Factors That Affect PERCIST-Defined Test-Retest Comparability

An Exploration of Feasibility in Routine Clinical Practice

Hui Yuan, MBBS,* Daniel King Hung Tong, FRACS, MS,† Varut Vardhanabhuti, FRCR,* and Pek-Lan Khong, FRCR, MD*

Purpose: The aim of this study was to evaluate the factors affecting the comparability of ^{18}F-FDG PET/CT scans using the PERSIST criteria for treatment response evaluation in a clinical PET/CT unit.

Patients and Methods: Patients diagnosed with esophageal cancer were assessed for treatment response by comparing 2 ^{18}F-FDG PET/CT scans, at baseline (PET 1) and 1 month after the end of induction chemoradiation (PET 2). According to the PERCIST recommendations, patients with mean SUV normalized by the lean body mass within reference volume of interest that changed less than 0.3 unit and less than 20% were deemed as comparable. Absolute differences of body weight, blood glucose level, activity of ^{18}F-FDG, and uptake time between the 2 scans were computed. Binary logistic regression was used to identify the predictive factors, and receiver operating characteristic curves were used for thresholds. $P < 0.05$ was considered statistically significant.

Results: Sixty-nine subjects were identified. The mean (SD) values at PET 0 and PET 2 were 5.9 (1.04) mmol/L and 6.2 (1.06) mmol/L ($P = 0.013$), 54.6 (10.0 kg) and 53.3 (10.3 kg) ($P = 0.013$), 7.7 (1.3 mCi) and 7.6 (1.5 mCi) ($P = 0.349$), as well as 74.2 (12.4) minutes and 73.0 (12.3) minutes ($P = 0.539$), for blood glucose level, body weight, injected activity, and uptake time, respectively. Seventeen (24.6%) failed to match the PERCIST-defined comparability criteria. Case-based discrepancies (mean [SD]) were 0.76 (0.62) mmol/L, 3.4 (2.9) kg, 0.8 (0.7) mCi, and 11.7 (9.8) minutes for blood glucose, body weight, injected activity, and uptake time, respectively, of which only uptake time significantly affected comparability ($P = 0.046$; odds ratio, 1.06; 95% confidence interval, 1.00–1.12), with a limit of 2.2-minute discrepancy identified as the requirement for 100% comparability.

Conclusions: Uptake time had the strongest effect on PERCIST-defined comparability. Therefore, for response assessment scans, reference to initial scans for determination of optimal uptake time is recommended.

Key Words: comparability, ^{18}F-FDG PET/CT, response evaluation, PERCIST

P ET/CT imaged with ^{18}F-FDG is widely utilized for the clinical management of a variety of malignancies.[1–4] Because neoadjuvant treatment is currently used as the de facto standard treatment in an increasing number of cancers, a procedure to accurately evaluate the treatment response quantitatively, especially in the early phase, is imperative, so that response-adaptive treatment may be tailored to the individual.[5–10] In view of this, PET/CT, with its ability to functionally detect metabolic changes before anatomical response occurs,[11] is a feasible method to evaluate early treatment response.[11] However, PET/CT-based functional quantitative parameters are dependent on multiple factors, such as blood glucose level, injected dose, and uptake time, and so on.[12,13] As a result, fundamental standardization is needed to ensure the necessary reproducibility so as to obtain comparable quantitative parameters between scans.[14,15]

