What FDG-PET response-assessment method best predicts survival after curative-intent chemoradiation in non-small cell lung cancer (NSCLC) EORTC, PERCIST, Peter Mac or Deauville criteria?


Published in:
The Journal of Nuclear Medicine

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright © 2018 by the Society of Nuclear Medicine and Molecular Imaging. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
What FDG-PET response-assessment method best predicts survival after curative-intent chemoradiation in non-small cell lung cancer (NSCLC): EORTC, PERCIST, Peter Mac or Deauville criteria?

Guy-Anne Turgeon¹, Amir Iravani², Tim Akhurst², Alexis Beaulieu², Jason W. Callahan², Mathias Bressel³, Aidan J. Cole¹,⁶, Sarah J. Everitt⁴,⁵, Shankar Siva¹,⁵, Rodney J. Hicks²,⁵, David L. Ball¹,⁵, Michael P. Mac Manus¹,⁵

¹ Division of Radiation Oncology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia
² Cancer Imaging, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia
³ Department of Biostatistics and Clinical Trials, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia
⁴ Radiation Therapy, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia
⁵ Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, VIC, Australia
⁶ Centre for Cancer Research and Cell Biology, Queen's University, Belfast, Northern Ireland.
**Corresponding and first author:**

Guy-Anne Turgeon MD, FRCPC (fellow)

305 Grattan Street, Melbourne

Victoria, 3000, Australia

Telephone: +61 3 8559 7720

Fax: +61 3 8559 7729

Email: guy-anne.turgeon@petermac.org

**Short running title:**  FDG-PET response criteria in NSCLC
Acknowledgement of research support:

The $^{18}$F-FLT/$^{18}$F-FDG study (ACTRN12611001283965) was supported by the National Health and Medical Research Council (APP1003895) and the Victorian Cancer Agency. The $^{68}$Ga-Ventilation/perfusion PET study (U1111-1138-4421) was supported by the National Health and Medical Research Council (APP1038399) and the Cancer Australia Priority-driven Collaborative Cancer Research Scheme (Project No.1060919).

Disclaimers:

There are no known conflicts of interest associated with this publication. GA Turgeon, A Iravani, T Akhurst, A Beaulieu, JW Callahan, M Bressel, AJ Cole, SJ Everitt and MP Mac Manus have no disclosure. DL Ball has an advisory role for Pfizer Australia. RJ Hicks has ownership interest in Telix Radiopharmaceuticals. S Siva has an advisory role for Astellas and Janssen pharmaceuticals, receives research funding from Varian Medical Systems and Merck-Sharp Dohme Pharmaceuticals as well as speakers’ bureau, and travel/accommodations/expenses from Bristol Bristol-Meyer-Squibb pharmaceuticals.
ABSTRACT

The optimal methodology for defining response with $^{18}$F-fluorodeoxyglucose positron emission tomography (FDG-PET) after curative-intent chemoradiation for non-small cell lung cancer (NSCLC) is unknown. We compared survival outcomes according to European Organization for Research and Treatment of Cancer (EORTC), Positron Emission tomography Response Criteria In Solid Tumors 1.0 (PERCIST), Peter Mac Metabolic Visual and Deauville criteria respectively. **Methods:** Three prospective trials of chemoradiation for NSCLC, involving baseline and post-treatment FDG-PET/Computer Tomography (CT) imaging, were conducted between 2004 and 2016. Responses were categorized as complete metabolic response (CMR), partial metabolic response, stable metabolic disease or progressive metabolic disease. Cox proportional hazard models and logrank tests assessed the impact of each response on overall survival (OS). **Results:** Eighty-seven patients underwent FDG-PET/CT before and after radical chemoradiation for NSCLC. Follow-up FDG-PET/CT scans were performed at a median of 89 days (IQR 79-93 days) after radiotherapy. Median follow-up and OS after PET response imaging were 49 months and 28 months respectively. Inter-observer agreements for EORTC, PERCIST, Peter Mac and Deauville were kappa 0.76, 0.76, 0.87 and 0.84, respectively. All four response criteria were significantly associated with OS. Peter Mac and Deauville showed better fit compared to EORTC and PERCIST, and distinguished better between CMR and non-CMR. **Conclusion:** All four response criteria were highly predictive of overall survival, but visual criteria showed greater inter-observer agreement and stronger discrimination between CMR and non-CMR, highlighting the importance of visual
assessment to recognize radiation pneumonitis, changes in lung configuration and patterns of response.

