

**Supplemental Table 1.** Comparison of repeatability (variance of relative test-retest difference) for different methods of normalization

Comparison			<i>p</i> -value
			Pitman-Morgan test
SUVmax	SUV	SUV <sub>AUC</sub>	0.036
	SUV	SUV <sub>TBR</sub>	< 0.001 **
	SUV <sub>AUC</sub>	SUV <sub>TBR</sub>	< 0.001 **
SUVpeak	SUV	SUV <sub>AUC</sub>	0.111
	SUV	SUV <sub>TBR</sub>	< 0.001 **
	SUV <sub>AUC</sub>	SUV <sub>TBR</sub>	< 0.001 **
SUVmean	SUV	SUV <sub>AUC</sub>	0.016
	SUV	SUV <sub>TBR</sub>	< 0.001 **
	SUV <sub>AUC</sub>	SUV <sub>TBR</sub>	0.001 *

*p*-values for pairwise comparisons of variances of (selected combinations of) measures. \* Significant at 5% level (Bonferroni-correction for 9 tests:  $p < 0.0056$ ). \*\* Significant at 1% level (Bonferroni-correction for 9 tests:  $p < 0.0011$ ).

**Supplemental Table 2.** Comparison of repeatability (variance of relative test-retest differences) between SUVmax, SUVpeak and SUVmean using the Pitman-Morgan test

Comparison			<i>p</i> -value
			Pitman-Morgan test
SUV	SUVmax	SUVpeak	0.356
	SUVmax	SUVmean	0.037
	SUVpeak	SUVmean	0.455
SUV <sub>AUC</sub>	SUVmax	SUVpeak	0.184
	SUVmax	SUVmean	0.081
	SUVpeak	SUVmean	0.955
SUV <sub>TBR</sub>	SUVmax	SUVpeak	0.732
	SUVmax	SUVmean	0.075
	SUVpeak	SUVmean	0.219

*p*-values for pairwise comparisons of variances of (selected combinations of) measures. Bonferroni correction for 9 tests performed by setting significance levels at 5% ( $0.05/9 = 0.0056$ ) and at 1% ( $0.01/9 = 0.0011$ ).

**Supplemental Table 3.** Comparison of variance of relative test-retest differences between groups using Levene's test

		Type: (bone vs. lymph node)	Location: (intrathoracically vs. abdominal/pelvic)	Size: below median vs. median and above <sup>1</sup>
		<i>p</i> -value Lev test	<i>p</i> -value Lev test	<i>p</i> -value Lev test
SUVmax	SUV	0.754	0.944	0.092
	SUV <sub>AUC</sub>	0.247	0.204	0.003 *
	SUV <sub>TBR</sub>	0.082	0.793	0.076
SUVpeak	SUV	0.655	0.808	0.264
	SUV <sub>AUC</sub>	0.744	0.805	0.075
	SUV <sub>TBR</sub>	0.226	0.354	0.459
SUVmean	SUV	0.718	0.725	0.556
	SUV <sub>AUC</sub>	0.341	0.768	0.046
	SUV <sub>TBR</sub>	0.257	0.541	0.537

<sup>1</sup>median of the mean of the test and retest measurement; *Lev* Levene's; \* Significant at 5% level (Bonferroni-correction for 9 tests:  $p < 0.0056$ )

**Supplemental Table 4.** Relative test-retest difference (in percentage) of the semi-quantitative parameters as a function of uptake, metabolic tumor volume, metastatic tissue type and location

