Journal of Nuclear Medicine, published on May 25, 2017 as doi:10.2967/jnumed.116.189050

Exploring New Multimodal Quantitative Imaging Indices for the Assessment of Osseous Tumour Burden in Prostate Cancer using ^{68}Ga-PSMA-PET/CT

Marie Bieth[1,2*], Markus Krönke[1*], Robert Tauber[3], Marielena Dahlbender[3], Margitta Retz[3], Stephan G. Nekolla[1], Bjoern Menze[2], Tobias Maurer[3], Matthias Eiber[1,4#], Markus Schwaiger[1#]

[1]Department of Nuclear Medicine, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[2]Department of Informatics, Technical University of Munich, Munich, Germany

[3]Department of Urology, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[4]Department of Molecular and Medical Pharmacology, David Greffen School of Medicine at UCLA, University of California, Los Angeles, California, USA

*both authors contributed equally to this work; [#] joint senior authorship

Short running title: Quantitative indices for PSMA-PET/CT
Word count: 5123 words

Keywords: Image Processing, PET/CT, Imaging biomarker, Prostate cancer

Financial Disclosure:
This work was partially funded by the German ministry for education and research (Bundesministerium für Bildung und Forschung) under Grant Agreement No. 01IS12057, by the European Union Seventh Framework Program (FP7) under Grant Agreement No. 294582 ERC MUMI and by the German Research Society (DFG) under Grant agreement SFB 824.

Corresponding author:

Matthias Eiber, MD,

Department of Nuclear Medicine

Klinikum rechts der Isar der TU München

Ismaninger Str. 22

81675 Munich, Germany

mail: matthias.eiber@tum.de

phone: +49 89 4140 6085

fax: +49 89 4140 4950

First author:

Marie Bieth, PhD candidate,

mail: marie.bieth@tum.de

phone: +49 89 4140 4576

fax: +49 89 4140 4950

Email addresses:

MK: markus.kroenke@mri.tum.de

RT: r.tauber@tum.de

MD: m.dahlbender@gmail.com

MR: margitta.retz@tum.de

SN: stephan.nekolla@tum.de

BM: bjoern.menze@tum.de

TM: tobias.maurer@tum.de

MS: markus.schwaiger@tum.de

ME: matthias.eiber@tum.de

ABSTRACT

Positron-emission-tomography (PET) combined with computed-tomography (CT) and prostate-specific-membrane-antigen (PSMA) ligands has gained significant interest for staging prostate cancer (PC). In this study, we propose two multimodal quantitative indices as imaging biomarker for the assessment of osseous tumour burden using ^{68}Ga-PSMA-PET/CT and present preliminary clinical data. We define two Bone-PET-Indices (BPI) that incorporate anatomical information from CT and functional information from ^{68}Ga-PSMA-PET: BPI_{VOL} is the percental bone volume affected by tumour. BPI_{SUV} additionally considers the level of PSMA-expression. We describe a semi-automatic computation method based on segmentation of bones in CT and of lesions in PET. Data from 45 patients with castration-resistant PC and bone metastases during Radium-223-dichloride were retrospectively analysed. We evaluated the computational stability and reproducibility of the proposed indices, and explored their relation to the prostate-specific-antigen (PSA) blood value, the Bone-Scan-Index (BSI) and disease classification using the PET response criteria in solid tumours (PERCIST). On the technical side, BPI_{VOL} and BPI_{SUV} showed an inter-observer maximum difference of 3.5% and their computation took only a few minutes. On the clinical side, BPI_{VOL} and BPI_{SUV} showed significant correlations with BSI ($r=0.76$ and 0.74 respectively, $p<0.001$) and PSA-values ($r=0.57$ and 0.54 respectively, $p<0.01$). When comparing the proposed indices against expert rating using PERCIST, BPI_{VOL} and BPI_{SUV} showed better agreement than BSI, indicating their potential for objective response evaluation. We propose the evaluation of BPI_{VOL} and BPI_{SUV} as imaging biomarkers for ^{68}Ga-PSMA-PET/CT in a prospective study exploring their potential for outcome prediction in patients with bone metastases from PC.

