



Editorial on “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?”

Joan Fledelius

Department of Nuclear Medicine, Herning Hospital, Herning, Denmark

Correspondence to: Joan Fledelius. Department of Nuclear Medicine, Herning Hospital, Gl. Landevej 62, 7400 Herning, Denmark.

Email: Joan.Fledelius@rm.dk.

Provenance: This is an invited Editorial commissioned by the Section Editor Jun Zhou (Department of Nuclear Medicine, Zhongshan Hospital, Fudan University, Shanghai, China).

Comment on: Turgeon GA, Irvani A, Akhurst T, *et al.* 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? *J Nucl Med* 2018. [Epub ahead of print].

Submitted Sep 13, 2018. Accepted for publication Sep 25, 2018.

doi: 10.21037/jtd.2018.09.136

View this article at: <http://dx.doi.org/10.21037/jtd.2018.09.136>

The treatment of non-small cell lung cancer (NSCLC) is fortunately an expanding field. Many new treatment regimens are emerging, with second, third, and fourth line therapies. This inadvertently results in a need for a reliable, early evaluation of the treatment response, during or at the end of treatment, enabling a more personalized treatment planning. Treatment response with FDG-PET/CT is an excellent candidate since the metabolic response evaluated with this modality precedes the anatomical response traditionally measured with CT scans and has been shown to be a promising tool in NSCLC (1-4). The variation in how to measure and report a response with FDG-PET/CT is large, rendering the comparison of studies on this subject very difficult.

The response is measured either visually or semi-quantitatively. The EORTC criteria and the PERCIST are the two major guidelines for evaluating response semi-quantitatively (5,6) and the Peter Mac criteria or some variation of this is often used for visual evaluation (2). Both the EORTC and PERCIST evaluate the “hottest lesion” response, and the visual evaluation is a total disease evaluation. Independent on the method chosen for evaluation, rigorous attention to the comparability of the compared scans is of utmost importance (7-10).

The ideal method for evaluation of response to treatment should result in distinctly different categories of response resulting in significantly different survival, ideally both overall survival and relapse- or progression-free survival. The method should be readily usable and reproducible, allowing for identification of a subpopulation that has a continuous response and no relapse, and another group of patients with a high risk of early relapse. The first group will need no further treatment, and perhaps fewer follow-up visits, but the latter group will need closer attention and perhaps early or immediate retreatment.

In a recently published manuscript in *Journal of Nuclear Medicine*, Turgeon and the esteemed response evaluation group from Melbourne elegantly compared two visual evaluation methods (the Peter Mac and slightly modified Deauville criteria) and two semi-quantitative methods (the EORTC and PERCIST) according to inter-observer variability and their ability to provide the best categorization for prediction of survival (11). They evaluate the response after chemoradiation therapy with curative intent in a total of 87 patients from three different prospective trial carried out between 2004 and 2016.

They report higher kappa values for the visual evaluation methods than for the semi quantitative methods 0.87 (95%

CI: 0.75–0.99) for Peter Mac and slightly lower but similar values for the Deauville method compared to 0.76 (95% CI: 0.62–0.89) for PERCIST, and nearly exactly the same for the EORTC criteria. This is interesting, since the visual evaluation methods is generally thought to be the most subjective method. According to the Peter Mac criteria a partial metabolic response is defined as “Any appreciable reduction of FDG uptake intensity of target tumor or reduction in tumor volume/extent and residual FDG uptake within target tumor greater than the mediastinum”, “appreciable reduction” being the subjective part of the definition. Whereas the definition of a partial metabolic response according to PERCIST is: “Decrease of target lesion SULpeak by $\geq 30\%$ and at least 0.8 SULpeak units’ difference and no increase $>30\%$ in SULpeak or size of target and non-target lesions”. The very high kappa values they provide show us that it is possible to exceed the more objective methods such as PERCIST. This should be seen in the light of the fact that the Peter Mac method was developed at their institution and may very well have provided them with the experience and expertise needed for this high level of agreement.

The major discrepancies between PERCIST and the Peter Mac method in the Melbourne study is that the visual evaluation identifies more complete responders and fewer stable disease patients than PERCIST, now this must result from a very sensitive “filter” or interpretation of “appreciable reduction” and could very well reflect the evaluators clinical experience. This is an important (of many) advantage of visual evaluation incorporating the evaluators knowledge and experience and implies the flexibility of the visual evaluation.