Key Words: FDG-PET/CT; NSCLC; PERCIST; EORTC; Peter Mac; Deauville criteria
INTRODUCTION

Imaging with $^{18}$F-fluorodeoxyglucose positron emission tomography (FDG-PET) plays an integral role in multidisciplinary treatment decision-making, at diagnosis, restaging or therapeutic response assessment, in patients with non-small cell lung cancer (NSCLC). Reliable early response assessment after curative-intent therapy would not only identify patients with persistent or progressive disease who require further treatment but also could avoid overtreatment of patients who may already be cured. Structural imaging, including computer tomography (CT) or magnetic resonance imaging, has significant limitations after treatment with curative-intent radiotherapy or chemoradiation (chemoRT) especially in lung malignancies. Tumors may be obscured by atelectasis, pneumonitis or fibrosis. In addition, tumor masses often regress gradually over several months, and residual opacity or scarring may persist, mandating serial measurements to assess response. FDG-PET/CT can overcome many of the limitations of structural imaging in response-assessment but the optimum method for characterising response to predict overall survival is not established.

Although survival remains poor in locally advanced NSCLC treated with chemoRT (1), the addition of adjuvant immunotherapy in stage III NSCLC has lately demonstrated a significant increase in progression-free survival (2). For decision-making concerning treatment intensification with immunotherapy, targeted therapies, localized radiotherapy dose escalation or even salvage surgery (3), a personalized risk-adapted approach, based on accurate early assessment of response after chemoradiation, would be logical.
The necessity for standardization of FDG-PET reporting is well recognized and different groups have developed criteria to categorize FDG-PET response assessment, with the objective of increasing both intra- and inter-observer reproducibility. Two semi-quantitative response criteria have been proposed, the European Organization for Research and Treatment of Cancer (EORTC) (4) and the Positron Emission tomography Response Criteria in Solid Tumors 1.0 (PERCIST) (5,6) criteria. The first visual response criteria (Peter Mac) for categorising PET responses after radiotherapy were reported in 2003 (7). Deauville visual response criteria (8-10) were developed specifically for lymphoma, but to our knowledge have never been studied in NSCLC. The semi-quantitative approaches (EORTC, PERCIST) are often assumed to be more accurate and reproducible than visual qualitative interpretations (Peter Mac, Deauville). However, semi-quantitative methods can be labour intensive for daily clinical practice and have not been systematically compared with rigorous visual criteria.

In this study, we compared the EORTC, PERCIST, Peter Mac and Deauville criteria for NSCLC response assessment after radical radiotherapy or chemoRT to discover which method is best able to predict overall survival (OS).

MATERIALS AND METHODS

Patients

Between 2004 and 2016, three prospective trials at our institution enrolled NSCLC patients treated with definitive radiotherapy or chemoRT. Seventy-six patients were enrolled in a PET-planning study (11) (Peter Mac protocol 03/55), 60 patients in the $^{18}$F-
FLT/$^{18}$F-FDG study (12) (Australian Clinical Trials Registry: ACTRN12611001283965) and 60 patients in $^{68}$Ga-Ventilation/perfusion PET study (13) (Universal Trial Number: U1111-1138-4421). All patients were aged ≥18 years, had an Eastern Cooperative Oncology Group performance status of 0-2 and had a histologically or cytologically confirmed NSCLC. All cases were reviewed in a multi-disciplinary lung tumor board. Follow-up schedules were uniform in these patient cohorts, with reviews every 3 months for 2 years and every 6 months until 5 years. Patients eligible for this analysis had undergone pre- and post-treatment FDG-PET/CT imaging, the latter acquired between 1.5 and 4 months after radiotherapy. Patients with distant metastatic disease, local recurrence or treatment after complete surgical resection were ineligible. The institutional review board, the Peter MacCallum Cancer Clinical Research and Ethics Committee approved this retrospective study and waived the requirement to obtain informed consent. All patients had previously provided written informed consent for their respective prospective studies.

