Mol Imaging Biol (2019) 21:1210–1219
DOI: 10.1007/s11307-019-01342-5

Published Online: 8 March 2019

RESEARCH ARTICLE

How Often Do We Fail to Classify the Treatment Response with [^{18}F]FDG PET/CT Acquired on Different Scanners? Data from Clinical Oncological Practice Using an Automatic Tool for SUV Harmonization

Maria Vittoria Mattoli,[1] Maria Lucia Calcagni,[1,2] Silvia Taralli,[1] Luca Indovina,[3] Bruce S. Spottiswoode,[4] Alessandro Giordano[1,2]

[1]*Nuclear Medicine Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Largo F. Vito 1, 00168, Rome, Italy*
[2]*Nuclear Medicine Institute, Università Cattolica del Sacro Cuore, Rome, Italy*
[3]*Medical Physics Unit, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy*
[4]*Siemens Medical Solutions USA, Inc., Knoxville, TN, USA*

Abstract

Purpose: Tumor response evaluated by 2-deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) positron emission tomography/computed tomography (PET/CT) with standardized uptake value (SUV) is questionable when pre- and post-treatment PET/CT are acquired on different scanners. The aims of our study, performed in oncological patients who underwent pre- and post-treatment [^{18}F]FDG PET/CT on different scanners, were (1) to evaluate whether EQ•PET, a proprietary SUV inter-exams harmonization tool, modifies the EORTC tumor response classification and (2) to assess which classification (harmonized and non-harmonized) better predicts clinical outcome.

Procedures: We retrospectively identified 95 PET pairs (pre- and post-treatment) performed on different scanners (Biograph mCT, Siemens; GEMINI GXL, Philips) in 73 oncological patients (52F; 57.8 ± 16.3 years). An 8-mm Gaussian filter was applied for the Biograph protocol to meet the EANM/EARL harmonization standard; no filter was needed for GXL. SUVmax and SUVmaxEQ of the same target lesion in the pre- and post-treatment PET/CT were noted. For each PET pair, the metabolic response classification (responder/non-responder), derived from combining the EORTC response categories, was evaluated twice (with and without harmonization). In discordant cases, the association of each metabolic response classification with final clinical response assessment and survival data (2-year disease-free survival, DFS) was assessed.

Results: On Biograph, SUVmaxEQ of all target lesions was significantly lower ($p = 0.001$) than SUVmax (8.5 ± 6.8 *vs* 12.5 ± 9.6; − 38.6 %). A discordance between the two metabolic response classifications (harmonized and non-harmonized) was found in 19/95 (20 %) PET pairs. In this subgroup ($n = 19$; mean follow-up, 33.9 ± 9 months), responders according to harmonized classification ($n = 9$) had longer DFS (47.5 months, 88.9 %) than responders ($n = 10$) according

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11307-019-01342-5) contains supplementary material, which is available to authorized users.

Correspondence to: Maria Mattoli; *e-mail:* mvittoriamattoli@yahoo.it

to non-harmonized classification (26.3 months, 50.0 %; $p = 0.01$). Moreover, harmonized classification showed a better association with final clinical response assessment (17/19 PET pairs).

Conclusions: The harmonized metabolic response classification is more associated with the final clinical response assessment, and it is able to better predict the DFS than the non-harmonized classification. EQ•PET is a useful harmonization tool for evaluating metabolic tumor response using different PET/CT scanners, also in different departments or for multicenter studies.

Key words: PET, [^{18}F]FDG, EORTC, Therapy response, Harmonization, EQ•PET

Introduction

2-Deoxy-2-[^{18}F]fluoro-D-glucose ([^{18}F]FDG) positron emission tomography/computed tomography (PET/CT) is a well-established technique for treatment response evaluation in oncological patients. Nowadays, several long-term therapeutic approaches are available, including radiotherapy, various chemotherapy lines, and biological drugs. Repeated PET/CT scans are required in oncological patients for accurate and timely evaluation of the disease status after treatment.

To this aim, the visual analysis of [^{18}F]FDG PET/CT images is the first and paramount used method in clinical practice. It is considered reliable when a complete resolution of tumor uptake or the appearance of new lesions is clearly evident at post-treatment PET/CT compared with baseline PET/CT. However, when a tumor uptake is still present after treatment and subtle changes have occurred, the response evaluation based on visual analysis becomes tricky and the image interpretation could be different among nuclear physicians. In this scenario, a semi-quantitative images analysis enables more objective evaluation of changes in tumor activity after treatment, improving the nuclear physician's reporting confidence and the inter-reader agreement.

Standardized uptake value (SUV) is the most widely used semi-quantitative measure of tumor activity. Two different criteria, based on the degree of SUV change after treatment, have been proposed for evaluating metabolic tumor response: the European Organization for Research and Treatment of Cancer 1999 (EORTC) criteria [1] with a 25 % threshold variation in maximum SUV (SUVmax) among the classes of response, and the PET Response Criteria in Solid Tumors (PERCIST) criteria [2] with a 30 % threshold variation in SUVpeak normalized to lean body mass (SULpeak). Several studies have been compared with these two criteria highlighting a good agreement, similar responses, and clinical outcomes [3–6].

The reliability of metabolic response assessment based on SUV is questionable when pre- and post-treatment examinations are acquired on different scanners, a scenario that is increasingly frequent in clinical practice, given the widespread variety of centers equipped with different PET/CT scanner models [7–10]. Indeed, it is well known that SUV measure is affected by several biological, technical, and physical factors. In particular, the differences in scanner hardware, partial volume effect, and acquisition and reconstruction protocols introduce significant measurement variations [11–15].

The EANM/EARL accreditation program includes a harmonization strategy aiming to make semi-quantitative PET parameters comparable across different PET systems, which is especially relevant for multicenter trials or centers equipped with multiple PET systems [15]. This harmonization strategy has recently been shown to be feasible and achievable [16, 17], even for modern systems with point spread function (PSF) correction [16, 18], although this approach requires two reconstructions: one for optimal lesion detection and one for harmonized quantification [19].