The visual and semi-quantitative methods all identify the same 17 patients with metabolic progressive disease and poor survival on account of new FDG-avid lesions outside the field of radiation treatment, this is hardly surprising since this is included in all methods as a criterion for progressive disease. The response was evaluated between 47 and 123 days after radiotherapy (with a median of 89 days), no infield progression was found at this time, which would possibly have been evaluated differently by to the four methods used.

A comparison of the four methods showed very high kappa values between the two visual methods and between the two semi-quantitative methods, but lower kappa values when comparing visual and semiquantitative methods, this is not surprising and is possibly a result of overall disease burden versus single lesion evaluation. In this study,

there might be an element of interobserver variation in the latter type of comparison, since the visual evaluations were performed by another pair of observers than the semi-quantitative evaluations, unfortunately. The visual evaluation methods showed a stronger association with overall survival than the semiquantitative methods, and a higher difference in 2-year OS for complete metabolic response (CMR) than non-CMR of 85% (95% CI: 73–100%) versus 44% (95% CI: 32–60%) for Peter Mac and 76% (95% CI: 60–97%) versus 50% (95% CI: 38–65%) for PERCIST, suggesting a higher clinical relevance.

Other groups, including ours have found that semi-quantitative evaluation provides a high agreement between observers and in our study, we found it to be higher than for the visual methods (12–14). We studied the interrater agreement for evaluating response after 2 cycles of induction chemotherapy, prior to radiotherapy in a similar population and found that the agreement was higher with PERCIST (Fleiss kappa for 8 observers): 0.76 (95% CI: 0.71–0.81) than with the Peter Mac method: 0.60 (95% CI: 0.55–0.64). We achieved Cohens kappa values (pairwise comparison) between 0.60 and 0.88 for PERCIST and between 0.50 and 0.76 for the Peter Mac method. Besides the difference in the timepoint for response evaluation (early versus late) another difference from the Melbourne group was, that ours was less experienced in response evaluation at the time of the study.

In a study on very early response evaluation in erlotinib treated patients after only 7–10 days of treatment we compared the general PERCIST method percentage (%) change in SULmax to among others % change in TLG and to visual evaluation. We found that the visual evaluation was at least as good at predicting the CT response after 3 months of treatment as the best performing semi-quantitative methods, in this study the total lesion glycolysis, even in this very early setting (15). In the same population in line with the recent Melbourne study we found a clear correlation between response categories and survival using the Peter Mac visual method, but a sensitive cut off level using % change in TLG showed the strongest correlation to both overall and progression free survival (16).

The Melbourne study is of great interest because visual evaluation is used worldwide, and at my institution is the method used in the daily clinic. It has some major advantages over the semi-quantitative methods, among others it allows for incorporation of the evaluators experience and knowledge of the disease, it is flexible, it is

a total disease burden evaluation and is often quicker than semi-quantitative analysis. It would be very interesting to evaluate whether the very high inter-observer agreement can be reproduced across institution, if this could be established the use of clearly described visual criteria such as the Peter Mac method would be appealing because of the quick overall disease evaluation it can provide.

The overall disease evaluation using total lesion glycolysis can be cumbersome, particularly in the advanced disease setting. It would be important to do some initial evaluations to compare and adjust the visual “filter” or interpretation of “appreciable decrease or increase” before using the Peter Mac method in studies including several centers.

It is important to recognize that different methods could be optimal for different settings, it seems reasonable to assume that one size does not fit all. The 30% change in SULpeak as defined as the limit for both response and progression by PERCIST could be optimal at a certain point during treatment, it is simple to use, but perhaps the limit should be altered according to the timing of the response evaluation, this complicates the already confusing area of response evaluation.

It all comes down to the sensitivity of the “filter” for change in FDG-uptake is applied whether it is a visual or a semi-quantitative filter. In the post treatment setting it is the distinction between complete metabolic response and partial metabolic response that is the determining factor for a method to successfully identifying a lasting, continuous response, whereas in an earlier evaluation during treatment it is likely to be the distinction between partial metabolic responders and stable metabolic disease, or even metabolic progressive disease and stable metabolic disease that should be focused on.

The post radiation treatment setting provides an additional complication in the form of post-radiation inflammatory changes, increasing the background FDG-uptake to a varying degree in the tumor surroundings, this is likely to be easier to interpret with the more flexible visual evaluation incorporating the experience of the evaluator, and could be an important contributor to the superiority of the visual evaluation in this setting. When evaluating response to chemotherapy, be it early or late in the course of treatment, this inflammatory response is not as big an issue.

