Test–Retest Variability in Lesion SUV and Lesion SUR in ^{18}F-FDG PET: An Analysis of Data from Two Prospective Multicenter Trials

Frank Hofheinz[1], Ivayla Apostolova[2], Liane Oehme[3], Jörg Kotzerke[3], and Jörg van den Hoff[1,3]

[1]Helmholtz-Zentrum Dresden-Rossendorf, PET Center, Institute of Radiopharmaceutical Cancer Research, Dresden, Germany; [2]Klinik für Radiologie Nuklearmedizin, Universitätsklinikum Magdeburg A.ö.R., Magdeburg, Germany; and [3]Department of Nuclear Medicine, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany

Quantitative assessment of radio- and chemotherapy response with ^{18}F-FDG whole-body PET has attracted increasing interest in recent years. In most published work, SUV has been used for this purpose. In the context of therapy response assessment, the reliability of lesion SUVs, notably their test–retest stability, thus becomes crucial. However, a recent study demonstrated substantial test–retest variability (TRV) in SUVs. The purpose of the present study was to investigate whether the tumor-to-blood SUV ratio (SUR) can improve TRV in tracer uptake. **Methods:** 73 patients with advanced non–small cell lung cancer from the prospective multicenter trials ACRIN 6678 (n = 34) and MK-0646-008 (n = 39) were included in this study. All patients underwent two ^{18}F-FDG PET/CT investigations on two different days (time difference, 3.6 ± 2.1 d; range, 1–7 d) before therapy. For each patient, up to 7 tumor lesions were evaluated. For each lesion, SUV_{max} and SUV_{peak} were determined. Blood SUV was determined as the mean value of a 3-dimensional aortic region of interest that was delineated on the attenuation CT image and transferred to the PET image. SURs were computed as the ratio of tumor SUV to blood SUV and were uptake time–corrected to 75 min after injection. TRV was quantified as 1.96 multiplied by the root-mean-square deviation of the fractional paired differences in SUV and SUR. The combined effect of blood normalization and uptake time correction was inspected by considering R_{TRV} (TRV_{SUR}/TRV_{SUV}), a ratio reflecting the reduction in the TRV in SUR relative to SUV. R_{TRV} was correlated with the group-averaged-value difference (δ) in CF_{mean} (δCF_{mean}) of the quantity $\delta CF = |CF - 1|$, where CF is the numeric factor that converts individual ratios of paired SUVs into corresponding SURs. This correlation analysis was performed by successively increasing a threshold value δCF_{min} and computing δCF_{mean} and R_{TRV} for the remaining subgroup of patients/lesions with $\delta CF \geq \delta CF_{min}$. **Results:** The group-averaged TRV_{SUV} and TRV_{SUR} were 32.1 and 29.0, respectively, which correspond to a reduction of variability in SUR by an R_{TRV} factor of 0.9 in comparison to SUV. This rather marginal improvement can be understood to be a consequence of the atypically low intrasubject variability in blood SUV and uptake time and the accordingly small δCF values in the investigated prospective study groups. In fact, subgroup analysis with increasing δCF_{min} thresholds revealed a pronounced negative correlation (Spearman $\rho = -0.99$, $P < 0.001$) between R_{TRV} and δCF_{mean}, where $R_{TRV} \approx 0.4$ in the δCF_{min} = 20% subgroup, corresponding to a more than 2-fold reduction of TRV_{SUR} compared with TRV_{SUV}. **Conclusion:** Variability in blood SUV and uptake time has been identified as a causal factor in the TRV in lesion SUV. Therefore, TRV in lesion uptake measurements can be reduced by replacing SUV with SUR as the uptake measure. The improvement becomes substantial for the level of variability in blood SUV and uptake time typically observed in the clinical context.

Key Words: PET; FDG; SUV; SUR; test–retest

J Nucl Med 2017; 58:1770–1775
DOI: 10.2967/jnumed.117.190736

Quantitative assessment of radio- and chemotherapy response with ^{18}F-FDG whole-body PET has attracted increasing interest in recent years (*1–5*). For practical reasons (unavailability of dynamic data, no arterial input function, ease of use) the SUV (tracer concentration normalized to injected dose per kilogram of body weight) has been the method used nearly exclusively for quantification of a lesion's tracer uptake in whole-body investigations. In the context of therapy response assessment, the reliability of lesion SUVs, notably their test–retest stability, thus becomes of crucial importance and has been addressed in several studies (*6–12*).