EQ·PET is a reference-based quantification technology developed by Siemens Healthineers that aims to enable comparable semi-quantitative measurements acquired on different scanners [12]. Specifically, EQ·PET provides a SUV harmonization across patient exams, regardless of scanner or reconstruction protocol used, and without the need to modify the preferred reconstruction protocol or reconstruct additional datasets. SUV is harmonized by applying a phantom-derived spatial filter optimized to align contrast recovery coefficients, either between scanners/reconstructions, or relative to a reference such as EANM specification [20]. The use of EQ·PET with the EANM specification has been shown to be equivalent to the EANM/EARL harmonization approach [21].

The utility of EQ·PET in the EANM/EARL-compliant SUV harmonization has been investigated in the metabolic tumor response classification using EORTC and PERCIST criteria: after the harmonization, agreement levels among different metabolic classifications (reference reconstruction scenario *vs* other scenarios involving reconstruction inconsistency) were found to be almost perfect [7, 8]. Comparing the two criteria, EORTC classification has resulted more influenced by different acquisition and reconstruction protocols than PERCIST classification [7, 16, 22]. Furthermore, the need for EARL-compliant SUV harmonization when using [^{18}F]FDG PET/CT as a prognosticator has been recently demonstrated in non-small-cell lung cancer patients [23].

To the best of our knowledge, up to now, it has not been demonstrated whether the metabolic classification of treatment response EANM-compliant harmonized with EQ·PET is consistent with final clinical response assessment and survival data. The aims of our study, performed in oncological patients who underwent PET/CT studies on different scanners for treatment response assessment were (1) to assess the amount of SUV change after the harmonization with EQ·PET; (2) to assess whether the harmonization modifies the EORTC response classification compared with non-harmonized SUV; and (3) to evaluate if discordant EORTC classification cases were present, which classification (harmonized and non-harmonized) was more strongly associated with the final clinical response assessment and with prognosis (disease-free survival, DFS).

Materials and Methods

Population

From a total of 467 oncological patients included in local PET/CT databases collected for treatment response evaluation (from 2012 to 2015 for research purpose), we retrospectively selected only those patients who performed two consecutive scans (named as a PET pair), before treatment (pre-PET) and after treatment (post-PET), on different scanner systems. Seventy-three out of 467 (15.6 %) patients (52F; mean age, 57.8 ± 16.3 years) were finally included. Tumor types were 32 (43.8 %) cervical cancers, 15 (20.5 %) lung cancers, 14 (19.3 %) colorectal cancers, 5 (6.8 %) lymphomas, 5 (6.8 %) breast cancers, 1 (1.4 %) pineal tumor, and 1 (1.4 %) osteosarcoma.

According to internal treatment protocols, types of treatment were chemotherapy followed by chemo-radiotherapy (cervical cancer), chemo-radiotherapy (colorectal cancer), induction chemo-radiotherapy followed by definitive chemo-radiotherapy (lung cancer), chemotherapy followed by radiotherapy (lymphoma), and chemotherapy and/or radiotherapy (breast cancers, pineal tumor, and osteosarcoma). Considering, for each selected patient, all PET pairs performed during the oncological patient's history and a single evaluated target lesion for each PET pairs, a total of 95 PET pairs (95 target lesions) were included in the study: 51/73 patients underwent one PET pair (51 PET pairs, 51 target lesions) and 22/73 underwent two PET pairs (44 PET pairs, 44 target lesions). The mean time interval between pre-PET and post-PET scans was 128 ± 77 days.

All scans were performed in our PET/CT center, with the data being used following patient written consent in accordance with the ethical standard of our institution.

PET/CT Protocols

The PET/CT scanner systems were Biograph mCT (Siemens Medical Solutions USA, Inc., Knoxville, TN) and GEMINI GXL (Philips Medical System, Cleveland, OH). The PET pair was represented by (1) pre-PET acquired on Biograph and post-PET acquired on GXL (Bio-GXL group) and (2) pre-PET acquired on GXL and post-PET acquired on Biograph (GXL-Bio group). The details of the PET/CT protocols are provided in the Supplementary Material (see Suppl. Table S1).

EQ·PET Application and PET/CT Data Analysis

SUVmax, defined as the hottest standard uptake value obtained in a volume of interest (VOI), was evaluated before and after its harmonization to the EANM standard [20], by applying an optimized Gaussian filter (EQ·PET filter) [24] and obtaining the SUVmax equalized (SUVmaxEQ). For the Biograph mCT system, an EQ·PET Gaussian filter parameter of 8 mm was applied, resulting in a correction of SUVmax values (SUVmaxEQ). For the GXL system, no EQ·PET Gaussian filter was needed (the contrast recovery coefficient was already well aligned with the EANM reference), resulting in a non-correction of SUVmax. Note that the SUVmaxEQ could alternatively be obtained by using a second reconstruction with the Gaussian filter integrated into the standard post filter, as per the EARL recommendation. The details of the recovery curve values before and after the application of the Gaussian filter for the two scanner systems are provided in the Supplementary Material—harmonization to EARL limit.

SUVmax and SUVmaxEQ of the same target lesion in the pre- and post-treatment PET/CT scans, as well as the percentage change of SUVmax (ΔSUVmax) and SUVmaxEQ (ΔSUVmaxEQ) between the pre- and post-treatment scans, were noted (see Supplementary Material for PET/CT data analysis description).

PET/CT Response Assessment

The metabolic tumor response was evaluated according to EORTC criteria [1], since it is based on SUVmax variation of the same target lesion, at baseline and post-treatment scans. The EORTC criteria identifying four categories of response based on percentage changes of SUVmax between pre-PET and post-PET (ΔSUV):

- Complete metabolic response (CMR): a complete resolution of [^{18}F]FDG uptake within the tumor volume so that it was indistinguishable from surrounding normal tissue;
- Partial metabolic response (PMR): a reduction in SUV > 25 %;
- Stable metabolic disease (SMD): an increase in SUV < 25 % or a reduction in SUV < 25 % and no visible increase in the extent of [^{18}F]FDG uptake (< 20 % in the longest dimension).
- Progressive metabolic disease (PMD): an increase in SUV > 25 % within the tumor region, a visible increase in the extent of [^{18}F]FDG tumor uptake (> 20 % in the longest

dimension) or the appearance of new [^{18}F]FDG uptake lesions.