INTRODUCTION

Patients with PC, even when treated early, can later develop recurrence with cancer spread to other structures, in particular the bones. Both for primary tumour and recurrence, bone scintigraphy, CT and PET/CT are commonly used for staging. For bone scintigraphy, the BSI *(1)* and its automatic computation method *(2)* can be used for quantitative analysis. However, in the absence of SPECT or SPECT/CT, bone scintigraphy is a two-dimensional modality which lacks detailed anatomical information, has suboptimal specificity and cannot reveal lymph node and visceral metastases. Due to its high detection rates and superb specificity for PC lesions ^{68}Ga-PSMA-PET/CT is increasingly used for staging of bone metastases instead of bone scintigraphy *(3,4)*. So far, for cross-sectional imaging only qualitative to semi-quantitative methods for assessment of tumour burden and response such as RECIST *(5)* for CT and PERCIST *(6,7)* for PET/CT have been developed. Additional important limitation for PC staging are: 1) RECIST is suboptimal for bone metastases staging as osteoblastic metastases without extraosseous tumour involvement are mainly regarded as non-measurable and 2) PERCIST has only been evaluated for ^{18}F-FDG-PET, which is rarely used in these patients.

Therefore, a comprehensive quantitative imaging biomarker which is capable of measuring the whole-body tumour burden, exploring the potential of a ^{68}Ga-PSMA, is an unmet clinical need. A first approach for the quantitative assessment of ^{18}F-Fluoride-PET has been developed by Etchebere et al. *(8)*. However, it is unimodal and neglects the anatomical information contained in the CT image, therefore impeding inter-patients comparability by e.g. not taking into account differences in patient size.

In this work, our aim was to define multimodal quantitative imaging indices using hybrid information from a ^{68}Ga-PSMA PET/CT scan. In a first step towards full body quantification, we focused on bone tumour burden. We also developed a method to compute the indices automatically with possible manual corrections, so that they can be easily implemented in clinical practice. In addition, we have performed preliminary clinical analyses applying this method to a retrospective cohort of PC patients with bone metastases that underwent ^{68}Ga-PSMA PET/CT under Radium-223-dichlorid and compared the results to a reference standard of clinical expert reading, BSI and serum PSA.

MATERIAL AND METHODS

Bone PET Index

Definition. BSI is defined as the percentage of skeletal mass affected by tumour calculated on a bone scintigram *(1)*. Because bone scintigraphy is intrinsically 2D and lacks detailed anatomical information, a standard weighting of bones is incorporated in the calculation. The commercial software EXINI Bone ^{BSI} (EXINI Diagnostics AB, Lund Sweden) allows for automated calculation of the BSI. Inspired by this definition, we defined two new multimodal imaging indices for PET/CT: BPI_{VOL} as the percental bone volume (including bone marrow) affected by tumour and BPI_{SUV} additionally considering the target expression measured by the average standardised uptake value (SUV). In both indices, the anatomical information was extracted from the CT image while the functional information was extracted from the PET image, making them intrinsically multimodal. They were calculated as follows, where *N* is the number of pixels *p* that belong to bone metastases:

$$BPI_{VOL} = 100 \times \frac{Bone\ metastases\ volume}{Skeleton\ volume} \qquad \text{(no unit)}$$

$$SUV_{mean} = \frac{1}{N} \sum_{\substack{p\ in\ bone \\ metastases}} SUV(p) \qquad \text{(unit: g/mL)}$$

$$BPI_{SUV} = BPI_{VOL} \times SUV_{mean}/100 \qquad \text{(unit: g/mL)}$$

Notably, both BPI_{VOL} differ from the metabolic tumor volume (MTV) and BPI_{SUV} from the total lesion glycosis (TLG) due to normalization by skeleton volume. Contrarily to the calculation of BSI, no standard weighting of the bones is needed because patient-specific anatomical information from the CT image is used instead. In PET/CT, depending on the type of cancer,

regularly only the trunk and not the whole body is imaged. Therefore, arms and legs as well as part of the head were excluded from the computation. To achieve a standardised calculation of the BPI that accounted for variation in the field of view and allowed for inter- and intra-patient comparison, only the slices between the bottom of the ischium (easily recognized on CT) and the caudal edge of the sub-lingual gland (easily recognized due to glandular uptake in PET) were considered. Of note, in the computation of BSI by EXINI Bone^{BSI}, the forearms and lower legs are excluded as well.

