^{18}F-FDG PET as a Surrogate Biomarker in Non–Small Cell Lung Cancer Treated with Erlotinib: Newly Identified Lesions Are More Informative Than Standardized Uptake Value

Thomas Bengtsson[1], Rodney J. Hicks[2], Amy Peterson[3], and Ruediger E. Port[4]

[1]Biostatistics, Genentech Inc., South San Francisco, California; [2]Centre for Cancer Imaging, Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia; [3]Early Clinical Development, Genentech Inc., South San Francisco, California; and [4]Early Development PKPD, Genentech Inc., South San Francisco, California

This study assesses the predictive value of ^{18}F-FDG PET for overall survival in lung cancer patients treated with a targeted drug. **Methods:** ^{18}F-FDG PET was performed in 125 second- or third-line non–small cell lung cancer (NSCLC) patients with a baseline Eastern Cooperative Oncology Group performance status less than 3 before treatment with erlotinib (150 mg daily) and 2 wk into treatment. The predictive value of ^{18}F-FDG PET, clinical parameters, and epithelial growth factor receptor (EGFR) mutation status for survival duration was evaluated by fitting accelerated failure time models. **Results:** New lesions on PET at 2 wk, EGFR mutation status, performance status, and baseline tumor burden were independent and significant predictors of overall survival. Reduction of maximum standardized uptake value by at least 35% was predictive of survival only when EGFR mutation status was not accounted for. **Conclusion:** ^{18}F-FDG PET in second- or third-line NSCLC patients at 2 wk after starting treatment with erlotinib carries information about overall survival. Parametric survival modeling enables a quantitative assessment of the predictive value of ^{18}F-FDG PET in the context of clinical and laboratory information. New-lesion status by ^{18}F-FDG PET at 2 wk is a potential surrogate biomarker for survival in NSCLC.

Key Words: ^{18}F-FDG PET; CT; NSCLC; overall survival; survival modeling; positron emission tomography

PET using ^{18}F-FDG, especially combined with CT (1), is an established tool for the diagnosis and staging of a variety of cancer types (1). There is also increasing evidence that ^{18}F-FDG PET can assess therapeutic response earlier than CT (2), enabling modification of ineffective therapy and potentially improving therapeutic outcomes. This utility hinges on a reliable and validated link between early ^{18}F-FDG PET response and improved clinical outcome. Although such

evidence is growing for several tumors, including lung cancer (3), and with treatments including radiotherapy and chemotherapy (4,5), it may not hold uniformly across indications and therapies. Here, using data from 2 global multicenter, phase II studies of second- or third-line non–small cell lung cancer (NSCLC) patients treated with a targeted agent, we investigate the extent to which survival duration can be predicted with ^{18}F-FDG PET information obtained at baseline and early in treatment.

If a predictive link can be demonstrated, ^{18}F-FDG PET may serve as a noninvasive biomarker for survival in NSCLC. In this context, ^{18}F-FDG PET responses may serve as a criterion for moving novel anticancer therapies forward to more costly phases of development or to stop the development of ineffective compounds. This may be particularly important for molecularly targeted therapies, some of which may have cytostatic effects and therefore not lead to conventional radiologic response. Effects of targeted therapies often depend on expression levels and mutations of receptor and signaling proteins (6). Thus, although changes on ^{18}F-FDG PET can occur within hours or days, well before significant cell loss, these may merely indicate effects on membrane transport or metabolism of glucose rather than cell kill (7), and it is not clear whether such responses are also predictive of survival.

While accounting for patient heterogeneity due to clinical and mutation status, we used multivariate survival analysis based on the accelerated failure time (AFT) models (8,9) to relate changes observed on ^{18}F-FDG PET with survival duration. In similar analyses of NSCLC (10,11) and colorectal cancer patients (12), AFT models were recently applied to assess the predictive value of CT tumor size measurements for survival. For both tumor types, these studies found tumor size before treatment, change of tumor size at 7 (12) or 8 wk into treatment (11), and the Eastern Cooperative Oncology Group (ECOG) performance status before treatment to be significant predictors of survival.

Because our analyses were performed on single-arm data, it was not possible to discern between predictive and prognostic effects. Therefore, throughout this article, our use of the term *predictive* refers primarily to a variable's

statistical ability to inform about survival duration, independent of possible treatment effects. Moreover, therapeutic effects due to drug treatment are not discernable from our work.

The overall goal of our analyses was to determine whether ^{18}F-FDG PET changes after 2 wk of treatment with a targeted drug can predict survival in NSCLC patients.

MATERIALS AND METHODS

Patients and Treatment

One hundred thirty-six patients with refractory or recurrent NSCLC after second- or third-line treatment were studied in the following 2 international, multicenter trials: a single-arm study of erlotinib (OSI3926g (*13,14*)) and in the erlotinib arm of a 2-arm study comparing an antibody to the MET receptor plus erlotinib with erlotinib alone (OAM4558g (*15*)). All patients received erlotinib at a dose of 150 mg daily (orally) until occurrence of disease progression or severe side effects. Epithelial growth factor receptor (EGFR) mutation status was available for 100 patients. Mileshkin et al. (*14*) and Spigel et al. (*15*) provide more complete descriptions of the design of the 2 studies.

