amongst experts and between expert and local readers was measured. RATHL and PET0 stage were concordant in 938 (80%) patients. PET-CT upstaged 159 (14%) and downstaged 74 (6%) patients. Upstaging by extranodal disease in bone marrow (92), lung (11) or multiple sites (12) on PET-CT accounted for most discrepancies. Follow-up of discrepant findings confirmed the PET characterisation of lesions in the vast majority. Five patients were upstaged by marrow biopsy; 7 by contrast-enhanced CT in bowel and/or liver or spleen. PET2 agreement amongst experts (140 scans) with kappa (95% CI) of 0.84 (0.76 – 0.91) was very good and between experts and local readers (300 scans) at 0.77 (0.68-0.86) was good. These results confirm PET-CT as the modern standard for staging HL and that response assessment using Deauville criteria is robust enabling translation of RATHL results into clinical practice.
Introduction

Positron Emission Tomography (PET) and PET-CT, using 2-deoxy-2-[fluorine-18]fluoro-D-glucose (FDG), has been extensively used for imaging patients with Hodgkin Lymphoma (HL) \(^1\text{-}^5\). International guidelines recently recommended that PET-CT be used for routine staging of FDG-avid lymphomas and for response assessment using a five-point scale (5-PS), the so-called ‘Deauville criteria’ \(^6\text{,}^7\).

PET-CT was preferred for staging due to improved accuracy compared with CT and as a baseline for subsequent response assessment \(^6\). Contrast enhanced CT (ceCT) may be required if accurate nodal measurement is needed e.g. in clinical trials, assessment of bowel involvement, compression/thrombosis of central vessels and for radiation planning \(^7\). Direct comparison between ceCT and PET-CT when the CT component is performed as a low-dose unenhanced scan, suggests that higher dose ceCT has no impact on lymphoma management \(^8\text{,}^9\). Further, changes in FDG uptake are more relevant than changes in nodal size for response assessment \(^6\). Despite this, both ceCT and PET-CT are frequently performed at diagnosis, with added cost and radiation exposure.

PET is reported to alter stage compared with CT \(^3\text{,}^{10}\), but most publications are retrospective and used stand-alone PET \(^11\text{-}^{16}\). Previous publications compared imaging techniques but do not report the impact of imaging on the more relevant final clinical stage, inclusive of bone marrow biopsy and some publications included patients with HL and non-Hodgkin lymphoma (NHL) \(^17\text{-}^{21}\).

Response assessment with the 5-PS has been reported by a number of investigators to have good interobserver agreement in a training set of 50 patients \(^22\), in a study of 260 advanced
HL patients with international expert readers\textsuperscript{23} and in paediatric HL\textsuperscript{24}. The 5-PS also improves the positive predictive value of PET compared with previous International Harmonisation Criteria\textsuperscript{25}. These studies, however, were all retrospective and the 5-PS was not used to direct therapy.

The aims of this study were to determine in a large cohort of patients with advanced HL, within a prospectively acquired clinical trial:

1. the difference in staging when unenhanced PET-CT is used in place of standard assessment with clinical examination, ceCT, and bone marrow biopsy and,
2. the agreement amongst experts and between local readers using the 5-PS to adapt treatment in ‘real time’.

Methods

Patients were registered in the ‘Response Adapted Therapy in Advanced Hodgkin Lymphoma’ (RATHL) study (www.cancer.gov/clinicaltrials, reference NCT00678327) and gave written informed consent in accordance with the Declaration of Helsinki. The study had Human Ethics Committee approval in all participating countries.