Semi-automatic Computation Method. For computation of BPI_{VOL}, SUV_{mean} and BPI_{SUV}, a precise segmentation of the skeleton in CT, and of bone metastases in PET were necessary. This could have been done manually with appropriate software, but would have been time-consuming. We proposed instead an automatic method with possible manual corrections that we have implemented using the programming language Python *(9)*.

Our tool read images in DICOM format. PET and CT were affinely registered using the information contained in the DICOM headers. This was possible because both images were acquired on the same scanner during the same session. The bed was automatically removed from the CT by simple morphological operations.

On CT, the skeleton can be segmented by using its density in Hounsfield Units, which is higher than that of soft-tissue and air. Following the first two steps of the method of Kang et al. *(10)*, we used global thresholding of the CT image followed by local adaptive thresholding to obtain a first bone mask. Then we applied a supplementary morphological hole closing to obtain a segmentation of the bones and bone marrow. In case of heavy calcification or in case of artefact resulting, e.g., from implants, manual corrections were possible: corrections could be applied either with a brush, or by removing in one click the whole “bone” from a slice. Details on the computation method are given in a supplementary file.

^{68}Ga-PSMA usually does not exhibit unspecific uptake in the bones and bone marrow. Therefore, regions of the skeleton with increased uptake can be considered as bone metastases. Thus, the lesions were segmented by using a SUV-threshold on PET and restricting the result to the skeleton segmented from the CT image (Fig. 1). This use of anatomical information avoided manual removal of normal uptake sites (e.g. bladder, kidneys) as proposed in *(8)*. Even though CT and PET are acquired consecutively on the same scanner, e.g. breathing can cause misalignment in the region of the ribs. As the liver and spleen show high physiological uptake of ^{68}Ga-PSMA, projection of the ribs in CT on liver and spleen in PET can potentially generate false positives. Using our current tool, such false positives had to be manually corrected.

The final segmentation depended on the SUV-value used to threshold the PET as described in the previous paragraph. Rather than choosing an arbitrary threshold, we therefore developed a method to choose the threshold based on the maximum tolerated false positive rate using negative training patients. For each patient, the lowest SUV-threshold resulting in the given false positive rate was automatically computed. BPI_{VOL} values obtained during this search are shown for one training patient in Fig. 2. Using the maximum of the thresholds computed for all training patients ensured that the false positive rate did not exceed the maximum for any training patient. The effects of choosing different maximum tolerated false positive rates on BPI_{VOL} and BPI_{SUV} are presented in the Results section.

Patients cohort

Between January 2014 and February 2016, 45 patients with metastatic-castration resistant PC (mCRPC) with bone metastases but no organ or relevant lymph node (>3 cm) metastases

underwent a Radium-223-dichloride therapy (Xofigo®, Bayer Healthcare, Leverkusen, Germany) at our institution. Mean age of the patients was 71 (±8 years). All patients received therapy at a dose of 50 kBq/kg per therapy cycle in monthly intervals with up to 6 cycles. 15 patients that underwent a ^{68}Ga-PSMA PET/CT at our institution and were regarded as negative for bone metastases by an experienced nuclear medicine physician were also retrospectively randomly selected to serve as negative training cohort for the method.

All patients underwent ^{68}Ga-PSMA PET/CT within 4 weeks prior to initiation of Radium-223-dichloride therapy. Thirty-one patients also underwent bone scintigraphy. Thirty-two patients underwent additional ^{68}Ga-PSMA PET/CT and 22 of them bone scintigraphy 3 to 6 months after the first scan. Table 1 shows a summary of data available for all patients.

The institutional review board (IRB) of the Technical University Munich approved the retrospective analysis (permit 5665/13) and all subjects signed a written informed consent for anonymised evaluation and publication of their data.

Data acquisition

^{68}Ga-PSMA PET/CT was obtained approximately 54 minutes (± 8 min, range: 43-88 min) after injection of mean 133 MBq (379 MBq, range: 52-239 MBq) ^{68}Ga-labelled HBED-CC. A diagnostic CT scan was performed in the portal venous phase after intravenous injection of contrast agent (Imeron 300). Immediately after the CT, the PET scan was acquired with 6-8 bed positions (3-5 minutes per bed position). PET was reconstructed using ordered-subset expectation-maximisation with point spread function and time-of-flight information (3 iterations, 21 subsets) and corrected for normalisation, attenuation, scatter, randoms and decay. The transaxial pixel size

was 4.07 mm for PET and 1.52 mm for CT and the slice thickness was 5 mm for both. Tc99m-HDP whole-body bone scintigraphy was performed in planar imaging mode with an acquisition time of 1 minute/10 cm body height. Activity was body weight-adjusted (9 MBq/kg) and injected 3 hours before imaging.

