Radiotherapy and Oncology 130 (2019) 39–45

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original article

Tumor to cervical spinal cord standardized uptake ratio (SUR) improves the reproducibility of ^{18}F-FDG-PET based tumor segmentation in head and neck squamous cell carcinoma in a multicenter setting

Sven van den Bosch [a,*], Tim Dijkema [a], Marielle E.P. Philippens [b], Chris H.J. Terhaard [b], Frank J.P. Hoebers [c], Johannes H.A.M. Kaanders [a], Wim J.G. Oyen [d,e]

[a] Department of Radiation Oncology, Radboud University Medical Center, Nijmegen; [b] Department of Radiation Oncology, University Medical Center Utrecht; [c] Department of Radiation Oncology (MAASTRO), Research Institute GROW, Maastricht University; [d] Department of Nuclear Medicine, Radboud University Medical Center, Nijmegen, The Netherlands; and [e] The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust, London, United Kingdom

ARTICLE INFO

Article history:
Received 19 February 2018
Received in revised form 15 June 2018
Accepted 27 June 2018
Available online 10 July 2018

Keywords:
Standardized uptake ratio
SUR
FDG-PET
Segmentation
Head and neck cancer
Head and neck squamous cell carcinoma

ABSTRACT

Background: In quantitative FDG-PET data analysis, normalization of the standardized uptake value (SUV) with an internal image-derived standard improves its reproducibility. In this study, the cervical spinal cord is proposed as an internal standard that is within the field of view of the radiotherapy planning PET/CT-scan in head and neck cancer. The aim is to evaluate if the tumor to cervical spinal cord standardized uptake ratio (SUR) can improve the reproducibility of a model to determine the metabolic tumor volume (MTV) on FDG-PET/CT in a multicenter setting.

Materials and methods: Ninety-five radiotherapy planning FDG-PET/CT-scans of patients with head and neck cancer were analyzed using the Bland–Altman method to evaluate differences in FDG-uptake in the cervical spinal cord and the mediastinal blood pool. Non-linear regression analysis was used to determine the optimal MTV using the gross tumor volume (GTV) as ground truth and a spatial overlap-index as statistical validation metric. Reproducibility was evaluated using the Bland-Altman method and external validation was performed in an independent dataset consisting of 62 patients.

Results: Bland–Altman's analyses demonstrated equivalence of FDG-uptake in the mediastinal blood pool and the cervical spinal cord. Reproducibility of the models improved when using SUR instead of SUV. These results were confirmed in the validation cohort.

Conclusion: The use of the tumor to cervical spinal cord SUR instead of SUV improves the reproducibility of a model to determine the MTV on FDG-PET/CT in a multicenter setting. This study indicates that SUR may be preferred over SUV based approaches.

In recent years, ^{18}F-fluorodeoxy-D-glucose positron emission tomography/computed tomography (FDG-PET/CT) imaging has an emerging role in staging, radiation treatment planning and treatment response assessment in head and neck squamous cell carcinoma [1].

The standardized uptake value (SUV) is the most commonly used (semi-)quantitative parameter for analysis of oncologic FDG-PET imaging [2]. However, many small physiological and technical factors can add up to considerable variations in SUV, up to 50% or more [3]. As a consequence of significant variations in image acquisition and reconstruction techniques, non-standardized quantitative results are not interchangeable between institutions [4]. Because of this poor reproducibility of the SUV, its usefulness has been heavily criticized [5].

To ensure the interchangeability of quantitative data between institutions, procedural guidelines aiming to harmonize acquisition and reconstruction of oncologic FDG-PET imaging were developed [2,6,7]. The reproducibility of quantitative data can be further improved by normalization of SUV recoveries with an internal, image-derived standard, also known as 'tumor-to-background ratio' [8]. Commonly used internal standards are the liver and the mediastinal blood pool [9,10]. Recently it was shown that an image derived input function using the (aortic) blood pool, was highly correlated with the arterial plasma time-activity input function to estimate the metabolic rate of glucose [11,12]. Additionally, it was demonstrated that the tumor to (aortic) blood ratio, referred to as the 'standardized uptake ratio (SUR)', has a higher correlation

* Corresponding author at: Radboud University Medical Center, Department of Radiation Oncology, Route 874, P.O. Box 9101, Nijmegen 6500 HB, The Netherlands.
E-mail address: Sven.vandenBosch@radboudumc.nl (S. van den Bosch).

https://doi.org/10.1016/j.radonc.2018.06.037
0167-8140/

to the metabolic rate of glucose than SUV [13]. For this reason it was suggested that SUR should be considered to replace SUV [13].

In radiotherapy for head and neck cancer, however, the mediastinal blood pool is usually not within the field of view of the radiotherapy planning PET/CT-scan. An alternative structure located in the head and neck area that may be used as an internal standard is the cervical spinal cord. Literature reports that FDG-uptake in the cervical spinal cord has both little inter-patient variability and is stable over time [14,15].

The first objective of this study was to determine if FDG-uptake in the cervical spinal cord and the mediastinal blood pool are equivalent and can be used interchangeably. The second objective was to investigate if the use of the tumor to cervical spinal cord SUR instead of SUV can improve the reproducibility of a model to determine the metabolic tumor volume (MTV) on FDG-PET/CT in a multicenter cohort of patients with head and neck squamous cell carcinoma.

