

Original Article

How to study optimal timing of PET/CT for monitoring of cancer treatment

Werner Vach¹, Poul Flemming Højlund-Carlsen², Barbara Malene Fischer³, Oke Gerke², Wolfgang Weber⁴

¹Clinical Epidemiology, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Stefan-Meier-Str. 26, D-79104 Freiburg, Germany ²Department of Nuclear Medicine, Odense University Hospital, Sdr. Boulevard 29, DK-5000 Odense C, Denmark ³Department of Clinical Physiology and Nuclear Medicine, Hvidovre Hospital, Copenhagen, Kettegård Alle 30, DK-2650 Hvidovre, Denmark ⁴Department of Nuclear Medicine, University Medical Center Freiburg, Hugstetterstr. 55, D-79106 Freiburg, Germany

Received June 5, 2011; accepted July 5, 2011; Epub July 20, 2011; Published August 15, 2011

Abstract: Purpose: The use of PET/CT for monitoring treatment response in cancer patients after chemo- or radiotherapy is a very promising approach to optimize cancer treatment. However, the timing of the PET/CT-based evaluation of reduction in viable tumor tissue is a crucial question. We investigated how to plan and analyze studies to optimize this timing. Methods: General considerations about studying the optimal timing are given and four fundamental steps are illustrated using data from a published study. Results: The optimal timing should be examined by optimizing the schedule with respect to predicting the overall individual time course we can observe in the case of dense measurements. The optimal timing needs not to and should not be studied by optimizing the association with the prognosis of the patient. Conclusions: The optimal timing should be examined in specific 'schedule optimizing studies'. These should be clearly distinguished from studies evaluating the prognostic value of a reduction in viable tumor tissue.

Keywords: cancer, response evaluation, prognostic value, optimal schedule

Introduction

The use of PET/CT for monitoring treatment response in cancer patients after chemo- or radiotherapy is a very promising approach to optimize cancer treatment [1–4] and already in use in some selected areas [5, 6]. The basic idea is to use PET/CT to judge the reduction in viable tumor tissue from the amount of glucose uptake and to continue with the current therapy, if a sufficient reduction is observed, but to adapt the dose or change to an alternative therapy, if an insufficient reduction is observed. In this way, molecular imaging criteria such as the recently developed Positron Emission Tomography Response Criteria in Solid Tumors (PERCIST) [7] allow response to be measured through assessment of metabolic activity rather than by recording of a decrease in anatomic size which is the conventional way to do it [8, 9]. Standardized uptake values (SUV) obtained by following a standardized measurement protocol prior to the start of therapy and at some pre-

specified time points during or after the therapy are the most promising option to come to objective, reproducible and reliable decisions on the reduction in viable tumor tissue [10], especially as differences of SUV measurements in the same patient are robust against the dependence of standardized uptake values on patient related factors [11].

However, the timing of the PET/CT-based evaluation of a reduction in viable tumor tissue is a crucial question [4, 12]. It is obvious that at least one measurement prior to therapy is necessary in order to have a baseline value. The question of how many follow-up measurements should be made during or after therapy and when they should be placed in time is more difficult to answer for several reasons:

1. SUV measurements are affected by some non-negligible measurement error [13–15]. Consequently, one has to wait long enough to be able to distinguish between random fluctua-

Optimal timing of PET/CT for monitoring of cancer treatment

tions of SUV measurements and a true reduction [16].

2. It may take some time for the therapy to affect the tumor, and hence a reduction may occur only after some onset time.

3. Reduction in viable tumor tissue may be a temporary phenomenon, and the tumor may start to grow again.

Additional complications appear, if one wants also to regard a deceleration in tumor growth without reaching a reduction as a sign of response. As tumors grow at different pace in different patients, at least two baseline measurements with sufficient interval are needed to judge the specific pace in a particular patient. However, this may at times be difficult to obtain since therapy should not be unduly postponed.

Although the question of optimal timing is known as a crucial one, there are actually very few studies addressing explicitly this question. The typical study on the value of PET/CT for the evaluation of treatment response is based on one follow-up measurement at a certain time point in addition to a baseline measurement and a comparison with some clinical outcome measure like survival. It is, therefore, impossible to learn from such a single study something about the optimal time point for response evaluation. A comparison of studies of this kind may allow some insight using meta-analytic techniques, but this approach is limited by the fact that it is hard to find a pair of studies differing only by the timing of follow-up measurements. Only few studies have included two follow-up measurements (e.g. [10, 17, 18]) or even more [19]. It is the purpose of this paper to present some general considerations about how to conduct and analyze a study to determine the optimal timing of PET/CT-based response evaluations and to illustrate these considerations using data from a small published study [19].