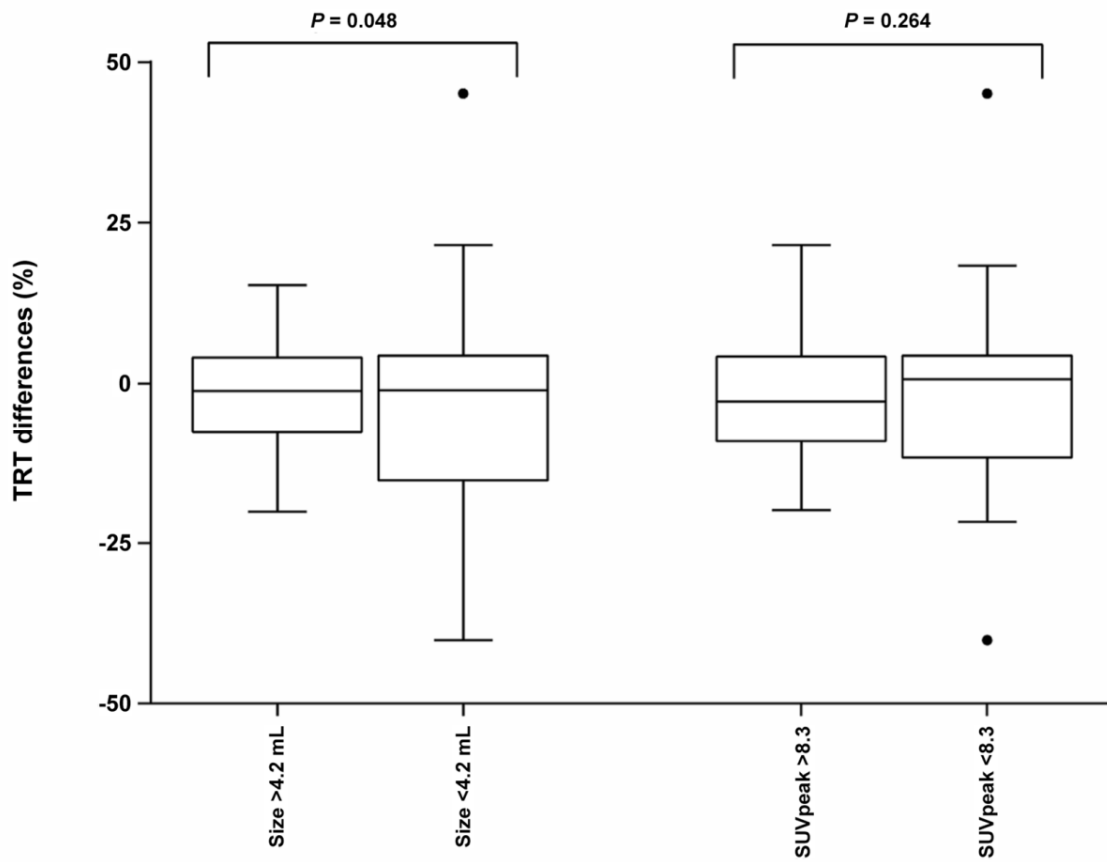
RC (%)	SUVpeak	SUVpeak <sub>AUC</sub>	SUVpeak <sub>TBR</sub>	MTV	TLCU	TLCU <sub>AUC</sub>	TLCU <sub>TBR</sub>
All	24	28	45	36	33	31	51
SUVpeak >8.3	19	23	41	34	34	30	50
SUVpeak <8.3	28	31	47	36	31	32	51
MTV >4.2 cm <sup>3</sup>	23	28	45	29	25	24	41
MTV <4.2 cm <sup>3</sup>	25	24	44	45	42	40	62
Bone	24	27	47	34	31	30	52
Lymph node	23	28	36	36	36	32	48
Intrathoracically	17	23	39	37	36	40	48
Abdominal/pelvic	24	27	45	34	32	29	51

*RC* repeatability coefficients; *SUV* standardized uptake value; *AUC* area under the curve of blood activity concentration; *TBR* blood concentration 40 minutes after injection; *MTV* metabolic tumor volume; *TLCU* total lesion choline uptake.

