Supplemental Table 1: Study site scanning protocol standards, with comparison to UPICT (*1*) standards.

Protocol feature	**Study sites**	**UPICT** (section in (*1*) supplement)
Site accreditation	See Materials and Methods and (*2*)	(see 2.2, 7.2)
Fasting	≥ 6 hours	Target: ≥ 6 hours Acceptable: ≥ 4 hours (4.1.1)
No vigorous exercise	≥ 24 hours	Target: ≥ 24 hours Acceptable: ≥ 6 hours (4.1.3)
Fasting blood glucose (non-diabetic patient)	≤ 175 mg/dL (one with 182 mg/dLwas scanned)	Target: <150 mg/dL Acceptable ≤ 200 mg/dL (4.2.2)
FDG injected dose	259-407 MBq recommended 287-566 MBq actual	370-740 MBq (United States) May be tailored by patient weight, 2D vs 3D scanning, acquisition time per bed position, % bed overlap (5.2)
Other preparation	Void 45 min post-injection	Adequate hydration (4.1.2, 4.2.3) Void 5-10 min prior to image acquisition (4.2.3) Shallow breathing or respiratory gating (7.1.1)
Uptake time	60 min ± 10 min post-injection	Target: 60 min post-injection Acceptable: 55-75 min post-injection (5.3)
Patient position	Arms up preferred	Arms up preferred (7.2.1)
Scan direction	Inferior to superior, CT attenuation before PET acquisition	Inferior to superior, CT attenuation before PET acquisition (7.1.1)
Voxel size	4 - 5.47 mm	Target/Acceptable: Reconstructed voxel size of 3-4 mm in all 3 dimensions Ideal: voxel size as small as possible with current technology (7.1.2)
Imaging data acquisition	2-7 fields-of-view	Typically 6 bed positions, 2-4 min per bed position (3D acquisition) or 3-8 min (2D acquisition); may be lowered for higher injected dose, lower body weight, higher bed overlap, etc. (7.2)
Reconstruction	Iterative	Iterative (7.3)
Shared features: clock synchronization, accurate patient weight, confirmation of patient preparation, record exact dose injected and residual dose in syringe, avoid/record extravasation		

Supplemental Table 2: Scan and lesion characteristics

	Same site/scanner (n=10 patients, n=51 lesions)	**Same institution, different scanner** (n=2 patients, n=34 lesions)	**Different site/scanner** (n=11 patients, n=77 lesions)	**Total** (n=23 patients, n=162 lesions)
Lesion site	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Bone	38 (75%)	26 (76%)	18 (23%)	82 (51%)
Soft tissue	13 (25%)	8 (24%)	59 (77%)	80 (49%)
	median (range)	**median (values)**	**median (range)**	**median (range)**
Injection dose (scan 1), MBq	368 (354-396)	393 (387,399)	379 (315-566)	379 (315-566)
Injection dose (scan 2), MBq	369 (305-381)	371 (367,375)	356 (287-401)	367 (287-401)
Absolute difference in injection dose, MBq	17 (1-58)	22 (20,24)	44 (2-247)	22 (1-247)
Blood glucose (scan 1), mg/dL*	92.5 (78-104)	110.5 (110,111)	100 (82-182)	93 (78-182)
Blood glucose (scan 2), mg/dL*	89.5 (81-97)	100.5 (98,103)	98.3 (78-143)	94 (78-143)
Absolute difference in blood glucose, mg/dL	5.8 (1-19)	10 (7,13)	18 (0.3-39)	13 (0.3-39)
Weight (scan 1), kg	75.6 (49-93.3)	114.1 (96.3,132)	78.6 (60.2-129)	76.2 (49-132)
Weight (scan 2), kg	74.8 (49-94.9)	115.5 (97.7,133.2)	78.6 (57.6-129)	76.2 (49-133.2)
Absolute difference in weight, kg	0.3 (0.0-1.6)	1.3 (1.2,1.4)	0.8 (0.0-4.8)	0.5 (0.0-4.8)
Uptake time (scan 1), min	60 (60-64)	60 (60,60)	60 (54-66)	60 (54-66)
Uptake time (scan 2), min	61 (58-70)	60 (60,60)	60 (56-70)	60 (56-70)
Absolute difference in uptake time, min	0.5 (0-6)	0 (0,0)	3 (0-11)	2 (0-11)
Liver SUVmean (avg of 2 scans)	2.4 (1.9-2.9)	2.8 (2.4,3.1)	2.4 (2.0-2.9)	2.4 (1.9-3.1)
Liver SUVmean absolute difference	0.1 (0.0-0.3)	0.2 (0.1,0.3)	0.2 (0.0-0.5)	0.1 (0.0-0.5)
Liver SUVmean absolute value of percentage difference	5.6 (1.8-13.5)	7.2 (2.9,11.4)	5.9 (0.5-23.7)	5.9 (0.5-23.7)
Tumor SUVmax (avg of 2 scans)	4.3 (1.0-18.2)	5.7 (4.4,16.8)	5.1 (2.0-28.8)	5.1 (1.0-28.8)
Tumor SUVmax absolute difference	0.4 (0.0-2.3)	0.4 (0.0,2.2)	0.8 (0.1-19.1)	0.6 (0.0-19.1)
Tumor SUVmax absolute value of percentage difference	8.6 (0.1-48.8)	5.8 (0.2,36.7)	17.4 (1.3-97.0)	11.9 (0.1-97.0)