i) Staging

Patients underwent clinical assessment, ceCT of neck, thorax, abdomen, pelvis and bone marrow biopsy to assess stage. RATHL included patients with stages IIB-IV and stage IIA with adverse features. Patients also underwent a PET-CT scan with low-dose unenhanced CT at staging (PETO). ceCT and PET-CT scans were performed within 28 days of enrolment. PET-CT scans were acquired around 60 minutes after the intravenous injection of 350-550MBq FDG. The ceCT scans at diagnosis were reported by the radiologist at the recruiting centre and
clinical assessment, CT findings and bone marrow biopsy were used to assign the final RATHL stage by the treating clinician on the case report form (CRF). In individual cases, ultrasound or MRI were also used for staging. Inclusion in the study was based on the RATHL stage not PET-CT. The PET-CT stage was assigned by central readers at core labs in UK, Italy, Sweden, Denmark and Australia without knowledge of RATHL stage, marrow biopsy or patient outcome. Causes for discrepancy in stage between PET-CT and other modalities were assessed with reference to the imaging reports, bone marrow biopsy results and by observing the changes that occurred with treatment on PET scans performed during chemotherapy. ceCT scans were not re-reviewed centrally. PET-CT scans were performed at multiple international centres using standardised methods for acquisition and quality control.  

ii) Response assessment

PET-CT was repeated after two cycles of doxorubicin, bleomycin, vinblastine, dacarbazine (ABVD) chemotherapy (PET2). All scans were PET-CT and stand-alone PET scans were not permitted in the trial. PET2 scans were performed 9-13 days after day 15 of cycle 2. PET2 was scored using the 5-PS, according to the level of the most intense residual FDG uptake at involved sites on PET0 as follows:

1. no uptake
2. uptake ≤ mediastinum
3. uptake > mediastinum but ≤ liver
4. uptake moderately higher than liver
5. uptake markedly higher than liver and/or new lesions

X new areas of uptake unlikely to be related to lymphoma
Scores 1, 2, 3 were regarded as ‘negative’ and scores 4, 5 as ‘positive’. Score 5 was regarded as uptake > 3 times the maximum SUV in normal liver.

Patients with negative PET2 were randomised to continue with four cycles of ABVD or have de-escalation of treatment with four cycles of AVD. Patients with positive PET2 had escalated treatment with bleomycin, etoposide, doxorubicin, cyclophosphamide, vincristine, procarbazine, prednisone (BEACOPP)-escalated or BEACOPP-14, according to the treating centre preference. A third PET-CT (PET3) was performed 9-13 days after day 8 of cycle 3 of BEACOPP-escalated or 2-6 days after day 8 of cycle 4 of BEACOPP-14 to ensure effectiveness of therapy. It was left to the treating physician’s discretion whether patients were offered salvage therapy if PET3 remained positive. An end-of treatment PET scan was not mandated. Patients with a negative PET2 or PET3 scan were not recommended to receive consolidation radiotherapy as routine, although local investigators had discretion to use radiotherapy if they felt it necessary.

PET2 was scored at the core labs within 72 hours and this score was used to direct treatment. To assess the level of agreement, readers from all core labs read the same paired PET0 and PET2 scans from: 1) a training set of 50 patients \(^{22}\) 2) the first 10 patients scored at each core lab and 3) a further 10 patients scored at each core lab during the trial. Readers at ‘local’ PET Centres were given the option to score scans. Levels of agreement were measured between central (core lab) readers and between local and central readers, using non-weighted kappa statistics\(^{27}\) [Stata version 12.1; Stata Corp, Texas]. Using this threshold, uptake in lesions higher than normal liver uptake resulted in treatment escalation. Agreement was also measured regarding score 1,2 as negative and score 3,4,5
as positive, as this threshold has been used as a benchmark for de-escalation in trials involving patients with HL\textsuperscript{1,28,29}.

Kappa values between 0.81-1.00 indicate very good agreement, 0.61-0.80 good agreement, 0.41-0.60 moderate agreement, 0.21-0.4 fair agreement\textsuperscript{30}.

Results

i) Staging

1214 patients were registered from 2008 - 2012. 1171 baseline PET-CT scans were available for staging assessment which was performed retrospectively (Figure 1).