We defined the binary *metabolic response classification* by combining the four EORTC response categories as follows: responders (R), the PET pairs showing complete (CMR) or partial metabolic response (PMR) and non-responders (NR), the PET pairs showing stable (SMD) or progressive metabolic disease (PMD). Each PET pair was classified twice, using ΔSUVmax (non-harmonized) and ΔSUVmaxEQ (harmonized). In cases of discordance between the two metabolic response classifications (non-harmonized and harmonized), the two metabolic response classifications were correlated on "short time" with the final clinical response assessment, and with the 2-year disease-free survival (DFS), defined as the time from the first PET scan (pre-PET) to local or distant relapse, based on computed tomography, magnetic resonance, ultrasound, PET/CT, and histopathological data, on a follow-up rhythm ranging from 3 to 6 months.

Statistical Analysis

Data were analyzed using the MedCalc Statistical Software version 16.8.4 (MedCalc Software bvba, Ostend, Belgium) with statistical significance set at $p < 0.05$. The results were calculated with 95 % confidence intervals (CI). The mean values are reported with standard deviation (SD). Since the data were found to be normally distributed by Shapiro–Wilk test, Student's paired t test was used to compare the SUV values at different time points. The relationship between SUVmax and SUVmaxEQ was assessed using Bland–Altman plots. The disease-free survival (DFS) was assessed using the Kaplan–Meier survival curves, with log-rank testing for p value calculation.

Results

SUVmaxEQ Vs SUVmax (on Biograph Scanner)

On Biograph scanner, SUVmaxEQ was significantly (p = 0.001) lower than SUVmax in all 95 target lesions (both pre-PET and post-PET): 8.5 ± 6.8 *vs* 12.5 ± 9.6 (Fig. 1), with a mean percentage difference of 38.6 % (95%CI 35.5–41.6; range 9.5–67.6 %). This trend was also confirmed when considering separately the target lesions acquired on pre-PET (n = 41) or on post-PET (n = 54): 12.2 ± 7.3 *vs* 17.5 ± 10.0 (p = 0.007) and 5.8 ± 4.8. *vs* 8.8 ± 7.2 (p = 0.012), respectively. These data are evident in the Bland–Altman plots in Fig. 2. Among the 95 analyzed lesions, Bland–Altman plot identified 5/95 (5 %) outliers for which the mean difference percentage between SUVmax and SUVmaxEQ was outside the limit of the upper confidence interval (range 54–57 %). The outliner lesions were 1 lung nodule, 1 pleural thickening, and 3 lymph nodes, with an overall mean SUVmax of 12.6.

EORTC Response Classification

Analyzing all PET pairs (n = 95), the EORTC response criteria based on ΔSUVmax (non-harmonized classification) showed 20/95 (21.1 %) with CMR, 46/95 (48.5 %) with PMR, 13/95 (13.6 %) with SMD, and 16/95 (16.8 %) with PMD. The corresponding EORTC response criteria based on ΔSUVmaxEQ (harmonized classification) showed 20/95 (21.1 %) PET pairs with CMR, 45/95 (47.3 %) with PMR, 19/95 (20.0 %) with SMD, and 11/95 (11.6 %) with PMD (Fig. 3). Table 1 shows the number and the percentage of PET pairs according to the two EORTC classifications (non-harmonized and harmonized) in the two groups of PET pairs: GXL-Bio (n = 54) and Bio-GXL (n = 41).

A discordance between the two EORTC response classifications (non-harmonized and harmonized) was observed in 25/95 (26.3 %) PET pairs, as showed in Table 2. In particular, when applying the EQ·PET harmonization, the EORTC response classification shifted toward a more favorable category (from PMD to SMD or from SMD to PMR) in 15/25 (60 %) PET pairs, belonging to GXL-Bio group, and shifted toward a less favorable category (from PMR to SMD or PMD) in 10/25 (40 %), belonging to Bio-GXL group.

Metabolic Response Classification (Responder/ Non-Responders)

According to the metabolic classification based on ΔSUVmax (non-harmonized classification), 66/95 (69.4 %) PET pairs were classified as responders and 29/95 (30.6 %) as non-responders. According to the metabolic classification based on ΔSUVmaxEQ (harmonized classification), 65/95 (68.4 %) PET pairs were classified as responders and 30/95 (31.6 %) as non-responders.

A discordance between the two metabolic response classifications (non-harmonized and harmonized) was observed in 19/95 (20 %) PET pairs (see Table 3). Each of these 19 PET pairs corresponded to a single patient. In these 19 patients, the harmonized classification was concordant with the final clinical response assessment in 17/19 (89.5 %) patients and discordant in only 2/19 (Table 4); the 2-year DFS rate (mean follow-up, 33.9 ± 9 months) was 68.4 % and the mean DFS time was 36.8 months (95%CI 29.29–44.31). Patients classified as responders on the harmonized classification (n = 9) had longer disease-free survival (47.5 months, 88.9 %) than patients classified as responders (n = 10) on the non-harmonized classification (26.3 months, 50.0 %; log-rank test, p = 0.01), as shown in Fig. 4. When considering the harmonized classification, at the end of the observational period, the overall DFS rate was 57.8 % (11/19 patients): 8/9 patients metabolically classified as responders were disease-

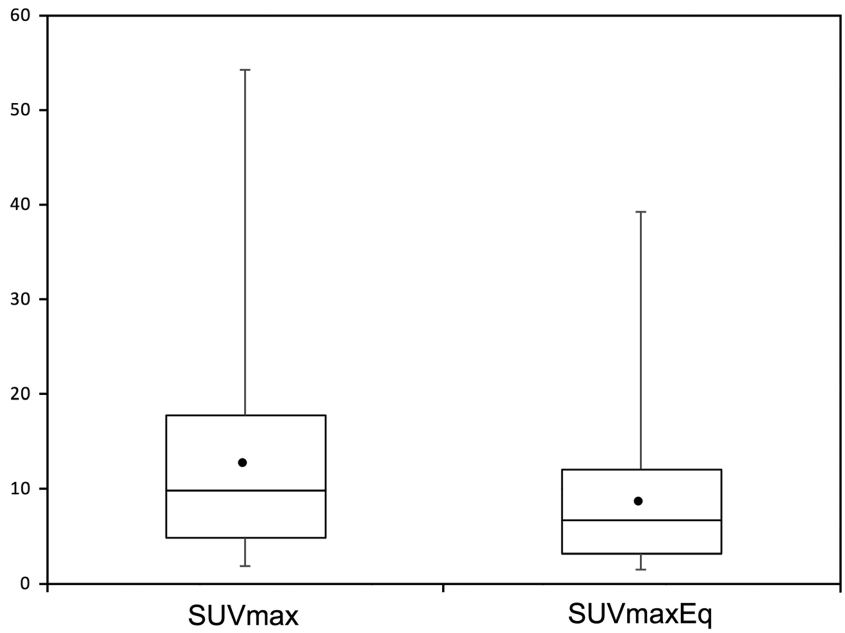

Fig. 1. xSUVmax and SUVmaxEQ of all target lesions ($n = 95$) acquired on Biograph mCT scanner.

progression free, and 7/10 patients metabolically classified as non-responders developed disease progression.