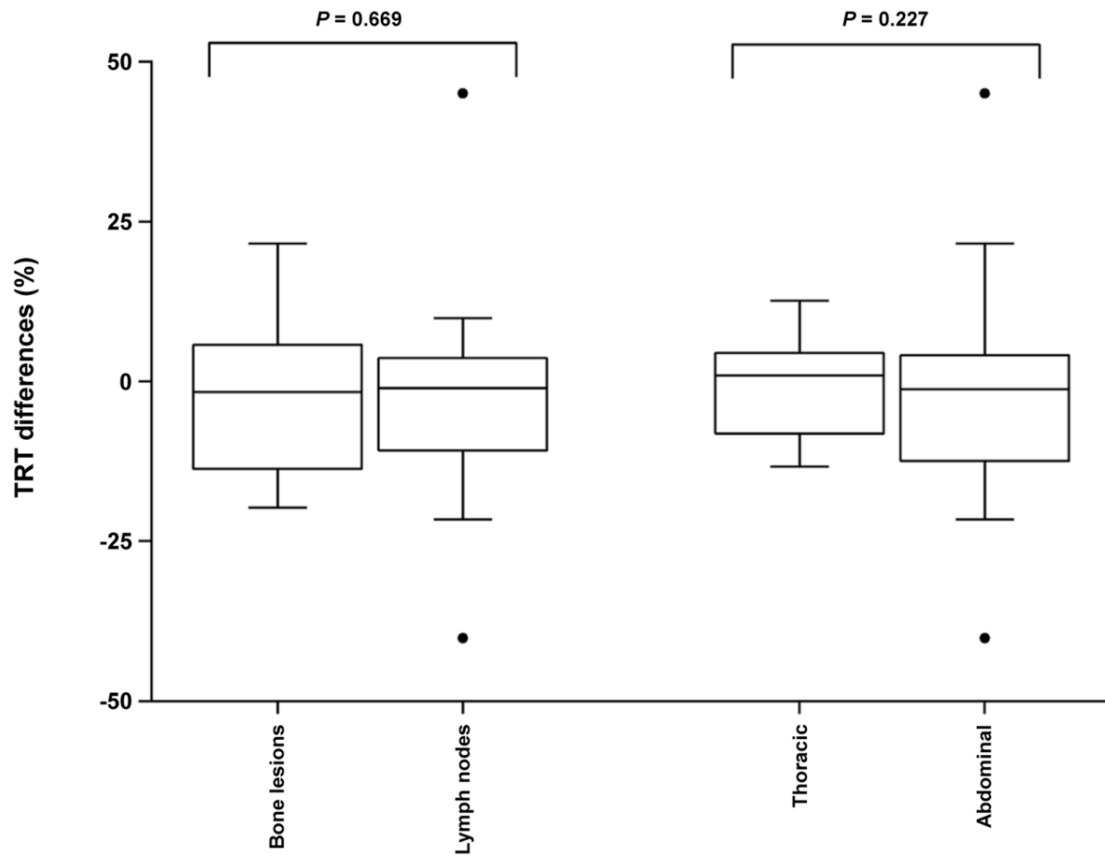
**Supplemental Table 5.** Test-retest difference of the semi-quantitative parameters as a function of uptake, metabolic tumor volume, metastatic tissue type and location

RC (test-retest)	SUV <sub>peak</sub>	SUV <sub>peak</sub> <sub>AUC</sub>	SUV <sub>peak</sub> <sub>TBR</sub>	MTV	TLCU	TLCU <sub>AUC</sub>	TLCU <sub>TBR</sub>
All	1.9	1.4	5.9	3.2	21	11	52
SUV <sub>peak</sub> >8.3	2.3	1.7	7.5	3.9	28	13	61
SUV <sub>peak</sub> <8.3	1.4	1.0	3.7	2.3	10	8.4	42
MTV >4.2 cm <sup>3</sup>	2.0	1.6	6.9	3.9	27	14	62
MTV <4.2 cm <sup>3</sup>	1.8	0.9	2.0	1.3	4.8	2.7	12
Bone	1.6	1.1	5.3	2.8	17	10	57
Lymph node	2.5	1.8	6.6	3.5	25	10	40
Intrathoracically	1.3	0.8	4.8	2.7	16	8.7	27
Abdominal/pelvic	2.0	1.5	6.1	3.3	23	11	51