In a recent study (*13*), Weber et al. demonstrated substantial test–retest variability (TRV) in SUVs even under well-standardized conditions regarding data acquisition and data evaluation. The authors considered several possible causes for the observed variability (body weight, age, clinical stage, blood glucose levels, location and number of lesions), but none of these turned out to be actually operative. The obvious consequence of the apparently unavoidable SUV TRV is that rather high thresholds have to be used to conclude that a true change in tumor tracer uptake related to therapy or disease progression has occurred. Obviously, it would be desirable to reduce the inherent TRV by identifying and correcting at least some of its causes.

In an earlier publication (*14*), we suggested two further factors not mentioned in the study of Weber et al. (*13*) that might explain at least part of the observed variability, namely interscan variation in arterial blood SUV (and the SUV scale of the whole arterial input function) and variability in tracer uptake time before scanning. Although uptake time variability has already been recognized as possibly contributing to SUV TRV (*11*), variability

Received Feb. 7, 2017; revision accepted Apr. 14, 2017.
For correspondence or reprints contact: Frank Hofheinz, PET Center, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstrasse 400, D-01328 Dresden, Germany.
E-mail: f.hofheinz@hzdr.de
Published online May 4, 2017.

in arterial blood SUV does not yet seem to have been considered. More importantly, if our assumption is correct, the recently introduced uptake time–corrected tumor-to-blood SUV ratio (SUR) (*15,16*) would represent a means of quantitatively accounting for these two sources of TRV and should therefore exhibit a lower TRV than SUV (which in turn could explain the superiority of SUR over SUV regarding prognostic value as demonstrated in initial clinical studies (*17–19*)). The purpose of the present study was to compare the TRV of both uptake measures.

MATERIALS AND METHODS

Patient Group and PET Imaging

In this study, 73 patients with advanced (stage III or IV) non–small cell lung cancer were included. All patients underwent two ^{18}F-FDG PET/CT scans on two different days (time difference, 3.6 ± 2.1 d; range, 1–7 d) before therapy. The scan started an average of 61 ± 7 min (range, 48–100 min) after injection. The difference in starting time between the two scans averaged −0.4 ± 3.6 min (range, −11 to 7 min). The included data are part of two prospective multicenter trials conducted by the American College of Radiology Imaging Network (ACRIN 6678, NCT00424138, $n = 34$) and by Merck & Co. Inc. (MK-0646-008, NCT00729742, $n = 39$ of a total of 40; data from one patient were not available). Two examples are shown in Figure 1. PET/CT images were acquired in accordance with National Cancer Institute guidelines (*20*). Further details on the patient groups and the PET imaging have been previously published (*13*). The institutional review board of each participating site approved the study, and all subjects gave written informed consent for future research use of trial data and images as part of the original consent process.

Image Analysis

Region-of-interest (ROI) definition and ROI analyses were performed using ROVER, version 3.0.21 (ABX).

In the PET images, the metabolically active part of the lung lesion with the highest uptake and up to 6 additional lesions (at arbitrary locations in the field of view) were delineated by an automatic algorithm based on adaptive thresholding taking the local background into account (*21,22*). Lesions in the immediate vicinity of the hot bladder and lesions smaller than 1 cm^3 were excluded. Altogether, 236 lesion ROIs were delineated (lung, 162; liver, 14; bone, 37; other locations, 23). For all ROIs, the alignment of PET and attenuation CT was visually inspected. ROIs showing a mismatch between PET and attenuation CT (with substantial parts of the ^{18}F-FDG uptake outside the morphologic lesion boundary as measured in the attenuation CT data) were excluded. This was the case for 21 of 236 ROIs (all of them pulmonary lesions). For the remaining ROIs/lesions ($n = 215$), SUV_{max} and SUV_{peak} were computed. To avoid partial-volume–induced bias, only lesions larger than 1.5 cm^3 were included in SUV_{peak} and SUR_{peak} analysis ($n = 210$).