Materials and methods

Patient selection

For the primary analysis, a total of 95 patients who received definitive radiotherapy for head and neck squamous cell carcinoma were eligible for this retrospective analysis. Inclusion criteria were the acquisition of a radiotherapy planning FDG-PET/CT-scan on which the arc of the aorta was within the field of view of the PET/CT-scan. For the secondary analysis, an additional independent cohort of 62 patients was included to validate the results of the primary analysis. In total, 157 patients were included and treated between 2013 and 2017 in three tertiary head and neck clinics in the Netherlands. Patient and tumor characteristics are shown in Table 1. A flow diagram of the different steps of the analysis in this study is provided as Supplementary data.

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.radonc.2018.06.037.

Table 1
Patient and tumor characteristics.

Characteristics	Primary dataset No. of patients (n = 95)	Validation dataset No. of patients (n = 62)
Median age at diagnosis (years)	64.0 (IQR: 58.0–68.6)	62.5 (IQR: 56.9–67.4)
Gender distribution		
Males	64 (67.4%)	45 (72.6%)
Females	31 (32.6%)	17 (27.4%)
Primary tumor site		
Nasopharynx	3 (3.2%)	0 (0.0%)
Oropharynx	52 (54.7%)	35 (56.5%)
Larynx	28 (29.5%)	19 (30.6%)
Hypopharynx	12 (12.6%)	8 (12.9%)
Stage		
II	19 (20.0%)	12 (19.4%)
III	27 (28.4%)	21 (33.9%)
IV	49 (51.6%)	29 (46.8%)
T-classification		
T1	2 (2.1%)	3 (4.8%)
T2	37 (38.9%)	22 (35.5%)
T3	32 (33.7%)	21 (33.9%)
T4	24 (25.3%)	16 (25.8%)
N-classification		
N0	37 (38.9%)	33 (53.2%)
N1	16 (16.8%)	8 (12.9%)
N2a	4 (4.2%)	5 (8.1%)
N2b	24 (25.3%)	10 (16.1%)
N2c	14 (14.7%)	6 (9.7%)

Abbreviations: IQR = inter quartile range.

This analysis was conducted according to the ethical principles for medical research involving human subjects as stated in the Declaration of Helsinki and in the ICH Good Clinical Practice guidelines. This study was exempt from approval by an ethics committee and the need to obtain informed consent, because of the retrospective character of this study and the processing of anonymized data only.

Image acquisition and reconstruction

Contrast enhanced radiotherapy planning FDG-PET/CT-scans were acquired in one session on EARL accredited state-of-the-art integrated PET/CT-scanners (Siemens BioGraph 40 mCT or TruePoint). Patients were immobilized using a custom-made head, neck and shoulders mask. Patients were kept fasted for at least 6 h before ^{18}F-FDG injection. Image acquisition was performed 60 min after ^{18}F-FDG injection applying the European Association of Nuclear Medicine (EANM) procedural guidelines for tumor PET imaging v1.0 [2]. Because the EARL accreditation program comprises extensive image- and calibration quality control and (cross)-calibrations between institutions, no additional calibrations were performed in the context of this study. Technical characteristics of PET/CT image acquisition and reconstruction are provided as Supplementary data. Because PET and CT imaging were acquired in one session using an immobilization mask, PET and CT were co registered 'as scanned' (thus without need for further translations or rotations). PET image series were not resampled for analysis. Analyses of imaging were performed in Pinnacle^3 v9.710 (Philips Medical Systems).

Delineation of internal standards

The cervical spinal cord was delineated by drawing a circular region of interest (Ø 8 mm) in the center of the cervical spinal cord on consecutive CT-slices, from the level of the cranial border of the hyoid bone down to the caudal border of the inferior thyroid notch (Fig. 1). The mediastinal blood pool was delineated by drawing a circular region of interest (Ø 15 mm) in the center of the arc of the aorta on 4 consecutive CT-slices. Delineation close to intravascular calcifications or to the vascular wall was avoided. Mean

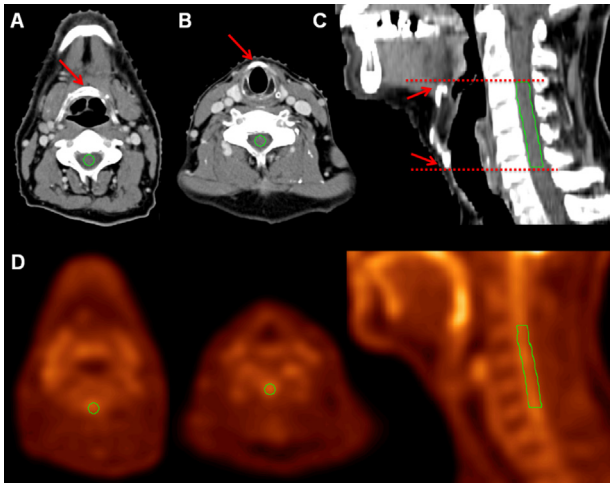

Fig. 1. Delineation of the cervical spinal cord. The cervical spinal cord is delineated on the CT-scan by drawing a circular region of interest (Ø 8 mm) in the center of the cervical spinal cord, starting from the level of the (A) cranial border of the hyoid bone through the (B) caudal border of the inferior thyroid notch. View of the region of interest in the sagittal plane (C). Projection of the cervical spinal cord on the FDG-PET-scan (D). Abbreviations: FDG = ^{18}F-fluorodeoxy-D-glucose; PET = Positron Emission Tomography; CT = Computed Tomography.

