Supplemental Table 1. Eligibility Criteria

Inclusion Criteria
1. Diagnosis of breast cancer bone metastasis by one of the following:
(a) Bone biopsy confirming breast cancer metastasis
(b) Clear diagnostic findings of bone metastasis on at least two imaging modalities (bone scan, CT, MRI, or FDG PET/CT)
2. Plan to start new systemic therapy (endocrine or single-agent chemotherapy) for breast cancer bone metastases
3. Willing and able to undergo serial PET imaging studies
4. Informed consent
Exclusion Criteria
1. History of malignancy other than breast cancer or non-melanoma skin cancer
2. Uncontrolled diabetes mellitus
3. Pregnancy
4. Inability to tolerate scanning (e.g. - claustrophobia, severe pain)
5. Other severe systemic illness that would impact treatment compliance or survival

Supplemental Table 2. Imaging Parameters

Characteristic	GE ADVANCE FDG*, NaF†	GE Discovery STE FDG, NaF
Slice thickness (mm)	4.30	3.27
Reconstruction diameter (mm)	550	700
Array size (pixels)	128x128	128x128
Attenuation (min)/low dose CT	3,3	120 kVp, auto mA
Bed Duration (min/field-of-view)	7,5	7,5
Reconstruction Method	2D FBP	2D FBP
Convolution kernel (filter, mm)	Hanning,12	Hanning,12
Scatter correction method	Convolution Subtraction	Convolution Subtraction

*Fasting for at least 6 hours was required for FDG-PET scans.
†Dynamic imaging for NaF-PET (data not shown) started at the beginning of a one-minute infusion and was followed by the torso survey over the same axial extent as the FDG-PET scan.

Calculations for standardized uptake value (SUV) and lean body mass corrected uptake (SUL) are as follows:

$$SUV = \frac{Tissue\ Activity\ (kBq/cc)}{\frac{Injected\ dose\ (MBq)}{Body\ Weight\ (kg)}} \quad SUL = \frac{Tissue\ Activity\ (kBq/cc)}{\frac{Injected\ dose\ (MBq)}{lean\ body\ mass\ (kg)}}$$

Lean body mass calculation is from James, W. B. T. Research on Obesity: a report of the DHSS/MRC group. Her Majesty's Stationary Office. London 94p. 1976.

Supplemental Table 3: Event characteristics. Response to therapy was determined clinically from the time of study entry.

	N=28 (%)	Median (months)	Range (months)
Time to skeletal related event (tSRE)	17 (61%)	8.3	0.26-86.5
Time-to-Progression (TTP)	26 (93%)	5.8	2.1-29.5
Overall Survival (OS)	25 (89%)	35.1	6.1-87.3

Supplemental Table 4: Time intervals between scans.

	Mean	Range
Time between FDG scans (N=24)	4.3 months	1.6-7.2 months
Time between NaF scans (N=24)	4.3 months	2.1-8.1 months
Time between NaF and FDG scan1	10 days	1-32 days
Time between NaF and FDG scan2	20 days	1-75 days
Time between FDG scan1 and 2 (N=28)	5.2 months	1.6-25.8 months

Supplemental Table 5: Univariate analysis of normal bone corrected SUVmax (SUVmax_c) NaF-PET parameters and clinical endpoints. Normal bone SUVmean was subtracted from the lesion SUVmax to correct for normal bone uptake in each lesion. The borderline significance of increased SUVmax NaF uptake on OS does not persist with normal bone correction.

	tSRE				TTP				OS			
NaF SUVmax_c (N=24)	*HR*	*P*	R^2	*C*	*HR*	*P*	R^2	*C*	*HR*	*P*	R^2	*C*
Index Lesion												
$SUV_{max}1$	1.23	0.372	0.032	0.576	1.25	0.291	0.044	0.571	1.07	0.736	0.005	0.483
$SUV_{max}2$	1.20	0.348	0.033	0.623	1.27	0.280	0.043	0.549	0.866	0.514	0.019	0.602
% difference	1.24	0.462	0.021	0.504	0.93	0.758	0.004	0.527	0.685	0.237	0.058	0.647