General considerations

When considering the use of PET/CT for response evaluation via monitoring the reduction in viable tumor tissue, it is important to distinguish between two different scientific questions:

1. Is PET/CT applied according to a predefined

time schedule capable of detecting and quantifying such a reduction?

2. Is a reduction in viable tumor tissue a good prognostic marker for the long term outcome of the patient, and, hence, a clinically valuable marker of treatment response?

The importance of distinguishing between these two questions stems from the fact that the answers to these two questions can be quite different. It may be that SUV measurements are perfect in assessing the reduction, but that there are good reasons that the reduction is not a perfect prognostic marker, as a reduction in viable tumor tissue does not tell us much about existing metastases, the degree of malignancy or the metastatic potential of the tumor. Or it may be that reduction in viable tissue is in fact the essential key to the long term prognosis of a patient, but that a high measurement error of SUV measurements or a high inter-individual variation in the reduction processes over time makes an assessment of the actual reduction in a patient difficult. Consequently, it is in general a poor idea to try to find an optimal schedule for a PET/CT-based evaluation by correlating the reduction observed directly with some clinical outcome, as we are mixing these two questions. It is wiser to address the two questions separately by different analytic approaches, although the necessary data collection may be combined.

The first question requires access to data on the development of the amount of viable tumor tissue under treatment. The most obvious type of data is series of SUV measurements obtained prior to the start of therapy and at several predefined time points during and/or after therapy in a well-defined subgroup of patients. This allows studying the typical individual patterns of the reduction process, e.g., whether it is linear in each patient or whether it follows some growth curve patterns, whether there is some onset prior to start of the reduction, or whether the reduction may stop and change to growth again. If one can – despite some variation in the individual patterns – agree upon some quantification of the essential aspect in the reduction processes (e.g., a slope or the reaching of a certain level) based on the complete series of measurements, then one can start to investigate the ability of different schedules of the SUV measurements to approximate this target measure. For example, one can investigate at which

Optimal timing of PET/CT for monitoring of cancer treatment

Table 1. Mean tumor SUV measurements and survival in 15 patients with inoperable non-small cell lung cancer

| Patient | SUV in week | | | | | | | Survival (in weeks) |
|---------|-------------|------|-----|-----|------|-----|------|---------------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1 | 6.9 | 7.2 | 7.2 | 5.2 | 5.6 | 5.2 | 5.8 | 76 |
| 2 | 2.9 | 3.4 | 3.0 | 2.9 | 3.1 | 3.0 | 2.7 | >87 |
| 3 | 6.5 | 7.7 | 6.6 | 5.6 | 4.3 | 2.7 | 2.4 | >87 |
| 4 | 12.3 | 10.8 | | | | | | |
| 5 | 7.1 | 6.6 | 5.8 | 6.4 | 6.0 | 5.7 | 6.0 | 23 |
| 6 | 8.9 | 8.5 | 8.6 | 9.2 | 8.5 | 8.4 | 10.7 | 8 |
| 7 | 9.2 | 9.0 | 8.5 | 9.0 | 10.0 | 9.1 | 10.0 | 19 |
| 8 | 9.1 | 9.6 | 7.4 | 8.8 | 7.7 | 9.6 | 9.9 | 30 |
| 9 | 8.1 | 10.9 | 7.1 | 6.7 | 7.7 | 7.2 | 6.7 | 35 |
| 10 | 6.3 | 6.2 | 6.4 | 5.6 | 5.0 | 5.1 | 5.9 | >44 |
| 11 | 7.2 | 5.8 | 6.0 | 6.5 | 6.7 | 6.8 | 7.2 | 20 |
| 12 | 7.7 | 7.7 | 7.0 | 7.0 | 6.6 | 6.2 | 6.4 | 25 |
| 13 | 9.8 | 8.8 | 7.6 | 6.7 | 7.4 | 6.3 | 7.2 | >41 |
| 14 | 7.8 | 5.6 | 7.6 | 7.7 | 7.6 | 6.8 | 7.8 | 15 |
| 15 | 8.6 | 9.2 | 9.4 | 9.3 | 8.9 | 8.3 | 9.0 | 19 |
| 16 | 9.7 | 9.2 | 8.6 | 8.2 | 7.4 | 6.9 | 7.0 | >11 |

Patient 4 withdrew from the study and is not included in the analysis. Reprinted by permission of the Society of Nuclear Medicine from: Nahmias C, Hanna WT, Wahl LM, et al. Time Course of Early Response to Chemotherapy in Non-Small Cell Lung Cancer Patients with ¹⁸F-FDG PET/CT. *J Nucl Med.* 2007; 48(5): 744-751.

week the difference compared to baseline approximates the target measure to a sufficient degree. We refer to such a study as a 'schedule optimizing' (SO) study.