FDG-uptake values were used for analysis. Both regions of interest approximated a volume of ±2 cc.

Gross tumor volume delineation

The gross tumor volumes (GTV) of the primary tumor and each nodal metastases were manually delineated by consensus on the radiotherapy planning CT-scans by an expert panel of three experienced radiation oncologists specialized in head and neck cancer. CT-based delineation was performed according to current clinical protocols using information from clinical examination and co-registered diagnostic magnetic resonance imaging (MRI) in case of oropharyngeal cancers. Necrotic lymph nodes with irrefutably disturbed FDG-distribution were excluded from analysis.

Metabolic tumor volume

The threshold (percentage of maximum FDG-uptake) for the MTV having the best spatial overlap with the manually delineated GTV was determined for all lesions in all patients. For this purpose, a total of 18 MTVs were created for each lesion, using thresholds ranging from 10% to 95% of the maximum FDG-uptake with 5% intervals. For each MTV, a contour of the overlapping region with the manually delineated GTV was created. The volume and

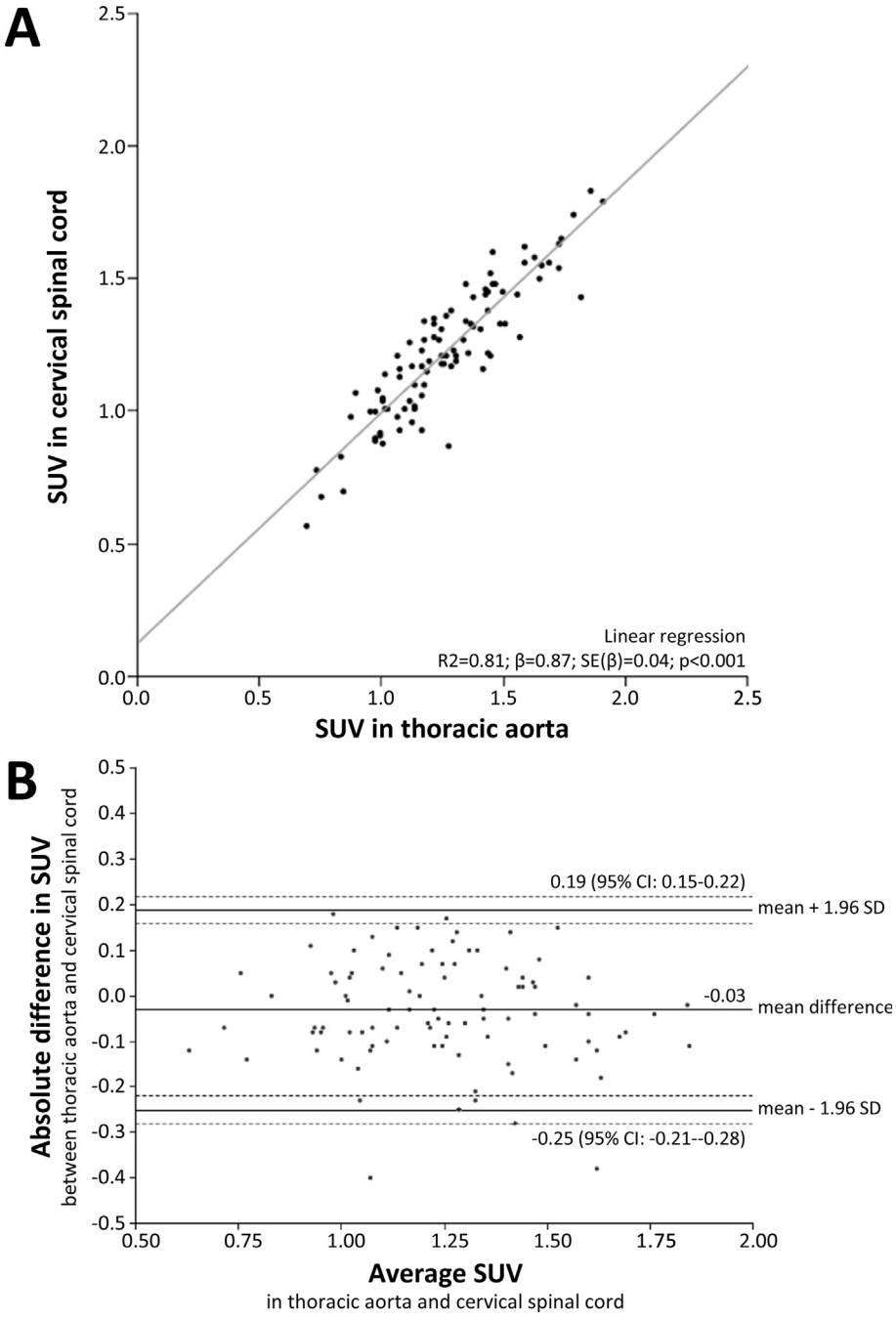

Fig. 2. FDG-uptake in the mediastinal blood pool and cervical spinal cord. (A) Scatter plot showing the relationship between the SUV of the mediastinal blood pool and the cervical spinal cord. (B) Bland–Altman's plot showing the differences in SUV between the mediastinal blood pool and the cervical spinal cord. Abbreviations: SUV = Standardized Uptake Value; SD = Standard Deviation; CI = Confidence Interval.

maximum FDG-uptake of the GTV, MTVs and the overlapping regions were exported for analysis.