Data and statistical analysis

To validate our new tool and the introduced BPI, we performed a reproducibility analysis. Randomly selected datasets of 10 mCRPC patients prior to application of Radium-223-dichloride were analysed by two trained observers applying manual corrections independently. The reproducibility threshold was then defined as the maximum absolute difference observed between both observers for each index. For BSI, Anand et al. defined the reproducibility threshold as 0.30 *(11)*.

For response assessment, BPI was compared to BSI and PERCIST by an experienced reader as well as PSA. BSI was computed from the bone scintigraphy images using the commercially available EXINI Bone^{BSI}. Response by PERCIST was evaluated by an experienced physician using recently published criteria *(6,7)* and criteria were adapted for ^{68}Ga-PSMA similarly to a recent work *(12)*. In brief, peak SUV value was measured in one to five target lesions and the appearance of new lesions was investigated.

As PERCIST is not quantitative, but only classifies the patient in progressive metabolic, stable metabolic disease or partial metabolic response, we also defined these categories for BPI and BSI using the respective reproducibility threshold: a change of magnitude smaller than the reproducibility threshold was considered as stable metabolic disease, an increase in value larger than the reproducibility threshold was considered as progressive disease and a decrease in value

larger than the reproducibility threshold was considered as partial metabolic response. Moreover, two separate analyses based on PERCIST criteria were performed: 1) metastatic status based on all types of target lesions (including potential new lymph node and visceral metastases, as prescribed by the criteria) and 2) metastatic status based on bone involvement only (to allow for direct comparison with BPI).

For comparing quantitative methods (i.e. BPI, BSI and PSA), we used the Pearson coefficient r. For all tests, a p-value smaller than 0.05 was considered significant.

RESULTS

Technical validation

Bone Segmentation and Manual Corrections. After manual correction, mean bone volume of the 45 treated patients was 4,184 cm³ (±503 cm³, range: 3,327-5,739 cm³). For 32 patients with two sequential ^{68}Ga-PSMA PET/CT, mean difference in computed bone volume between two scans was 66 cm³ (± 61cm³) with a maximum of 270 cm³. After bones had been segmented and manually corrected, false positives (e.g. in the rib cage) were corrected by an expert reader. On 54 scans from the patient cohort, an average of 3.8 cm³ false positives per patient had to be manually corrected. This represents an average difference in BPI_{VOL} of 0.0008 per patient.

Selection of SUV-Threshold for Lesion Segmentation. The SUV-threshold lesion segmentation was determined by using 15 negative training patients. For each patient, we computed the threshold that resulted in a BPI_{VOL} of 0.1 and 1 (equal to 0.1% and 1% of false positive voxels) respectively. Corresponding SUV-thresholds for all negative training patients were in the range of 1.15 to 1.95 (mean: 1.42) for a false positive threshold 1 and of 1.7 to 2.65 (mean: 2.06) for 0.1, respectively. Fig. 2 shows the different BPI_{VOL} values obtained for different thresholds for one negative training patient.

SUV-Threshold Influence on Lesion Segmentation. Based on these results, thresholds of 1.5 (mean value obtained for the scenario of 1% false positive results) and 3 (conservative approach ensuring a maximum of 0.1% false positive lesions in all patients) were used for the initial analysis of the baseline ^{68}Ga-PSMA PET/CT of all 45 patients. There was a strong correlation between the BPI_{SUV} values obtained with these two thresholds ($r=0.99$; $p<0.001$, Fig. 3). This also held for

BPI_{VOL} and SUV_{mean} (r=0.95 for both; p<0.001). Due to the high correlation between both values, we chose a threshold of 3 to ensure a high specificity of the BPI, with less of 0.1 of BPI_{VOL} being related to false positives. All following results were computed using a SUV-threshold of 3.