**Treatment Policy**

All patients were treated to 50-60 Gy using three-dimensional conformal or intensity modulated radiotherapy planning according to institutional guidelines. Patients underwent staging, simulation and treatment with arms raised and breathing freely. Tumor motion was accounted for using the FDG-PET/CT planning scan (11) or with a four-dimensional planning CT scan. Concomitant chemotherapy was administered, either cisplatin/etoposide, or carboplatin/paclitaxel.
PET Scanning Acquisition and Processing

All FDG-PET/CT scans were acquired on an integrated PET/CT scanner including GE Discovery LS, GE STE (GE Medical Systems Milwaukee, WI, USA) or Biograph 16 (Siemens Medical Solutions, Erlangen, Germany). Each baseline and post-treatment FDG-PET/CT were performed on the same scanner with uniform protocol. Patients fasted for at least 6 hours and had blood glucose measurements before administration of 4.2 MBq per kilogram of 18F-FDG. The emission scan commenced 60 to 70 minutes later.

Assessment of Treatment Response

All semi-quantitative analyses and qualitative assessments were performed using MIM software (MIM 5.4.4; MIM Software, Cleveland, OH). For each patient, the four FDG-PET/CT response criteria were reported retrospectively blinded to outcome. Two readers (AI and TA) assessed Peter Mac and Deauville response criteria and two (JC and AB) assessed EORTC and PERCIST criteria. Readers were paired into groups based on their years of experience in FDG-PET/CT response assessment (>10 years or <5 years). Responses to therapy were categorized as complete metabolic response (CMR), partial metabolic response (PMR), stable metabolic disease (SMD) or progressive metabolic disease (PMD) as described in the online Supplemental Table 4 and as illustrated in Figure 1.

Inter-observer agreement between the two respective observers were calculated for each criterion. For the EORTC and PERCIST criteria all cases were read independently. For the Peter Mac and Deauville criteria, AI and TA assessed together the
first 19 cases to discuss the interpretation of the Deauville criteria as they were not usually employed outside of lymphoma. The remaining cases were assessed independently. For the criteria comparison, all discrepant cases were discussed between the two respective observers and a consensus classification was reached.

**Statistical Methods**

Overall survival was measured from the date of post-treatment FDG-PET/CT to the date of death. Patients alive at the last contact had their survival censored at that date. Kaplan-Meier methods were used to describe the survival curves for each of the four response criteria and also grouped as CMR vs. non-CMR. Cox proportional hazard models were used to estimate the hazard ratios, calculate the c-statistic, Akaike information criterion and $r^2$ and assess the impact of each criteria on overall survival. Univariable and multivariable results adjusting for sex, histology, age, performance status, stage, weight loss and treatment were provided. Cohen’s Kappa was used to assess the paired concordance and to assess the inter-observer agreement of each criteria. Fisher’s exact test was used to compare the CMR rate of responses assessed prior or after 90 days after the last day of radiotherapy. All statistical analyses were performed in R 3.2.3 (R Core Team, 2015). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria).

**RESULTS**

**Patients**
There were 87 patients eligible for analysis: 8 from the PET-planning study, 47 from the $^{18}$F-FLT/$^{18}$F-FDG study and 32 from the GALLIPET-VQRT study. Baseline and treatment characteristics are described in Table 1. Supplemental Table 5 presents the patients' baseline glucose levels for the pre- and post-FDG-PET/CT. In the PET-planning study, response assessment with FDG-PET/CT was optional and only 21 patients were potentially eligible for analysis. However, only 8 patients' images could be retrieved from PACS archives. In the $^{18}$F-FLT/$^{18}$F-FDG and GALLIPET-VQRT studies, the post-treatment FDG-PET/CT was mandatory. For these studies patients were ineligible as follows: no post-treatment FDG-PET/CT (n=30), M1 disease at baseline (n=4), radiation for recurrent disease (n=5) and one case each of small-cell-lung cancer and adjuvant radiotherapy for microscopic disease. All patients completed their planned irradiation.

Follow-up FDG-PET/CT scans were performed at a median of 89 days (IQR 79-93 days and full range 47-123 days) after radiotherapy. The CMR rate using the EORTC criteria did not differ (p=0.64) from patients who had their post-treatment FDG-PET/CT within 47-89 days (CMR = 25%) and within 90-132 days (CMR = 30%). There was also no difference when response was assessed using the other three criteria (results not shown).