The method of classification errors (CE) was used as a statistical validation metric to evaluate spatial overlap quality of the manually delineated GTV with each of the MTVs [16]. An important advantage of the CE method is that it does not only take volume into account but also the spatial position and shape of the contours due to the use of both false-positive and false-negative volumes (Eq. (1)). The CE can range from 0 to infinite, in which a lower CE implies better spatial overlap. The threshold used for the 'optimal MTV' (having the lowest CE) was registered as the 'optimal threshold'.

$$\text{Classifications Errors (CE)} = \frac{\text{false positive volume} + \text{false negative volume}}{\text{volume of gross tumor volume}} \quad (1)$$

Statistical analysis

All analyses were performed using SPSS v22 for Windows (IBM Corporation). Data characterized by normal distribution were expressed as means with 95% confidence interval (95% CI). Parameters not normally distributed were expressed as median with the interquartile range (IQR). In case of a normal distribution, comparison of means was performed by applying the Student *T*-test for paired- or unpaired data, as appropriate. Two tailed *p*-values ≤ 0.05 were considered to be significant.

Evaluation of internal standards was performed in the primary dataset. The bias (mean difference) and the 'limits of agreement' (95% CI of the bias) were calculated to evaluate for acceptable agreement between the SUV in the mediastinal blood pool and cervical spinal cord. To evaluate the magnitude of differences, the absolute differences between SUVs were plotted against the average SUV of the mediastinal blood pool and the cervical spinal cord according to the Bland–Altman method [17].

Models of the relationship between the maximum FDG-uptake of a lesion (for both SUV_{max} and SUR_{max}) and the optimal threshold for the MTV (having the best spatial overlap with the GTV) were acquired using non-linear least squares regression analysis based on data of the primary dataset. An inverse model [$y = a + (b/x)$], a power model [$y = a \cdot x^b$] and a sigmoidal model [$y = e^{a + (b/x)}$] were used to fit the data and the model having the best fit was used for analysis. The tumor to cervical spinal cord standardized uptake ratio (SUR_{max}) was calculated using Eq. (2).

$$SUR_{max} = \frac{SUV_{max} \text{ of lesion}}{SUV_{mean} \text{ of cervical spinal cord}} \quad (2)$$

Reproducibility of the models was evaluated using the Bland–Altman method by assessing the differences in threshold for and volume of the MTV that was predicted by the model and the empirically determined optimal one, using SUV_{max} and SUR_{max} [17]. The validation dataset was used to re-evaluate the reproducibility of the models that were developed based on the primary dataset.

Results

Internal standards were evaluated in the primary dataset. The mean SUV in the mediastinal blood pool was 1.26 (95% CI: 0.75–1.77) and 1.23 (95% CI: 0.74–1.72) in the cervical spinal cord. There were no inter-institutional differences of SUV in the mediastinal blood pool ($p > 0.37$) or in the cervical spinal cord ($p > 0.65$). The SUVs in both structures were strongly correlated with a Pearson correlation coefficient of 0.90 ($p < 0.001$) (Fig. 2A). A Bland–Altman plot shows the differences in SUV between the mediastinal blood pool and the cervical spinal cord (Fig. 2B). There were no inter-institutional differences in absolute differences of SUV between the mediastinal blood pool and the cervical spinal cord ($p > 0.14$).

For both datasets, GTVs were delineated on CT by consensus and MTVs were created in all patients. The MTV having the best spatial overlap (i.e. having the lowest CE) with the GTV could be determined for all lesions. Characteristics of contours are shown in Table 2.

Based on the primary dataset, non-linear least squares regression analysis was used to create models describing the relationship between SUV_{max}/SUR_{max} and the optimal threshold for the MTV (having the best spatial overlap with the GTV). The power model fitted the data best. The model using SUV_{max} is described by Eq. (3) (Fig. 3A) and the model using SUR_{max} is described by Eq. (4) (Fig. 3B).

$$\text{Threshold}\,(\%\,SUV_{max}) = 119.83 \cdot (SUV_{max})^{-0.66} \quad (3)$$

$$\text{Threshold}\,(\%\,SUR_{max}) = 116.93 \cdot (SUR_{max})^{-0.75} \quad (4)$$

Bland–Altman's plots show the differences in threshold and volume of the MTVs that were predicted by the model when using SUV versus SUR and the empirically determined optimal ones for the primary dataset (Fig. 3) and the validation dataset (Fig. 4). In the primary dataset, the 'limits of agreement' were significantly

Table 2
Delineation and isocontouring.