To date, 2 criteria have been recommended for quantitative PET/CT-based response evaluation, the European Organization for Research and Treatment of Cancer Recommendations criteria[16] and the PET Response Criteria in Solid Tumors 1.0 (PERCIST) criteria.[17] Both of them require standardization of technical factors, that is, injected dose, uptake time, scan time, and glucose level, to assure comparability for the assessment of SUV normalized by the lean body mass (SUL) values at designated lesions. For any trial, the comparability should be verified before response evaluation. To achieve this comparability, ideally, there should be zero discrepancy in these parameters between scans. However, this may not be easily implemented or controllable in clinical practice.[18] Practically, a case with change in mean SUL in the reference volume of interests (VOIs) placed in the liver (3-cm sphere in right lobe, when liver is disease free) or descending aorta (1×2 cm cylinder, when liver is involved) of less than 20% (and 0.3 SUL mean units) at baseline and follow-up scans is still considered reasonably comparable, rendering minor discrepancies in the technical parameters acceptable for the response assessment.[17] In light of this, we aim to evaluate the effect of the discrepancies of these factors on comparability. Thus, the extent of discrepancy that can be allowed to produce adequate comparability can be determined in a clinical practice.

PATIENTS AND METHODS

Patient Recruitment

This is a retrospective review of consecutive patients with pathologically proven esophageal cancer who underwent baseline (PET 1) and response assessment (PET 2) scans using standardized treatment protocol between 2008 and 2013 in a single center. Because of the fact that only anonymized technical data were investigated, the need for written informed consent for data acquisition was approved by the institutional review board to be waived. Patients with blood glucose level higher than 10 mmol/L were excluded from this cohort.

PET/CT Examinations

All scans were performed with the same scanner model (64-MDCT, Discovery VCT; GE Healthcare, Bio-Sciences Corp, New Jersey). ^{18}F-FDG was administrated intravenously with an activity adjusted by body weight (0.13 mCi/kg) after a minimal fasting of 6 hours on the condition that the blood glucose level was tested to be lower than 10 mmol/L. Patients were fully hydrated with

TABLE 1. Mean Values of Explored Factors at PET 1 and PET 2

Factors	Mean (SD) at PET 1	Mean (SD) at PET 2	*P*
Blood glucose level, mmol/L	5.9 (1.04)	6.2 (1.06)	0.013*
Body weight, kg	54.6 (10.0)	53.3 (10.3)	0.013*
Injected activity, mCi	7.7 (1.3)	7.6 (1.5)	0.349
Uptake time, min	74.2 (12.4)	73.0 (12.3)	0.539
SUL_{ref}	1.8 (0.3)	1.8 (0.3)	0.646

*Statistical significance ($P < 0.05$).

500 mL water unless physically intolerant, for example, from dysphagia, during the 60-minute uptake time. Each acquisition with coverage from the base of skull to the upper one third of thighs comprised a CT scan (120 kVp; 200–400 mA; 0.5 seconds per CT rotation; pitch, 0.984:1; 2.5 mm intervals; with CT contrast medium injected at the dose of 1.5 mL/kg when necessary) followed by a PET scan (2 minutes and 30 seconds per bed position and 6 bed positions per case) in the hybrid scanner. Ordered-subset expectation maximization iterative reconstruction algorithm (14 subsets and 2 iterations) was used for CT-based attenuation correction for PET images, and the resultant images were fused with CT images for subsequent viewing (Advanced Workstation ADW 4.3; GE Healthcare Bio-Sciences Corp).

Image Analyzing and Statistical Analyzing

SUVs were standardized by SUL. Liver function status was identified according to the clinical records and imaging records. Spherical VOIs with diameter of 3 cm were placed in the right lobe of the liver. For subjects with metastasis in the liver (n = 1), cylindrical VOIs were placed in the descending aorta (1 × 2 cm). The mean value and SD of the SUL within these VOIs were recorded as SUL_{ref} and SD_{ref}, respectively. For response evaluation, only those with a discrepancy of SUL_{ref} less than 0.3 unit and less than 20% were considered as "comparable."