*In some cases, 3 glucose measurements were made from the same blood sample and the average was recorded.

Supplemental Table 3: Additional covariates predicting test-retest SUVmax differences. Random intercept (subject) linear mixed effects models, with each predictor added to 2-group class (same institution vs. different institution and scanner). Fitted values at 25% and 75% percentile for continuous predictors. Patient and scanner factors did not appear to substantially affect magnitude of test-retest differences, in part because of rigorous control of factors such as uptake time.

Additional predictor	**Fitted percent test-retest difference in tumor SUVmax for same institution ***	**95% confidence interval**	**Wald p-value**
SITE			0.07
Bone	7%	(5%, 9%)	
Soft tissue	10%	(6%, 14%)	
BMI (scan 1)			0.52
Obese	8%	(6%, 12%)	
Not obese	7%	(5%, 10%)	
DAYS BETWEEN SCANS			0.92
2 days between scans	8%	(5%, 12%)	
14 days between scans	7%	(5%, 11%)	
	Fitted test-retest tumor SUVmax ratio for same institution†	**95% confidence interval**	**Wald p-value**
DIFFERENCE IN BLOOD GLUCOSE			0.08
1 mg/dL higher at scan 2	0.997	(0.994, 1.000)	
15 mg/dL higher at scan 2	0.956	(0.908, 1.006)	
DIFFERENCE IN UPTAKE TIME			0.55
1 min later at scan 1	1.005	(0.990, 1.020)	
10 min later at scan 1	1.046	(0.901, 1.215)	
DIFFERENCE IN LIVER SUVmean			0.93
6% lower at scan 2	0.949	(0.874, 1.030)	
4% higher at scan 1	0.946	(0.866, 1.034)	

* Outcome is log(absolute percent difference + 1)