Characteristics	Primary dataset No. of patients ($n = 95$)	Validation dataset No. of patients ($n = 62$)
Primary tumors		
-No. of lesions	96	62
-Volume GTV* (cc)	15.4 (IQR: 7.6–28.5)	10.6 (IQR: 6.6–17.3)
-Volume PET** (cc)	12.7 (IQR: 5.7–21.4)	8.6 (IQR: 4.9–15.2)
-SUR_{max}	6.7 (IQR: 4.6–9.0)	6.2 (IQR: 4.4–8.0)
-Threshold ($\%_{max}$)	30 (IQR: 20–40)	30 (IQR: 25–35)
-Optimal CE	0.35 (IQR: 0.26–0.43)	0.32 (IQR: 0.21–0.45)
Nodal metastases		
-No. of lesions	133	80
-Volume GTV* (cc)	2.4 (IQR: 1.4–5.4)	2.1 (IQR: 1.0–4.8)
-Volume PET** (cc)	1.5 (IQR: 0.8–3.9)	1.3 (IQR: 0.7–3.8)
-SUR_{max}	2.4 (IQR: 1.8–4.3)	2.5 (IQR: 1.9–3.6)
-Threshold ($\%_{max}$)	60 (IQR: 40–75)	60 (IQR: 40–70)
-Optimal CE	0.55 (IQR: 0.36–0.79)	0.55 (IQR: 0.40–0.75)

Abbreviations: GTV = gross tumor volume; PET = positron emission tomography; CE = classification errors; IQR = inter quartile range; CT = computed tomography; FDG = fluorodeoxyglucose; SUR = standardized uptake ratio.

* Consensus GTV delineation on radiotherapy planning CT-scan.