There was agreement between the RATHL stage and the PET-CT stage in 938 (80%) patients. 159 patients (14%) were upstaged and 74 patients (6%) were downstaged by PET-CT (Table 1). The main reason for upstaging was detection of extranodal disease (Table 2), most commonly in bone marrow (Figure 2). Upstaging due to nodal involvement also occurred, mostly below the diaphragm (Table 2). Reasons for downstaging included enlarged nodes and/or spleen which were not FDG-avid and extranodal sites with abnormal morphology but no FDG uptake (Table 2).

The PET2 scans of patients with discrepant staging findings were compared with the PET0 scan (Supplementary Table 1). At PET2, FDG uptake at sites that resulted in upstaging, decreased in parallel with other sites of disease during treatment in all cases (Table 2). Twenty patients had extranodal lesions on CT that were not FDG-avid. Five of these twenty patients had lesions that did not change on treatment (Figure 3) and were considered unlikely to represent lymphoma (1 adrenal adenoma, 3 lung nodule/s, 1 bone lesion). One patient had a 37mm cavitating lung nodule on CT, which enlarged from 5mm seven days...
prior on PET-CT, and was probably inflammatory. Six patients had indeterminate lung
nodules and one patient had lobar consolidation that all resolved; 3 patients had pleural
effusions that resolved – in all these cases the changes may have been reactive,
inflammatory or related to lymphoma. There were four patients - small bowel (1), liver
lesions (2) and 1 patient with bowel and liver lesions - where ceCT was considered more
likely to indicate the correct stage. Three patients had splenic lesions on ceCT not seen on
PET. Bone marrow involvement was missed on PET-CT but identified on biopsy in five
patients.

There were 21 cases where review of the reports suggested that the same imaging findings
had been differently interpreted by the local radiologist and the core lab PET CT reader or
the local radiologist and the treating clinician (Table 2).

ii) Reporting of early response assessment

1123 PET2 scans were assessable in the trial (Figure 1). 223 PET2 scans were performed in
the PET Centre of one of the core labs (Figure 1). Local readers scored 300 of the remaining
900 scans (33%). 140 PET2 scans were scored by all readers at the five core labs.

When the liver threshold was used, there was agreement between core labs that a scan was
‘negative’ (score 1-3) or ‘positive’ (score 4 or 5) in 122/140 scans; Kappa (95% CI): 0.84 (0.76
– 0.91) indicating very good agreement. There was agreement between central and local
readers in 276/300 scans; Kappa (95% CI): 0.77 (0.68 – 0.86) (Table 3A) indicating good
agreement.

When the mediastinal threshold was used, there was agreement between core labs that a
scan was ‘negative’ (score 1 or 2) or ‘positive’ (score 3-5) in 81/140 scans; Kappa (95% CI):
0.58 (0.50 – 0.66) indicating moderate agreement. There was agreement between central and local readers in 249/305 scans, Kappa (95% CI): 0.64 (0.55 – 0.73) (Table 3B) indicating good agreement.

Identical scores ranging from 1-5 were given in 169/300 PET2 scans by central and local readers (Table 3C). When assigning scans as positive or negative, 15 PET2 scans scored by local readers as ‘positive’ were deemed ‘negative’ by central review whereas 9 PET2 scans scored by local readers as ‘negative’ were deemed ‘positive’ by central review using the liver threshold that determined treatment.

161 PET3 scans were performed, 26 at the PET Centre of one of the core labs (Figure 1). Local readers scored 47 of the remaining PET3 scans (36%). Using the liver threshold, there was agreement amongst central and local readers that a scan was ‘negative’ or ‘positive’ in 45/47 scans; Kappa (95% CI): 0.91 (0.78 – 1.00) indicating very good agreement. For the mediastinal threshold, there was agreement in 39/47 scans; Kappa (95% CI): 0.61 (0.45 – 0.87) indicating good agreement.

Identical scores ranging from 1-5 were given in 24/47 PET3 scans by central and local readers using the liver threshold.