Blood glucose level, injection time, body weight, and tracer activities were recorded from the patient's notes. Time of acquisition was read from the workstation directly. Uptake time was defined as the time interval between time of injection and acquisition. The mean values at PET 1 and PET 2 of these aforementioned factors were calculated and compared using paired *t* test. Case-based discrepancies in blood glucose level, body weight, tracer activity, and uptake time between PET 1 and PET 2 were calculated. Their absolute values were taken as continuous variables, and their contribution toward PERCIST defined "comparability" were consequently analyzed using univariable binary logistic regression, respectively. Predictive discrepancies were further analyzed with multivariable binary logistic regression. Receiver operating characteristic curves were used to identify desirable thresholds for predictive factors for the identification of maximal comparable cases while allowing affordable clinical tolerance. All aforementioned statistical analyses were performed using SPSS (version 21; IBM SPSS statistics, IBM).

RESULTS

A total of 69 subjects were eligible, comprising 12 females and 57 males. The median age was 66 years (range, 21–78 years; SD, 11.2 years). Technical factors, such as the blood glucose level, body weight, injected activity, and uptake time at PET 1 and PET 2 were tabulated in Table 1. The mean (SD) SUL_{ref} values at 2 time points were 1.8 (0.3) and 1.8 (0.3) ($P = 0.646$).

Among the 69 subjects, the scans of 17 subjects (24.3%) were not eligible for response evaluation based on the PERCIST standard. The mean absolute values of case-based discrepancies between PET 1 and PET 2 for these 4 technical factors were tabulated in Table 2. They were included as variables in univariate analysis using logistic regression. Except for uptake time ($P = 0.046$; odds ratio, 1.06; 95% confidence interval, 1.00–1.12), none of the explored factors was statistically significant to affect the comparability (detailed values in Table 2). Multivariable logistic analysis involving those 4 factors generated a similar result, with only uptake time included in the equation ($P = 0.046$; odds ratio, 1.06; 95% confidence interval, 1.00–1.12), whereas others were excluded from the equation (forward: conditional method).

Receiver operating characteristic curve was used for the purpose of finding the threshold for the comparability (Fig. 1). Exclusion rates of samples with differences of uptake time higher or less than a specific threshold were also plotted, with the corresponding cumulative rates of exclusion also tabulated (Table 3). Accordingly, 2.2 minutes was determined as the threshold in uptake time discrepancy to achieve 100% comparability.

DISCUSSION

The aims of this research were to identify the factors that have the strongest impact on comparability for PERCIST-based response evaluation in routine clinical practice and to determine the impact of the discrepancies in these factors on comparability. We have identified that uptake time is the factor with the strongest effect on comparability in PERCIST-based response evaluation in our cohort, which is consistent with the study by Kuruva et al[19] who found the SUL_{mean} of the liver in their cohort to be solely dependent on uptake time and independent on individual variances including sex, age, and blood glucose level.

The SUL of the liver and blood pool will decrease, whereas in tumors, it will generally increase over time after injection of tracers; thus, FDG uptake is time dependent.[17,20] Therefore, standardization of uptake time is of key importance for PERCIST-based evaluation. To acquire 100% comparability, discrepancy of up to only 2.2 minutes in each subject is allowed, which is a stringent requirement in clinical practice. The results of our study also illustrate

TABLE 2. Case-Based Discrepancies and Their Corresponding Odds Ratios

Factors	Case-Based Discrepancies, Mean (SD)	Odds Ratio (95% CI)	*P*
Blood glucose level, mol/L	0.76 (0.62)	1.12 (0.47–2.68)	0.798
Body weight, kg	3.4 (2.9)	1.13 (0.94–1.36)	0.194
Injected activity, mCi	0.8 (0.7)	1.00 (0.98–1.02)	0.993
Uptake time, min	11.7 (9.8)	1.06 (1.00–1.12)	0.046*

*Statistical significance ($P < 0.05$).
CI, confidence interval.

the expected percentage of comparability based on the degree of discrepancy in uptake time, and therefore in the planning of prospective studies, this may be taken into account depending on the set level of accepted dropout rate balanced with feasibility in the clinical setting.