For the second question we need in addition to series of SUV measurements data on some clinical outcome, e.g., survival, and then we can investigate the prognostic value of different definitions of reduction by standard procedures for the evaluation of prognostic factors [20]. We refer to such a study as a 'prognosis evaluation' (PE) study.

Material

To illustrate our general considerations, we used a published data set [19] providing data on 15 patients with inoperable non-small cell lung cancer (NSCLC). SUV measurements were made prior to start of the chemotherapy and weekly for the next 6 weeks. In addition, data on the survival of the patients were given. The data used in this paper is shown in **Table 1**.

Methods

In SO studies as well as PE studies the first step is to investigate the observed patterns in the reduction processes in the single patients. This should always start with a visual inspection of

the patterns, which may give rise to a parametric model covering the observed patterns. Allowing the parameters, for example the slope, to vary across patients allows disentangling random fluctuations of the data from true heterogeneity across patients using random effect models.

The basic step in an SO study is the comparison of different schedules with respect to the ability to approximate a target measure of the reduction. This target measure should result from the investigation of the individual patterns and has to be defined for each patient using all data. Correspondingly, a surrogate measure has to be defined for each schedule using only the data available according to the schedule. A first approach to measure the ability of a given schedule to approximate the target measure would be to consider the Pearson correlation coefficient r between the surrogate measure and the target measure. However, as we are also interested in knowing the absolute value of the accuracy of the approximation, it is more appropriate to estimate the magnitude of the error term in regressing the target measure on the surrogate measure, and, hence, to report the estimated standard deviation s of the error term. A prediction of the target measure from the surrogate measure has roughly a 95% prediction interval of $\pm 2s$, and, consequently, the value of s allows

us to judge in a simple manner whether the approximation is sufficient.

The basic step in a PE study is to investigate and compare the prognostic value of different definitions of the degree of reduction. We will base such a comparison on a visualization of the association of the degree of reduction with the survival using scatter plots and by computing hazard ratios in a Cox model after standardizing the variables measuring the degree of reduction to mean 0 and standard deviation 1, such that hazard ratios are directly comparable.

We will also illustrate the inadequateness of a direct comparison of different schedules with the clinical outcome.

Results

Analysis of individual reduction processes

A visual inspection of the individual series of SUV measurements (**Figure 1**) suggests a roughly linear pattern in all patients, disturbed only by some measurement error. Indeed, a quadratic model improves the fit of a linear

model significantly only for one patient (no. 13, $p=0.033$). Hence, the assumption of a linear development in each patient is a reasonable working hypothesis. Consequently, we will use in the following the slope β of a fitted regression line (cf. **Figure 1**) as the essential measure of the true degree of reduction in each patient. This 6-week-slope can be interpreted as the change in SUV per week, and negative values indicate a reduction and positive values a continuing increase in viable tumor tissue.

In **Figure 1** we can observe some patients with a slight weekly increase of the SUV values, many patients with a nearly constant level and some with a very distinct reduction in glucose uptake. This heterogeneity was confirmed by a linear random effects model, indicating highly significant differences of the 6-week-slope across the patients ($p<0.001$).

Choice of optimal schedule

If we take the individual 6-week-slopes as the target measure for the reduction in viable tumor tissue, we can now study the precision of different schedules of PET/CT measurements to ap-