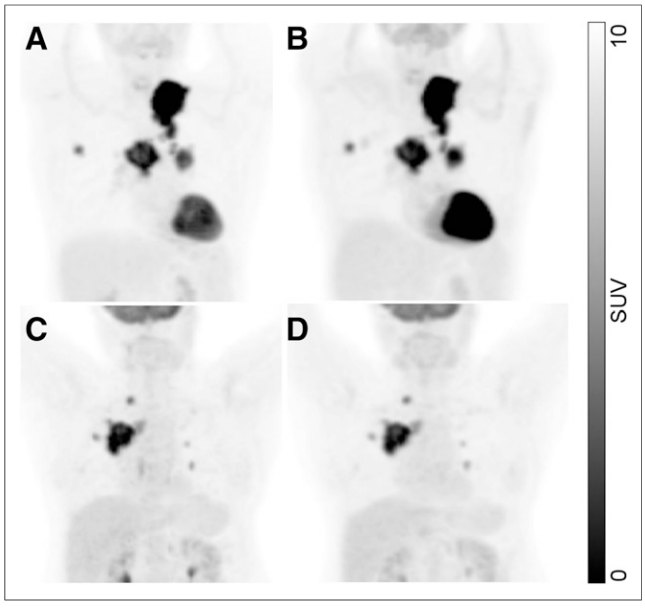

FIGURE 1. Maximum-intensity projections of image data from 2 patients: first scan (A and C) and second scan (B and D).

Arterial blood SUV was determined by defining a roughly cylindric aortic ROI in the attenuation CT data; this ROI was then transferred to the PET data. To reduce partial-volume effects, a concentric safety margin was used in the transaxial planes, centering the ROI in the aorta. Planes showing high tracer uptake near the aorta (pathologic or otherwise) were excluded. The aortic ROI was positioned in the descending aorta, and a minimum ROI volume of 5 cm^3 was ensured. Blood SUV was computed as SUV_{mean} in this aortic ROI.

Lesion SUR was then computed as the uptake-time–corrected ratio of lesion SUV to blood SUV. Uptake time correction to $T_0 = 75$ min after injection was performed as described previously (*16*). $T_0 = 75$ min was chosen as being close to the average actual lesion measurement time, which naturally is somewhat larger than the mean scanning start time of 61 min.

A value of zero was assumed for the apparent volume of distribution (i.e., $V_r = 0$ was used in the correction formula) for reasons discussed previously (*18*). The uptake-time–corrected SUR is then given by

$$\mathrm{SUR} = \frac{T_0}{T} \times \frac{\text{lesion SUV}(T)}{\text{blood SUV}(T)} = \frac{\text{lesion SUV}(T)}{\mathrm{T} \times \text{blood SUV}(T)} \times T_0, \quad \text{Eq. 1}$$

where T is the actual time of measuring the lesion uptake in the respective scan. For each lesion, T was estimated by linear interpolation between the scanning time of the first transaxial plane (T_{start}) and that of the last plane (T_{start} + total scan duration D_{total}), according to

$$T = T_{start} + \frac{S-1}{N-1} \times D_{total}, \quad \text{Eq. 2}$$

where N is the total number of transaxial planes in the image volume and S is the number of the transaxial plane in which the lesion center is located. The worst-case inaccuracy of T is approximately half the acquisition duration per bed position (typically 1–2 min), which is perfectly acceptable for our purposes.

The intersubject stability of blood SUV was described by the SD of the pooled distribution of the blood SUVs from scans 1 and 2. The intrasubject stability of blood SUV was described by the SD of the distribution of paired difference between the second and first scans (Δ blood SUV = blood SUV_2 − blood SUV_1).

Test–retest variability in lesion SUV was assessed considering the distribution of fractional paired δ:

$$\delta\mathrm{SUV} = \frac{\mathrm{SUV}_2 - \mathrm{SUV}_1}{0.5 \times (\mathrm{SUV}_1 + \mathrm{SUV}_2)}. \quad \text{Eq. 3}$$

δSUV was computed for each individual lesion. To derive a quantitative TRV measure, we used the root-mean-square deviation (RMS) of δSUV:

$$RMS_{\delta SUV} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \delta SUV_i^2}, \quad \text{Eq. 4}$$

where N is the number of ROIs/lesions included in the evaluation. For a dataset with zero sample mean, RMS coincides with the sample SD. We chose RMS rather than SD as the variability measure because SD exhibits large statistical errors for small sample sizes. To the extent that the underlying distribution can be approximated by a gaussian with zero mean, RMS is a more accurate estimate of the distribution's true SD than is SD. The TRV in lesion SUV was defined as

$$TRV_{SUV} = 1.96 \times RMS_{\delta SUV}, \quad \text{Eq. 5}$$

and an analogous procedure was used to compute TRV_{SUR}. To the extent that δSUV is gaussian-distributed with mean zero, [−TRV, +TRV] represents the 95% confidence interval of the fractional paired differences.