On the other hand, our result does not necessarily mean that the other factors will not affect the SUL. The reason why there were no statistically predictive factors could be attributed to the fact that, in our cohort, the routine clinical control for these factors is stringent enough. Nevertheless, our results suggest that the impact of uptake time is the largest in the scenario of routine clinical practice.

With regard to blood glucose level, it is known that this factor affects the glucose metabolism in organs, such as brain, muscle, liver, and adipose tissue, although not linearly.[21] Again, uptake time was found to play vital role: when uptake time was set at 60 (10) minutes, a coefficient correlation of 0.25 was reported between SUL_{ref} and blood glucose level; however, if the uptake time was increased to 90 (10) minutes, the correlation became substantially stronger, with a coefficient correlation of 0.73, indicating that liver uptake might be better correlated with blood glucose only when the uptake time is prolonged.[22] In our cohort, scans were performed with uptake time of 60 to 85 minutes, correlations between liver uptake and blood glucose level might not be strong enough to produce statistical significance. Plus, in our clinical routine, a cutoff level of blood glucose level was applied, and consequently, blood glucose level variance in our cohort was well controlled, resulting in a statistically undiscernible effect on SUL_{ref}. It may be possible that this factor will become significant in a substantially larger cohort or if the variance in blood glucose is large with longer uptake times adopted. Nevertheless, our findings support the fact that blood glucose variation in a typical clinical cohort is narrow enough to meet the PERCIST standard. Potentially, with a larger cohort, a mathematical method could be developed to correct the variance caused by undesirably controlled uptake time.
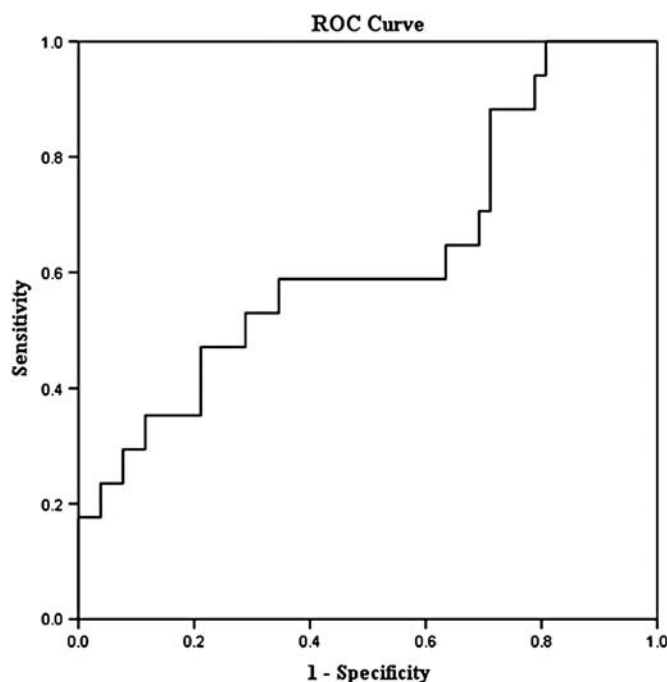
TABLE 3. Samples With Different Discrepancies in Uptake Times and the Corresponding Exclusion Rates

Discrepancies in Uptake Time/Min, Range	Exclusion Rate if All Cases Are in the Range
0–2.2	0%
0–4.3	15%
0–5.1	24%
>20.9	40%
>23.03	50%
>31.13	75%
>32.05	100%

Other factors, such as treatment regimen (eg, those affecting liver function), body weight, and injected activity, are also contributing factors to variance of SUL in organs, theoretically. Even the placement of reference VOI could theoretically affect the PERCIST comparability due to the major dependence on measurements of reference VOI and heterogeneity of tracer distribution within liver parenchyma. However, results of a study has shown that intersite variance of SUL_{mean} of reference VOIs is negligible.[23] In our cohort, the time interval between the 2 scans was relatively short, within about 3 months. As a result, the body weight, and thus injected activity differences that is based on body weight, between the scans was not large enough to be of significant impact. In cancer patients, it is not unexpected that body weight may decrease, as in our cohort where the average body weight decreased from a mean of 54.6 kg at PET 1 to 53.3 kg at PET 2 ($P = 0.013$).[24–28] Although the decrease in body weight was statistically significant, it is more likely that this is a result from loss of adipose tissue rather than lean body mass in such a short time frame of 3 months.[29] As a result, SUL, which is standardized by lean body mass, is not significantly affected by the loss of body weight.