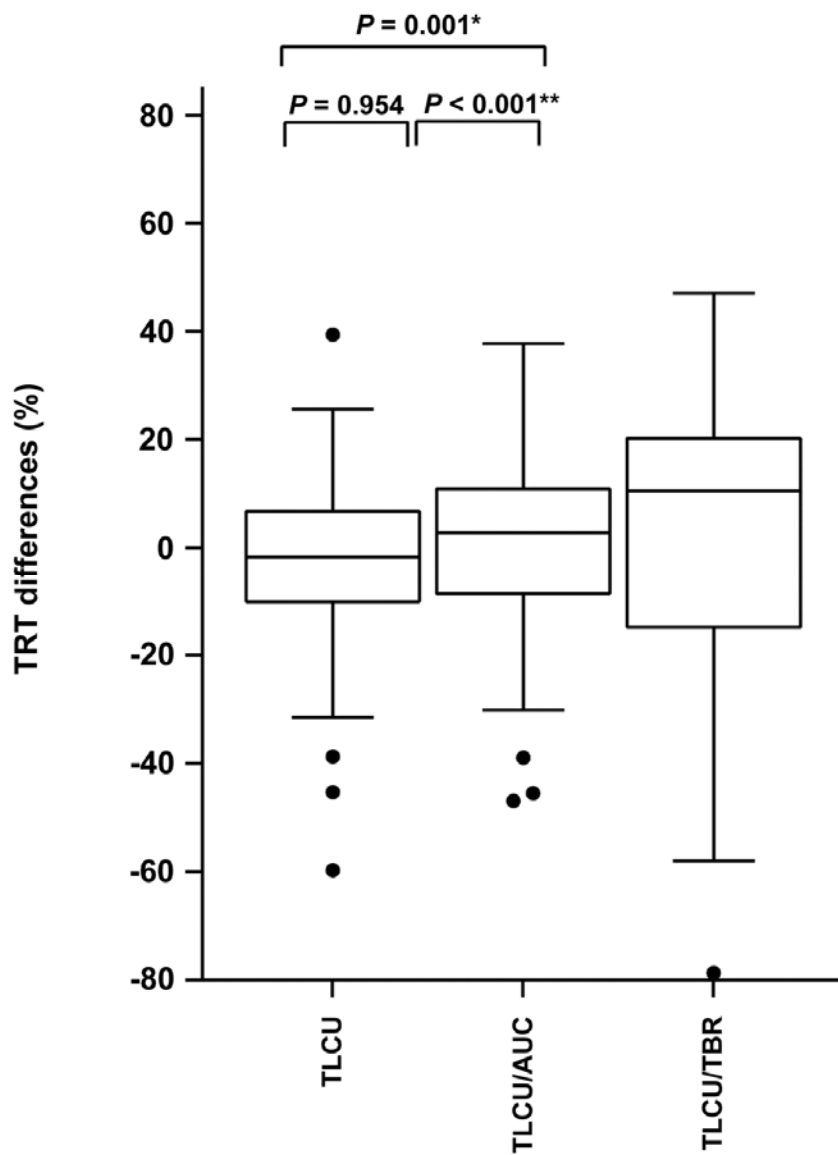
*RC* repeatability coefficients; *SUV* standardized uptake value; *AUC* area under the curve of blood activity concentration; *TBR* blood concentration 40 minutes after injection; *MTV* metabolic tumor volume; *TLCU* total lesion choline uptake.