The combined influence of blood normalization and uptake time correction on the TRV was inspected by considering the ratio $R_{TRV} = TRV_{SUR}/TRV_{SUV}$, which represents the relative change in the TRV in SUR in comparison to that in SUV. R_{TRV} was correlated with the group-averaged-value δCF_{mean} of the quantity $\delta CF = |CF - 1|$ (and tested by Spearman rank correlation), where CF is the numeric factor that converts individual ratios of paired SUVs into corresponding SURs (as is immediately obvious from Eq. 1):

$$CF = \frac{T_1 \times \text{blood SUV}_1}{T_2 \times \text{blood SUV}_2}. \quad \text{Eq. 6}$$

This correlation analysis was performed by successively increasing a threshold value δCF_{min} (starting from zero) and restricting the computation of δCF_{mean} and R_{TRV} to the respective subgroup of patients/lesions with $\delta CF \geq \delta CF_{min}$. Differences in TRV_{SUV} and TRV_{SUR} were tested for significance in all groups and subgroups using a 2-tailed F test.

RESULTS

Group averages of SUV_{max} and SUR_{max} in the first scan were 9.7 ± 5.8 (range, 2.4–54.7) and 6.7 ± 3.8 (range, 1.9–32.7), respectively. Group averages of SUV_{max} and SUR_{max} in the second scan were 9.6 ± 5.9 (range, 2.4–54) and 6.8 ± 4.0 (range, 1.5–35.8), respectively. Paired differences in SUV_{max} and SUR_{max} averaged 0.1 ± 1.7 (range, −7.0–6.6) and −0.1 ± 1.0 (range, −3.8 –2.5), respectively. The group averages did not differ significantly between scans 1 and 2 according to paired Wilcoxon testing (SUV_{max}, $P = 0.87$; SUR_{max}, $P = 0.79$).

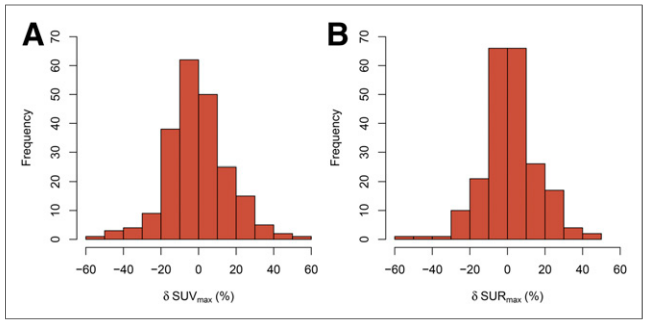

FIGURE 2. Histogram of fractional paired differences in lesion SUV_{max} (A) and lesion SUR_{max} (B).

The mean and SD (i.e., intersubject variation) of blood SUV did not differ significantly in the first and second scans (mean, $P = 0.54$; SD, $P = 0.3$) and was 1.57 ± 0.33 in the pooled data. Intersubject variability was about twice as large as intrasubject variability, that is, the SD of the paired-differences distribution (mean ± SD, 0.03 ± 0.18).

In Figure 2, the histograms of the fractional paired differences δSUV_{max} and δSUR_{max} are shown. Both distributions deviate somewhat from normal distributions (Kolmogorov–Smirnov test). Notably, the δSUR_{max} histogram is somewhat narrower than the δSUV_{max} histogram, and there is a more pronounced peak of small deviations (below ±10%). However, the RMS deviation from zero (the chosen measure of TRV) is only slightly reduced for SUR in comparison to SUV, by a factor R_{TRV} of 0.9. The first row in Table 1 also shows the similar result found when the peak rather than the maximum values of the respective uptake parameter are used.