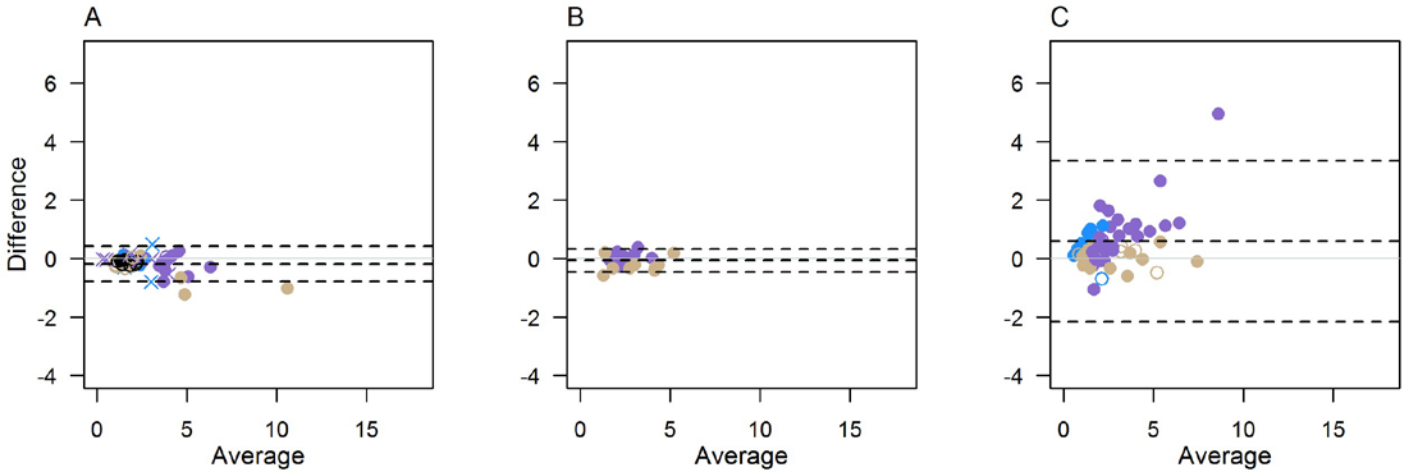
† Outcome is (log(SUVmax1) – log(SUVmax2)) without absolute value, since directionality as well as magnitude is part of the relationship between outcome and predictor

Supplemental Table 4: SUVmax within-subject coefficient of variation (wCV), repeatability coefficient (RC) and 95% confidence intervals, calculated using the same methodology as prior studies (*3,4*).

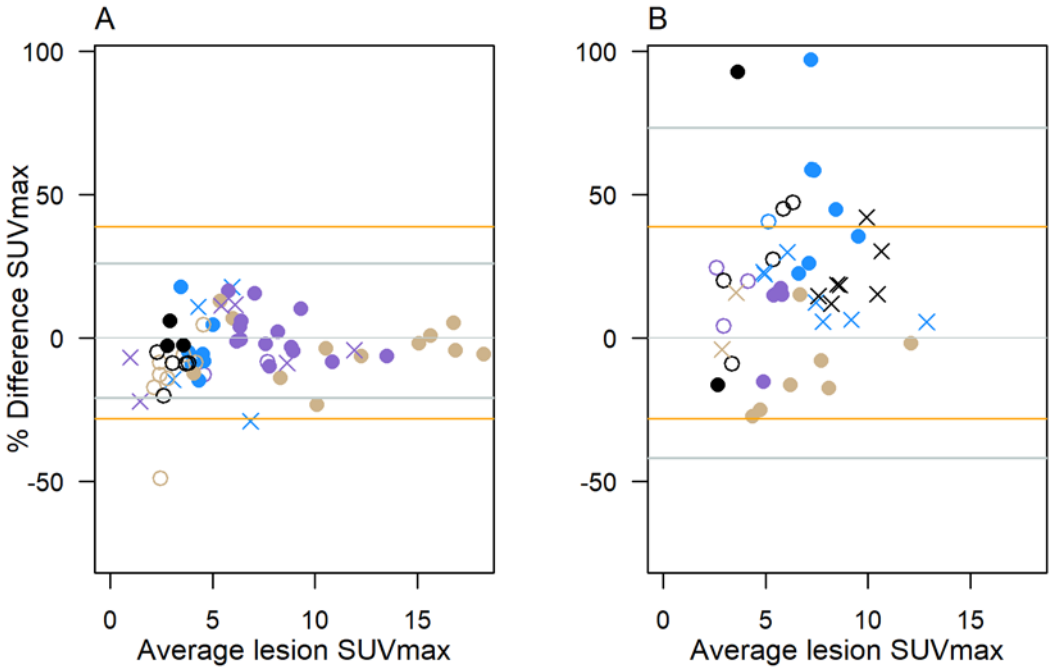
Difference score	N	wCV*	Lower 95% RC†	Upper 95% RC†	95% CI for lower RC‡	95% CI for upper RC‡
7 most FDG-avid lesions§	109	19%	-38%	62%	-43% to -33%	50% to 76%
Same institution only	61	9%	-21%	26%	-25% to -17%	20% to 33%
Different institution only	48	22%	-42%	73%	-50% to -34%	51% to 99%
Single most FDG-avid lesion at first scan	23	22%	-42%	72%	-50% to -32%	47% to 102%
Average of 7 most FDG-avid lesions	23	17%	-36%	55%	-45% to -25%	32% to 82%