** FDG-PET based isocontour having the best spatial overlap with the CT-based GTV.

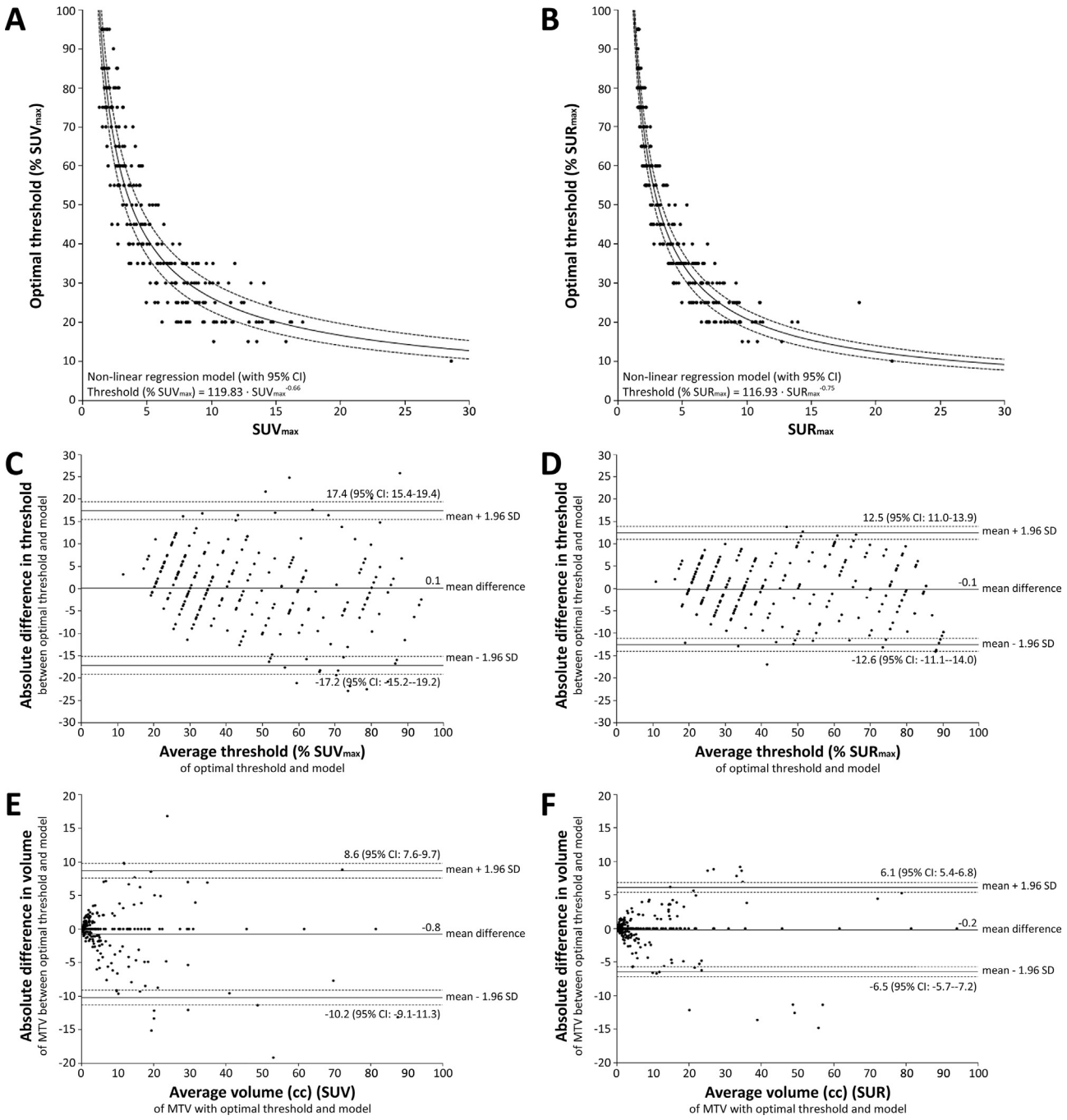

Fig. 3. Reproducibility of the models in the primary dataset (A + B). Non-linear regression models describing the relationship between the threshold for the optimal MTV (having the best spatial overlap with the GTV) and SUV_{max} (A) and SUR_{max} (B). (C + D) Bland–Altman's plot showing the differences in threshold between the threshold for the optimal MTV and the one predicted by the models using SUV_{max} (C) and SUR_{max} (D) in the primary dataset. (E + F) Bland–Altman's plot showing the differences in volume of the MTV between the optimal MTV and the one predicted by de models using SUV_{max} (E) and SUR_{max} (F) in the primary dataset. Abbreviations: MTV = Metabolic Tumor Volume; GTV = Gross Tumor Volume; SUV = Standardized Uptake Value; SUR = Tumor to cervical spinal cord standardized uptake ratio; SD = Standard Deviation; CI = Confidence Interval.

narrower when using SUR as compared to SUV and decreased from 35% to 25% ($p = 0.04$) for threshold (Fig. 3C, D) and from 18.8 cc to 12.6 cc for volume ($p < 0.01$) (Fig. 3E, F). Linear regression analysis showed no relevant dependency between absolute differences and average thresholds and volumes when using SUV or SUR.

In the secondary validation dataset, the 'limit of agreement' was significantly narrower for threshold when using SUR instead of SUV and decreased from 32% to 23% ($p = 0.05$) (Fig. 4A, B). For volume, the 'limit of agreement' decreased from 12.1 cc to 9.6 cc when using SUR instead of SUV, but did not reach statistical significance ($p = 0.13$) (Fig. 4C, D). Linear regression analysis showed a negative dependency between absolute difference and average threshold and volume when using SUV, indicating a systematic underestimation of the MTV using the SUV-based model (Fig. 4A, C). No relevant dependencies were observed for threshold or volume when using SUR.

Discussion

It is well established that the reproducibility of quantitative FDG-PET data can be improved by normalization of SUV recoveries