The small intrasubject variability in blood SUV together with the well-standardized uptake time corresponds to mostly small δCF values (mean ± SD, 7.1% ± 8.0%; 95% confidence interval, 0.4%–35.7% [Fig. 3]). This in turn explains the rather small difference between TRV_{SUV} and TRV_{SUR} in the full study group. On the other hand, subgroup analysis using δCF_{min} values of 5%, 10%, 20%, and 30% as lower thresholds demonstrates that TRV_{SUV} increases with increasing δCF_{min} threshold whereas TRV_{SUR} stays approximately constant (Table 1), ultimately leading to a more than 2-fold reduction ($R_{TRV} \approx 0.4$) in the TRV in TRV_{SUR} compared with TRV_{SUV} in the $\delta CF_{min} = 20\%$ subgroup. This behavior is demonstrated in detail in Figure 4. Figure 4A shows TRV as a function of δCF_{mean} in the respective subgroup (the δCF_{min} threshold was successively set to all values in the sorted list of δCF values actually occurring in the data). Figure

TABLE 1
Percentage TRV

		Maximum					Peak				
δCF_{min}	δCF_{mean}	*n*	TRV_{SUV}	TRV_{SUR}	R_{TRV}	*P*	*n*	TRV_{SUV}	TRV_{SUR}	R_{TRV}	*P*
All data	7.1%	215	32.1	29.0	0.90	0.060	210	33.6	30.7	0.91	0.085
5%	12.5%	102	35.2	28.8	0.82	0.016	99	36.1	29.6	0.82	0.023
10%	19.6%	43	43.2	30.3	0.70	0.007	40	42.1	29.0	0.69	0.01
20%	34.7%	10	61.6	24.5	0.40	0.008	8	57.4	24.0	0.42	0.005
30%	41.8%	6	65.5	25.4	0.39	0.54	4	60.0	24.0	0.40	0.53

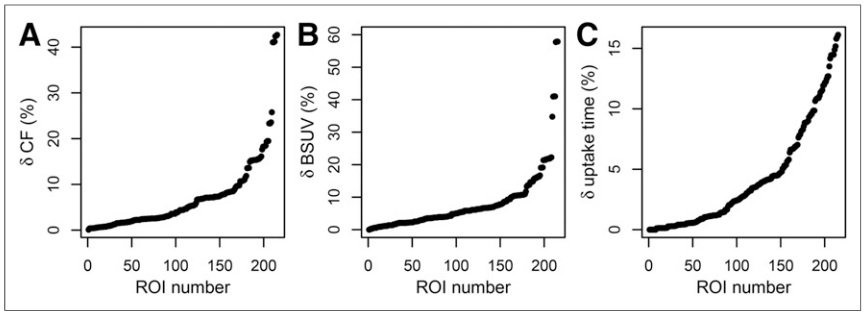

FIGURE 3. (A) Deviation from unity, δCF = |CF − 1|, of SUV-to-SUR conversion factor (CF) for all ROIs/lesions. (B and C) Fractional difference in blood SUV (BSUV) (B) and uptake time (C) between first and second scans.

4B demonstrates the pronounced correlation between R_{TRV} (the reduction in TRV in SUR relative to SUV) and δCF_{mean} (Spearman $\rho = -0.99$, $P < 0.001$).

DISCUSSION

In this work, we investigated whether the TRV in ^{18}F-FDG uptake measurements in tumor lesions can be reduced when SUR rather than SUV is used as the quantitative uptake measure. Our investigation had two major findings.

In the group as a whole, SUR exhibited a slightly reduced TRV in comparison to SUV (by a factor of 0.9 [Table 1, first row]). This finding can be understood to be a direct consequence of an overall unexpectedly low intrasubject variation in blood SUV between the two scans in combination with a well-standardized uptake time T in this patient group. This led to a conversion factor CF—relating the SUV retest–test ratios to the respective SURs—that, on average, did not deviate much from unity.

On the other hand, in subgroups of patients for whom the change in the factor blood SUV(T) × T between the two scans was larger (corresponding to larger values of δCF), SUR exhibited increasingly better test–retest stability than SUV: the magnitude of the improvement (described by R_{TRV}) correlated strongly with the threshold δCF_{min} chosen for subgroup selection (and also with the subgroup average δCF_{mean}) (Table 1; Fig. 4). The improved test–retest behavior of SUR might ultimately be explained by the fact that SUR is a much better surrogate than SUV for the metabolic rate of ^{18}F-FDG accumulation in the tumor, as was demonstrated previously (*15,18*).