In addition, we separately analyzed the most numerous tumor type subgroup of our series, represented by the 32/73 (43.8 %) patients with cervical cancer who underwent a total of 39/95 (41.0 %) PET pairs (two PET pairs in 7 patients). In line with the data observed in the overall study population, in the 32 patients with only one PET pair each (all performed to evaluate the efficacy of chemotherapy), we found a discordance between the two metabolic classifications (non-harmonized and harmonized) in 6/32 (18.7 %) PET pairs. When analyzing these 6 discordant cases, a concordance between the harmonized metabolic classification and the final clinical response assessment was found in 5/6 (83.3 %), whereas the non-harmonized metabolic classification was concordant in only 1/6 cases. Figure 5 shows a representative case of EQ·PET application in clinical practice.

Discussion

Our study evaluates the potential clinical impact of the use of EANM SUV harmonization for the treatment response assessment in oncological patients imaged with different PET/CT scanner systems; for the first time, the clinical response assessment and the survival data are used as the "gold standard" in discordant metabolic response classification cases.

From our data, the harmonization of the SUVmax values acquired on the Biograph system to the EANM reference resulted in a mean SUVmax reduction of 38 % for all target lesions. This data is in line with the results of a previous prospective multicenter study [7, 8] that applied EQ·PET to the reconstruction used for optimal tumor detection (PSF or PSF ± time-of-flight (TOF)), comparing its results with that obtained with ordered subset expectation maximization (OSEM) reconstruction. From these reported data, it is

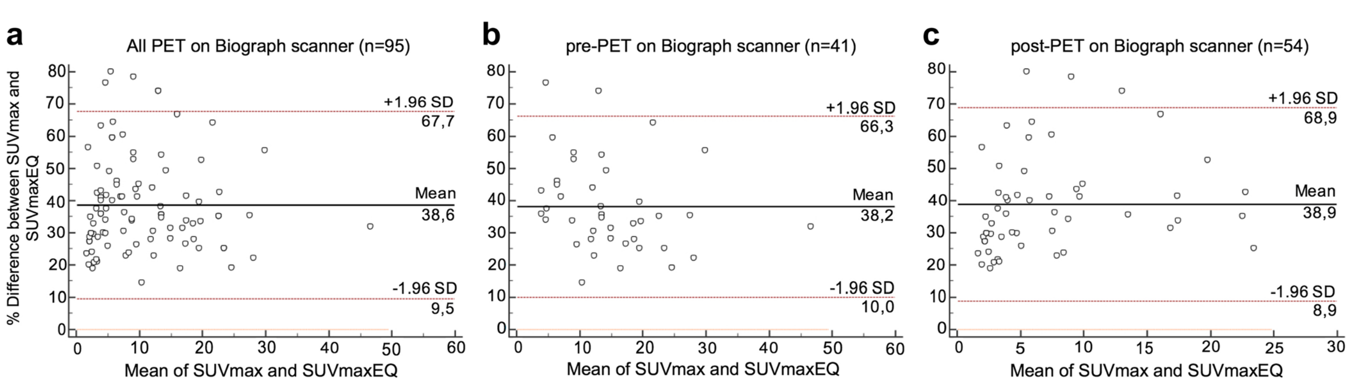

Fig. 2. Bland–Altman plots showing the relationship between SUVmax and SUVmaxEQ in all target lesions acquired on Biograph mCT scanner (**a**) in target lesions acquired on Biograph mCT scanner as pre-PET (**b**), and in target lesions acquired on Biograph mCT scanner as post-PET (**c**).

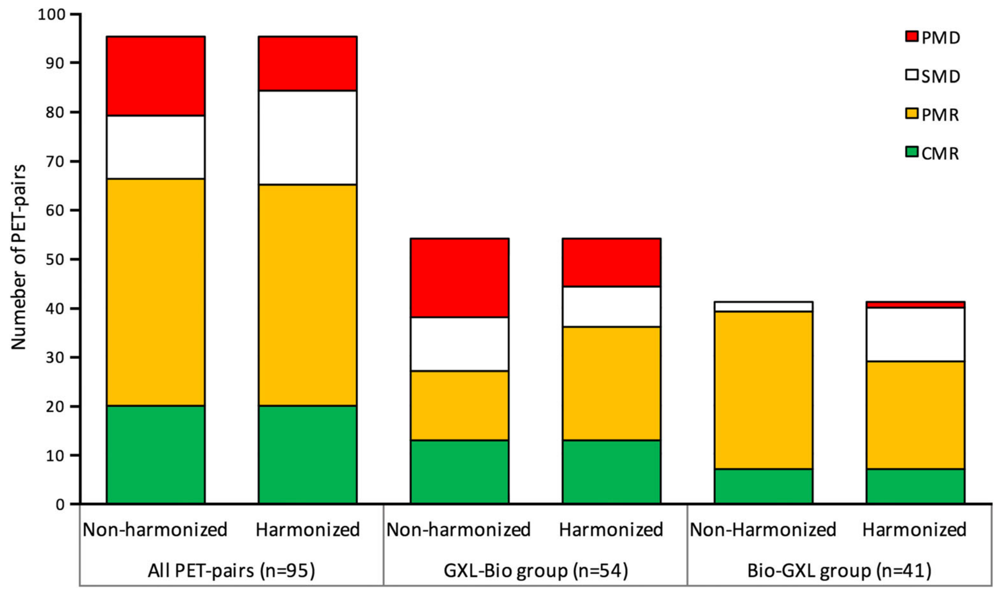

Fig. 3. EORTC non-harmonized classification and harmonized classification after application of EQ·PET methodology for all PET pairs, for the GXL-Bio group and for the Bio-GXL group. PMD, progressive metabolic disease; SMD, stable metabolic disease; PMR, partial metabolic response; CMR, complete metabolic response.

possible to deduce a mean reduction of 35 % of SUVmax between PSF ± TOF, comparable with our Biograph system, and OSEM reconstruction, used as the reference standard and comparable with our GXL system [7].