Reproducibility. Comparison from two independent observers using 10 randomly selected datasets showed a nearly perfect correlation (r=0.999; p<0.001 for both). The maximum observed percentage differences between both observers were 3.5 % for BPI_{SUV} and 2.2% for BPI_{VOL}. The maximum absolute difference was 0.055 for BPI_{SUV} and 0.37 for BPI_{VOL}. Rounding up these values, we defined 0.06 and 0.4 as reproducibility thresholds for the respective index. Note that a wide range of disease was present in the analysed patients (range of BPI_{SUV}: 0.09-3.39, BPI_{VOL}: 1.53-38.05).

Quantification using BPI_{VOL}, SUV_{mean} and BPI_{SUV}

BPI_{VOL}, SUV_{mean} and BPI_{SUV} before and after therapy. The average values of BPI_{VOL}, SUV_{mean} and BPI_{SUV} before therapy were 19.5, 8.3 and 1.59 respectively. After therapy, the average values were 26.0, 7.7 and 1.99. This represented changes of +33%, -7% and +25% respectively.

Correlation between BPI_{SUV} and BPI_{VOL} . BPI_{VOL} and BPI_{SUV} for all 45 mCRPC patients before therapy were strongly correlated (r=0.89, p<0.001, Fig. 4A). The percentage changes of BPI_{VOL} and BPI_{SUV} during therapy were very strongly correlated (r=0.97, p<0.001, Fig. 4B).

Correlation of BPI to clinical parameters

At baseline, BPI_{VOL} and BPI_{SUV} showed a moderate and significant correlation with BSI ($r=0.76$ and 0.74 respectively, $p<0.001$, Fig. 4C). There was a tendency to a stronger correlation with PSA-value for BPI_{VOL} and BPI_{SUV} ($r=0.57$ and 0.54 respectively, $p<0.01$) than for BSI ($r=0.49$, $p<0.01$).

A moderate correlation between change of BPI_{VOL} and BPI_{SUV} and percentage change of PSA-value after treatment was observed ($r=0.70$; $p<0.01$). There was no correlation of change in BSI with percentage change in PSA-value ($r=0.24$; $p=0.32$).

When compared to PERCIST for all type of target lesions (Table 2), BPI_{VOL}, BPI_{SUV} and BSI showed agreement for 65.6% (21/32), 68.7% (22/32) and 57.9% (11/19) of patients and opposite results for 25.0% (8/32), 15.6% (5/32) and 21.1% (4/19) respectively. When compared to PERCIST for bone involvement only (Table 2), BPI_{VOL}, BPI_{SUV} and BSI showed agreement for 62.5% (20/32), 68.7% (22/32) and 63.2% (12/19) of patients and opposite results for 12.5% (4/32), 6.2% (2/32) and 10.5% (2/19) respectively. An exemplary case of a patient showing divergent results with considerable decrease of PSMA-expression under therapy indicating response in BPI_{VOL} and BPI_{SUV} but progressive disease due to new lesions in PERCIST is given in Fig. 5. Notably, cases in which results from one method indicated stable disease but the other method indicated progression or response were not counted as opposite results.

DISCUSSION

In this study, we have described the BPI as new quantitative multimodal imaging indices for the assessment of bone metastases in ^{68}Ga-PSMA-PET/CT using an automatic computation method. We have shown that, on our patient cohort, small differences in SUV thresholds and small misalignments between PET and CT did not substantially influence the results. Different observers also obtained very similar values of the indices, showing a good inter-observer reproducibility. A small amount of manual correction was still necessary, especially due to calcifications and endoprothesis. The computation of both indexes with our software took only a few minutes, including manual corrections of skeleton segmentation and false positives. In contrast, a complete manual segmentation of a whole-body dataset would take several hours even for an experienced physician, since the skeleton and not only the lesions have to be segmented.

From a clinical perspective, our preliminary data indicated that BPI holds potential for quantitative response assessment. This was documented by a high correlation of BPI with BSI and PSA in mCRPC patients, and reasonable prediction of tumour response compared to PERCIST. We are aware, that these results only allow for a first estimation of the potential clinical usability due to the known limitations of theses comparators. Therefore, future prospective clinical studies using more objective endpoints (overall survival, radiographic progression-free-survival, skeletal adverse events) are necessary to fully investigate the potential of the proposed quantitative biomarkers for response prediction.