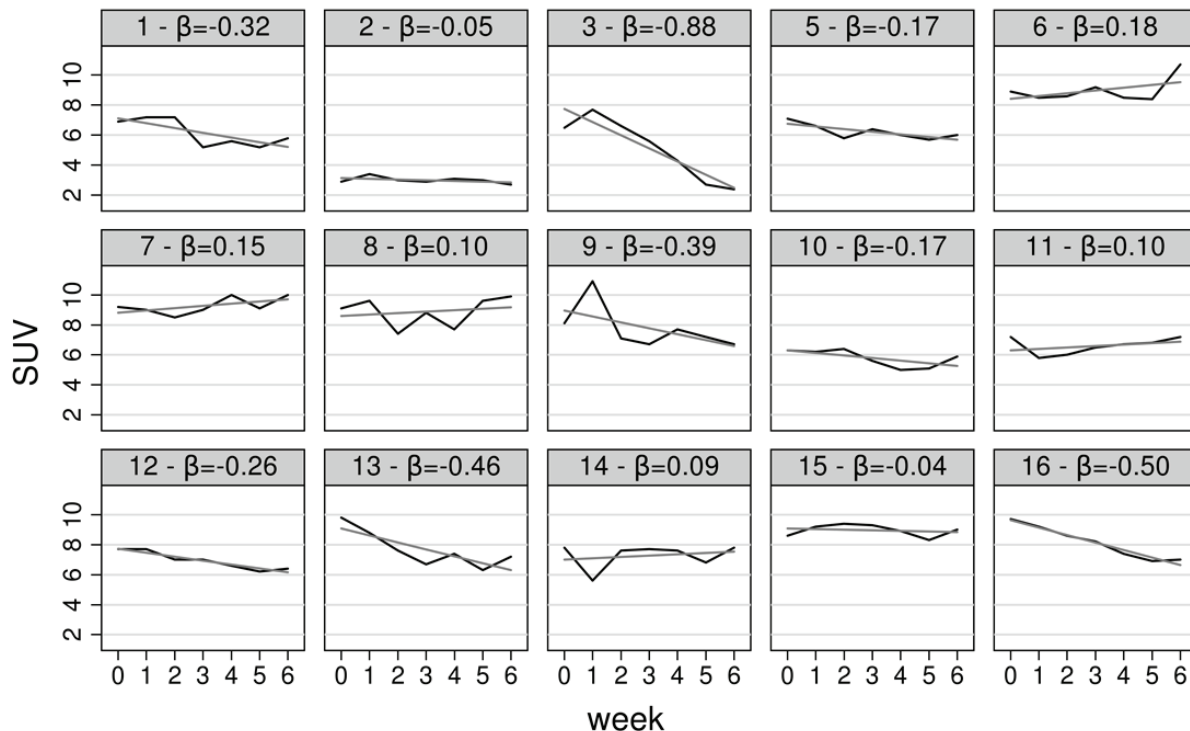


Figure 1. The individual series of SUV measurements for all 15 patients (black lines). The fitted regression curves are shown as grey lines. β denotes the estimated slope of the regression line.

Optimal timing of PET/CT for monitoring of cancer treatment

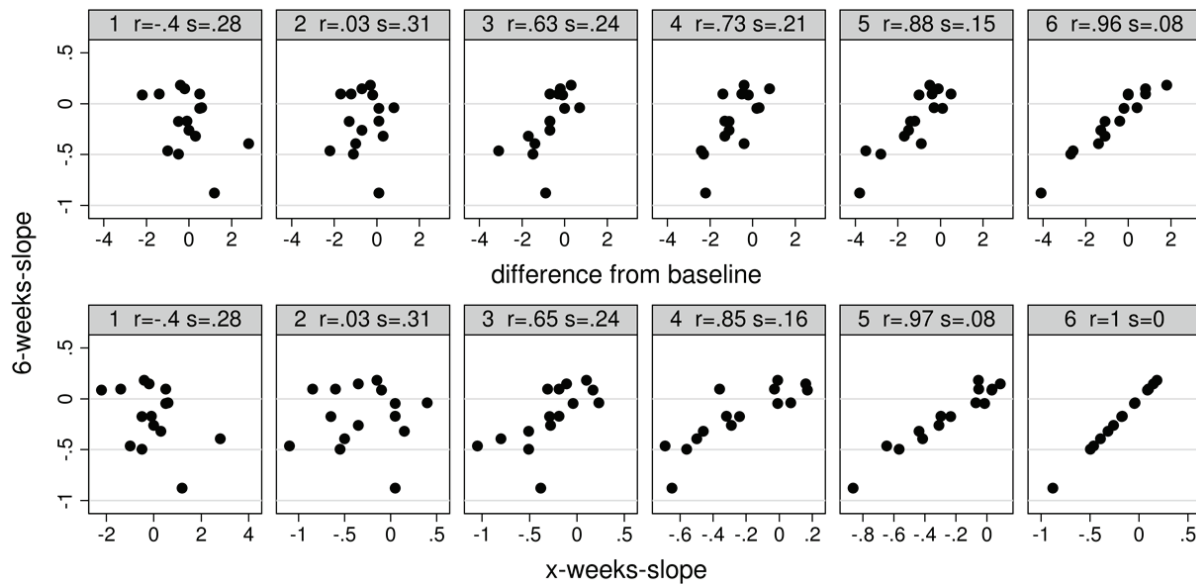


Figure 2. Comparison of different schedules with respect to approximating the 6-week-slope. Upper panel: 6-week-slope vs. difference from baseline for each week from week 1 to week 6. Lower-panel: 6-week-slope vs. x-week-slope for x from 1 to 6. The x-week-slope is the slope based on the weekly measurements up to week x. r denotes the Pearson correlation coefficient, s denotes the standard deviation of the error term.