Regarding Bland–Altman analysis comparing SUVmax and SUVmaxEQ, there were few outliers (5 %), in which the mean difference percentage was outside the upper confidence interval, and these occurred in lesions with relatively high SUVmax value and mainly for small structures as lymph nodes. A possible explanation of this result could be a combination of high SUVmax values, not excluding the small dimensions of the lesions. Regarding this latter aspect, although it has been suggested that the effect of the harmonization tool on SUV values could be linked also to the lesion volume (being more pronounced in small size lesions), it has been demonstrated that the EQ·PET technology is not affected by the tumor size [22].

Table 1. The number and percentage of PET pairs according to the two EORTC classifications (non-harmonized and harmonized) in the GXL-Bio and Bio-GXL groups

EORTC	GXL-Bio group ($n = 54$)				Bio-GXL group ($n = 41$)			
	Non-harmonized		Harmonized		Non-harmonized		Harmonized	
	N	%	*N*	%	*N*	%	*N*	%
CMR	13	24.1	13	24.1	7	17.1	7	17.1
PMR	14	25.9	23	42.6	32	78.0	22	53.7
SMD	11	20.4	8	14.8	2	4.9	11	26.8
PMD	16	29.6	10	18.5	0	0	1	2.4
Total	54	100	54	100	41	100	41	100

CMR, complete metabolic response; *PMR*, partial metabolic response; *SMD*, stable metabolic disease; *PMD*, progressive metabolic disease

In our study, the harmonization modified the EORTC response classification in more than 25 % of PET pairs. This result is in line with that reported by Lasnon *et al*. and Quak *et al*. [7, 8] who investigated the impact of reconstruction inconsistences on tumor response classification and the utility of EQ·PET tool in the harmonization of data. The authors found discordant classifications in up to 20 % [8] and 31 % of the cases [7], respectively, when evaluating the impact of different reconstructions (PSF ± TOF and OSEM) on EORTC response classifications. This considerable rate of modified response classifications clearly underlines the problem of the reliability of SUV comparison in the classification of tumor response, especially in a clinical practice running two scanners, where the implications of an incorrect exam interpretation imply serious repercussions on the subsequent patients' clinical management. In addition, our data suggest that the use of EQ·PET is not only able to harmonize the quantitative information of PET data regardless of the reconstruction/scanner used but also results in a response classification which better predicts the clinical outcome than when using non-harmonized SUVs.

When comparing the non-harmonized and the harmonized metabolic classifications, we found some differences depending whether Biograph was used as pre-PET or post-PET scan. Applying the harmonization when the Biograph was used as pre-PET (Bio-GXL group), we observed an increase of more unfavorable response categories (stable or progression disease): patients in partial metabolic response at non-harmonized classification shifted toward stable or, even, to progressive disease and patients in stable disease shifted toward progression disease. This finding can be explained considering that EQ·PET consistently reduces the pre-treatment SUVmax; consequently, the SUV interval range

Table 2. Discordance between the non-harmonized and harmonized EORTC classifications in overall cases, in the GXL-Bio and Bio-GXL groups

EORTC	All discordant cases ($n = 25$)		GXL-Bio group ($n = 15$)		Bio-GXL group ($n = 10$)	
	Non-harmonized	Harmonized	Non-harmonized	Harmonized	Non-harmonized	Harmonized
CMR	0	0	0	0	0	0
PMR	10	9	0	9	10	0
SMD	9	15	9	6	0	9
PMD	6	1	6	0	0	1

CMR, complete metabolic response; *PMR*, partial metabolic response; *SMD*, stable metabolic disease; *PMD*, progressive metabolic disease

between baseline and post-therapy is reduced, potentially resulting in a greater amount of patients classifiable as non-responders. When applying the harmonization when Biograph was used for post-PET (GXL-Bio group), we observed an increase of more favorable response categories (stable disease or partial response): patients in progression disease at non-harmonized classification shifted toward stable disease and patients in stable disease shifted toward partial response. This finding can be explained considering that, in this case, EQ·PET consistently reduces the post-therapy SUVmax; consequently, the SUV interval range between pre- and post-therapy is increased, resulting in a greater amount of patients classifiable as responders. To no surprise, the harmonization did not impact on patients classified as complete metabolic response; they did not shift to a different category in either PET pair groups (Bio-GXL or GXL-Bio). Indeed, to classify an exam as CMR, it is enough that the tumor activity evident at post-treatment PET/CT is visually indistinguishable from surrounding normal tissue, regardless of the SUV measurements. In the study conducted by Quak *et al*. [8] that evaluated the capability of EQ·PET to harmonize metabolic response classification across different reconstructions, the authors found a similar trend when EQ·PET was applied in two reconstruction scenarios: $OSEM_{PET1}/PSF \pm TOF_{PET2}$ ($n =$ 17, down-classification), comparable to our GXL-Bio group and $PSF \pm TOF_{PET1}/OSEM_{PET2}$ ($n = 13$, up-classification), comparable to our Bio-GXL-group.