Supplemental Table 6: Additional statistical analysis. Two-sample comparisons were performed for 3 SUVmax variables [scan1, scan2, and unit difference (scan2-scan1)] to test for differences in FDG-PET and NaF-PET information. The unit difference is the mean SUVmax in each scan for up to 5 lesions. To reduce sensitivity to distributional assumptions, 2-sample Wilcoxon tests were used along with t-tests. **A.** SUVmax variables were tested for consistency in the patients with ER+ disease who had chemotherapy (hormone-refractory subgroups). Means for the two groups (No Chemo vs Chemo) are given in parentheses in the t-test column. Means at scan1 and scan2 were not significantly different for patients with vs without chemotherapy. The difference between scans (scan2-scan1) was marginally different between groups for FDG (p=0.05; means -2.60 and 0.29; Wilcoxon p=0.14). **B.** Patients who are within 15 weeks between their baseline scan and scan2: Means are in parentheses (within 15 weeks vs over 15 weeks groups). For both FDG-PET and NaF-PET 8 patients had less than 15 weeks from scan1 to scan2. FDG SUVmax was significantly different for patients who were scanned within the 15 week period vs those who were not (p=0.02; Wilcoxon p=0.04). No other variables showed differences for FDG-PET or NaF-PET. **C.** SUVmax variables were tested for patients with predominately lytic vs predominately sclerotic or mixed; two patients had type unknown). There were no significant differences for any of the three PET variables.
Cox modeling in the subgroups does not indicate substantial differences in the results for FDG-PET or NaF-PET. The limited sample size does not justify detailed presentation of these results.

A. ER+ with or without chemotherapy				
	FDG (N=25; 8 had chemo)		NaF (N=22; 6 had chemo)	
	t-test: p-value (means)	Wilcoxon p-value	t-test: p-value (means)	Wilcoxon p-value
$SUV_{max}1$	0.42 (10.90, 9.08)	0.93	0.39 (36.29, 30.92)	0.69
$SUV_{max}2$	0.30 (6.92, 8.93)	0.34	0.20 (32.67,27.43)	0.37
Unit Difference	0.05 (-2.60, 0.29)	0.14	0.97 (-3.32,-3.43)	0.69
B.Time from scan 1 to scan 2 < 15 weeks (y/n)				
	FDG (N=28; 8 <15wks)		NaF (N=24; 8 <15wks)	
	t-test: p-value (means)	Wilcoxon p-value	t-test: p-value (means)	Wilcoxon p-value
$SUV_{max}1$	0.02 (6.73,11.31)	0.04	0.97 (34.24, 34.43)	0.65
$SUV_{max}2$	0.24 (5.74, 7.68)	0.14	0.40 (28.77, 32.28)	0.98
Unit Difference	0.17 (-0.41, -2.29)	0.53	0.37 (-5.34, -2.28)	0.11
C.Lesion type				
	FDG (N=26; 11 lytic)		NaF (N=22; 10 lytic)	
	t-test: p-value (means)	Wilcoxon p-value	t-test: p-value (means)	Wilcoxon p-value
$SUV_{max}1$	0.33 (11.64, 9.21)	0.13	0.31 (38.82, 31.69)	0.46
$SUV_{max}2$	0.34 (8.03, 6.46)	0.41	0.26 (34.69, 27.94)	0.46
Unit Difference	0.46 (-2.75, -1.26)	0.96	0.62 (-2.51, -4.72)	0.87

Supplemental Table 7: Response by modified PERCIST measures. Modified PERCIST criteria allowed lesions 1.5 times the liver SULpeak; change in normal liver/aorta not within 20% and .3 SUL; uptake time between scans >15 min; dose between scans >20%; scans done on different scanners.

Response Criteria	N=28 (%)	Median tSRE (95% CI)	Median TTP (95% CI)	Median OS (95% CI)
Complete Response (CR) all lesion uptake <liver SULpeak. No new lesions	1 (4%)	29.7 mos (NA-NA)	29.4 mos (NA-NA)	35.0 mos (NA-NA)
Partial Response (PR) any lesion change <30% and at least 0.8 SUL difference. No new lesions	6 (21%)	NA (14.9-NA)	11.6 mos (5.8-NA)	33.2 mos (15.5-NA)
Stable Disease (SD) any lesion change not less than or greater than 30% and 0.8 SUL. No new lesions.	6 (21%)	47.6 mos (32.2-NA)	12.0 mos (4.5-NA)	66.5 mos (48.5-NA)
Progressive Disease (PD) Any lesion change >30% and 0.8 SUL or any new lesion*	11 (39%)	4.6 mos (4.1-NA)	3.8 mos (3.5-NA)	25 mos (18.5-NA)
Unevaluable no lesions above liver threshold or normal liver/aorta not available	4 (14%)	60.7 mos (60.7-NA)	12.1 mos (2.1-NA)	64.1 mos (11.6-NA)

* 7 patients with progressive disease had new lesions at follow-up.

Supplemental Figure 1: Multivariable analysis. Kaplan-Meier survival patterns for high and low (greater/less than median) risk patients. The top panel variables include SUVmax of a single index lesion plus the mean change in SUVmax for up to 5 lesions. The bottom panel variables include SULpeak of a single index lesion plus the mean change in SULpeak of up to 5 lesions. Column **A** is tSRE, **B** is TTP and **C** is OS. Risk is identified by Multivariate Cox Regression Analysis for each endpoint. Log-rank tests are used to calculate p-values for the difference between survival curves.