Independent of the choice of method it is of utmost importance to pay strict attention to the standardization of the patient conditions and the comparability of the

baseline and follow-up scans. The definitions suggested by the PERCIST group are the most comprehensive, and to my experience not too difficult to comply to. It is often mentioned that it is of high importance to do this when evaluating semi-quantitatively, but I would argue that is equally important for visual evaluation, particularly when evaluating with a sensitive “filter”, the evaluators interpretations is equally dependent on comparable conditions for the baseline and follow-up scan.

Perhaps a practice of reporting a “standard FDG-response evaluation Table 1” (similar to the classic oncologist Table 1 summarizing the patient and treatment characteristics) could be adopted including the variations in standardization parameters such as: Injected activity, Glucose level, time between injection and scan, and the adherence to PERCIST or similar guidelines, including time between treatment and scan would also be beneficial. This way it would be easier for all to evaluate the validity of the comparison between scans.

Acknowledgements

None.

Footnote

Conflicts of Interest: The author has no conflicts of interest to declare.

References

1. Langer NH, Christensen TN, Langer SW, et al. PET/CT in therapy evaluation of patients with lung cancer. *Expert Rev Anticancer Ther* 2014;14:595-620.
2. Mac Manus MP, Hicks RJ, Matthews JP, et al. Positron emission tomography is superior to computed tomography scanning for response-assessment after radical radiotherapy or chemoradiotherapy in patients with non-small-cell lung cancer. *J Clin Oncol* 2003;21:1285-92.
3. Mac Manus MP, Hicks RJ. The role of positron emission tomography/computed tomography in radiation therapy planning for patients with lung cancer. *Semin Nucl Med* 2012;42:308-19.
4. Michaelis LC, Ratain MJ. Measuring response in a post-RECIST world: from black and white to shades of grey. *Nat Rev Cancer* 2006;6:409-14.
5. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [18F]-

- fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer* 1999;35:1773-82.
6. Wahl RL, Jacene H, Kasamon Y, et al. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 2009;50 Suppl 1:122S-50S.
 7. Shankar LK, Hoffman JM, Bacharach S, et al. Consensus recommendations for the use of 18F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute Trials. *J Nucl Med* 2006;47:1059-66.
 8. Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med* 2009;50 Suppl 1:11S-20S.
 9. Kramer GM, Frings V, Hoetjes N, et al. Repeatability of Quantitative Whole-Body 18F-FDG PET/CT Uptake Measures as Function of Uptake Interval and Lesion Selection in Non-Small Cell Lung Cancer Patients. *J Nucl Med* 2016;57:1343-9.
 10. Munk OL, Tolbod LP, Hansen SB, et al. Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative. *EJNMMI Phys* 2017;4:5.
 11. Turgeon GA, Iravani A, Akhurst T, et al. 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? *J Nucl Med* 2018. [Epub ahead of print].
 12. Fledelius J, Khalil A, Hjorthaug K, et al. Inter-observer agreement improves with PERCIST 1.0 as opposed to qualitative evaluation in non-small cell lung cancer patients evaluated with F-18-FDG PET/CT early in the course of chemo-radiotherapy. *EJNMMI Res* 2016;6:71.
 13. Jacene HA, Leboulleux S, Baba S, et al. Assessment of interobserver reproducibility in quantitative 18F-FDG PET and CT measurements of tumor response to therapy. *J Nucl Med* 2009;50:1760-9.
 14. Benz MR, Herrmann K, Walter F, et al. (18)F-FDG PET/CT for monitoring treatment responses to the epidermal growth factor receptor inhibitor erlotinib. *J Nucl Med* 2011;52:1684-9.
 15. Fledelius J, Winther-Larsen A, Khalil AA, et al. 18F-FDG PET/CT for Very Early Response Evaluation Predicts CT Response in Erlotinib-Treated Non-Small Cell Lung Cancer Patients: A Comparison of Assessment Methods. *J Nucl Med* 2017;58:1931-7.
 16. Fledelius J, Winther-Larsen A, Khalil AA, et al. Assessment of very early response evaluation with 18F-FDG-PET/CT predicts survival in erlotinib treated NSCLC patients-A comparison of methods. *Am J Nucl Med Mol Imaging* 2018;8:50-61.

Cite this article as: Fledelius J. Editorial on “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?”. *J Thorac Dis* 2018;10(Suppl 33):S4076-S4079. doi: 10.21037/jtd.2018.09.136