Repeatability of Quantitative Whole-Body ^{18}F-FDG PET/CT Uptake Measures in Patients with Non–Small Cell Lung Cancer: Dynamic Versus Test–Retest Design

TO THE EDITOR: In the September issue of *The Journal of Nuclear Medicine,* Kramer et al. comprehensively investigated the repeatability (R) of various quantitative ^{18}F-FDG uptake metrics in lung cancer patients, including SUV_{max}, SUV_{mean}, SUV_{peak}, metabolically active tumor volume (MATV), and total lesion glycolysis (TLG) (*1*). A test–retest study was performed within 3 d involving double-baseline whole-body ^{18}F-FDG PET/CT at 60 and 90 min after injection. The results were compared with those of some previously published studies on various oncologic diseases, such as ovarian cancer, non–small cell lung cancer, esophageal cancer, and liver metastases for TLG in particular.

However, a previous study on TLG variability in lung cancer patients—a study involving SUV_{mean} and MATV—was not included in the comparison of Kramer et al. (*2*). That study assessed R differently; instead of a test–retest acquisition within 3 d over a total of 60 lesions, as used by Kramer et al., a dynamic acquisition involving 10 frames within 60–110 min after injection over 13 lesions was used. Moreover, SUV_{peak} R was obtained from a further dynamic study involving 20 lung cancer lesions (*3*). We thought it would of interest to compare the two methods—that is, dynamic and test–retest.

The values for dynamic R obtained within 60–110 min after injection versus test–retest R obtained at 90 min after injection for all lesions in the study of Kramer et al. (R is reproducibility coefficient in their Table 3) are as follows: 19.6 versus 23.3%, 14.1 versus 17.8%, 13.2 versus 15.8%, 31.6 versus 23.7%, and 36.4 versus 30.7% for SUV_{max}, SUV_{mean}, SUV_{peak}, TLG, and MATV (95% reliability), respectively (*1–3*). Although the parameter range and the "50% of SUV_{max}" thresholding method (used for assessing SUV_{mean}, MATV, and hence TLG) were not exactly alike, the dynamic R percentages are consistent with the test–retest R percentages, as much as can be determined considering that R estimate uncertainty was not provided.

Let us further compare the two designs. The dynamic design involved ten 2.5-min frames leading to a ±12.5-min time window around a mean uptake time, in comparison with the test–retest design, which reported a maximal range of 7 min for uptake time (scan 1 in Table 1 (*1*)). We suggest that reducing the acquisition time for bed positions—that is, less than the 2.5 min that is possible with modern PET/CT systems—and reducing the number of dynamic frames may bridge the two designs. In the framework of assessing response to treatment, it is noteworthy that the dynamic design does not take into account some origins of parameter variability such as changes in plasma glucose level (within the reference range), injected dose, and differences in uptake time. However, Kramer et al. reported that glucose correction does not improve R performance (and even deteriorates it) and that the relative uncertainty about the injected dose is usually very low (*1*). Regarding differences in uptake time, Kramer et al. showed that the correction proposed by van den Hoff et al. significantly reduced differences between 60-min and 90-min data (*4*). We would like to emphasize, first, that the time window of the dynamic design includes a variability of ±12.5 min around a mean uptake time, which may take into account differences in uptake time usually met in current clinical practice, and second, that simply removing the ^{18}F physical decay correction can reduce differences between 60-min and 90-min SUVs (*5*). Finally, in comparison with the test–retest design, the dynamic design involving several frames reduces the number of lesions to be investigated (and hence the number of patients to be recruited) in order to reliably determine R, since for the same number of lesions, the greater the number of dynamic frames the lower the R estimate uncertainty. We thus suggest that the dynamic design takes into consideration both the patient radiation dose and a busy clinical practice.

In conclusion, the comprehensive study of Kramer et al. about the R value for various ^{18}F-FDG uptake measures will be useful to nuclear physicians in their current practice (*1*). That study was achieved using a test–retest design, and we would like to emphasize that the tool box for assessing measurement uncertainty in quantitative PET imaging fortunately offers various designs, each with its own pros and cons. An alternative dynamic design is available that may be particularly suitable when the role of technical parameters in this uncertainty is investigated in an arbitrary PET/CT system. Nevertheless, whatever the design, it clearly appears that guidelines should recommend that any quantitative outcome be accompanied by its measurement uncertainty, which should be specifically determined for each PET/CT system as soon as commissioning.

REFERENCES

1. 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.
2. Laffon E, de Clermont H, Lamare F, Marthan R. Variability of total lesion glycolysis by ^{18}F-FDG-positive tissue thresholding in lung lesion cancer. *J Nucl Med Technol.* 2013;41:186–191.
3. Laffon E, Burger IA, Lamare F, de Clermont H, Marthan R. SUVpeak performance in lung cancer: comparison to average SUV from 40 hottest voxels. *J Nucl Med.* 2016;57:85–88.
4. 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.
5. Laffon E, de Clermont H, Marthan R. A method of adjusting SUV for injection-acquisition time differences in ^{18}F-FDG PET imaging. *Eur Radiol.* 2011;21:2417–2424.

Eric Laffon*
Roger Marthan
[*]Hôpital du Haut-Lévèque
avenue de Magellan
33604 Pessac, France
E-mail: elaffon@u-bordeaux2.fr

Published online Dec. 8, 2017.
DOI: 10.2967/jnumed.116.186676

REPLY: We thank Drs. Thie, Laffon, and Marthan for their response on our paper (*1*). In this study, we tried to comprehensively investigate the repeatability of various whole-body ^{18}F-FDG uptake metrics and assess the influence of several correction methods to normalize ^{18}F-FDG uptake.