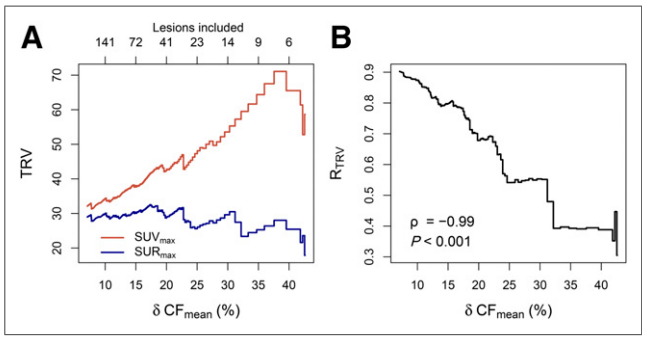

FIGURE 4. TRV_{SUV} and TRV_{SUR} (A) and R_{TRV} (B) achieved with SUR as function of δCF_{mean} in remaining subgroup (subgroup sizes indicated at top of A).

Regarding the extent of the observed SUV TRV, our results are in full agreement with those of Weber et al. (*13*), who analyzed the same study group (but obviously not exactly the same set of lesions): the 95% confidence interval of the "repeatability coefficient" as defined by Weber et al. is −29% to 40% for SUV_{max} in the present study, which is close to the −28% to 39% reported by Weber et al. The agreement extends to the observation that there was no notable difference in the TRV in SUV_{max} and SUV_{peak}, as well as confirming the finding that no relevant difference between ROI-based and patient-averaged evaluation could be detected (data not shown). In another study, Kramer et al. investigated a group of 11 non–small cell lung cancer patients (*12*) who also underwent two PET investigations on two different days without intervening therapy. The investigators reported a repeatability coefficient (defined as 1.96 × SD of δSUV) of less than 15% for SUV_{max} when the evaluation was restricted to lesions that satisfied the PERCIST criterion (*23*) and of less than 10% when per-patient averages over these lesions were used. The apparent contradiction with our corresponding result (repeatability coefficient, 32.2) is resolved by noting that inclusion of all evaluated lesions in the study of Kramer et al. (*12*) leads to a repeatability coefficient of 26.6. The remaining small difference might be related to the fact that the data of Kramer et al. (*12*) were acquired with a single scanner rather than in a multicenter setting.

Although the present work has identified the combined influence of variable blood SUV and variable uptake time before scanning as a causal factor in the observed substantial TRV in lesion SUV, a switch from SUV-based to SUR-based evaluation would not be of much practical relevance if the group-averaged results obtained in the present investigation were directly applicable to the clinical setting. However, the present study group was quite atypical of the usual clinical situation.

For one, the remarkably low intrasubject variation in blood SUV (SD, 0.18) in the present study was not in accord with two previous investigations (one from our group (*24*) and the other from Boktor et al. (*25*)), both of which found much larger intrasubject variations (SDs of 0.32 and 0.42, respectively). On the other hand, the observed intersubject variability in blood SUV was in complete agreement between the present and the two former investigations (SDs of 0.33 vs. 0.36 and 0.38).

A possible reason for reduced intrasubject blood SUV variability might be that in contrast to the two former investigations (*24,25*), no therapeutic intervention took place between the two scans in the present study. It clearly is conceivable that therapeutic intervention affects and modifies systemic ^{18}F-FDG kinetics, and we hypothesize that this was the cause of the higher intrasubject blood SUV variability in the previous studies. If this conjecture turns out to be correct, it would imply that lesion SUV TRV (or, more precisely, spurious contributions to an observed SUV change that are unrelated to a real [treatment-related] effect) must be expected to be higher during therapy response assessment than was the case in the present study group. At

the same time, SUR variability should be unaltered compared with the present study since the SUR approach corrects for blood SUV variability.

The present study group was also atypical of the usual clinical situation in that uptake time differences between paired scans were small. In contrast, uptake times in clinical routine can vary substantially between scans, as has been demonstrated repeatedly (e.g., 64 ± 14 min [range, 20–90 min] (*26*); 69 ± 25 min [range, 21–143 min] (*27*)). This, too, will increase the TRV in SUV in the clinical setting compared with the results of the present investigation, and the uptake time correction included into SUR computation could be expected to account for this effect as well.

To estimate the typical magnitude of δCF to be expected in clinical routine, we reanalyzed our previously published data (*24*), where the variability in uptake time was approximately as reported in two previous publications (*26,27*) and the variability in blood SUV was comparable to blood SUV as reported in two other previous publications (*24,25*). This analysis yielded a δCF_{mean} of 21.4% ± 19.1%, with a 95% confidence interval of 1.0%–58.0%, instead of the δCF_{mean} of 7.1% ± 8% and 95% confidence interval of 0.4%–35.7% found in the present study group. The SUV TRV of about 30% observed in the present and previous studies in our view thus has to be considered a best case that is not representative of the clinical situation (in which distinctly larger spurious variations in SUV will occur).