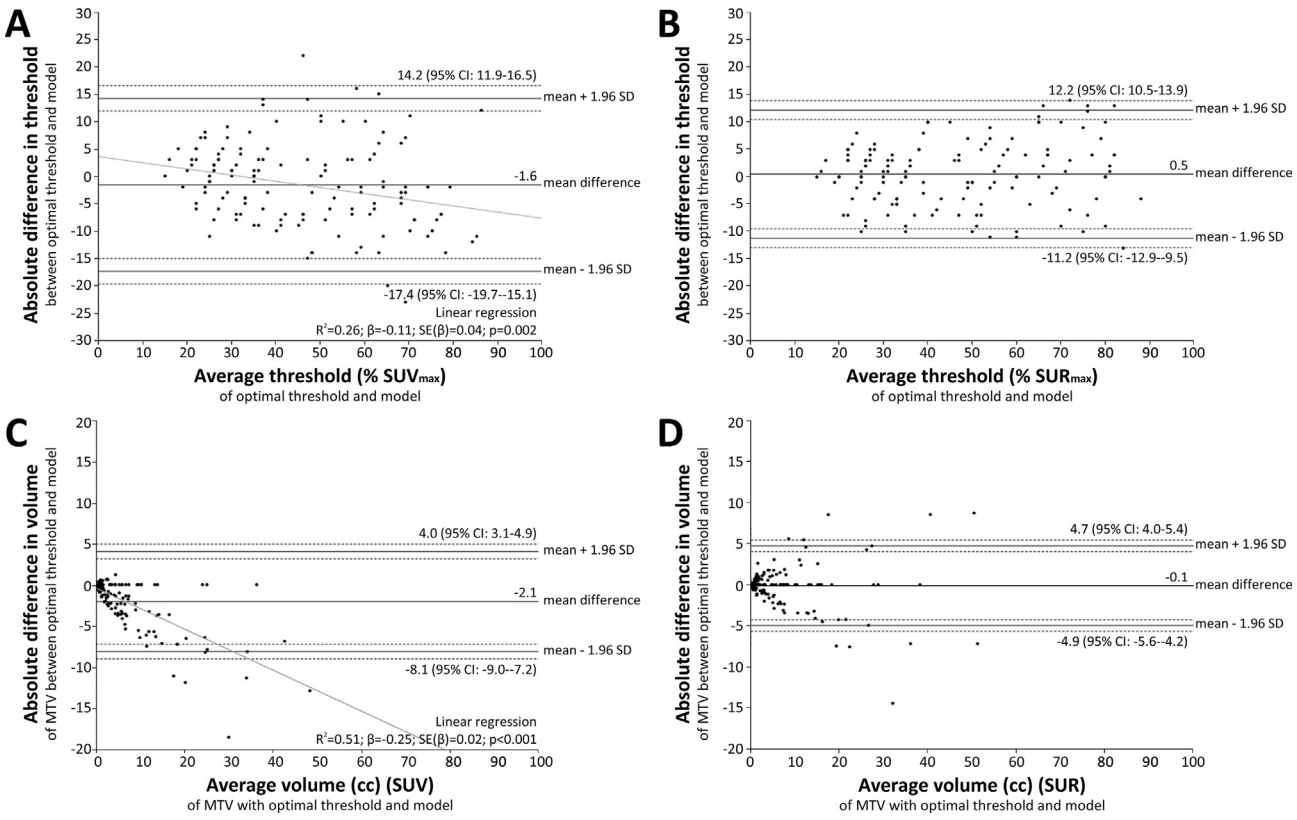

Fig. 4. Reproducibility of the models in the validation dataset (A + B). Bland–Altman's plot showing the differences in threshold between the threshold for the optimal MTV and the one predicted by the models using SUV_{max} (A) and SUR_{max} (B) in the validation dataset. (C + D) Bland–Altman's plot showing the differences in volume of the MTV between the optimal MTV and the one predicted by de models using SUV_{max} (C) and SUR_{max} (D) in the validation dataset. Abbreviations: MTV = Metabolic Tumor Volume; SUV = Standardized Uptake Value; SUR = Tumor to cervical spinal cord standardized uptake ratio; SD = Standard Deviation; CI = Confidence Interval.

with an internal, image-derived standard [13,18,19]. In the current study, it was demonstrated that FDG-uptake within the cervical spinal cord and the mediastinal blood pool are equivalent and can be used interchangeably. The current study also shows that the use of the tumor to cervical spinal cord SUR instead of SUV improves the reproducibility of a model to determine the MTV on FDG-PET/CT in a large multicenter cohort of patients with head and neck squamous cell carcinoma. Furthermore, improvement of reproducibility of the model was also shown in an independent multicenter validation dataset when using SUR instead of SUV. As such, the standardized methods described in this manuscript are currently investigated prospectively in a clinical trial [20].

Commonly used internal image-derived standards such as the liver and the mediastinal blood pool are not ideal for use in radiotherapy for head and neck cancer because these structures are usually not within the field of view of the radiation treatment planning PET/CT-scan [9,10]. For this reason, there is need for an alternative structure in the head and neck area that can be used as internal image-derived standard. Daisne et al. investigated the lateral muscular massif of the neck as an internal image-derived standard for a source-to-background (SBR) based segmentation algorithm in head and neck squamous cell carcinoma [21,22]. However, the use of muscle as internal image-derived standard in FDG-PET imaging is less ideal, because muscular contraction due to patient discomfort or anxiety can result in an increased physiological FDG-uptake in skeletal muscles [23]. This can introduce uncontrollable variations in FDG-uptake and can result in a poor inter-patient reproducibility of the internal standard. In contrast, FDG-uptake in the cervical spinal cord has both little inter-patient variability, is stable over time and is not influenced by muscular activity [14,15]. The current analysis shows that the cervical spinal cord has equivalent FDG-uptake as the mediastinal blood pool. As such, the cervical spinal cord is ideal to be used as internal image-derived standard in FDG-PET imaging of the head and neck area.

The next step in this analysis was to investigate if the use of the tumor to cervical spinal cord SUR instead of SUV may improve the reproducibility of quantitative FDG-PET data. This was done by creating a model to determine the MTV on FDG-PET/CT.

Unique in the development of this model is the use of 'real life' clinical FDG-PET/CT imaging data of patients with head and neck squamous cell carcinoma. This has advantages over phantom-based experiments, which use symmetrical volumes with homogeneous activity and a sharp demarcation from the background activity [22]. Obviously, 'real life' tumors can have complex shapes with heterogeneous distribution of FDG-uptake and can have variable background activity. For these reasons, models based on phantom experiments may have a limited performance in 'real life' clinical scenarios. Conversely, the advantage of phantom-based experiments is a known 'ground truth'. For this analysis, we chose the CT-based delineation of the GTVs as a practical approximation of 'ground truth' for the development of the model. Special care was taken to eliminate inter-observer variability. First, delineation of the GTVs was performed by the consensus of three radiation oncologists specialized in head and neck cancer. Second, the optimal threshold for the MTV having the best spatial overlap with the GTVs was determined using a spatial overlap-index as statistical validation metric.

A strength of the current analysis is that it was conducted in a multi-center setting, which is quite challenging for evaluation of quantitative FDG-PET imaging. A total of 95 radiation treatment

planning FDG-PET/CT-scans from 3 tertiary head and neck clinics were used for the primary analysis. Consequently, the FDG-PET/CT-scans were acquired on 3 different PET/CT-scanners and different reconstruction methods were used. It was demonstrated that the reproducibility of the model to determine the MTV on FDG-PET/CT can be significantly improved using the tumor to cervical spinal cord SUR instead of SUV. Analysis of an independent multicenter validation dataset consisting of 62 radiation treatment planning FDG-PET/CT-scans confirms an improved reproducibility of the model when using SUR instead of SUV for threshold. For volume, a similar trend was observed, but did not reach statistical significance. This can be explained by better compliance with the EANM procedural guidelines regarding image acquisition in the validation dataset, having a more accurate FDG incubation time compared to the primary dataset (Supplementary data). Most important is the systematic underestimation of the MTV that was observed when using the SUV-based model in the validation dataset while this was not the case for the SUR-based model. This again suggests that the use of SUR is more robust than SUV is, because it is less susceptible for small variations in image acquisition.