For exploratory purposes, OSI3926g and OAM4558g prospectively collected ^{18}F-FDG PET data at baseline and early in treatment. Eleven patients either dropped out for clinical reasons (e.g., due to withdrawing consent, progressive disease, adverse event, or death) or did not have complete imaging data, leaving 125 patients (Table 1) who received treatment for at least 2 wk and who had ^{18}F-FDG–avid PET scans at baseline.

Survival duration was defined from commencement of erlotinib treatment. Median survival in the OSI3926g and OAM4558g studies was 7.2 and 7.8 mo, respectively, whereas the time since initial NSCLC diagnosis was 10.3 and 11.8 mo. At the time of each study's closure, 34 cases were censored. Marginal distributions of several demographic (age, sex, smoking status) and clinical variables (performance status, EGFR mutation status, histology, survival duration) between the 2 studies were all strikingly similar in both studies, justifying their combination. Except for viable tissue sample availability, which was higher in OAM4558g because of a mandatory tissue sampling requirement, there were no statistically significant differences in the baseline variables described in Table 1 at the 0.05 level of significance between the 2 studies.

Imaging Acquisition and Quantitation

All 125 patients underwent PET/CT using low-dose, unenhanced CT. As multicenter trials, a range of scanners from different vendors was used but serial scanning was performed on the same scanner, prospectively qualified by an imaging core laboratory, and was acquired according to an imaging charter designed to comply with guidelines from the National Cancer Institute on the use of PET for response evaluation (*16*). Baseline scans were to be obtained within 14 d before treatment initiation, and early-response scans were targeted for day 14 after the start of treatment, with a window for days 11–17 (for OSI3926g) and days 10–14 (OAM4558g). The observed mean for the number of days elapsed between baseline and follow-up scans was 22.3 (range, 12–35 d) and 20.1 (range, 13–42 d), respectively, for OSI3926g and OAM4558g. Overall, 7 patients violated the imaging charter requirement with respect to scan day specifications.

Both trials specified a fasting time of at least 4 h. Audited imaging compliance parameters included ^{18}F-FDG uptake time, administered activity, scanning direction and arm position, and pre–^{18}F-FDG blood glucose levels. A high level of compliance with this charter within OSI3926g has previously been documented (*13*). The compliance with the imaging charter for OAM4558g was comparable to that of OSI3926g (*17*). Accordingly, it is reasonable to assume that the consistency of imaging methodology was good in

TABLE 1
Baseline Characteristics of Patients in OSI3926g and OAM4558g

Characteristic	OSI3926g ($n = 65$)	OAM4558g ($n = 60$)
Age (y)		
Mean	62.7	62.1
SD	9.2	10.6
Women	43.1%	39.7%
Race or ethnic group		
Asian	4.6%	0%
Other	95.4%	100%
Smoking status		
Never	20.0%	19.0%
Previous	12.3%	12.7%
Current	67.7%	68.3%
Baseline ECOG status		
0	24.6%	30.2%
1	61.5%	66.7%
2	13.8%	3.2%
EGFR mutation	6.2%	9.5%
Viable tissue sample	69.2%	88.9%
Histologic type		
Adenocarcinoma	69.2%	63.2%
Squamous cell carcinoma	18.5%	25.5%
Median time since initial diagnosis (mo)	10.3	11.8
Median survival (mo)	7.2	7.8

Except for available tumor tissue, no baseline variable is statistically significantly different between the 2 studies at 0.05 level.

individual patients within and between both trials. Two separate core laboratories, masked to clinical details, performed analysis of the PET results. For the OSI3926g trial, target lesions were selected first for significant ^{18}F-FDG uptake (compared with adjacent background) and second for suitability for response evaluation on the unenhanced CT. The target lesion selection procedure in OAM4558g was the same but supported by diagnostic CT if available (otherwise by coacquired CT). Lesions with extensive necrosis at baseline, as indicated by central photopenia, were generally not selected. Newly detected lesions were defined as regions of interest that were not sufficiently ^{18}F-FDG–avid at baseline to be definite sites of disease but subsequently determined to be above the diagnostic threshold at the week-2 scan. Radiologic responses were evaluated by different core laboratories on the basis of diagnostic CT with intravenous contrast at day 56 of treatment.

The average number of target lesions per patient in the 2 trials was 2.8 (OSI3926g) and 2.7 (OAM4558g), with 1–5 lesions in both studies. More than one target lesion was present in 72.3% (OSI3926g) and 73.3% (OAM4558g) of patients. The mean of baseline average maximum standardized uptake value (SUV_{max}) across both trials was not significantly different ($P = 0.195$) at 6.2 (SD, 2.5) for OSI3926g and 7.6 (SD, 3.5) for OAM4558g. The ranges were 1.7–68.2 and 1.4–76.5, respectively. Notably, EGFR mutation status did not influence these baseline statistics. As measured by coacquired baseline CT, the mean sum of longest dimensions among target lesions on OSI3926g was 8.0 cm (SD, 4.3), with a range of 1.2–19.2 cm. The corresponding figures for OAM4558g were 7.8 cm (SD, 4.7), with a range of 1.1–18.6 cm.