**Discussion**

Our study is the first to compare PET-CT staging with the established standard of clinical assessment, contrast-enhanced CT and bone marrow biopsy stage in a large cohort of patients with advanced HL in an international trial. PET-CT altered staging in 20% of patients compared with the standard approach, which is at variance with earlier reports that
suggested stage change occurred more often in patients with early stage rather than advanced disease \textsuperscript{10,20}. Upstaging occurred more frequently than downstaging with extranodal disease accounting for 74\% of the upstaged scans, mostly due to an increased sensitivity of PET for detecting bone marrow involvement.

Baseline and response scans were compared and the findings were correlated with other imaging, where available, to determine the etiology of lesions that accounted for the discrepancy in stage. This supported the notion that PET-CT stage is more accurate than CT in the majority of cases. There were only 4 patients with organ involvement and 3 with probable splenic involvement identified on ceCT that were missed on PET-CT. Bone marrow biopsy identified involvement in 5 cases (0.4\%) where there was a normal marrow appearance on PET-CT confirming that routine bone marrow biopsy is no longer required\textsuperscript{7,31}.

We cannot determine if staging by PET-CT impacted on management because patients were registered in the trial based on having at least stage IIA disease with adverse risk factors on other imaging and marrow biopsy. Upstaging by PET-CT would have been unlikely to impact management as patients had already been assessed as requiring full course chemotherapy. Similarly, downstaging from stage 4 to stage 3 would not have impacted on treatment. PET-CT, however, downstaged 56 patients to stage 2 and 1 patient to stage 1 (5\% of the study population) and this could potentially influence treatment choices.

Our results are in keeping with earlier reports \textsuperscript{11,13-16} that PET stages HL patients differently to ceCT in approximately 15 - 30\% of cases. Early reports had fewer patients, were retrospective and published prior to the widespread introduction of PET-CT \textsuperscript{11-16}. More recently Hutchings et al\textsuperscript{3} compared PET with CT staging prospectively in 99 patients with HL, (61 with PET-CT) and Rigacci in 186 patients (56 with PET-CT)\textsuperscript{10}. Upstaging, especially by
identification of extranodal sites, occurred more frequently than downstaging, similar to our findings. Lesions were also overlooked on CT and more easily identified on PET.

The ‘truth’ as to whether lesions identified only on PET represent lymphoma is difficult to determine as biopsy of discrepant lesions is rarely performed. Correlative imaging, treatment response, follow-up, biopsy, or a combination, usually serves as a reference standard. Young et al, in 49 HL patients, verified stage by laparotomy in 11 patients and biopsy of discordant lesions in the remainder. PET stage was correct in 26 upstaged patients and a single patient downstaged compared to CT. Taken together, these studies demonstrated improved sensitivity and similar specificity using PET compared to CT.

Changes in management were reported in 12/186 patients by Rigacci et al. Ten patients were upstaged from limited to advanced disease and 2 patients had an increase in the radiotherapy field using PET. Hutchings et al reported that 7 patients upstaged by PET to advanced disease were treated for limited disease but only 1 patient developed progressive lymphoma. Eighteen patients had disease progression overall, suggesting that understaging by CT probably did not adversely impact on outcome. The same group later reported, however, that the routine use of PET-CT in their clinical practice resulted in stage migration, with a higher risk of progression associated with focal FDG-avid skeletal lesions. Bone marrow involvement was the most common reason for discrepancy in stage between PET and CT in our series. Munker et al also reported significantly more treatment failures among patients staged as I or II on CT yet III or IV on PET compared with patients who were stage I or II by both CT and PET. More accurate delineation of disease is thus likely to be of benefit for patient management and for radiation planning.
International guidelines have recommended that PET-CT should be used for staging of FDG-avid lymphomas\(^7\), because it is more accurate than CT and because a staging scan improves the accuracy of response assessment\(^6\). The role of ceCT in staging is still debated although the international guidelines concluded that ceCT had limited application\(^7\) as it rarely altered staging or management.\(^8,9\) Our study confirms this conclusion; organ involvement was detected on ce-CT but not PET-CT in 4 patients and splenic lesions in 3 patients, leading to stage change. On the other hand PET-CT detected extranodal disease in 118 patients and splenic FDG-avid foci in 7 patients and stage was changed. Further the effective dose associated with ceCT is approximately 16mSv, which is similar to the combined dose from PET with lower dose unenhanced CT. Avoiding ceCT for staging would thus reduce a patient’s radiation dose by 50%. Although ceCT may be required for planning radiotherapy, in our study, less than 10% of patients required this\(^7\).