Compared to the imaging biomarkers FTV_{10} representing the volume of fluoride bone metastases and TLF_{10} representing the total fluoride metastatic uptake introduced by Etchebere et al. *(8)*, calculation of BPI_{VOL} and BPI_{SUV} was an intrinsically multimodal approach. TLF_{10} and FTV_{10} exclusively rely on the information from PET and no correction is possible with regard to

the patient size. In contrast, for BPI_{VOL} and BPI_{SUV}, the skeleton volume based on additional information from CT was included as well, thus also allowing for inter-patient comparison.

In a first step, we demonstrated that the bone volume using the newly introduced tool was reproducible between different scans of the same patient (mean difference of less than 2%). The small discrepancy can be explained by slightly different positions of the patient in the scanner. The absolute values obtained in our study (mean 4,184±503 cm^3) were in the expected range of a reference human skeleton with an estimated total volume (including bone marrow, averaged for both sexes) of 7,700 cm^3 *(13)* of which an average of 50.3 % was included in our segmentation *(14)*.

In a second step, both imaging indices BPI_{VOL} and BPI_{SUV} proved highly reproducible with a maximum inter-observer difference of 3.5%. From a practical point of view, the false positive correction was done with only a few clicks and the full image segmentation with manual corrections only took a few minutes, which makes it usable in clinical practice. The limited influence of different SUV-thresholds for computation underlined the robustness for clinical analyses. For the quantitative read-out of both indices, we finally chose a conservative approach (SUV-cut-off of 3) to ensure that less than 0.1 is due to false positives.

In principle, the two introduced indices BPI_{VOL} and BPI_{SUV} were highly correlated. Nevertheless, it is important to mention that the information they provide is not completely equivalent, since BPI_{SUV} also took into account the level of expression of PSMA. The percentage changes of BPI_{VOL} and BPI_{SUV} during therapy were highly correlated. Interestingly, while the average values of BPI_{VOL} and BPI_{SUV} increased during therapy, the average value of SUV_{mean} decreased. This showed that BPI_{VOL}, SUV_{mean} and BPI_{SUV} provided different information.

Notably, some clear outliers in the comparison between BPI and BSI (Fig. 6) were observed. It has to be respected that bone scintigraphy and ^{68}Ga-PSMA PET/CT image two different biological processes: bone scintigraphy displays the reactive changes of the tumour on the skeleton *(15)*, while ^{68}Ga-PSMA-PET directly shows the intensity of PSMA-expression on viable tumour cells. Thus, no absolute equivalence of BSI and BPI could be expected. Other explaining factors were the higher sensitivity of ^{68}Ga-PSMA PET/CT for bone metastases *(16)*, the delay after which BS usually shows changes as well as the "flare" phenomenon *(15,17).*

Compared to standardized imaging response evaluation using PERCIST results for all type of target lesion, BPI_{SUV} showed a higher agreement than BSI and also BPI_{VOL}. This was not unexpected as BPI_{SUV} considers also quantitative values, as does PERCIST. An even better correlation with a lower number of patients showing divergent results between PERCIST and both indices was achieved when they were compared to PERCIST results based only on bone involvement. This clearly implied the need of further adaptation of BPI as quantitative PET-imaging index also for the assessment soft-tissue tumour burden. A further argument for the potential clinical value of BPI compared to conventional methods was the statistically significant substantial correlation of BPI with percentage change of PSA-value during therapy ($r=0.70$, $p<0.01$) compared to BSI which showed no correlation ($r=0.24$, $p=0.32$).

Our study had several limitation. First, it was based on retrospective data analysis. Second, after the baseline scan prior to Radium-223-dichloride follow-up scans for assessment of therapy response were either after three or six cycles. Nevertheless, the aim of the study was primarily the introduction of new quantitative imaging biomarker and only secondly to establish first preliminary correlation to clinical data. The influence of previous lines of treatment was not assessed, which

could potentially impact signals derived from both ^{68}Ga-PSMA PET and bone scan *(18,19)*. These confounding factors should be investigated in future studies encompassing larger patient cohorts. Another limitation is the fact that no respiratory gating was used which potentially would minimize the need for manual correction

CONCLUSION

We have introduced BPI_{SUV} and BPI_{VOL} as new multimodal quantitative imaging indices for ^{68}Ga-PSMA-PET/CT, representing a robust tool for quantitative assessment of osseous tumour burden. We have shown that their automatic computation (with mininal manual corrections) is feasible and highly reproducible on a retrospective cohort of mCRPC patients. Finally, our results demonstrated that BPI_{VOL} and BPI_{SUV} provide clinically meaningful information when correlated to PERCIST, BSI and PSA-value. However, the full clinical value for e.g. predicting patient outcome has to be investigated in future prospective studies.