In the absence of guidance on the optimal methodology to assess therapeutic response, we adopted a pragmatic approach of using the arithmetic mean of the percentage change in individual lesions identified prospectively on the baseline study. This approach is as near as possible to standard radiologic assessment wherein target lesions are selected and followed prospectively to assess response but identification of new lesions is deemed to indicate progressive disease. PET Response Criteria in Solid Tumors (PERCIST) (2) had not been published when this trial commenced.

Multivariate Survival Analysis

The Kaplan–Meier (18) procedure was used to provide nonparametric estimates of median survival and to produce descriptive plots of the survival function. To quantify and test whether changes in the tumor glucose uptake were related to length of survival, we used the AFT model. The AFT model is a commonly used tool in survival analysis (8,9) and relates the log of survival duration to a set of explanatory variables through a linear parametric form.

The AFT model is specified as follows, with T representing the length of survival in months and $X_1, \ldots, X_p$ denoting observations on p explanatory variables:

$$\log T = \alpha_0 + \alpha_1 X_1 + \alpha_1 X_2 + \ldots + \alpha_p X_p + \epsilon,$$

where $\alpha_1, \ldots, \alpha_p$ are parameters linking the predictor variables to log T, and α_0 is the mean log-survival when all predictors are identically zero. Further, in the equation, ϵ represents a noise term that describes sources of variability in log T that remains unaccounted for by the predictor variables. Parameter estimation requires a further assumption on the distribution of the noise term (e.g., normal, Weibull), and similarly to Wang et al. (11), we assume that ϵ is zero-mean gaussian distributed with variance σ^2. The AFT model easily handles censored data.

On the basis of the information obtained by ^{18}F-FDG PET, we evaluated the following AFT model:

$$\log T = \alpha_0 + \alpha_1 \times PS + \alpha_2 \times SUV_{base} + \alpha_3 \times \delta SUV_{wk2} + \alpha_4 \times NL_{wk2}^{PET} + \alpha_5 \times EGFR + \epsilon,$$

where PS is ECOG performance status at baseline (0, 1, or 2), SUV_{base} represents tumor burden as measured by ^{18}F-FDG PET at baseline (see the "Results" section), δSUV_{wk2} is the average percentage change across lesions of SUV_{max} at 2 wk, NL_{wk2}^{PET} indicates the appearance of new lesions at 2 wk (day 14), and EGFR represents EGFR mutation status.

We initially restricted our analyses to evaluating the predictive value of average percentage change in SUV_{max} expressed on a linear scale. We later extended the definition of δSUV_{wk2} to be based on that of the best-performing lesion (in percentage change from baseline), the worst-performing lesion, and the criteria given by PERCIST (2). The model can also handle thresholds; for example, if δSUV_{wk2} is less than -25%, as specified by the PET partial-response guidelines of the European Organisation for Research and Treatment of Cancer, one sets the explanatory variable associated with δSUV_{wk2} equal to 1 (and zero otherwise). However, the actual threshold to determine a response may vary across indications and therapies, and we defer to the "Discussion" section consideration of this type of model that includes estimation of the threshold.

The model was fitted in a forward fashion by first evaluating the prognostic variables PS and SUV_{base}, followed by an evaluation of the predictor variables δSUV_{wk2} and NL_{wk2}^{PET}. Although a known prognostic variable for patients treated with erlotinib, EGFR mutation status was available only in a subset of 100 (of 125) patients, and this variable was therefore entered last. For statistical inference, we relied on P values, and to describe model goodness of fit, we used the percentage of variability in observed survival duration as explained by the model, that is, R^2.

RESULTS

The Kaplan–Meier procedure gave an estimated median survival of 7.6 mo, with a 95% confidence interval from 6.9 to 9.2 mo (Fig. 1A). A lognormal AFT model without explanatory covariates was also fitted to the data, and the resulting parametric estimate of the survival function is plotted.

Baseline PS is clearly associated with survival duration ($P < 4.5 \times 10^{-4}$), with an R^2 value of 9.2% (Fig. 1B, which depicts Kaplan–Meier and AFT estimates of the survival functions for patients with baseline PS values of 0 [$n = 34$], 1 [$n = 80$], and 2 [$n = 11$]). To model the effect of baseline tumor burden (i.e., SUV_{base}), we evaluated both the sum of SUV_{max} across lesions and the simple metric given by the total number of target lesions. Because these variables give statistically indistinguishable (and significant) results, we let SUV_{base} be defined by the total number of target lesions. Both baseline ECOG performance status ($P < 8.8 \times 10^{-4}$) and the number of target lesions ($P < 1 \times 10^{-3}$) were statistically significant, with a resultant model R^2 of 16.4%. We next evaluated the predictive value of observing changes in ^{18}F-FDG PET.