We evaluated whether the harmonized metabolic treatment response classification is consistent with the final clinical response assessment. Combining the four EORTC categories in the binary metabolic classification (responder/non-responder), we observed 19 discordant PET pairs (20 %). In these discordant cases, when compared with the non-harmonized metabolic classification, the harmonized classification showed a greater and almost complete concordance with the final clinical response assessment (17/19). In particular, the harmonized corrected classification was not in line with the clinical response assessment in only two patients (#2 and #19). In the first patient (case #2 with cervical cancer), the pre-PET/CT acquired on Biograph scanner showed intense uptake in the cervical target lesion (SUVmax1 = 21.9, SUVmaxEQ = 17.0); the post-treatment PET/CT acquired on GXL scanner (at the end of first-line chemotherapy) showed a reduction, but still persistence, of the activity of the target lesion (SUVmax2 = 15.2): patient was classified as responder on the non-harmonized classification (ΔSUVmax = − 30.4 %, PMR) and as non-responder on the harmonized classification (ΔSUVmaxEQ = − 10.5 %, SMD). The post-treatment pelvic MR showed the persistence of a portion of tumor tissue at the cervix, reporting a partial response, in line with the non-harmonized evaluation. In the second patient (case #19, rectal cancer), the staging PET/CT acquired on GXL scanner showed intense uptake in the rectal target lesion (SUVmax1 = 10.7), and moderate uptake in the regional lymph nodes. Post-treatment PET/CT acquired on the Biograph scanner (after first-line chemotherapy) showed persistent activity in the rectal lesion (SUVmax2 = 8.9; SUVmaxEQ = 5.8) and no residual activity in the lymph nodes; according to the non-harmonized classification, the patient was classified as a non-responder (ΔSUVmax = − 17.1 %, SMD), whereas as a responder on the harmonized classification (ΔSUVmaxEQ = − 45.6 %, PMR). The subsequent radiological examinations showed an increase of the rectal lesion, in line with the non-harmonized evaluation of no-response. With these exceptions, the overall high concordance with the final clinical response assessment is of crucial importance since a significant number of patients could have been correctly classified if EQ·PET methodology had been used.

Table 3. Discordance between the non-harmonized and harmonized metabolic response classifications (R/NR) in overall cases, in the GXL-Bio and Bio-GXL groups

	All discordant cases ($n = 19$)		GXL-Bio group ($n = 9$)		Bio-GXL group ($n = 10$)	
	Non-harmonized	Harmonized	Non-harmonized	Harmonized	Non-harmonized	Harmonized
R	10	9	0	9	10	0
NR	9	10	9	0	0	10

R, responder; *NR*, non-responder

Table 4. PET pairs ($n = 19$) showing a discordance between the metabolic response classifications (non-harmonized and harmonized) along with the final clinical response assessment

Case	Primary tumor	Group	EORTC		R/NR		Final clinical response assessment	Concordance R/NR harmonized *vs* final clinical response assessment
			Non-harmonized	Harmonized	Non-harmonized	Harmonized		
1	Cervical cancer	GXL-Bio	SMD	PMR	NR	R	R	Y
2	Cervical cancer	Bio-GXL	PMR	SMD	R	NR	R	N
3	Cervical cancer	Bio-GXL	PMR	SMD	R	NR	NR	Y
4	Cervical cancer	GXL-Bio	SMD	PMR	NR	R	R	Y
5	Cervical cancer	GXL-Bio	SMD	PMR	NR	R	R	Y
6	Cervical cancer	Bio-GXL	PMR	SMD	R	NR	NR	Y
7	Lymphoma	GXL-Bio	SMD	PMR	NR	R	R	Y
8	Lymphoma	Bio-GXL	PMR	SMD	R	NR	NR	Y
9	NSCLC	Bio-GXL	PMR	SMD	R	NR	NR	Y
10	NSCLC	GXL-Bio	SMD	PMR	NR	R	R	Y
11	NSCLC	GXL-Bio	SMD	PMR	NR	R	R	Y
12	NSCLC	GXL-Bio	SMD	PMR	NR	R	R	Y
13	NSCLC	Bio-GXL	PMR	SMD	R	NR	NR	Y
14	Rectal cancer	Bio-GXL	PMR	SMD	R	NR	NR	Y
15	Rectal cancer	GXL-Bio	SMD	PMR	NR	R	R	Y
16	Rectal cancer	Bio-GXL	PMR	SMD	R	NR	NR	Y
17	Rectal cancer	Bio-GXL	PMR	SMD	R	NR	NR	Y
18	Rectal cancer	Bio-GXL	PMR	PMD	R	NR	NR	Y
19	Rectal cancer	GXL-Bio	SMD	PMR	NR	R	NR	N

NSCLC, non-small-cell lung cancer; *CMR*, complete metabolic response; *PMR*, partial metabolic response; *SMD*, stable metabolic disease; *PMD*, progressive metabolic disease; *R*, responder; *NR*, non-responder; *Y*, yes; *N*, no

Furthermore, when evaluating the potential prognostic value of both harmonized and non-harmonized classification, the harmonized classification showed a high and better association with the disease-free survival, compared with the non-harmonized classification.

These findings suggest that the harmonization of PET data could have important implications in the perspective of personalized medicine, on both subsequent diagnostic and therapeutic management. It is important to note that the same results could be achieved by employing a separate EARL-compliant reconstruction instead of the EQ·PET software approach [15], but this increases reconstruction time, data storage requirements, and measurement time [19].

There are several possible advantages to implementing EANM harmonization in clinical practice: (1) in the management of the tomographs' daily lists in a center running two (or more) different scanners, SUV harmonization makes it less critical to ensure that the post-treatment PET is acquired on the same scanner as the baseline examination; this simplifies the workflow and reduces strain in cases during scanner downtime; (2) patients may not need to return to the same center where a baseline PET/CT has been performed, but rather choose a more geographically convenient department/hospital; (3) advantage for the nuclear physician by increasing confidence in the reporting a post-treatment PET that has been acquired with a different PET systems than which used for baseline PET/CT; and (4) advantage for the multicenter studies.

The main limitations of our study were its retrospective nature and the relatively small size of the population. It is important to note that in centers running two or more PET/CT systems, as in ours, every effort is made to ensure that the post-treatment PET/CT is acquired on the same scanner as the baseline PET/CT, limiting cases like these as much as possible. A possible bias derives from the inclusion of patients with a variety of primary tumors and different treatment types; nevertheless, when analyzing our results in patients with cervical tumor, which represents the most numerous subgroup of patients with homogeneous characteristics (for both tumor and treatment type), we found a substantially equal rate of discordant cases and concordance between the harmonized classification and the clinical response assessment compared with those derived from the overall population.

Prospective and multicenter studies with larger samples of oncological patients are needed to confirm these preliminary data and in order to better evaluate the usefulness of the EQ PET software in the harmonization of semi-quantitative PET data.