PET-CT is recommended for response assessment in FDG-avid lymphomas using the 5-PS\(^34\). The 5-PS has good interobserver agreement\(^22,24,25,35-37\) and is predictive of patient outcomes\(^29,35,38-40\), especially in advanced HL\(^37,41\). An advantage of the 5-PS is that the threshold used to define complete metabolic response can be altered according to the clinical context or research question\(^34\). In the RATHL design where patients with a ‘positive’ scan received escalation from ABVD to BEACOPP, investigators preferred to use the liver threshold to avoid the risk of over-treating patients. In some trials designed to explore de-escalation strategies, a lower mediastinal threshold has been used to avoid the risk of under-treating patients\(^1,28,29\). Yahalom expressed concern that the good agreement reported amongst expert readers may not be reproducible in the community setting, in
particular, if treatment was adapted on the basis of the result. Our study demonstrated similar levels of agreement amongst experts and local readers, with good or very good agreement for the liver but lower moderate to good agreement for the mediastinal threshold. In RATHL the outcome for patients with a ‘negative’ scan was not influenced by the PET2 score and our observations suggest that the liver is likely to be a more reproducible threshold, although readers may possibly have paid closer attention to the decision to assign a score of 4 rather than 3, knowing this would result in treatment escalation. The better agreement using the liver may be explained by the higher uptake in liver than mediastinum, so the contrast between lesions with low-grade uptake and the reference region is easier to appreciate. In addition, there is more uniform uptake in the liver than the mediastinum where uptake can be heterogeneous with focal uptake in the vessel wall. Standardisation of patient preparation and scanning are important to ensure homogenous uptake in the liver. The agreement for scoring at PET3 was good, which has not previously been reported, although the numbers assessed were small. This is reassuring as BEACOPP chemotherapy can be associated with diffuse bone marrow uptake, which might have made interpretation more challenging.

The main limitations to our study were that it was not possible for ceCT scans to be re-reviewed alongside PET-CT scans and we could not measure the impact of PET-CT on management. Bias may have occurred in the scoring of response scans by local readers as it was optional to score scans, although readers were evenly split between academic and non-academic institutions.
Conclusion

Staging of HL patients in this large prospective study confirms that an important proportion will be staged differently using PET-CT compared to clinical assessment, CT and bone marrow biopsy. When discordance occurs in the imaging stage, PET-CT is usually more accurate than CT, which may impact on management. These findings support the move to a modern standard using PET-CT for staging and suggests that in the vast majority of cases ceCT is not required.

Good agreement between local and expert readers indicates that the 5-PS is robust for assessing response when standardised PET protocols are used and it works effectively in the community setting and in clinical trials. The final results of the RATHL trial will determine if response adaptation using the 5-PS is successful at improving outcomes in advanced HL. In the meantime, our results strengthen the application of the 5-PS as the optimal method for response assessment in HL.