A scatterplot of the week-2 percentage change from baseline in SUV_{max} versus survival duration (in log scale; Fig. 2A) shows that large relative decreases in SUV_{max}
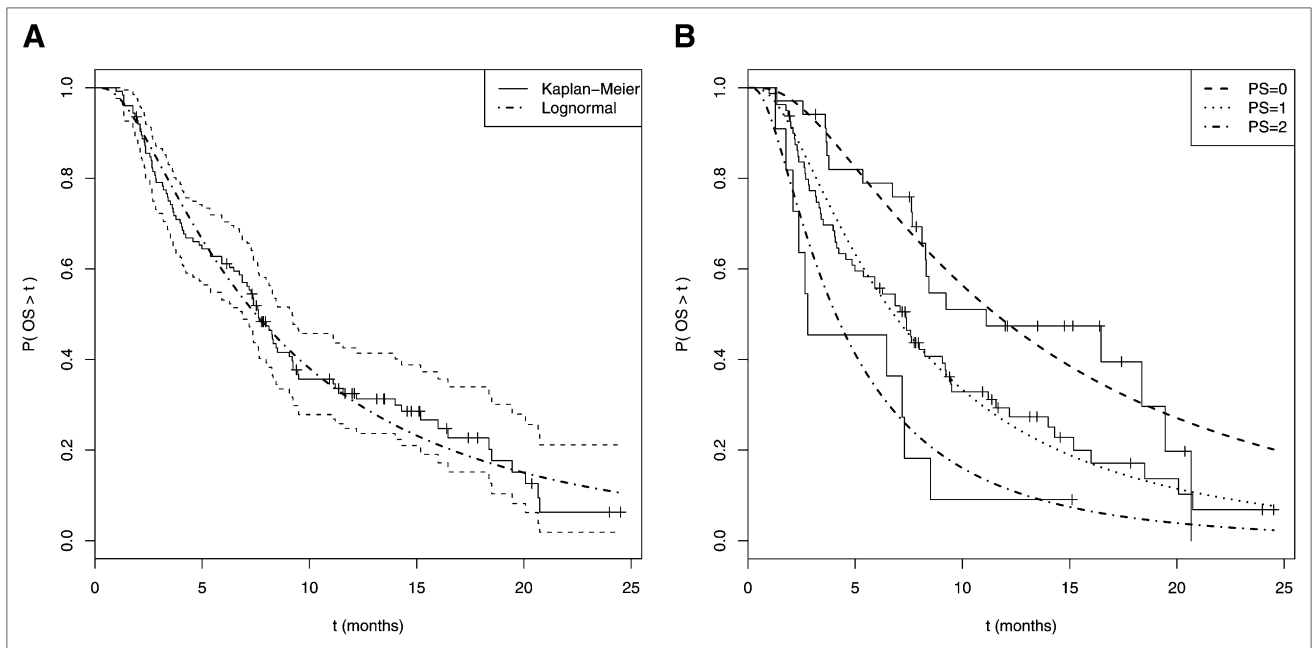
FIGURE 1. (A) Estimates of survival function for patients who were ^{18}F-FDG PET–evaluable at day 14. Solid line represents Kaplan–Meier estimate, with thin dashed lines denoting 95% pointwise confidence limits. Smooth, dash-dotted line is estimated survival function based on AFT-model. (B) Estimates of survival function by *PS* values (0, 1, and 2). Solid lines give Kaplan–Meier estimates, and smooth lines depict corresponding estimates based on AFT model.

appeared to be associated with longer survival times. Although the data are noisy, when evaluated as a linear relationship (i.e., as a correlation), the association between SUV_{wk2} and survival duration is statistically significant ($P < 0.028$). The week-2 percentage change in SUV_{max} also remained significant when added to the prognostic model ($P < 0.017$), but the unique contribution of SUV_{wk2} to R^2 was a low 3.3%. Thus, for these data, only substantial decreases in SUV_{max} were
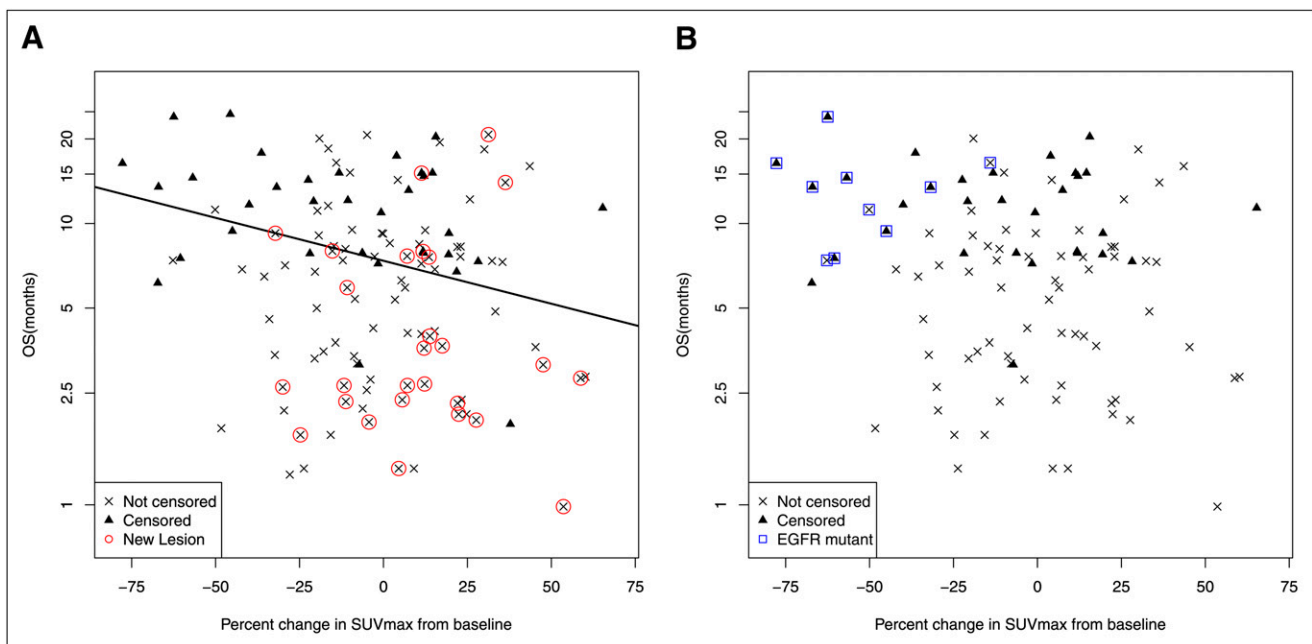

