**­SUPPLEMENTARY MATERIAL**

Estrogen receptor binding (FES PET) and glycolytic activity (FDG PET) predict progression-free survival on endocrine therapy in patients with ER+ breast cancer

Brenda F. Kurland1, Lanell M. Peterson2, Jean H. Lee2, Erin K. Schubert3, Erin R. Currin4, Jeanne M. Link5, Kenneth A. Krohn2, David A. Mankoff3, Hannah M. Linden4

1. Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA

2. Department of Radiology, University of Washington, Seattle, WA

3. Department of Radiology, University of Pennsylvania, Philadelphia, PA

4. Division of Medical Oncology, University of Washington, Seattle, WA

5. Department of Diagnostic Radiology, Oregon Health & Science University, Portland, OR

Corresponding author:

Brenda F. Kurland, PhD

Research Associate Professor, Biostatistics

University of Pittsburgh &

University of Pittsburgh Cancer Institute

email: bfk10@pitt.edu

**Supplementary results:** *Agreement of imaging measures of endocrine resistance*

To assess agreement among patient-level summary measures, quantitative measures were compared using graphical displays and the intraclass correlation coefficient. Qualitative measures (and classification according to predefined cutpoints) were compared using summary tables and kappa statistics. **Table S1** summarizes concordance of the predefined classifiers. More than half (51/90) were classified as not resistant to endocrine therapy according to all 4 classifiers, and only 5 (6%) were classified as likely nonresponders by all 4 classifiers. Kappa measures of agreement were modest, ranging for pairs of classifiers from 0.22 (for any low FES/FDG and the qualitative measure) to 0.61 (for FES dynSUV and FES SULmean3 classifiers). Pairwise agreement of quantitative measures was examined by the intraclass correlation coefficient (**Table S2**) and ranged from 81% (95% CI 73%-87%) for FES dynSUV and FES SULmean3, to 15% (4%-47%) for FES dynSUV and FES/FDG ratio3. There were no consistent trends to explain patterns for lack of concordance based on patient or disease characteristics (age, number of lesions, number of months between breast cancer diagnosis and FES PET scan). Descriptive plots for agreement are shown in **Figure S1**.

**Supplementary results:** *Sensitivity analyses for variable selection by recursive partitioning*

Recursive partitioning identified three groups with distinct patterns of progression-free survival (PFS). We examined whether different predictors would be selected for overall survival (OS) and clinical benefit. When all predictors were considered, number of lesions (5 or fewer vs >5) was the first split selected for OS, but no variables consistently improved the tenfold cross-validated prediction error. When only imaging variables were considered, FDG PET SULmax was selected as a classifier for overall survival, with a nearly identical cutpoint as for PFS (2.29). Recursive partitioning predicting clinical benefit (PFS6) selected three terminal nodes nearly identical to classifications for continuous PFS: FDG SULmax < 2.34, and a dynSUV cutpoint of 1.47 for higher FDG SULmax.

Finally, we explored models predicting PFS by only FES SUVmax, FES SULmax, and FDG SUVmax, because maximum values are the current clinical standard and clinicians are familiar with relative values of FDG SUVmax but not FDG SULmax. Classifications were similar as for the model in Table 3 (23 cases with FDG SUVmax < 3.4; 53 cases with high FDG SUVmax and FES SULmax ≥1.4; 8 cases with high FDG SUVmax and FES SULmax <1.4) (**Figure S3**), but were not retained after tenfold cross-validation.

An additional sensitivity analysis for recursive partitioning with PFS as outcome was to examine possible period effects by analyzing data from earlier and later scans separately, with August 2004 as the date for dividing into two groups of equal size. Analysis of both subsets, including pruning based on cross-validation, resulted in similar classification to the full dataset. First, FDG SULmax with a cutpoint of about 2.2 was used to identify the 20-30% of patients with indolent disease by FDG PET. Among the remaining patients, an FES SULmean of 1.42 (earlier scans) or an FES SULmax of 5.51 (later scans) was the cutpoint to identify patients with high FES uptake and longer PFS on endocrine therapy (16/29 of higher-FDG patients for earlier scans, 7/33 for later scans). We hypothesize that known changes in recruitment strategies and changes in standard treatments were likely responsible for these differences, rather than unidentified differences in scanner hardware or protocols. Although precise thresholds for FES PET differed in the data period subgroups, we emphasize that recursive partitioning in two non-overlapping datasets both selected similar classification (first FDG PET SULmax, then FES PET SULmean or SULmax) from 12 potential predictors of PFS (2 FDG PET measures, 4 FES PET measures, 1 FES/FDG ratio, age, PgR, number of lesions, presence/absence of visceral disease, and number of prior chemotherapy regimens for metastatic disease).