Acknowledgements

The authors are grateful to Cancer Research UK (CRUK/07/033), the Associazione Angela Serra for Cancer Research, Modena, Italy, Larvik kreftforening, Norway and Cancer Australia’s Priority Driven Collaborative Cancer Research Scheme who provided funding for the RATHL trial. The authors are also grateful to Cancer Research UK, the National Institute for Health Research in England and the Departments of Health for Scotland, Wales and Northern Ireland (C19631/A16091 – MRPTADR) for funding the UK NCRI PET Core Lab. The enthusiastic support of PET reviewers, PET Centres, Trial Investigators, Patients and their families is also gratefully acknowledged.
Author Contributions

SFB, PWJ designed the research; all authors performed research; SFB, AAK, THR, LP, MF, SL, JR, JT, AFossa, LB, DM, FD’A, DS, PS, PWJ collected data; SFB, AAK, A Franceschetto, MJF, HA, EB, KH analyzed and interpreted data; AAK performed statistical analysis; all authors wrote and approved the manuscript.

Conflict of Interest

Peter Johnson - paid consultancies for Takeda and Bristol-Myers Squibb

All other authors have no conflicts of interest.
References


42. Yahalom J. Chemotherapy only in early-stage Hodgkin lymphoma: more relapses but "same" (or possibly worse) survival--reconsidering the misguided trend to omit radiotherapy. *Curr Hematol Malig Rep.* 2014;9(3):212-216.


Table 1 - level of agreement between RATHL stage (assigned by clinical assessment, ceCT and bone marrow biopsy) and the baseline PET-CT.

<table>
<thead>
<tr>
<th>PET-CT</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>406</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>38</td>
<td>240</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>47</td>
<td>74</td>
<td>292</td>
</tr>
</tbody>
</table>
Table 2 - Reasons for upstaging and downstaging according to PET-CT compared with the RATHL stage.

<table>
<thead>
<tr>
<th>Upstaging by PET-CT due to</th>
<th>N (% of upstaged pts)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extranodal disease on PET-CT in the following sites:</strong></td>
<td>118 (74.2)</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>92</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
</tr>
<tr>
<td>Liver</td>
<td>2</td>
</tr>
<tr>
<td>Pleura</td>
<td>1</td>
</tr>
<tr>
<td>Multiple organs</td>
<td>12</td>
</tr>
<tr>
<td><strong>Nodal disease:</strong></td>
<td>35 (22.0)</td>
</tr>
<tr>
<td>Normal sized nodes that were FDG avid below the diaphragm</td>
<td>20</td>
</tr>
<tr>
<td>Normal sized nodes that were FDG avid above the diaphragm</td>
<td>7</td>
</tr>
<tr>
<td>Splenic FDG avid foci</td>
<td>7</td>
</tr>
<tr>
<td>Both of the above</td>
<td>1</td>
</tr>
<tr>
<td><strong>Difference in opinion:</strong></td>
<td>6 (3.8)</td>
</tr>
<tr>
<td>Imaging findings interpreted differently by the local radiologist and the core lab PET CT reader or the treating clinician</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Downstaging by PET-CT due to</th>
<th>N (% of downstaged pts)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extranodal disease:</strong></td>
<td>25 (33.8)</td>
</tr>
<tr>
<td>Lung</td>
<td>11</td>
</tr>
<tr>
<td>Bone marrow</td>
<td>5</td>
</tr>
<tr>
<td>Pleura</td>
<td>3</td>
</tr>
<tr>
<td>Bone</td>
<td>1</td>
</tr>
<tr>
<td>Adrenal gland</td>
<td>1</td>
</tr>
<tr>
<td>Liver</td>
<td>2</td>
</tr>
<tr>
<td>Small bowel</td>
<td>1</td>
</tr>
<tr>
<td>Multiple organs</td>
<td>1</td>
</tr>
<tr>
<td><strong>Nodal disease:</strong></td>
<td>34 (45.9)</td>
</tr>
<tr>
<td>Enlarged nodes that were not FDG avid below the diaphragm</td>
<td>21</td>
</tr>
<tr>
<td>Enlarged nodes that were not FDG avid above the diaphragm</td>
<td>1</td>
</tr>
<tr>
<td>Splenomegaly on CT, normal FDG uptake</td>
<td>8</td>
</tr>
<tr>
<td>Splenic lesion/s on ceCT not FDG avid</td>
<td>3</td>
</tr>
<tr>
<td>Splenomegaly on CT &amp; enlarged nodes not FDG avid</td>
<td>1</td>
</tr>
<tr>
<td><strong>Difference in opinion:</strong></td>
<td>15 (20.3)</td>
</tr>
<tr>
<td>Imaging findings interpreted differently by the local radiologist and the core lab PET CT reader or the local radiologist and the treating clinician</td>
<td>15</td>
</tr>
</tbody>
</table>
Table 3 - Agreement between ‘central’ and ‘local’ readers at PET2 using liver threshold (A) mediastinal threshold (B) and scores 1-5 (C)