FIGURE 2. Percentage change in SUV_{max} between baseline and week 2 vs. survival duration (mo). Censored observations are denoted by triangles, and Xs denote actual event times. (A) Best-fitting linear AFT model. Patients with new lesions are overlaid by red circle. (B) Data subset with known mutation status. Patients with EGFR mutations are overlaid by blue box.

reliably associated with favorable survival, yet such changes occur mostly in EGFR mutant tumors.

New lesions (by ^{18}F-FDG PET) were detected in 27 (of 125) patients on the day-14 scan. In contrast to changes in SUV_{max}, new-lesion status by ^{18}F-FDG PET was a highly significant predictor of survival duration ($P < 2.2 \times 10^{-5}$). Most patients with shorter survival have new lesions (Fig. 2A), whereas most patients with longer survival do not exhibit new lesions. The parameter estimate associated with NL_{wk2}^{PET} is -0.75 (Table 2) and indicates a significant survival deficit for patients with new lesions by ^{18}F-FDG PET. We see also that there were no substantial decreases in SUV_{max} among patients with new lesions. Therefore, when NL_{wk2}^{PET} was added as a predictor, SUV_{wk2} was no longer significant ($P = 0.20$). The unique contribution of NL_{wk2}^{PET} to R^2 (in addition to PS and SUV_{base}) was 10.6%. We next accounted for EGFR mutation status in the analysis.

Figure 2B is based only on the subset of patients (100/125) with known mutation status and demonstrates the effect of EGFR mutant tumors on survival duration. As can be seen, all (known) EGFR-mutant patients have a long survival duration (relative to the median), and most exhibit large reductions in SUV_{max}. To estimate the independent effect of mutation status on survival, we add the variable EGFR to the model that already includes PS, SUV_{base}, and NL_{wk2}^{PET}. For this model, all variables are statistically significant ($PS, P < 0.012; SUV_{base}, P < 0.047; NL_{wk2}^{PET}, P < 4.4 \times 10^{-4}$; and EGFR, $P < 4.9 \times 10^{-3}$), with parameter estimates as reported in Table 2. Although this model is not fitted on the same sample as the previous models, refitting the prognostic and predictive models on the subset with 100 patients does not substantially alter the estimates or their significance.

It is plausible that the average percentage change in SUV_{max} across lesions does not adequately capture clinical benefit and that perhaps a metric that focuses on the best-performing (or worst–performing) lesion would do better in terms of predicting overall survival (OS). To test this possibility, we defined SUV_{wk2} by the minimum percentage change across lesions (i.e., the lesion with the largest decrease). However, this definition failed to predict OS even when entered by

itself into the model ($P = 0.35$) or when added to the predictive model that already included PS, SUV_{base}, and NL_{wk2}^{PET} ($P = 0.19$) or to the model that also included EGFR status ($P = 0.81$). Interestingly, the worst-performing lesion performed adequately by itself ($P < 0.0024$), with an R^2 of 6.8%, but its predictive power was covered by the other variables in the predictive models ($P = 0.25$ and 0.95). We also evaluated a definition based on the PERCIST criteria in which SUV_{wk2} is based on the percentage change between the most intense lesion at baseline and at day 14, but this approach did not provide an improvement in predictive power either when included by itself in the model ($P < 0.035$) or when added to the predictive models ($P = 0.22$ and 0.84).

Returning to the final predictive model of Table 2, we note that NL remained highly statistically significant in the model that included PS, SUV_{base}, and EGFR mutation status. Thus, new-lesion status is a disease characteristic that is strongly related to survival but is not reflected by the baseline characteristics as measured by PS and SUV_{base} and is also not uniquely attributable to EGFR mutation status. In a limited patient sample such as this, this result would be difficult to demonstrate without the use of a multivariate modeling approach to control for confounding factors.

To illustrate the estimated model that includes performance, new-lesion, and EGFR mutation status, Figure 3 shows the estimated survival functions for EGFR wild-type patients with PS values of 0 (Fig. 3A) and 1 (Fig. 3B), with and without new lesions. Controlling for EGFR mutation status, the figures show that the appearance of a new lesion shifts the survival curve to the left and that this result holds independently of PS. Overlaid Kaplan–Meier curves validate the AFT model fits (Fig. 1). As seen in both plots, the median survival was approximately halved for patients with new lesions at day 14. Accordingly, when expressed using the proportional hazards model, lack of a newly detected lesion at day 14 was associated with a hazard ratio of approximately 0.45, or alternatively, greater than 2 when defined by appearance of new lesions. There are not enough data to perform this particular validation for a PS of 2 or across values of SUV_{base}.