**Inter-Observer Agreement**

The EORTC and PERCIST inter-observer agreement was assessed on all 87 patients. The weighted Kappa for EORTC, PERCIST, Peter Mac and Deauville were respectively 0.76 (95% CI, 0.63-0.89), 0.76 (95% CI, 0.62-0.89), 0.87 (95% CI, 0.75-0.99)
and 0.84 (95% CI, 0.70-0.91). All subsequent analyses were based on the consensus decisions.

**Inter-Criteria Agreement**

All scans were evaluated except in one case where PERCIST response was not assessable due to missing patient data. For EORTC, PERCIST, Peter Mac and Deauville criteria, CMR/PMR/SMD were reported in 24/37/9, 24/38/7, 30/39/1 and 31/38/1 patients, respectively. In all criteria, PMD was reported in 17 patients with new lesions outside of the treatment fields.

Both semi-quantitative criteria (EORTC and PERCIST) and both qualitative response criteria (Peter Mac and Deauville) showed almost perfect agreement with each other with kappa values of respectively 0.95 (95% CI, 0.89-1.00) and 0.98 (95% CI, 0.95-1.00). The agreements between the semi-quantitative and the qualitative criteria were lower (Table 2). When EORTC/PERCIST were discordant with the Deauville/Peter Mac (Figure 2), they underestimated the visual criteria response in all but one case.

**Overall Survival by Response**

The estimated median follow-up was 49 months and the median survival, calculated from the date of the follow-up FDG-PET/CT, was 28 months. All four response criteria were associated with overall survival (Table 3). Overall, the Peter Mac and the Deauville criteria showed stronger associations compared to EORTC and PERCIST
(higher c-statistic, higher \( r^2 \) and lower Akaike information criterion, see Supplemental Table 6). Overall survival for each response criteria is shown in Figure 3.

The predicted 2-year OS for CMR versus non-CMR was respectively: 76% (95%, CI 60-97%) versus 51% (95%, CI 39-66%) for EORTC, 76% (95%, CI 60-97%) versus 50% (95%, CI 38-65%) for PERCIST, 85% (95%, CI 73-100%) and 44% (95%, CI 32-60%) for Peter Mac criteria and 82% (95%, CI 69-98%) and 45% (95%, CI 33-61%) for Deauville criteria.

**DISCUSSION**

The intention of definitive radiotherapy/chemoRT in unresectable NSCLC is permanent eradication of disease. Ideally, therapeutic response assessment should reproducibly identify a group of patients with an excellent prognosis, many of whom are likely to be cured, and also to identify those with partial responses or progressive disease who may benefit from early intervention with additional therapies. In this setting the reliability of the assessment of CMR is critical. Although CMR patients may still harbor subclinical disease and may benefit from additional therapy, it will be important to avoid overtreatment in a group where many are already cured.

To harmonize FDG-PET/CT reporting and determine which FDG-PET/CT response criteria is most predictive for NSCLC, two semi-quantitative (EORTC and PERCIST) and two qualitative (Peter Mac and Deauville) response criteria were compared. Reassuringly, all four response criteria showed highly significant associations
with overall survival. These findings are consistent with our previous 2003 report, that qualitative interpretation of metabolic response at a single time-point early after curative-intent radiotherapy/chemoRT provides powerful predictive information, stratifying patients into groups with widely differing survival probabilities (7,14). An early post-treatment FDG-PET scan was more powerfully predictive of OS that CT imaging, stage or performance status.

Prognostic stratification using FDG-PET/CT response has been explored in different settings in NSCLC. Two pathological validation studies performed in patients treated with trimodality approaches demonstrated the association between maximum standardized uptake value and pathologic response after neoadjuvant treatment (3,15). After single modality palliative chemotherapy, the two semi-quantitative PET response criteria (EORTC and PERCIST) were shown to be more sensitive and accurate than RECIST 1.1 for the detection of an early therapeutic response (16,17). Fledelius et al. applied PERCIST and Peter Mac response criteria following induction chemotherapy and prior to radiotherapy in 21 NSCLC patients. Both criteria showed comparable results with a strong association with OS (18). The same group evaluated different visual and semi-quantitative methods as well as different response category cut-off points to evaluate the early treatment response of NSCLC after 7-10 days of erlotinib (19). In their analysis of 29 different methods and parameters total lesion glycolysis used in combination with PERCIST and Peter Mac visual criteria were the best methods for predicting treatment responses.
Metabolic response assessment with FDG-PET can yield false-positive results related to active inflammation, especially in the early post-radiotherapy phase (20,21). Serial imaging indicates that that inflammatory $^{18}$F-FDG uptake in normal tissues increases in the first few months after treatment rather than occurring early during radiotherapy (22).