<table>
<thead>
<tr>
<th></th>
<th>PET2 (using liver threshold)</th>
<th>PET2 (using mediastinal threshold)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Central</td>
<td>Central</td>
</tr>
<tr>
<td></td>
<td>Local Positive</td>
<td>Negative</td>
</tr>
<tr>
<td>Positive</td>
<td>54</td>
<td>15</td>
</tr>
<tr>
<td>Negative</td>
<td>9</td>
<td>222</td>
</tr>
</tbody>
</table>

C PET2 (using five scores)

<table>
<thead>
<tr>
<th></th>
<th>Central score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5</td>
</tr>
<tr>
<td>Local score</td>
<td>1  55  59  15  2  0</td>
</tr>
<tr>
<td></td>
<td>2  3  47  20  2  0</td>
</tr>
<tr>
<td></td>
<td>3  3  1  19  5  0</td>
</tr>
<tr>
<td></td>
<td>4  3  5  7  34  2</td>
</tr>
<tr>
<td></td>
<td>5  0  0  0  4  14</td>
</tr>
</tbody>
</table>
Legends

Figures

Figure 1
Consort flow diagram demonstrating progress of patients through the trial from baseline to PET3.

Figure 2
CT (left) PET (right) coronal images show a case upstaged by PET-CT.
Nodal and splenic involvement was reported on CT and interpreted as stage 3. The PET scan showed multifocal uptake in bone marrow upstaging to stage 4.

Figure 3
CT (left) PET (right) axial images show a case downstaged by PET-CT.
There was nodal disease in the mediastinum and a 22mm lung nodule in the left lung (arrowed), reported on CT as stage 4. The PET scan showed high uptake in lymph nodes but no FDG uptake in the lung nodule suggesting the nodule was unlikely to be due to lymphoma interpreted as stage 2 (Panel A). After treatment, there was resolution of uptake in lymph nodes but the lung nodule (arrowed) was unchanged (Panel B).
Registered
N = 1214

Excluded
Misdiagnosis, not HL, n = 3

Eligible for analysis
N = 1211

Baseline PET

Not assessable
N = 40
Baseline scans could not be retrieved
IV contrast used in CT component of PET-CT
n = 27
n = 13

Available
N = 1171
Core labs *
UK n = 926
Italy n = 128
AUS/NZ n = 82
Sweden n = 24
Denmark n = 11

PET scan not assessable due to scanning protocol violations
N = 67

No PET2 scan performed
N = 21

No PET 3 scan performed
N = 1050

PET 2
N = 1123

Performed at one of the Core Labs
N = 223

Central score given, Local readers chose not to score scans
N = 600

‘Central’ and ‘local’ scores available
N = 300
UK n = 163
Italy n = 98
AUS/NZ n = 39

PET 3
N = 161

Performed at one of the Core Labs
N = 26

Central score given, Local readers chose not to score scans
N = 88

‘Central’ and ‘local’ scores available
N = 47
UK n = 29
Italy n = 18

*UK Core Lab read scans from UK & Norway

Australia Core Lab read scans from Australia and New Zealand
PET-CT for staging & early response: results from 'Response Adapted Therapy in Advanced Hodgkin Lymphoma' (RATHL) (CRUK/07/033)