TABLE 2
Maximum-Likelihood Parameter Estimates Along with R^2 Values for Fitted AFT Models

Model	$E(\log T)$	$\hat{\sigma}$	R^2
Population	2.02	0.947	Not applicable
Prognostic (PS, SUV)	$2.91 - 0.47\ PS^* - 0.19 - SUV^\dagger_{base}$	0.855	16.4
Predictive (δSUV)	$2.86 - 0.46\ PS^* - 0.19\ SUV^\dagger_{base} - 0.67\ \delta SUV^\ddagger_{wk2}$	0.840	19.7
Predictive (NL)	$3.02 - 0.45\ PS^* - 0.18\ SUV^\dagger_{base} - 0.75\ NL_{wk2}^{PET\P}$	0.792	27.0
Predictive (NL, EGFR)	$2.75 - 0.40\ PS^\ddagger - 0.12\ SUV^\ddagger_{base} - 0.71\ NL_{wk2}^{PET\dagger} + 0.95\ EGFR^\dagger$	0.800	30.6

$^*P < 0.0001$.
$^\dagger P < 0.001$.
$^\ddagger P < 0.01$.
$^\P P < 0.05$.
Prognostic model includes PS and SUV_{base}. Additionally, the 3 predictive models include, respectively, $\delta SUV_{wk2}, NL^{PET}$, and NL^{PET} and EGFR.
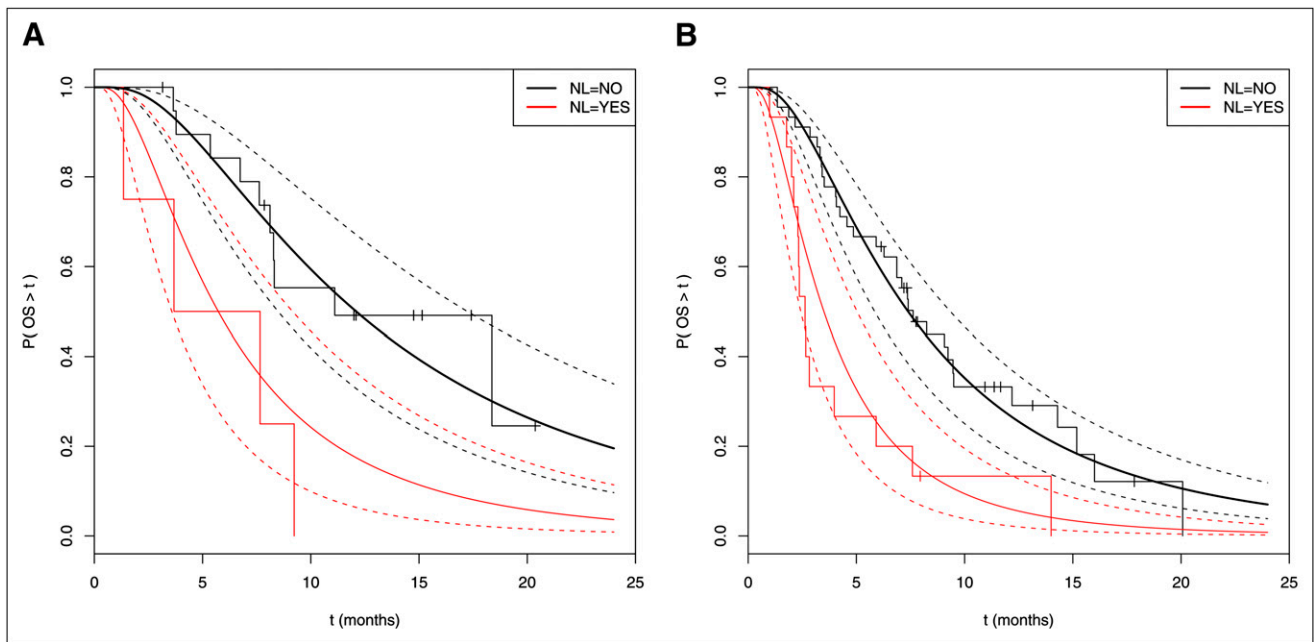
FIGURE 3. Estimated survival functions (smooth curves) for EGFR wild-type patients with $PS = 0$ (A) and $PS = 1$ (B), with and without new lesions (red and black curves, respectively). For lognormal AFT model, dotted lines show approximate 95% pointwise confidence intervals. Step functions give corresponding Kaplan–Meier estimates.

DISCUSSION

Our results indicate that the detection of new lesions by PET at early response assessment is a strong, independent predictor of OS in refractory NSCLC patients treated with an EGFR inhibitor. The results also indicate that observed reductions in SUV_{max} are informative about survival only when such changes are large, yet they fail to reach significance when EGFR mutation status is considered.

Although the development of new lesions within 2 wk of starting therapy may seem to be an unlikely event, recent work demonstrates that rapid progression can occur in NSCLC (*19*). Hence, it is plausible that new lesions on ^{18}F-FDG PET reflect aggressive biology, with new metastasis or growth of small lesions overcoming partial-volume effects within the steep component of the count-recovery curve leading to a significant increase in standardized uptake value (SUV) (*20*). By either mechanism, early disease progression would be expected to be associated with poor survival, consistent with our results.