proximate this slope. The simplest schedule would be one SUV measurement at a certain week in addition to the baseline measurement and to use the difference compared to baseline as the surrogate measure. Using the published data at hand, we can now study the association of this difference at each week with the 6-week-slope. The upper panel of **Figure 2** demonstrates that for the first two weeks we can obtain only a poor association with correlation coefficients below 0.1. Starting with week 3, the associations become acceptable, and especially after 5 weeks we have an error standard deviation of $s=0.15$, suggesting that we can predict the 6-week-slope with a 95% prediction interval of ± 0.3 . Taking into account that the patients with a probably clinically relevant reduction tend to have 6-week-slopes around -0.3 or less (cf. the next section), a prediction interval of such a size seems to be sufficiently small.

However, we can also study other schedules. In the lower panel of **Figure 2** we study the schedule of weekly measurements up to a certain week, using the slope based on the measurements up to this week as an approximation to the 6-week-slope. In comparison with the first schedule, we can observe no improvement for the first three weeks, but it looks like that we

can now obtain a precision in the approximation we have observed with the first schedule after 5 weeks already after 4 weeks.

Prognostic value of a reduction

On the left-hand side of **Figure 3** we study the relation between the reduction in viable tumor tissue expressed by the 6-week-slope and survival. We could observe that among the three patients with longest survival, we found the patient with highest reduction and the patient with the fifth highest reduction, but also one patient with nearly no reduction (patient no. 2). This patient had, however, the smallest baseline SUV value among all patients and, probably due to this reason, a good prognosis. Among the patients with a limited survival up to 40 weeks we can observe a surprisingly clear, linear relationship between the 6-week-slope and the survival time. However, as all these patients died within a few months, this relation is of limited clinical value.

In **Figure 1** we could observe a rather substantial variation of the baseline SUV values. This suggests that it might be possible to study a difference in prognostic value between the absolute change and the relative change in SUV

Optimal timing of PET/CT for monitoring of cancer treatment

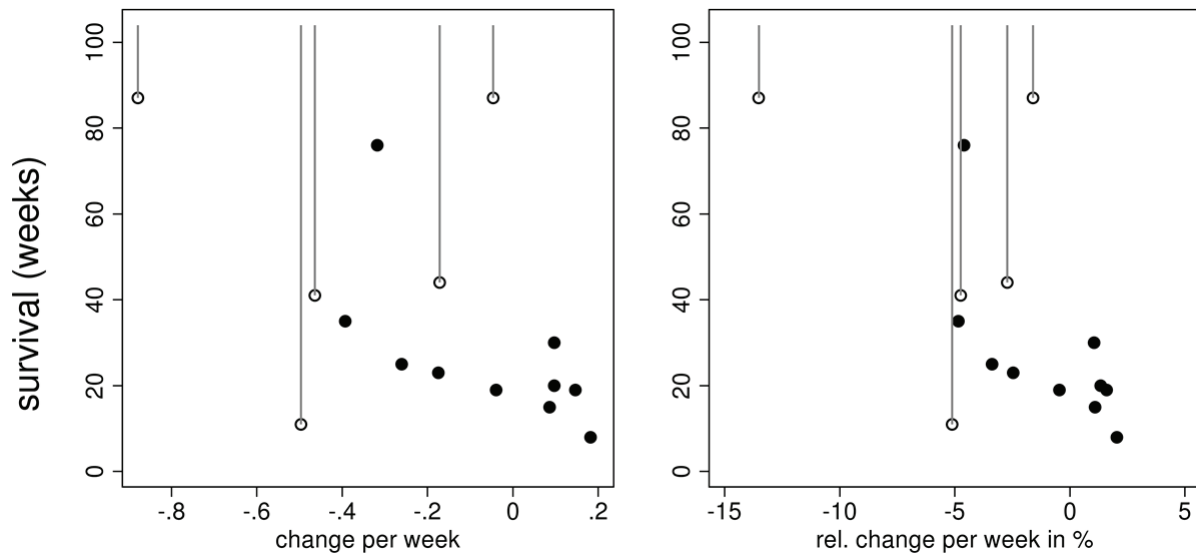


Figure 3. The relation between the reduction of SUV and survival. Censored survival times are shown as a circle. An additional line illustrates the range of potential true survival times. Censored survival times occurred in patients who were still alive at that time point, but could not be followed up. Left side: Absolute reduction, expressed as change in SUV per week. Right side: Relative reduction, i.e., change per week expressed as percentage of the baseline value.

values. However, correlating the relative change with the survival (right-hand side of **Figure 3**) we observe exactly the same pattern as with the absolute change, and also nearly identical hazard ratios in a Cox model (4.42 vs. 4.57). This is due to the fact that in spite of the substantial variation of the baseline values the absolute and relative changes are highly correlated with a Pearson correlation coefficient of 0.98.