Conclusions

In oncological clinical practice, EQ·PET is a potentially useful tool for evaluating metabolic tumor response regardless of the scanner used to monitor the treatment. Indeed, by harmonizing the semi-quantitative data with the EAMN/EARL standards, EQ·PET makes comparable two or more PET/CT examinations acquired on different scanners. Compared with the non-harmonized metabolic response classification, EQ·PET is able to provide a metabolic response classification which better predicts the final clinical response assessment and the prognosis, with

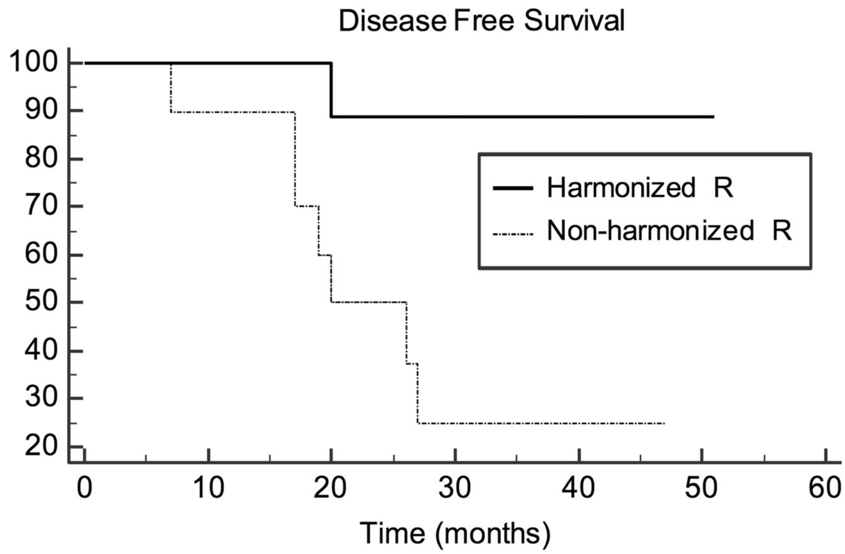

Fig. 4. Disease-free survival according to responder patients in the two metabolic classifications (harmonized and non-harmonized $p = 0.01$). R, responder.

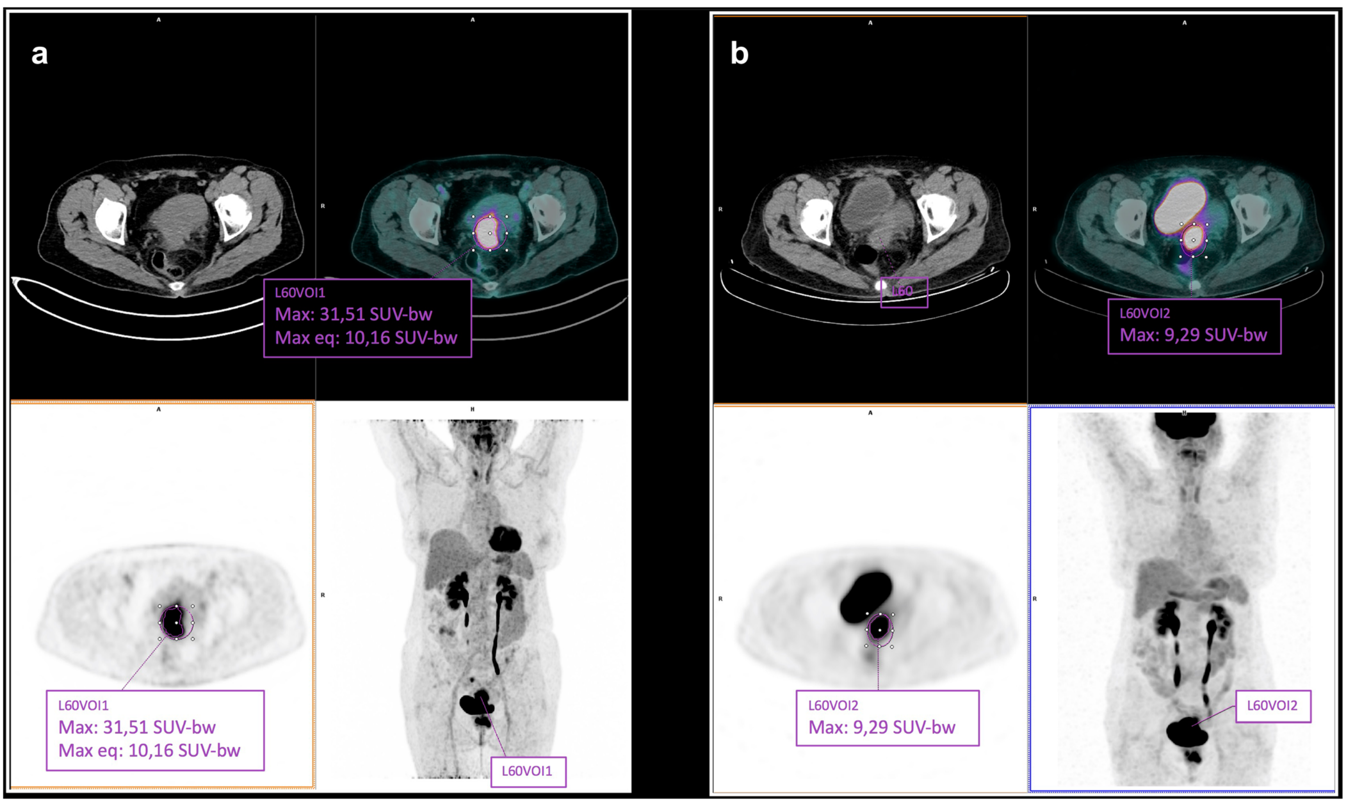

Fig. 5. A representative case of a 68-year-old patient with locally advanced cervical cancer (squamous cell carcinoma—FIGO IIB). **a** Axial co-registered CT, PET, fused, and maximum intensity projection (MIP) [^{18}F]FDG PET/CT images acquired at staging on Biograph mCT scanner showing intense uptake in the cervix lesion (target) with no evidence of lymph nodal or distant metastases. **b** Axial co-registered CT, PET, fused, and MIP [^{18}F]FDG PET/CT images after concomitant chemo-radiotherapy on GXL scanner, showing the persistence of intense uptake at the target lesion. Semi-quantitative analysis demonstrated a reduction in SUVmax between the scans of 31.8 % (partial metabolic response according to EORTC criteria), classifying patient as a responder. Applying EQ•PET harmonization at baseline Biograph PET, the reduction in SUVmax was only 8 % (stable metabolic disease), changing the metabolic classification from a responder to a non-responder. A subsequent MR exam showed stable disease; the patient developed disease progression 8 months later.

potentially important implications in the perspective of personalized patient management.