New lesions were strikingly absent in patients with an SUV_{wk2} of less than −35% (Fig. 2), further suggesting that these lesions represent disease progression. This observation suggests that a comparison of new-lesion counts (by PET) between treatment groups may be a way of detecting drug effects in clinical trials. The value of this observation should be clarified by systematic inclusion of ^{18}F-FDG PET into placebo-controlled trials of molecularly targeted therapies. Moreover, our findings may not hold for patients treated with chemotherapy or in other cancers with more indolent behavior.

There is no clear-cut site predominance among the new lesions identified by ^{18}F-FDG PET. Moreover, there are no significant differences in age, sex, baseline ECOG performance status, smoking status, KRAS mutation, or histologic subtype for patients with and without newly detected lesions. The newly detected lesions had an overall mean SUV_{max} of 6.1 (range, 1.3–16) and were clearly detected above adjacent background tissue.

When modeled as a linear predictor (see the "Results" section), we found that changes in SUV_{max} were a poor predictor of survival. However, inspection of Figure 2 shows that a drop of SUV_{max} by more than 40% fairly consistently predicts survival beyond the overall median value. In contrast, there is no correlation between δSUV_{wk2} and survival when δSUV_{wk2} drops by less than 40% (or when it is positive). On the basis of these data patterns, the data were explored by fitting a threshold model in which an optimal value of −35% was estimated for δSUV_{wk2}, with a 95% confidence interval (based on the profile loglikelihood) from −50% to −30%. These results hold regardless of whether one accounts for the baseline prognostic variables and new-lesion status. Thus, in this setting, the relationship between changes in SUV_{max} and survival appears to be nonlinear, with a potentially larger response required for clinical significance than that needed after chemotherapy (e.g., −25%, as suggested by guidelines of the European Organisation for Research and Treatment of Cancer) (*2,21*).

For these reasons, a partial metabolic response by ^{18}F-FDG PET, defined by a change in average SUV_{max} below −25% and the absence of new lesions, is mainly informative about survival because of the strong effect due to *NL*. In fact, on the basis of a predefined cutoff of −25%, changes in SUV_{max} are not predictive of OS ($P = 0.093$). Our results differ from the primary report of one of the studies (*14*), likely because we have taken covariates into account.

The finding that EGFR mutation predicts improved survival (Table 2) is consistent with the recently completed phase III European Tarceva versus Chemotherapy study, which shows improved clinical response in this subpopulation (22). In our study, for the subset of patients in whom EGFR mutation status was available, this variable was able to replace a δSUV_{wk2} of -35% or less as a predictor, with a superior result. In accordance, Figure 2 suggests that a δSUV_{wk2} of -35% or less is strongly linked to EGFR mutation status.

One plausible reason why the average change in SUV_{max} is not a particularly sensitive predictor for OS may be the large degree of heterogeneity between lesions within patients. For example, among the 91 patients with at least 2 target lesions, the average range between the best- and worst-responding lesions (in terms of percentage change from baseline) is 38 percentage points. Moreover, 25% of these 91 patients have a range greater than 50 percentage points, and 5 patients have a range greater than 100 percentage points. However, as noted in the "Results" section, neither the single best-performing (or worst–performing) lesion nor the PERCIST criteria improve the predictive power of these models. To obtain a clinically useful metric, we speculate that one may need to require some degree of agreement across the within-patient lesion changes in addition to reduction in ^{18}F-FDG uptake.

Performance status before treatment (PS) was found to be an important predictor of OS (Table 2), confirming results reported for another NSCLC study involving treatment with targeted drugs and conventional chemotherapy (11). That study also identified the baseline sum of the longest dimensions as another significant predictor of survival. Here, similar to previously published studies (23,24), our study demonstrated the prognostic value of baseline ^{18}F-FDG PET. However, these studies did not account for performance status, and it has been noted that the prognostic value of baseline ^{18}F-FDG PET may also depend on histologic subtype (24).

The described analyses can also be performed using the semiparametric proportional hazards model (25,26). However, for ease of interpretation, and to enable a straightforward comparison with previous parameter estimates in NSCLC populations (11), we chose the AFT modeling framework. In addition, the parametric survival model enables simulation of hypothetical clinical trial outcomes (10). An excellent tutorial review of survival analysis in the context of cancer research, including a comparison of the AFT- and Cox-regression frameworks, is given in the study by Bradburn et al. (27).

CONCLUSION

According to this study of 125 NSCLC patients with residual disease after second- or third-line therapy with erlotinib, ^{18}F-FDG PET early after beginning treatment with a targeted drug can carry information about OS. Newly detected lesions appear to be more informative for OS than changes in SUV_{max}. Placebo-controlled clinical trials that include ^{18}F-FDG PET will have to clarify to what extent early ^{18}F-FDG PET responses to targeted drugs help identify drugs that prolong survival.