An inadequate strategy to search for an optimal schedule

If one decides to investigate the optimal timing of a single SUV measurement by directly considering the association of the difference in SUV to baseline with the clinical outcome survival, we observe the series of scatter plots shown in **Figure 4**. We can again observe no association for the first two weeks, and some association later, but it is hard to make any distinction among the choices of 3, 4, 5, or 6 weeks, as we observe a nearly constant degree of association. So, here it is impossible to come to a clear decision about the optimal schedule.

Discussion

In this paper we introduced the concept of SO studies. Such studies aim at determining an

optimal schedule for using PET/CT as a monitoring tool. Optimality refers to judging a reduction in SUV as early as possible with sufficient precision. Such studies should be clearly distinguished from PE studies, aiming to investigate the prognostic value of a reduction in SUV with respect to some clinical outcome measure and, hence, to establish the value of PET/CT as a tool to monitor treatment response.

We suggested some simple analysis strategies for both study types and illustrated them using a published data set. As this data set comprises only 15 patients, we were forced to examine only rather simple schedules with measurements at fixed time points in our illustration of a SO study. There can be no doubt that in the long run more flexible schedules will be of clinical importance allowing a sequential decision process. For example, a first follow-up SUV measurement may be scheduled at week x . If there are no signs for a reduction, the treatment is changed immediately. If there is a distinct reduction, the monitoring is stopped. For the remaining patients, a second SUV measurement is scheduled at week y and similar criteria are applied. Determining such a schedule in an SO study requires a larger number of patients to be included, as we have to determine optimal values for several cut points involved. We expect

Optimal timing of PET/CT for monitoring of cancer treatment

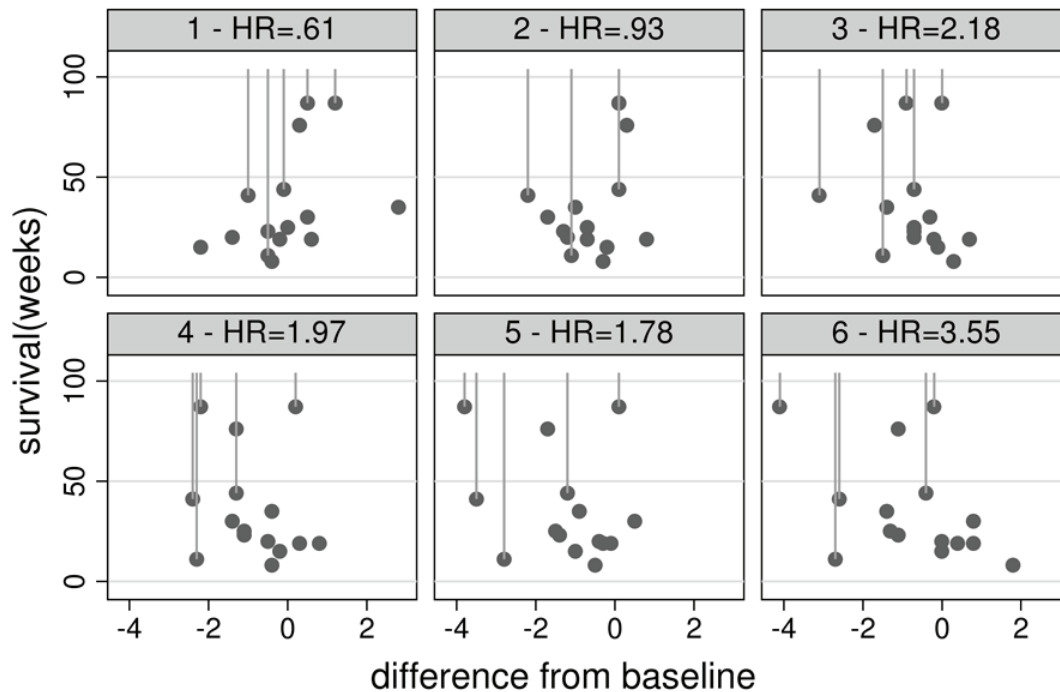


Figure 4. Comparison of different schedules with respect to the association with patient survival. The survival time is plotted versus the difference in SUV from baseline for each week from week 1 to week 6. Censored survival times are shown as a circle. An additional line illustrates the range of potential true survival times. HR is the hazard ratio from a Cox regression after standardizing the covariate 'difference from baseline' to mean 0 and standard deviation 1.

that with 50 patients a reliable estimation of such a procedure will be possible.