Regarding SUR, on the other hand, our results support the notion that its use does eliminate (or distinctly reduce) the adverse influence of blood SUV and uptake time variations on the test–retest stability of SUVs. Formally, this means that SUR compensates for two sources of systematic errors when true tracer uptake changes are being assessed. One could of course hypothesize that the concomitant increase in statistical error when SUR rather than SUV is used (caused by the residual uncertainties in the blood SUV and T entering the SUR computation in Eq. 1) exceeds the reduction of systematic errors. However, our results (both in previous investigations (*17,18*) and in the present one) demonstrate that this is not the case: in comparison to SUV, the statistical error of SUR is only modestly increased but the systematic error is much more pronouncedly reduced.

A limitation of this study was the decreasing size of the subgroups selected via increasing δCF_{min} thresholds (e.g., for only 10 lesions was δCF larger than 20%). Consequently, statistical evaluation at sufficiently high δCF_{min} thresholds (and the corresponding δCF_{mean} values) becomes increasingly unreliable. Nevertheless, we believe our results convincingly demonstrate a monotonous increase in TRV_{SUV} as a function of δCF_{mean} and a constant value of TRV_{SUR} over the whole range of δCF_{mean} (Fig. 4A). Such behavior can be interpreted as demonstrating the increasing influence of blood $SUV(T) \times T$ changes between successive scans on TRV_{SUV} whereas TRV_{SUR} remains unaffected. The constant (blood $SUV(T) \times T$ changes–independent) level of TRV_{SUR} thus would represent a residual variability (of about 25%–30%) caused by unidentified other factors. It would be desirable to confirm our findings in further investigations and to perform studies, prospective as well as retrospective, evaluating the prognostic value of SUR in comparison to SUV in different oncologic applications.

CONCLUSION

Variability in blood SUV and uptake time has been identified as a causal factor in TRV in lesion SUV. Therefore, TRV in lesion uptake measurements can be reduced by replacing SUV with SUR as the uptake measure. The improvement can be expected to be substantial at the level of variability in blood SUV and uptake time typically observed in the clinical context. Further studies will be necessary to investigate whether this improved test–retest behavior translates into an improved prognostic value for SUR in comparison to SUV.

DISCLOSURE

The image data used in this study were collected through funding support from the National Cancer Institute through the grants U01 CA079778 and U01 CA080098. No other potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENT

We thank the American College of Radiology Imaging Network (ACRIN) for providing the image data used in this study.

REFERENCES

1. Rebollo-Aguirre ÁC, Ramos-Font C, Villegas Portero R, Cook GJ, Llamas Elvira JM, Romero Tabares A. Is FDG-PET suitable for evaluating neoadjuvant therapy in non-small cell lung cancer? Evidence with systematic review of the literature. *J Surg Oncol.* 2010;101:486–494.
2. Skoura E, Datseris IE, Platis I, Oikonomopoulos G, Syrigos KN. Role of positron emission tomography in the early prediction of response to chemotherapy in patients with non–small-cell lung cancer. *Clin Lung Cancer.* 2012;13:181–187.
3. Plaxton NA, Brandon DC, Corey AS, et al. Characteristics and limitations of FDG PET/CT for imaging of squamous cell carcinoma of the head and neck: a comprehensive review of anatomy, metastatic pathways, and image findings. *AJR.* 2015;205:W519–W531.
4. Jeraj R, Bradshaw T. Simončič U. Molecular imaging to plan radiotherapy and evaluate its efficacy. *J Nucl Med.* 2015;56:1752–1765.
5. Coyle M, Kostakoglu L, Evens AM. The evolving role of response-adapted PET imaging in Hodgkin lymphoma. *Ther Adv Hematol.* 2016;7:108–125.
6. Minn H, Zasadny KR, Quint LE, Wahl RL. Lung cancer: reproducibility of quantitative measurements for evaluating 2-[F-18]-fluoro-2-deoxy-D-glucose uptake at PET. *Radiology.* 1995;196:167–173.
7. Nahmias C, Wahl LM. Reproducibility of standardized uptake value measurements determined by ^{18}F-FDG PET in malignant tumors. *J Nucl Med.* 2008;49:1804–1808.
8. Kumar V, Nath K, Berman CG, et al. Variance of standardized uptake values for FDG-PET/CT greater in clinical practice than under ideal study settings. *Clin Nucl Med.* 2013;38:175–182.
9. Rockall AG, Avril N, Lam R, et al. Repeatability of quantitative FDG-PET/CT and contrast-enhanced CT in recurrent ovarian carcinoma: test–retest measurements for tumor FDG uptake, diameter, and volume. *Clin Cancer Res.* 2014;20:2751–2760.
10. Rasmussen JH, Fischer BM, Aznar MC, et al. Reproducibility of ^{18}F-FDG PET uptake measurements in head and neck squamous cell carcinoma on both PET/CT and PET/MR. *Br J Radiol.* 2015;88:20140655.
11. Yuan H, Tong DKH, Vardhanabhuti V, Khong PL. Factors that affect PERCIST-defined test-retest comparability: an exploration of feasibility in routine clinical practice. *Clin Nucl Med.* 2015;40:941–944.
12. Kramer GM, Frings V, Hoetjes N, et al. Repeatability of quantitative whole body ^{18}F-FDG PET/CT uptake measures as function of uptake interval and lesion selection in non–small cell lung cancer patients. *J Nucl Med.* 2016;57:1343–1349.
13. Weber WA, Gatsonis CA, Mozley PD, et al. Repeatability of ^{18}F-FDG PET/CT in advanced non–small cell lung cancer: prospective assessment in 2 multicenter trials. *J Nucl Med.* 2015;56:1137–1143.