**Supplementary results:** SUV *quantification by reconstruction method (filtered back-projection vs. iterative reconstruction)*

Filtered back-projection (FBP) was maintained as the primary reconstruction method used during the period of the study. However, iterative reconstruction (specifically ordered-subset expectation maximization, or OSEM) was also applied to many of the scans, with the same reconstructed in-plane spatial resolution as for FBP. We examined these data to explore differences in SUV quantitation.

Previous comparisons of FBP and OSEM reconstruction methods for FDG PET SUV quantitation have yielded conflicting findings, likely due to additional differences in reconstruction filters, ROI definition, and other factors also related to the period of primary use ([*1*](#_ENREF_1)). Using ROIs identified from the OSEM reconstruction, Krak et al. ([*1*](#_ENREF_1)) found “no statistically significant differences” between FBP and OSEM FDG PET SUVs in 46 lesions in 27 breast and lung cancer patients.

For FDG PET SULmax we found a small but predictable shift toward higher values using OSEM values compared to FBP (**Figure S4** panel A). The association is well-described by a linear fit to the logged SULmax values, with a slope of 1.03 (roughly parallel to the dashed line of identity) and 98% of the variation in OSEM FDG SULmax explained by variation in FBP FDG SULmax (R-squared = 0.98). Based on this model, the FDG SULmax value of 2.2 from FBP reconstruction (Table 3, Figure 2) translates to a value of 2.5 using OSEM reconstruction. However, like other cutpoints reported, extensive validation in independent cohorts is required. For example, the relationship between FDG SULmax by FBP and OSEM reconstruction is not replicated in 22 patients with early breast cancer (**Figure S4** panel B), in which for most lesions there is lower FDG SULmax for OSEM. However, those tumors were smaller and with lower uptake, so the range is not comparable and partial volume effects were likely a factor in SUV quantitation.

For FES SULmean, agreement was excellent between FBP and OSEM reconstructions (**Figure S5**), with intercept of -0.04 and slope of 1.004 and R2=0.96 (without log transformation).

**REFERENCE**

**1.** Krak NC, Boellaard R, Hoekstra OS, Twisk JW, Hoekstra CJ, Lammertsma AA. Effects of ROI definition and reconstruction method on quantitative outcome and applicability in a response monitoring trial. *Eur J Nucl Med Mol Imaging.* 2005;**32**(3):294-301.

294-301.

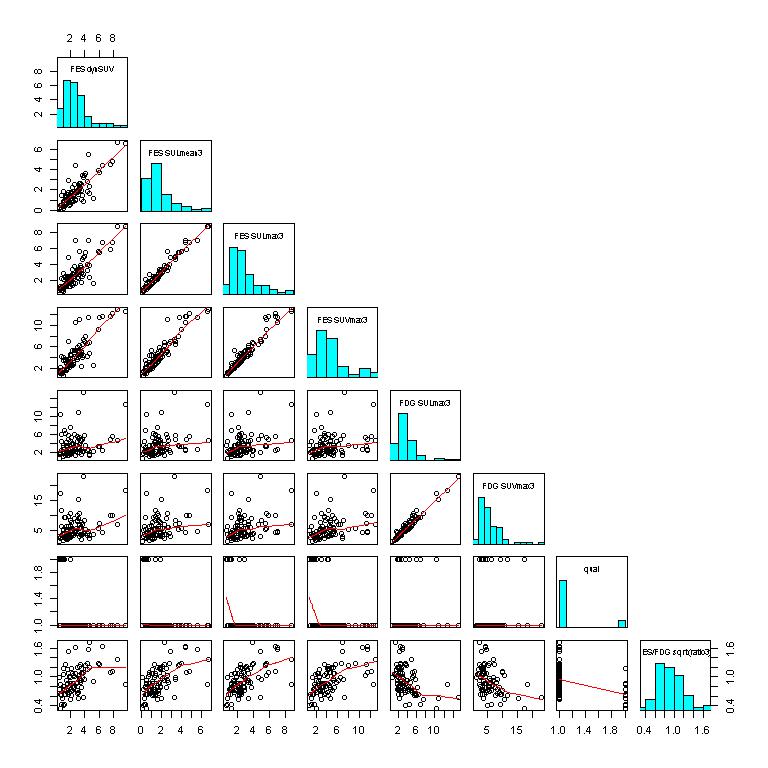
**Table S1.** Agreement of predefined classifiers (n=90).