The present study has limitations. To determine the SUR, delineation of the cervical spinal cord must be performed. This could potentially introduce new intra- and inter-observer variability effects. However, this equally applies to traditional internal standards (e.g. liver or mediastinal blood pool). To minimize observer variability, clear instructions for delineation of the cervical spinal cord are described in this manuscript and illustrated in Fig. 1. Another potential limitation of the current study is the lack of histological confirmation of the GTVs. However, the objective of this analysis was not to assess the absolute accuracy of the model to determine the MTV on FDG-PET/CT, but to investigate if the use of the tumor to cervical spinal cord SUR can improve the reproducibility of the model. In literature, there are only two series on head and neck cancer available using histological specimen as ground truth reference [21,24]. These series show a slight overestimation of the extent of disease for CT-based GTV delineation compared to the histological ground truth reference. Therefore, a potential pitfall of accepting the CT-based GTV as 'ground truth' for the model in the current analysis may be a systematic overestimation of the extent of disease.

In conclusion, this study demonstrates that FDG-uptake in the cervical spinal cord and the mediastinal blood pool is equivalent and can be used interchangeably. The use of the tumor to cervical spinal cord SUR instead of SUV improves the reproducibility of a model to determine the MTV on FDG-PET/CT in a multicenter setting. These results were confirmed in an independent multicenter validation cohort. This study indicates that SUR may be preferred over SUV based approaches.

Conflict of interest statement

All authors declare having no conflict of interest related to the content of this manuscript.

References

[1] Hoeben BA, Bussink J, Troost EG, Oyen WJ, Kaanders JH. Molecular PET imaging for biology-guided adaptive radiotherapy of head and neck cancer. Acta Oncol 2013;52:1257–71.
[2] Boellaard R, O'Doherty MJ, Weber WA, FDG PET and,, et al. 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. Standards for PET image acquisition and quantitative data analysis. J Nucl Med 2009;50:11S–20S.
[4] Westerterp M, Pruim J, Oyen W, et al. Quantification of FDG PET studies using standardised uptake values in multi-centre trials: effects of image reconstruction, resolution and ROI definition parameters. Eur J Nucl Med Mol Imaging 2007;34:392–404.
[5] Keyes Jr JW. SUV: standard uptake or silly useless value? J Nucl Med 1995;36:1836–9.
[6] 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 2014.
[7] Delbeke D, Coleman RE, Guiberteau MJ, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. J Nucl Med 2006;47:885–95.
[8] Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. J Nucl Med 2009;50:122S–50S.
[9] Boktor RR, Walker G, Stacey R, Gledhill S, Pitman AG. Reference range for intrapatient variability in blood-pool and liver SUV for 18F-FDG PET. J Nucl Med 2013;54:677–82.
[10] Paquet N, Albert A, Foidart J, Hustinx R. Within-patient variability of (18)F-FDG: standardized uptake values in normal tissues. J Nucl Med 2004;45:784–8.
[11] de Geus-Oei LF, Visser EP, Krabbe PF, et al. Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET studies. J Nucl Med 2006;47:945–9.
[12] Burger IA, Burger C, Berthold T, Buck A. Simplified quantification of FDG metabolism in tumors using the autoradiographic method is less dependent on the acquisition time than SUV. Nucl Med Biol 2011;38:835–41.
[13] 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.
[14] Chong A, Song HC, Byun BH, et al. Changes in (18)f-fluorodeoxyglucose uptake in the spinal cord in a healthy population on serial positron emission tomography/computed tomography. Chonnam Med J 2013;49:38–42.
[15] Do BH, Mari C, Tseng JR, Quon A, Rosenberg J, Biswal S. Pattern of 18F-FDG uptake in the spinal cord in patients with non-central nervous system malignancy. Spine (Phila Pa 1976) 2011;36:E1395–401.
[16] Hatt M, Lamare F, Boussion N, et al. Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET. Phys Med Biol 2007;52:3467–91.
[17] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
[18] van den Hoff J, Hofheinz F. Repeatability of tumor SUV quantification: the role of variable blood SUV. J Nucl Med 2015;56:1635–6.
[19] Weber WA, Gatsonis C, Siegel B. Reply: repeatability of tumor SUV quantification: the role of variable blood SUV. J Nucl Med 2015;56:1636.
[20] van den Bosch S, Dijkema T, Kunze-Busch MC, et al. Uniform FDG-PET guided GRAdient Dose prEscription to reduce late Radiation Toxicity (UPGRADE-RT): study protocol for a randomized clinical trial with dose reduction to the elective neck in head and neck squamous cell carcinoma. BMC Cancer 2017;17:208.
[21] Daisne JF, Duprez T, Weynand B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology 2004;233:93–100.
[22] Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V. Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol 2003;69:247–50.
[23] Shreve PD, Anzai Y, Wahl RL. Pitfalls in oncologic diagnosis with FDG PET imaging: physiologic and benign variants. Radiographics 1999;19:61–77. quiz 150-1.
[24] Ligtenberg H, Jager EA, Caldas-Magalhaes J, et al. Modality-specific target definition for laryngeal and hypopharyngeal cancer on FDG-PET, CT and MRI. Radiother Oncol 2017;123:63–70.