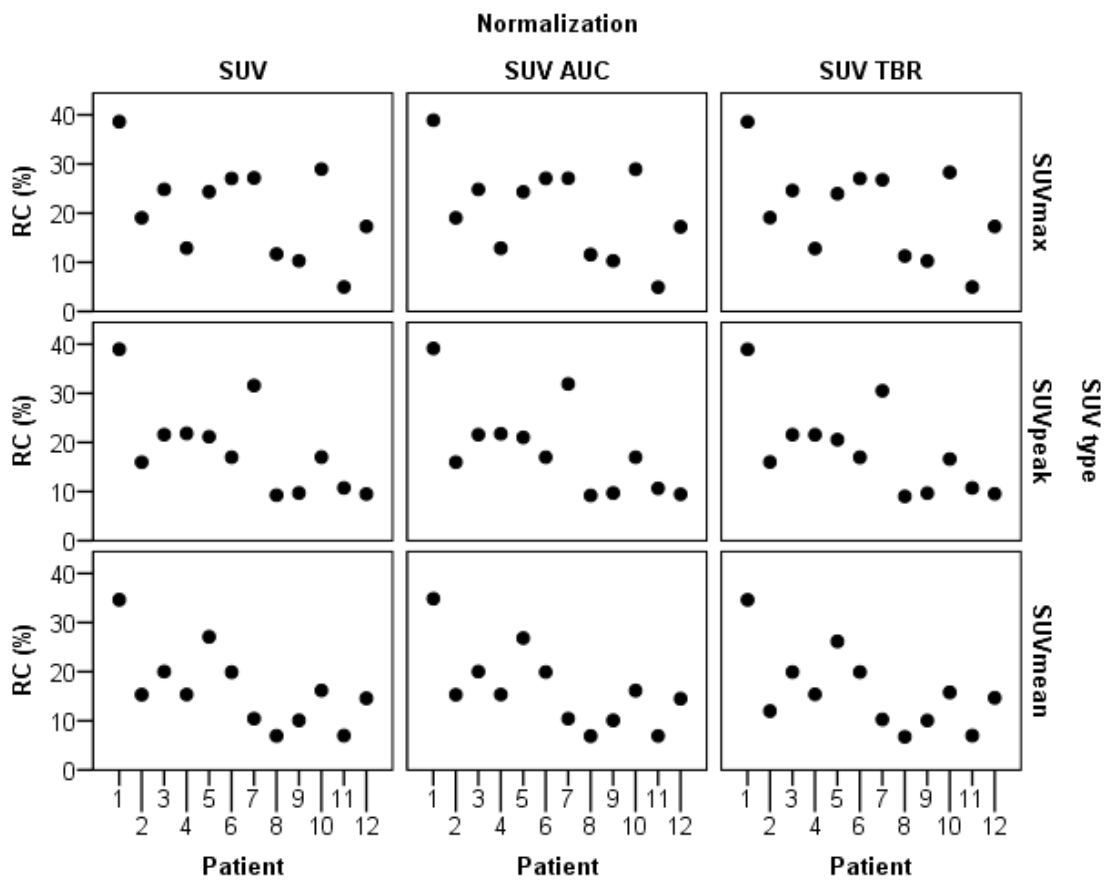
**Supplemental Figure 1.** Relative test-retest repeatability for metabolic volume (left) and SUV<sub>peak</sub> (right), respectively. Black dots represent outliers (more than 2 SD).