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Classifier** | **description** | **N (%)** |
| FES dynSUV > 1.5  FES SULmean3 > 1  no low FES/FDG  qualitative FES+ | Likely responder, by all 4 classifiers | 51 (57%) |
| FES dynSUV ≤ 1.5  FES SULmean3 ≤ 1  any low FES/FDG  qualitative FES- | Likely nonresponder, by all 4 classifiers | 5 (6%) |
| FES dynSUV ≤ 1.5  FES SULmean3 ≤ 1  no low FES/FDG  qualitative FES+ | Likely nonresponder by FES dynSUV,  FES SULmean3  Likely responder by FES/FDG, qualitative | 7 (8%) |
| FES dynSUV > 1.5  FES SULmean3 > 1  any low FES/FDG  qualitative FES+ | Likely responder by FES dynSUV,  FES SULmean3, qualitative  Likely nonresponder by FES/FDG | 6 (7%) |
| Other patterns | Other patterns | 21 (22%) |

**Table S2.** Pairwise intraclass correlation coefficients (ICCs) for selected quantitative measures\*.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Measure #1** | **Measure #2** | **ICC (95% confidence interval)** |
| FES dynSUV | FES SULmean3 | 0.81 (0.73, 0.87) |
| FES dynSUV | FES/FDG ratio3 | 0.15 (0.04, 0.47) |
| FES SULmean3 | FES/FDG ratio3 | 0.22 (0.08, 0.47) |
|  |  |  |
| FES SULmean3 | FES SULmax3 | 0.94 (0.91, 0.96) |
| \* See Table 1 footnotes for full descriptions of measures. N=90, N=88 for FES/FDG ratio3 | | |

**Figure S1**. Scatterplot of imaging parameter agreement.



**Figure S2.** Association of clinical benefit (progression-free survival > 6 months on endocrine therapy) with quantitative imaging biomarkers predicting endocrine response, defined in Table 1. **A.** FES dynSUV **B.** FES SULmean3 **C.** FES/FDG ratio3. ▬ is median, ◊ is mean. None are associated with PFS6 by logistic regression (N=76, Wald test p>0.10).

C:\Users\bfk10\Box Sync\Imaging\FES Consortium BC Pilot\Paper2-PFS\FigS2.TIFF

**Figure S3.** Results of analysis using recursive partitioning to predict PFS by selected imaging parameters (FES SUVmax3, FES SULmax3, and FDG SUVmax3)  **A.** Progression-free survival predicted by FES SULmax3 (cutpoint 1.4) and FDG SUVmax3 (cutpoint 3.4) (log-rank p<0.001 but classifiers not retained by tenfold cross-validation) **B.** Overall survival predicted by FES SULmax3 and FDG SUVmax3 (log-rank p=0.005).

C:\Users\bfk10\Box Sync\Imaging\FES Consortium BC Pilot\Paper2-PFS\Fig S3.TIFF

**Figure S4.** Association of log(FDG SULmax) with reconstruction of the same scan by filtered back-projection (X axis) or ordered-subset expectation maximization (Y axis). Solid line is best linear fit; dashed line is perfect agreement (intercept=0, slope=1). A) 36 lesions in 7 patients from this study. B) 22 lesions in 22 patients from pre- and post-therapy scans in a separate cohort of patients with early stage, operable breast cancer.

C:\Users\bfk10\Box Sync\Imaging\FES Consortium BC Pilot\Paper2-PFS\CCR revision\Fig S4.wmf

**Figure S5.** Association FES SULmean with reconstruction of the same scan by filtered back- projection (X axis) or ordered-subset expectation maximization (Y axis) for 93 lesions in 16 patients from this study. Solid line is best linear fit; dashed line is perfect agreement (intercept=0, slope=1). The second panel is the same data as the first, with restricted axis values to illustrate agreement in that range.

C:\Users\bfk10\Box Sync\Imaging\FES Consortium BC Pilot\Paper2-PFS\CCR revision\Fig S5.wmf