REFERENCES

1. Czernin J, Benz MR, Allen-Auerbach MS. PET/CT imaging: the incremental value of assessing the glucose metabolic phenotype and the structure of cancers in a single examination. *Eur J Radiol.* 2010;73:470–480.
2. 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;50(suppl 1):122S–150S.
3. Hicks RJ. Role of ^{18}F-FDG PET in assessment of response in non-small cell lung cancer. *J Nucl Med.* 2009;50(suppl 1):31S–42S.
4. 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;21:1285–1292.
5. Cachin F, Prince HM, Hogg A, Ware RE, Hicks RJ. Powerful prognostic stratification by [^{18}F] fluorodeoxyglucose positron emission tomography in patients with metastatic breast cancer treated with high-dose chemotherapy. *J Clin Oncol.* 2006;24:3026–3031.
6. Palazzo A, Iacovelli R, Cortesi E. Past, present and future of targeted therapy in solid tumors. *Curr Cancer Drug Targets.* 2010;10:433–461.
7. Contractor KB, Aboagye EO. Monitoring predominantly cytostatic treatment response with ^{18}F-FDG PET. *J Nucl Med.* 2009;50(suppl 1):97S–105S.
8. Klein J, Moeschberger M. *Survival Analysis.* 2nd ed. New York, NY: Springer; 2003.
9. Kalbfleisch J, Prentice R. *The Statistical Analysis of Failure Time Data.* Hoboken, NJ: John Wiley & Sons; 2002.
10. Claret L, Andre V, de Alwis D, Bruno R. Using change in tumor size as primary endpoint in oncology phase II studies. Available at: www.page-meeting.org/?abstract=1386. Accessed February 14, 2012.
11. Wang Y, Sung C, Dartois C, et al. Elucidation of relationship between tumor size and survival in non-small-cell lung cancer patients can aid early decision making in clinical drug development. *Clin Pharmacol Ther.* 2009;86:167–174.
12. Claret L, Girard P, Hoff PM, et al. Model-based prediction of phase III overall survival in colorectal cancer on the basis of phase II tumor dynamics. *J Clin Oncol.* 2009;27:4103–4108.
13. Binns D, Pirzkall A, Yu W, et al. OSI3926g Study Team. Compliance with PET acquisition protocols for therapeutic monitoring of erlotinib therapy in an international trial for patients with non-small cell lung cancer. *Eur J Nucl Med Mol Imaging.* 2011;38:642–650.
14. Mileshkin LR, Hughes BG, Mitchell P, et al. Changes in FDG- and FLT-PET imaging in patients with non-small cell lung cancer (NSCLC) following treatment with erlotinib (E). *Clin Cancer Res.* 2011;17:3304–3315.
15. Spigel DR, Ervin TJ, Ramlau R, et al. Final efficacy results from OAM4558g, a randomized phase II study evaluating MetMAb or placebo in combination with erlotinib in advanced NSCLC [abstract]. *J Clin Oncol.* 2011;29(suppl):7505.
16. Shankar L, Hoffman J, Bacharach S, et al. Consensus recommendations for the use of ^{18}F-FDG PET as an indicator of therapeutic response in patients in National Cancer Institute trials. *J Nucl Med.* 2006;47:1059–1066.

17. Patel P, Bothos J, Fredrickson J, Bengtsson T, Peterson A, de Crespigny A. Results from FDG-PET imaging in OAM4558g, a randomized, placebo-controlled, multicenter phase II trial of erlotinib $+/-$ metmab in second- and third-line NSCLC. Presented at the World Conference on Lung Cancer 2011; July 3–7, 2011; Amsterdam, The Netherlands.

18. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc.* 1958;53:457–481.

19. Everitt S, Herschtal A, Callahan J, et al. High rates of tumor growth and disease progression detected on serial pre- treatment fluorodeoxyglucose-positron emission tomography/computed tomography scans in radical radiotherapy candidates with nonsmall cell lung cancer. *Cancer.* 2010;116:5030–5037.

20. Hickeson M, Yun M, Matties A, et al. Use of a corrected standardized uptake value based on the lesion size on ct permits accurate characterization of lung nodules on FDG-PET. *Eur J Nucl Med Mol Imaging.* 2002;29:1328–1333.

21. 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. *Eur J Cancer.* 1999;35:1773–1782.

22. Rosell R, Gervais R, Vergnenegre A. Erlotinib versus chemotherapy (CT) in advanced non-small-cell lung cancer (NSCLC) patients (p) with epidermal growth factor receptor (EGFR) mutations: interim results of the European Tarceva versus Chemotherapy (EURTAC) phase III randomized trial [abstract]. *J Clin Oncol.* 2011;29(suppl):7503.

23. Berghmans T, Dusart M, Paesmans M, et al. Primary tumor standardized uptake value (SUVmax) measured on fluorodeoxyglucose positron emission tomography (FDG-PET) is of prognostic value for survival in non-small cell lung cancer (NSCLC): a systematic review and meta-analysis (MA) by the European lung cancer working party for the IASLC Lung Cancer Staging Project. *J Thorac Oncol.* 2008;3:6–12.

24. Um S, Kim H, Koh W, et al. Prognostic value of 18F-FDG uptake on positron emission tomography in patients with pathologic stage I non-small cell lung cancer. *J Thorac Oncol.* 2009;4:1331–1336.

25. Cox D. Regression models and life tables (with discussion). *JR Stat Soc.* 1972;34:187–220.

26. Cox D. Partial likelihood. *Biometrika.* 1975;62:269–276.

27. Bradburn M, Clark T, Love S, Altman D. Survival analysis part II: multivariate data analysis—an introduction to concepts and methods. *Br J Cancer.* 2003;89:431–436.