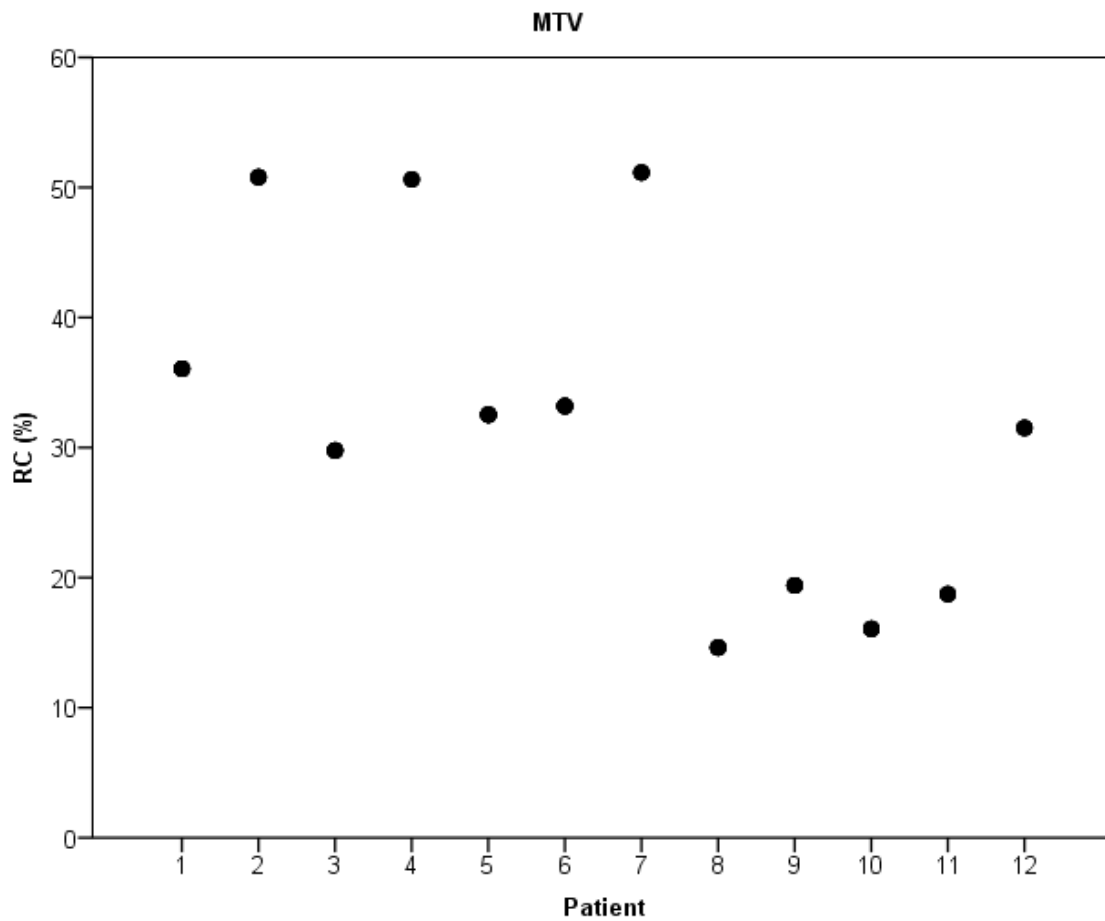
**Supplemental Figure 2.** Relative differences between test-retest data ( $SUV_{peak}$ ) and the type and anatomical localization of the metastatic lesions. Black dots represent outliers (more than 2 SD).



**Supplemental Figure 3.** Relative differences between test-retest data ( $SUV_{mean}$ ) and different types of normalized total lesion choline uptake (TLCU,  $TLCU_{AUC}$  and  $TLCU_{TBR}$ ). Black dots represent outliers (more than 2 SD). \*  $p$ -value significant at 5% level ( $p < 0.0056$ ) and \*\*  $p$ -value significant at 1% level ( $p < 0.0011$ ).