Fifteen patients are also insufficient to expect reliable results from a PE study. Especially if we want to differentiate between different definitions of the degree of reduction, we need larger sample sizes. With a sufficient number of patients measures for the prognostic accuracy like sensitivity and specificity in predicting one-year survival can be estimated in a reliable manner, which was not possible in our small example.

Another obvious shortcoming of our example is the limitation of SUV measurements to a period of 6 weeks. This does not allow for investigating of whether and how often a reduction in viable tumor tissue may stop and change to an increase again. The existence of such patterns and its predictability from the initial degree of reduction is highly relevant for defining optimal schedules and judging the prognostic value of reductions. Therefore, we strongly recommend planning a short term follow-up with dense measurements and a long term follow-up with

larger intervals between measurements for any SO study.

SO and PE studies are interacting. If we start with an SO study with dense measurements of the response process, we can already use the optimal schedule to minimize the number of PET/CT scans in the PE study, and, hence, reduce cost. It may be argued that selecting already one target measure in the SO study and optimizing the schedule for this target may limit the possible search for optimal measures of early response in the PE study, but this has to be balanced with the fact that PE studies require typically more patients than SO studies and, therefore, it is more costly to have also dense measurements in a PE study. Compared with PE studies, SO studies have – besides a smaller sample size – the basic advantage that it is not necessary to wait for the clinical outcome data and this makes them attractive to start with. Of course, it may happen that in an SO study it is difficult to make a choice among several possible target measures, and then one has to extend the study to a PE study by collect-

ing also clinical outcome data to determine the most prognostic target measure. Note that even if one agrees in a SO study on one target measure to be used for optimization, this does not exclude the possibility to investigate in a subsequent PE study different variants like an absolute or relative reduction with respect to their prognostic value. Although these considerations suggest to start always with an SO study, there are also some reasons to start with a PE study, as PE studies are often fully feasible only in the early phase of research. If there is already enough evidence to justify treatment alterations in the case of insufficient reductions, and if patients benefit from such an alteration on average, then a PE study will underestimate the prognostic value of the response evaluation. In contrast, SO studies can be still performed even if we allow treatment alterations in dependence on the observed reductions. This is in any case true if we wait with such a treatment alteration until the last SUV measurement, but also earlier alterations can be adjusted for in the analysis, as long as they appear according to a fixed, pre-specified rule. Whenever one decides to start immediately with a PE study, it will be a good idea to use the first patients enrolled to perform a SO study as a sub-study, which may allow applying the optimal schedule already in the remaining patients.

In any case, whenever both dense SUV measurements and clinical outcome data are available in a set of patients, one should never follow the temptation to use directly the outcome data in deciding on the optimal schedule, as this approach makes no use of the information provided by the complete series of SUV measurements and, hence, it is not very powerful, as also illustrated by our empirical results.

As the patterns of reduction in the viable tumor tissue over time depend both on tumor characteristics and type of treatment, it will be necessary to perform SO studies separately for each clinically relevant group of patients. The use of standardized protocols for SUV measurements should reduce possible center effects, but it can be wise to perform SO studies with two or three centers in order to demonstrate cross-center comparability. SO studies will also allow evaluating and optimizing the use of semi-automated systems for judging changes in SUV.

Our considerations point also to a general short-

coming in the current culture of studying the value of PET/CT for monitoring treatment response. As most studies in this area are based on one rather arbitrary schedule for PET/CT-based evaluations and a comparison with some clinical outcome measure, we do not know whether the often somewhat disappointing results are due to an inefficient assessment of the reduction in viable tumor tissue or the limited value of such a reduction as a prognostic marker. If SO studies and PE studies are performed separately in the future, as described in this paper, this shortcoming will disappear.

Finally, we should mention that our considerations may apply also to other attempts to assess response in cancer patients, as long as they are based on a quantitative measure [21].

Conclusion

The question of the optimal timing of PET/CT in monitoring of treatment response should be addressed in the future by specific 'schedule optimizing studies'. These may at times be conducted as part of PE studies provided the respective study types are analyzed independently. In any case, SO and PE studies should interact.