In their letter, Drs. Laffon and Marthan discuss the advantages and disadvantages of a dynamic and test–retest study design for the assessment of ^{18}F-FDG repeatability. We fully agree that measurement uncertainty of all quantitative PET metrics should be

determined. As stated, a dynamic design may be particularly suitable when assessing the role of statistical noise on variability of individual PET/CT systems. Yet, we would like to emphasize that a test–retest study is needed for the assessment of repeatability in a response evaluation setting in which patients are scanned at different occasions. A true test–retest design is the closest approximation of the clinical conditions met during response assessment and includes all sources of variability encountered in clinical practice, such as variability in injected activity, uptake time, physiologic status, patient repositioning, and breathing-induced artifacts. These results can be used to determine thresholds that are able to differentiate metabolic response and progression from intrinsic measurement variability of quantitative uptake metrics after the start of treatment.

We also want to point out that differences in ^{18}F-FDG uptake measures due to variation in uptake time are caused by differences in ^{18}F-FDG kinetics at 60 and 90 min after injection and not physical decay of ^{18}F (*2*). Omitting physical decay correction to correct for differences in uptake time between 2 scans falsely assumes that ^{18}F decay and ^{18}F-FDG kinetics are proportional. This uptake time correction method should therefore not be used in a longitudinal setting because of physiologic variations in ^{18}F-FDG kinetics.

In addition, we assessed the effect of several ^{18}F-FDG uptake normalization methods, including one for glucose correction, on repeatability. In the current cohort, all plasma glucose levels (4.5–7.1 mmol/L) were well within the recommended range and showed a low interscan variability (≤2.2 mmol/L) (*3*). The influence of competing endogenous glucose on ^{18}F-FDG uptake metrics was thus likely to be limited. However, by correcting tumor uptake for glucose correction a potential source of measurement variability is also introduced. This is supported by the finding that the median difference of repeated glucose level measurements in the same patient, using a calibrated device, was 0.2 mmol/L (0–0.8 mmol/L) in this study. We would therefore suggest that glucose correction should not be performed if glucose levels are within the reference range, as also noted by Dr. Thie in his letter. We would like to encourage Dr. Thie and colleagues to study the influence of other (more complex) glucose-correction methods on the repeatability of ^{18}F-FDG uptake metrics in a cohort with a higher variability in plasma glucose levels. This is of particular interest for metastatic diseases because a wide variety of tissues can be affected.

REFERENCES

1. 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.
2. Boellaard R, O'Doherty MJ, Weber WA, et al. FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. *Eur J Nucl Med Mol Imaging.* 2010;37:181–200.
3. Boellaard R, Delgado-Bolton R, Oyen WJ, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging.* 2015; 42:328–354.

Gerbrand Maria Kramer*
Otto S. Hoekstra
Ronald Boellaard
**VU University Medical Center*
P.O. Box 7057
1007 MB Amsterdam, The Netherlands
E-mail: ge.kramer@vumc.nl

Published online Jun. 15, 2017.
DOI: 10.2967/jnumed.117.195461

Regarding "Subjecting Radiologic Imaging to the Linear No-Threshold Hypothesis: A Non Sequitur of Non-Trivial Proportion"

TO THE EDITOR: I applaud the authors of the recent article by Siegel et al., "Subjecting Radiologic Imaging to the Linear No-Threshold Hypothesis: A Non Sequitur of Non-Trivial Proportion" (*1*), and *The Journal of Nuclear Medicine* for publishing this important review in such a prominent journal. The authors condense an enormous amount of scientific data and rigorous interpretation into a relatively small space. I hope this paper will help slow the disheartening impact of radiophobia that is sweeping our country and reducing the quality and use of radiologic imaging and consequently of medical care, as described so clearly in the article. The authors have raised the level of scientific discussion regarding the health effects of radiation. Any rebuttal to their cogent arguments needs to be on that same high scientific level. I think part of the wide acceptance of the linear, no-threshold theory is the unfamiliarity of most people with the widespread biologic phenomenon of the J-shaped curve, namely, that many things that are harmful at high doses are harmless, or even helpful, at low doses. The classic and best-studied example is that of alcohol on all-cause human mortality, nicely summarized in Figure 1 modified from Di Castelnuovo et al. (2), a metaanalysis involving more than 1 million subjects. The data in the red circle in Figure 1A are shown larger in Figure 1B, and clearly demonstrate the strikingly nonlinear relationship between alcohol and mortality at low doses. This corresponds to the area of contention in the radiation–cancer relationship (equivalent to several whole-body CT scans), for which there are no data—or rather, the data at such low radiation doses are so noisy that no reliable signal can be discerned above background. The data are much better for alcohol and show a relationship that could never be predicted from the high-dose data.

REFERENCES

1. Siegel JA, Pennington CW, Sacks B. Subjecting radiologic imaging to the linear no-threshold hypothesis: a non sequitur of non-trivial proportion. *J Nucl Med.* 2017;58:1–6.
2. Di Castelnuovo A, Costanzo S, Bagnardi V, Donati MB, Iacoviello L, de Gaetano G. Alcohol dosing and total mortality in men and women: an updated meta-analysis of 34 prospective studies. *Arch Intern Med.* 2006;166:2437–2445.

Ronald E. Fisher
Houston Methodist Hospital
6565 Fannin St.
Houston, Texas 77030
E-mail: rfisher@bcm.edu

Published online Apr. 27, 2017.
DOI: 10.2967/jnumed.117.193920