In our study, the two visual criteria showed stronger associations with OS than EORTC and PERCIST. Although many aspects of PERCIST have been improved compared to the EORTC criteria, rigid interpretation of the SUL values cannot account for post-radiotherapy inflammatory changes and alteration in the size of the region of interest can compromise assessment. Accurate measurement of SUL values, accurate assignment of region-of-interest and accurate registration of the images from a series of examinations, particularly in the presence of FDG-avid pneumonitis and evolving change in the lung configuration, can be challenging. In addition, the respiratory motion in the lung may superimpose inflammatory activity onto the adjacent residual mass on CT. Careful qualitative interrogation of PET images is critically important. PET physicians and radiologists must use their clinical expertise to interpret the distribution, shape and variations of the FDG activity. The qualitative visual assessments of Peter Mac and Deauville allow this flexible interpretation, which in this study has translated into a superior prediction of OS and critically superior distinction between CMR and non-CMR.

To account for post-treatment inflammation, the Hopkins group in their devised visual criteria incorporated an intermediate response group, defined by diffuse FDG
uptake above blood pool or liver activity (23). This group was classified as “probably inflammatory” and considered negative for malignancy for their final analysis. When this group was separately analysed for OS association, a distinct intermediate survival curve was delineated. Based on semi-quantitative response assessment this group could be categorised as PMR, SMD or even PMD with misleading prognostication (Figure 2a). In our study, the largest inter-criteria response migration was seen in SMD group, with the semi-quantitative response showing higher number of patients despite their survival curve being similar to CMR patients. This again underscores the importance of a meticulous visual assessment on sequential PET scans. Respiratory gating may help to minimize such discordance (24).

EORTC and PERCIST criteria have shown a kappa of 0.95 corresponding to almost perfect agreement with each other. A meta-analysis of 6 studies included 348 patients with solid tumors (13% with lung cancer) showed an almost perfect concordance between EORTC and PERCIST with a kappa of 0.946. Four additional studies confined to NSCLC (16,25-27) corroborate this excellent agreement (28).

The Peter Mac and Deauville responses were identical in all but one patient. This is unsurprising as these criteria are very similar and are intended to be simple and reproducible. For CMR, Peter Mac criteria requires the uptake to be equal or less than the reference tissue in which the baseline lesion is located. Due to very low background activity in the normal lung parenchyma, blood pool is the reference tissue in chest malignancies. Deauville, however, uses the organs of reference (blood pool or liver
activity) for comparison. In lymphoma clinical trials, Deauville score 1-3 (uptake below or equal to the liver, see Supplemental Table 4) on post-treatment PET are usually considered CMR (29). Adopting from lymphoma, in this study Deauville score of 3 (equal to liver) was considered CMR. In practice, the difference between metabolic activity in the blood pool and liver is rather small and may be difficult to differentiate visually, particularly in the presence of treatment induced inflammatory changes. Using liver as the organ of reference provides slightly higher visual range for the imaging specialist which may translate to easier interpretation and potentially less subjectivity. In addition, Deauville criteria has the advantage of widespread familiarity among imaging specialists. Nonetheless, prospective studies are needed to validate Deauville criteria in lung cancer and other solid tumour malignancies.

By pairing the two observers in each group based on their years of experience in PET response assessment, we aimed to minimise the inevitable inter-observer variability. All four criteria showed strong inter-observer agreement. The slightly higher kappa values of Peter Mac criteria may reflect the greater familiarity of the observers with these criteria. In a study by Fledelius et al. following two cycles of chemotherapy (no radiation) in 35 NSCLC patients, eight readers evaluated the response based on PERCIST and Peter Mac criteria (30). Both approaches showed strong inter-observer agreement but with higher overall agreement in PERCIST criteria. Subjective variability in inclusion of atelectatic changes in the region of interest for semi-quantitative analysis was mentioned as one the reason for variability of response assessment by PERCIST. This is likely to be a more significant problem after radiotherapy due to its local inflammatory effects.
Therefore, careful visual analysis is of paramount importance, regardless of the response criteria used.