It should be noted that our results may not be generalized to other cohorts. However, our study has highlighted the importance of stringent uptake time standardization between the 2 scans to ensure comparability. This is a single-center study using only 1 PET/CT scanner, which means that all scans were performed with the same hardware and standardization. Thus, other potential factors that may affect comparability such as differences in partial volume effect, method of reconstruction, machine models, scanning protocols, and so on were not studied in our cohort.

CONCLUSIONS

Standardization of uptake time is the most important factor that should be paid heed to in studies that utilize the PERCIST-defined standard for comparability in a routine clinical practice. For follow-up scans performed for the purpose of response assessment, reference to the uptake time used in the baseline scans is therefore recommended.

FIGURE 1. ROC with PERCIST-defined comparability as the binary classifier and differences in uptake time as the discrimination thresholds.

REFERENCES

1. Fukukita H, Suzuki K, Matsumoto K, et al. Japanese guideline for the oncology FDG-PET/CT data acquisition protocol: synopsis of Version 2.0. *Ann Nucl Med*. 2014;28:693–705.
2. Garcia Vicente AM. Recommendations for the use of PET-CT with ^{18}F-FDG in radiotherapy planning [in Spanish]. *Rev Esp Med Nucl Imagen Mol*. 2012; 31:364.
3. Giannopoulou C, Fragaki C. Positron emission tomography in evaluating the response to treatment of brain tumors, lymphomas and breast cancer [in Greek, Modern]. *Hell J Nucl Med*. 2006;9:117–125.
4. Gregoire V, Haustermans K, Geets X, et al. PET-based treatment planning in radiotherapy: a new standard? *J Nucl Med*. 2007;48(Suppl 1):68S–77S.

5. Francis AM, Sepesi B, Correa AM, et al. The influence of histopathologic tumor viability on long-term survival and recurrence rates following neoadjuvant therapy for esophageal adenocarcinoma. *Ann Surg*. 2013; 258:500–507.

6. Fujiwara Y, Yoshikawa R, Kamikonya N, et al. Trimodality therapy of esophagectomy plus neoadjuvant chemoradiotherapy improves the survival of clinical stage II/III esophageal squamous cell carcinoma patients. *Oncol Rep*. 2012;28:446–452.

7. Hartley A, Ho KF, Mcconkey C, et al. Pathological complete response following pre-operative chemoradiotherapy in rectal cancer: analysis of phase II/III trials. *Br J Radiol*. 2005;78:934–938.

8. Igidbashian L, Fortin B, Guertin L, et al. Outcome with neck dissection after chemoradiation for N3 head-and-neck squamous cell carcinoma. *Int J Radiat Oncol Biol Phys*. 2010;77:414–420.

9. Katakami N, Matsumoto H, Tomioka H, et al. Concurrent chemoradiotherapy followed by surgery for advanced stage III non-small cell lung cancer (NSCLC) [in Japanese]. *Gan To Kagaku Ryoho*. 1995;22:531–537.

10. Semrau S, Schmidt D, Lell M, et al. Results of chemoselection with short induction chemotherapy followed by chemoradiation or surgery in the treatment of functionally inoperable carcinomas of the pharynx and larynx. *Oral Oncol*. 2013;49:454–460.