Acknowledgments. We would like to acknowledge Dr. Margherita Milite and Dr. Antje Schulte for their contributions to the revised version of this manuscript.

Compliance with Ethical Standards

Conflict of Interest

Dr. Mattoli has received a research support from Siemens Molecular Imaging and Dr. Spottiswoode is a full-time employee of Siemens Healthineers.

Ethical Approval

For this type of study, formal consent is not required.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

References

1. Young H, Baum R, Cremerius U et al (1999) 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 35:1773–1782
2. Wahl RL, Jacene H, Kasamon Y, Lodge MA (2009) From RECIST to PERCIST: evolving considerations for PET response criteria in solid tumors. J Nucl Med 50(Suppl 1):122S–150S
3. Turgeon GA, Iravani A, Akhurst T, Beaulieu A, Callahan JW, Bressel M, Cole AJ, Everitt S, Siva S, Hicks RJ, Ball DL, Mac Manus MP (2018) What FDG-PET response-assessment method best predicts survival after curative-intent chemoradiation in non-small cell lung cancer (NSCLC): EORTC, PERCIST, Peter Mac or Deauville criteria? J Nucl Med. https://doi.org/10.2967/jnumed.118.214148
4. Kim JH (2016) Comparison of the EORTC criteria and PERCIST in solid tumors: a pooled analysis and review. Oncotarget 7:58105–58110
5. Pinker K, Riedl C, Weber WA (2017) Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. Eur J Nucl Med Mol Imaging 44(Suppl 1):55–66
6. Skougaard K, Nielsen D, Jensen BV, Hendel HW (2013) Comparison of EORTC criteria and PERCIST for PET/CT response evaluation of patients with metastatic colorectal cancer treated with irinotecan and cetuximab. J Nucl Med 54:1026–1031
7. Lasnon C, Quak E, Le Roux PY et al (2017) EORTC PET response criteria are more influenced by reconstruction inconsistencies than PERCIST but both benefit from the EARL harmonization program. EJNMMI Phys 4:17
8. Quak E, Le Roux PY, Lasnon C et al (2016) Does PET SUV harmonization affect PERCIST response classification? J Nucl Med 57:1699–1706
9. Enilorac B, Lasnon C, Nganoa C et al (2018) Does PET reconstruction method affect Deauville scoring in lymphoma patients? J Nucl Med 59:1167–1169
10. Barrington SF, Sulkin T, Forbes A, Johnson PWM (2018) All that glitters is not gold - new reconstruction methods using Deauville criteria for patient reporting. Eur J Nucl Med Mol Imaging 45:316–317
11. Boellaard R (2009) Standards for PET image acquisition and quantitative data analysis. J Nucl Med 50(Suppl 1):11S–20S
12. Kelly MD, Declerck JM (2011) SUVref: reducing reconstruction-dependent variation in PET SUV. EJNMMI Res 1:16
13. Lasnon C, Desmonts C, Quak E, Gervais R, Do P, Dubos-Arvis C, Aide N (2013) Harmonizing SUVs in multicentre trials when using different generation PET systems: prospective validation in non-small cell lung cancer patients. Eur J Nucl Med Mol Imaging 40:985–996
14. Keyes JW Jr (1995) SUV: standard uptake or silly useless value? J Nucl Med 36:1836–1839
15. Aide N, Lasnon C, Veit-Haibach P, Sera T, Sattler B, Boellaard R (2017) EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies. Eur J Nucl Med Mol Imaging 44(Suppl 1):17–31
16. Kaalep A, Sera T, Rijnsdorp S, Yaqub M, Talsma A, Lodge MA, Boellaard R (2018) Feasibility of state of the art PET/CT systems performance harmonisation. Eur J Nucl Med Mol Imaging 45:1344–1361
17. Kaalep A, Sera T, Oyen W, Krause BJ, Chiti A, Liu Y, Boellaard R (2018) EANM/EARL FDG-PET/CT accreditation - summary results from the first 200 accredited imaging systems. Eur J Nucl Med Mol Imaging 45:412–422
18. Ferretti A, Chondrogiannis S, Rampin L, Bellan E, Marzola MC, Grassetto G, Gusella S, Maffione AM, Gava M, Rubello D (2018) How to harmonize SUVs obtained by hybrid PET/CT scanners with and without point spread function correction. Phys Med Biol 63:235010
19. Rubello D, Colletti PM (2018) SUV harmonization between different hybrid PET/CT systems. Clin Nucl Med 43:811–814
20. Boellaard R, O'Doherty MJ, Weber WA et al (2010) FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 37:181–200
21. Lasnon C, Salomon T, Desmonts C, Dô P, Oulkhouir Y, Madelaine J, Aide N (2017) Generating harmonized SUV within the EANM EARL accreditation program: software approach versus EARL-compliant reconstruction. Ann Nucl Med 31:125–134
22. Quak E, Le Roux PY, Hofman MS et al (2015) Harmonizing FDG PET quantification while maintaining optimal lesion detection: prospective multicentre validation in 517 oncology patients. Eur J Nucl Med Mol Imaging 42:2072–2082
23. Houdu B, Lasnon C, Licaj I, Thomas G, Do P, Guizard AV, Desmonts C, Aide N (2018) Why harmonization is needed when using FDG PET/CT as a prognosticator: demonstration with EARL-compliant SUV as an independent prognostic factor in lung cancer. Eur J Nucl Med Mol Imaging 46:421–428. https://doi.org/10.1007/s00259-018-4151-8
24. Kelly M (2014) EQ•PET: achieving NEMA referenced SUV across technologies. White Paper. Siemens Healthcare Sector, Order No. A91MI-10410-1T-7600, Siemens AG