* Eq. 4 in (*3*); differs from Table 2 in (*5*) because values are log-transformed before the difference is calculated.
† Eq. 5 in (*3*) and (*4*)
‡ CI for limits of agreement (*6*)
§ linear mixed effects model with random intercept (*4*)

Supplemental Figure 1: Bland-Altman plots of **SULpeak** for repeated scans: **A)** 10 patients (51 lesions) with repeat scans using the same scanner; **B)** 2 patients (34 lesions) with repeat scans using different scanners from the same unit; **C)** 11 patients (77 lesions) with repeat scans using different scanners from different sites. Within each panel, plotting character/color is the same for multiple lesions in a single patient. Average difference and 95% limits of agreement are shown with dashed lines. The two lesions from a melanoma patient are not shown on panel C but contribute to limits of agreement calculations. Shown on the same scale as for SUVmax, absolute difference in lesion SULpeak was small for test-retest at the same institution, and larger for studies conducted with network sites. This SULpeak measure has the lean-body-mass correction, approximate ROI volume, and mean uptake (rather than uptake for hottest pixel) recommended by PERCIST 1.0, although the ROI shape is cubic instead of spherical. Fitted percent difference in repeat scans (comparable to Model 2 in Table 3) is 6% (4%-9%, 95% confidence interval) for the same scanner/model condition and 19% (12%-28%) for different manufacturer/institution.



Supplemental Figure 2: Percentage difference in SUVmax versus average SUVmax: **A)** 12 patients (85 lesions) with repeat scans using the same scanner or different scanners from the same unit (combined Fig. 1A and B data); **B)** 11 patients (77 lesions) using different scanners from different sites. Plotting character/color identifies multiple lesions in a single patient, as for Fig. 1. Solid grey lines (—) = estimated 95% repeatability coefficients (from Supplemental Table 4, centered around zero (*5*)). Orange lines (—) = consensus SUVmax 95% limits of same-scanner repeatability (*4,7*).



REFERENCES

1. Graham MM, Wahl RL, Hoffman JM, et al. Summary of the UPICT protocol for 18F-FDG PET/CT imaging in oncology clinical trials. *J Nucl Med.* 2015;56:955-961.

2. Byrd DW, Doot RK, Allberg KC, et al. Evaluation of cross-calibrated 68Ge/68Ga phantoms for assessing PET/CT measurement bias in oncology imaging for single- and multicenter trials. *Tomography.* 2016;2:353-360.

3. Velasquez LM, Boellaard R, Kollia G, et al. Repeatability of 18F-FDG PET in a multicenter phase I study of patients with advanced gastrointestinal malignancies. *J Nucl Med.* 2009;50:1646-1654.

4. Weber WA, Gatsonis CA, Mozley PD, et al. Repeatability of 18F-FDG PET/CT in advanced non-small cell lung cancer: Prospective assessment in 2 multicenter trials. *J Nucl Med.* 2015;56:1137-1143.

5. Lodge MA. Repeatability of SUV in oncologic (18)F-FDG PET. *J Nucl Med.* 2017;58:523-532.

6. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res.* 1999;8:135-160.

7. QIBA Profile: FDG-PET/CT as an imaging biomarker measuring response to cancer therapy version 1.13 (technically confirmed version). http://qibawiki.rsna.org/images/1/1f/QIBA_FDG-PET_Profile_v113.pdf. Updated November 18, 2016. Accessed January 29, 2018.