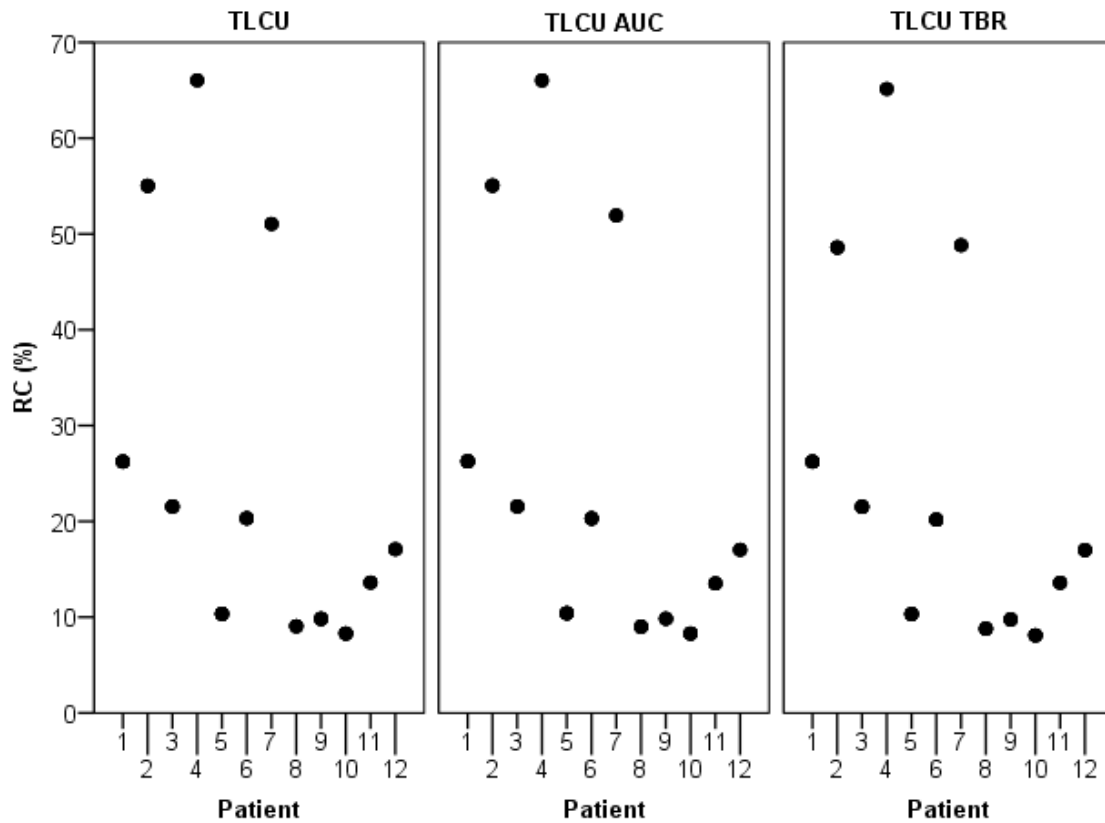


**Supplemental Figure 4.** Repeatability coefficients (RC; in percentage) for the individual patients, regarding the different types of standardized uptake value (SUVmax, SUVpeak and SUVmean) and their normalization (SUV, SUV<sub>AUC</sub> and SUV<sub>TBR</sub>).



**Supplemental Figure 5.** Repeatability coefficients (RC; in percentage) for the individual patients, regarding the metabolic tumor volume (MTV). Comparable RCs across the majority of subjects were observed. However, for 3 patients (2, 4 and 7) both MTV and total lesion choline uptake (TLCU) were larger than those of the other subjects. This is likely explained by difficulty in lesion segmentation, such that errors in MTV are also propagated into poorer repeatability for TLCU.





**Supplemental Figure 6.** Repeatability coefficients (RC; in percentage) for the individual patients, regarding the total lesion choline uptake (TLCU) and corresponding normalizations (TLCU,  $TLCU_{AUC}$  and  $TLCU_{TBR}$ ). Comparable RCs across the majority of subjects were observed. However, for 3 patients (2, 4 and 7) both metabolic tumor volume (MTV) and TLCU were larger than those of the other subjects. This is likely explained by difficulty in lesion segmentation, such that errors in MTV are also propagated into poorer repeatability for TLCU.