This study has some limitations that should be acknowledged, including the retrospective nature of the analysis and the long-time period over which scans were accrued. However, all PET scans were performed as part of prospective trials, harmonised protocols were applied within each, and all assessments were blinded to the eventual outcomes.

CONCLUSION

In patients with NSCLC treated with definitive chemoRT, qualitative and semi-quantitative FDG-PET/CT response criteria both provided powerful early, post-treatment predictive information. The Peter Mac and Deauville visual criteria, showed stronger associations compared to the two semi-quantitative criteria, EORTC and PERCIST. Regardless of the criteria used, careful and intelligent visual assessment of FDG-PET/CT images is of paramount importance due to commonly-occurring post-radiotherapy inflammatory changes.

ACKNOWLEDGEMENTS

The authors are grateful for the contributions of Lisa Selbie and Ann Officer our radiation oncology data managers and clinical trial colleagues.
REFERENCES


Figure 1: Computer Tomography (CT) (top left) and FDG-PET/CT fusion (top right) are baseline images and CT (bottom left) and FDG-PET/CT fusion (bottom right) are post-chemoradiotherapy images. Examples of the four response categories: A) complete metabolic response (CMR); B) partial metabolic response (PMR); C) stable metabolic disease (SMD) and D) progressive metabolic disease (PMD). Case A, B and D had consensus responses for the four criteria (EORTC, PERCIST, Peter Mac and Deauville). Case C was reported SMD by EORTC, PERCIST but PMR by Peter Mac and Deauville.
Figure 2: Computer Tomography (CT) (top left) and FDG-PET/CT fusion (top right) are baseline images and CT (bottom left) and FDG-PET/CT fusion (bottom right) are post-chemoradiotherapy images. Examples of response discrepancies between semi-quantitative and visual responses. Case A) demonstrates significant changes in the lung configuration and pneumonitis in the periphery of the residual lung lesion causing difficulty in assigning the region-of-interest and in the registration of pre- and post-treatment images. This was assessed PMR by EORTC and PERCIST, while following careful visual assessment, Peter Mac and Deauville assigned CMR. Case B) shows a significant decrease in size of the lesion but due to $<30\%$ reduction in the target $\text{SUL}_{\text{peak}}$ and $<15\%$ reduction in the region-of-interest $\text{SUV}_{\text{BSA}}$, this was assessed SMD by EORTC and PERCIST. Peter Mac and Deauville reported PMR based on the reduction of the extent of metabolic activity. Case C) shows minimal residual uptake overlapping the target lesion. Due to the overlapping uptake compared to the surrounding background, EORTC and PERCIST reported PMR. Peter Mac and Deauville reported CMR.
Figure 3: Overall survival by the four FDG-PET/CT response criteria: A) EORTC, B) PERCIST, C) Peter Mac and D) Deauville.

A

Overall survival (%)

Time since posttreatment PET (mo)

Number at risk

<table>
<thead>
<tr>
<th></th>
<th>CMR</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24</td>
<td>19</td>
<td>16</td>
<td>16</td>
<td>14</td>
<td>11</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>33</td>
<td>28</td>
<td>23</td>
<td>18</td>
<td>15</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Overall survival (%)

Time since posttreatment PET (mo)

Number at risk

<table>
<thead>
<tr>
<th>Group</th>
<th>24</th>
<th>19</th>
<th>16</th>
<th>16</th>
<th>14</th>
<th>11</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMR</td>
<td>24</td>
<td>19</td>
<td>16</td>
<td>16</td>
<td>14</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>PMR</td>
<td>38</td>
<td>34</td>
<td>28</td>
<td>23</td>
<td>18</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>SMD</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>PMD</td>
<td>17</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Overall survival (%) vs Time since posttreatment PET (mo)

Number at risk:

<table>
<thead>
<tr>
<th>Group</th>
<th>30</th>
<th>24</th>
<th>22</th>
<th>22</th>
<th>20</th>
<th>16</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PMR</td>
<td>39</td>
<td>35</td>
<td>27</td>
<td>22</td>
<td>16</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>SMD</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PMD</td>
<td>17</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 1: Patient and tumor characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
<td><strong>Total (%)</strong></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td>Median [range] 68 [46-86]</td>
</tr>
<tr>
<td>Sex</td>
<td>30 (35%)  Male 57 (65%)</td>
</tr>
<tr>
<td>ECOG</td>
<td>32 (37%)  52 (60%)  3 (3%)</td>
</tr>
<tr>
<td>Weight lost &gt;10% at dx</td>
<td>No 76 (87%)  Yes 11 (13%)</td>
</tr>
<tr>
<td>Stage at diagnosis</td>
<td>I 7 (8%)  II 10 (11%)  IIA 45 (52%)  IIIB 25 (29%)</td>
</tr>
<tr>
<td>Histology</td>
<td>Adenocarcinoma 36 (41%)  Large Cell 7 (8%)  NSCLC 13 (15%)  SCC 31 (36%)</td>
</tr>
<tr>
<td>Molecular testing</td>
<td>Complete (ALK, BRAF, EGFR &amp; KRAS) 19 (22%)  Incomplete 13 (15%)  Not done 55 (63%)</td>
</tr>
<tr>
<td>Mutation</td>
<td>ALK 1 (3%)  BRAF 1 (3%)  EGFR 2 (6%)  KRAS 6 (19%)  No known mutation 22 (69%)</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td>Concurrent 80 (92%)  No chemotherapy 7 (8%)</td>
</tr>
<tr>
<td>Chemotherapy type</td>
<td>Carboplatin-Etoposide 1 (1%)  Carboplatin-Paclitaxel 45 (56%)  Carboplatin only 3 (3%)  Cisplatin-Etoposide 31 (39%)</td>
</tr>
<tr>
<td>radiation dose</td>
<td>50-56 Gy 3 (3%)  60 Gy 84 (97%)</td>
</tr>
</tbody>
</table>

ECOG: Eastern Cooperative Oncology Group Performance Status; dx: diagnosis; SCC, squamous cell carcinoma.
Table 2: Inter-criteria agreement.

<table>
<thead>
<tr>
<th>Method</th>
<th>PERCIST Kappa (95% CI)</th>
<th>Deauville Kappa (95% CI)</th>
<th>Peter Mac Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EORTC</td>
<td>0.95 (0.89-1.00)</td>
<td>0.71 (0.59-0.83)</td>
<td>0.73 (0.61-0.85)</td>
</tr>
<tr>
<td>PERCIST</td>
<td>0.74 (0.62-0.86)</td>
<td>0.76 (0.64-0.87)</td>
<td></td>
</tr>
<tr>
<td>Deauville</td>
<td>0.98 (0.95-1.00)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kappa, weighted Cohen's Kappa ; CI, confidence interval.
Table 3: Overall survival model fit by response.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Level</th>
<th>N</th>
<th>HR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EORTC</td>
<td>Per category</td>
<td>87</td>
<td>1.6 (1.2-2.2)</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>CMR</td>
<td>24</td>
<td>1</td>
<td>0.197</td>
</tr>
<tr>
<td></td>
<td>Non-CMR</td>
<td>63</td>
<td>1.6 (0.8-3.5)</td>
<td></td>
</tr>
<tr>
<td>PERCIST</td>
<td>Per category</td>
<td>86</td>
<td>1.6 (1.2-2.2)</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>CMR</td>
<td>24</td>
<td>1</td>
<td>0.187</td>
</tr>
<tr>
<td></td>
<td>Non-CMR</td>
<td>62</td>
<td>1.6 (0.8-3.5)</td>
<td></td>
</tr>
<tr>
<td>Deauville</td>
<td>Per category</td>
<td>87</td>
<td>1.8 (1.3-2.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>CMR</td>
<td>31</td>
<td>1</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Non-CMR</td>
<td>56</td>
<td>2.3 (1.1-4.8)</td>
<td></td>
</tr>
<tr>
<td>Peter Mac</td>
<td>Per category</td>
<td>87</td>
<td>1.8 (1.3-2.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>CMR</td>
<td>30</td>
<td>1</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Non-CMR</td>
<td>57</td>
<td>2.6 (1.2-5.7)</td>
<td></td>
</tr>
</tbody>
</table>

N, number of patients; HR, hazards ratio; CI, confidence interval; CMR, complete metabolic response.