14. van den Hoff J, Hofheinz F. Repeatability of tumor SUV quantification: the role of variable blood SUV. *J Nucl Med.* 2015;56:1635–1636.
15. van den Hoff J, Oehme L, Schramm G, et al. The PET-derived tumor-to-blood standard uptake ratio (SUR) is superior to tumor SUV as a surrogate parameter of the metabolic rate of FDG. *EJNMMI Res.* 2013;3:77.
16. van den Hoff J, Lougovski A, Schramm G, et al. Correction of scan time dependence of standard uptake values in oncological PET. *EJNMMI Res.* 2014;4:18.
17. Bütof R, Hofheinz F, Zöphel K, et al. Prognostic value of pretherapeutic tumor-to-blood standardized uptake ratio in patients with esophageal carcinoma. *J Nucl Med.* 2015;56:1150–1156.
18. Hofheinz F, van den Hoff J, Steffen IG, et al. Comparative evaluation of SUV, tumor-to-blood standard uptake ratio (SUR), and dual time point measurements for assessment of the metabolic uptake rate in FDG PET. *EJNMMI Res.* 2016;6:53.
19. Shin S, Pak K, Kim IJ, Kim BS, Kim SJ. Prognostic value of tumor-to-blood standardized uptake ratio in patients with resectable non-small-cell lung cancer. *Nucl Med Mol Imaging.* November 7, 2016 [Epub ahead of print].
20. Shankar LK, Hoffman JM, 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.
21. Hofheinz F, Pötzsch C, Oehme L, et al. Automatic volume delineation in oncological PET: evaluation of a dedicated software tool and comparison with manual delineation in clinical data sets. *Nuklearmedizin.* 2012;51:9–16.
22. Hofheinz F, Langner J, Petr J, et al. An automatic method for accurate volume delineation of heterogeneous tumors in PET. *Med Phys.* 2013;40:082503.
23. Wahl RL, Jacene H, Kasamon Y, Lodge M. From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. *J Nucl Med.* 2009;50 (suppl):122S–150S.
24. Hofheinz F, Bütof R, Apostolova I, et al. An investigation of the relation between tumor-to-liver ratio (TLR) and tumor-to-blood standard uptake ratio (SUR) in oncological FDG PET. *EJNMMI Res.* 2016;6:19.
25. Boktor RR, Walker G, Stacey R, Gledhill S, Pitman AG. Reference range for intrapatient variability in blood-pool and liver SUV for ^{18}F-FDG PET. *J Nucl Med.* 2013;54:677–682.
26. Beyer T, Czernin J, Freudenberg LS. Variations in clinical PET/CT operations: results of an international survey of active PET/CT users. *J Nucl Med.* 2011;52:303–310.
27. Tahari AK, Wahl RL. Quantitative FDG PET/CT in the community: experience from interpretation of outside oncologic PET/CT exams in referred cancer patients. *J Med Imaging Radiat Oncol.* 2014;58:183–188.