11. Milano A, Perri F, Ciarmiello A, et al. Targeted-therapy and imaging response: a new paradigm for clinical evaluation? *Rev Recent Clin Trials*. 2011;6:259–265.

12. Kamibayashi T, Tsuchida T, Demura Y, et al. Reproducibility of semi-quantitative parameters in FDG-PET using two different PET scanners: influence of attenuation correction method and examination interval. *Mol Imaging Biol*. 2008;10:162–166.

13. Groheux D, Delord M, Rubello D, et al. Variation of liver SUV on ([18])FDG-PET/CT studies in women with breast cancer. *Clin Nucl Med*. 2013;38: 422–425.

14. Buckler AJ, Boellaard R. Standardization of quantitative imaging: the time is right, and ^{18}F-FDG PET/CT is a good place to start. *J Nucl Med*. 2011;52:171–172.

15. Michalski MH, Chen X. Molecular imaging in cancer treatment. *Eur J Nucl Med Mol Imaging*. 2011;38:358–377.

16. Young H, Baum R, Cremerius U, et al. Measurement of clinical and subclinical tumour response using [^{18}F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations. European Organization for Research and Treatment of Cancer (EORTC) PET Study Group. *Eur J Cancer*. 1999;35:1773–1782.

17. Wahl RL, Jacene H, Kasamon Y, et al. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009; 50(Suppl 1):122S–150S.

18. Fencl P, Belohlavek O, Harustiak T, et al. The analysis of factors affecting the threshold on repeated ^{18}F-FDG-PET/CT investigations measured by the PERCIST protocol in patients with esophageal carcinoma. *Nucl Med Commun*. 2012;33:1188–1194.

19. Kuruva M, Mittal BR, Abrar ML, et al. Multivariate analysis of various factors affecting background liver and mediastinal standardized uptake values. *Indian J Nucl Med*. 2012;27:20–23.

20. Hamberg LM, Hunter GJ, Alpert NM, et al. The dose uptake ratio as an index of glucose-metabolism—useful parameter or oversimplification? *J Nucl Med*. 1994;35:1308–1312.

21. Busing KA, Schonberg SO, Brade J, et al. Impact of blood glucose, diabetes, insulin, and obesity on standardized uptake values in tumors and healthy organs on ^{18}F-FDG PET/CT. *Nucl Med Biol*. 2013;40:206–213.

22. Kubota K, Watanabe H, Murata Y, et al. Effects of blood glucose level on FDG uptake by liver: a FDG-PET/CT study. *Nucl Med Biol*. 2011;38: 347–351.

23. Viner M, Mercier G, Hao F, et al. Liver SUL$_{mean}$ at FDG PET/CT: interreader agreement and impact of placement of volume of interest. *Radiology*. 2013;267:596–601.

24. Yucel B, Akkas EA, Okur Y, et al. The impact of radiotherapy on quality of life for cancer patients: a longitudinal study. *Support Care Cancer*. 2014;22: 2479–2487.

25. Ediebah DE, Galindo-Garre F, Uitdehaag BM, et al. Joint modeling of longitudinal health-related quality of life data and survival. In: *Qual Life Res*. 2014;24:795–804.

26. Giesinger JM, Wintner LM, Zabernigg A, et al. Assessing quality of life on the day of chemotherapy administration underestimates patients' true symptom burden. *BMC Cancer*. 2014;14:758.

27. Poulsen GM, Pedersen LL, Osterlind K, et al. Randomized trial of the effects of individual nutritional counseling in cancer patients. *Clin Nutr*. 2014;33: 749–753.

28. Hebuterne X, Lemarie E, Michallet M, et al. Prevalence of malnutrition and current use of nutrition support in patients with cancer. *JPEN J Parenter Enteral Nutr*. 2014;38:196–204.

29. Coupaye M, Bouillot JL, Poitou C, et al. Is lean body mass decreased after obesity treatment by adjustable gastric banding? *Obes Surg*. 2007;17: 427–433.