Address correspondence to: Dr. Werner Vach, University Medical Center Freiburg, Clinical Epidemiology, Institute of Medical Biometry and Medical Informatics, Stefan Meier Str. 26, D-79104 Freiburg, Tel: +49 761 203 6722, Fax: +49 761 203 6680, E-mail: wv@imbi.uni-freiburg.de

References

- [1] Juweid ME, Cheson BD. Positron-emission tomography and assessment of cancer therapy. *N Engl J Med* 2006; 354: 496–507.
- [2] Seam P, Juweid ME, Cheson BD. The role of FDG-PET scans in patients with lymphoma. *Blood* 2007; 110: 3507–3516.
- [3] Herrmann K, Krause BJ, Bundschuh RA, Dechow T, Schwaiger M. Monitoring response to therapeutic interventions in patients with cancer. *Semin Nucl Med* 2009; 39: 210–232.
- [4] Weber WA. Assessing tumor response to therapy. *J Nucl Med* 2009; 50, Supplement 1: 1S–10S.
- [5] Hillner BE, Siegel BA, Shields AF, Liu D, Gareen IF, Hanna L, Stine SH, Coleman RE. The impact of positron emission tomography (PET) on expected management during cancer treatment: findings of the National Oncologic PET Registry. *Cancer* 2009; 115: 410–418. Erra-

Optimal timing of PET/CT for monitoring of cancer treatment

- tum in: *Cancer* 2009; 115: 1133.
- [6] Høilund-Carlsen PF, Gerke O, Vilstrup MH, Nielsen AL, Thomassen A, Hess S, Høilund-Carlsen M, Vach W, Petersen H. PET/CT without capacity limitations: a Danish experience from a European perspective. *Eur Radiol* 2011; 21: 1277-1285.
- [7] Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving considerations for PET response criteria in solid tumors. *J Nucl Med* 2009 May; 50 Suppl 1: 122S-50
- [8] Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancey J, Arbuck S, Gwyther S, Mooney M, Rubinstein L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 2009; 45: 228-247.
- [9] Young H, Baum R, Cremerius U, Herholz K, Hoekstra O, Lammertsma AA, Pruim J, Price P. Measurement of clinical and subclinical tumour response using [¹⁸F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. *Eur J Cancer* 1999; 35: 1773-1782.
- [10] Hoekstra CJ, Stroobants SG, Smit EF, Vansteenkiste J, van Tinteren H, Postmus PE, Golding RP, Biesma B, Schramel FJHM, van Zandwijk N, Lammertsma AA, Hoekstra OS. Prognostic relevance of response evaluation using [¹⁸F]-2-fluoro-2-deoxy-D-glucose positron emission tomography in patients with locally advanced non-small-cell lung cancer. *J Clin Oncol* 2005; 23: 8362-8370.
- [11] Allen-Auerbach M, Weber WA. Measuring response with FDG-PET: Methodological aspects. *Oncologist* 2009; 14: 369-377.
- [12] Hicks RJ. Role of ¹⁸F-FDG PET in assessment of response in non small cell lung cancer. *J Nucl Med* 2009; 50: 31S-42S.
- [13] Minn H, Zasadny KR, Quint LE, Wahl RL. Lung-cancer - reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology* 1995; 196: 167-173.
- [14] Weber WA, Ziegler SI, Thodtman R, Hanauske AR, Schwaiger M. Reproducibility of metabolic measurements in malignant tumors using FDG PET. *J Nucl Med* 1999; 40: 1771-1777.
- [15] Nakamoto Y, Zasadny K, Minn H, Wahl RL. Reproducibility of common semi-quantitative parameters for evaluating lung cancer glucose metabolism with positron emission tomography using 2-deoxy-2-[¹⁸F]-fluoro-D-glucose. *Mol Imaging Biol* 2002; 4: 171-178.
- [16] Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer* 1976; 38: 388-394.
- [17] Findlay M, Young H, Cunningham D, Iveson A, Cronin B, Hickish T, Pratt B, Husband J, Flower M, Ott R. Noninvasive monitoring of tumor metabolism using fluorodeoxyglucose and positron emission tomography in colorectal cancer liver metastases: Correlation with tumor response to fluorouracil. *J Clin Oncol* 1996; 14: 700-708.
- [18] Römer W, Hanauske AR, Ziegler S, Thodtman R, Weber W, Fuchs C, Enne W, Herz M, Nerl C, Garbrecht M, Schwaiger M. Positron emission tomography in nonhodgkin's lymphoma: Assessment of chemotherapy with fluorodeoxyglucose. *Blood* 1998; 91: 4464-4471.
- [19] Nahmias C, Hanna WT, Wahl LM, Long MJ, Hubner KF, Townsend DW. Time course of early response to chemotherapy in non-small cell lung cancer patients with ¹⁸F-FDG PET/CT. *J Nucl Med* 2007; 48: 744-751.
- [20] Steyerberg EW: *Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating*. New York, Springer, 2008.
- [21] Verweij J, Therasse P, Eisenhauer E. Cancer clinical trial outcomes: Any progress in tumor-size assessment? *Eur J of Cancer* 2009; 45: 225-227.