REFERENCES

1. Imbriaco M, Larson SM, Yeung HW, et al. A new parameter for measuring metastatic bone involvement by prostate cancer: the Bone Scan Index. *Clin Cancer Res*. 1998;4:1765-1772.

2. Ulmert D, Kaboteh R, Fox JJ, et al. A novel automated platform for quantifying the extent of skeletal tumour involvement in prostate cancer patients using the bone scan index. *Eur Urol*. 2012;62:78-84.

3. Afshar-Oromieh A, Malcher A, Eder M, et al. PET imaging with a [68Ga]gallium-labelled PSMA ligand for the diagnosis of prostate cancer: biodistribution in humans and first evaluation of tumour lesions. *Eur J Nucl Med Mol Imaging*. 2013;40:486-495.

4. Afshar-Oromieh A, Zechmann CM, Malcher A, et al. Comparison of PET imaging with a (68)Ga-labelled PSMA ligand and (18)F-choline-based PET/CT for the diagnosis of recurrent prostate cancer. *Eur J Nucl Med Mol Imaging*. 2014;41:11-20.

5. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45:228-247.

6. 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 Suppl 1:122S--50S.

7. O JH, Lodge MA, Wahl RL. Practical PERCIST: A Simplified Guide to PET Response Criteria in Solid Tumors 1.0. *Radiology*. 2016;280:576-584.

8. Etchebehere EC, Araujo JC, Fox PS, Swanston NM, Macapinlac HA, Rohren EM. Prognostic Factors in Patients Treated with 223Ra: The Role of Skeletal Tumor Burden on Baseline 18F-Fluoride PET/CT in Predicting Overall Survival. *J Nucl Med*. 2015;56:1177-1184.

9. Python Software Foundation. Python Language Reference, version 2.7. http://www.python.org.

10. Kang Y, Engelke K, Kalender W a. A new accurate and precise 3-D segmentation method for skeletal structures in volumetric CT data. *IEEE Trans Med Imaging*. 2003;22:586-598.

11. Anand A, Morris MJ, Kaboteh R, et al. Analytic Validation of the Automated Bone Scan Index as an Imaging Biomarker to Standardize Quantitative Changes in Bone Scans of Patients with Metastatic Prostate Cancer. *J Nucl Med*. 2016;57:41-45.

12. Yadav MP, Ballal S, Tripathi M, et al. 177Lu-DKFZ-PSMA-617 therapy in metastatic castration resistant prostate cancer: safety, efficacy, and quality of life assessment. *Eur J Nucl Med Mol Imaging*. 2017;44:81-91.

13. International Commission on Radiological Protection. Report of the task group on Reference Man: ICRP Publication 23. *Ann ICRP*. 1975;4((3-4)).

14. Valentin J. Basic anatomical and physiological data for use in radiological protection: reference values: ICRP Publication 89. *Ann ICRP*. 2002;32:1-277.

15. Messiou C, Cook G, DeSouza NM. Imaging metastatic bone disease from carcinoma of the prostate. *Br J Cancer*. 2009;101:1225-1232.

16. Pyka T, Okamoto S, Dahlbender M, et al. Comparison of bone scintigraphy and (68)Ga-PSMA PET for skeletal staging in prostate cancer. *Eur J Nucl Med Mol Imaging*. 2016;43:2114-2121.

17. Ulmert D, Solnes L, Thorek DL. Contemporary approaches for imaging skeletal metastasis. *Bone Res*. 2015;3:15024.

18. Eiber M, Maurer T, Souvatzoglou M, et al. Evaluation of Hybrid ^{68}Ga-PSMA Ligand PET/CT in 248 Patients with Biochemical Recurrence After Radical Prostatectomy. *J Nucl Med*. 2015;56:668-674.

19. Afshar-Oromieh A, Avtzi E, Giesel FL, et al. The diagnostic value of PET/CT imaging with the (68)Ga-labelled PSMA ligand HBED-CC in the diagnosis of recurrent prostate cancer. *Eur J Nucl Med Mol Imaging*. 2015;42:197-209.

Figures

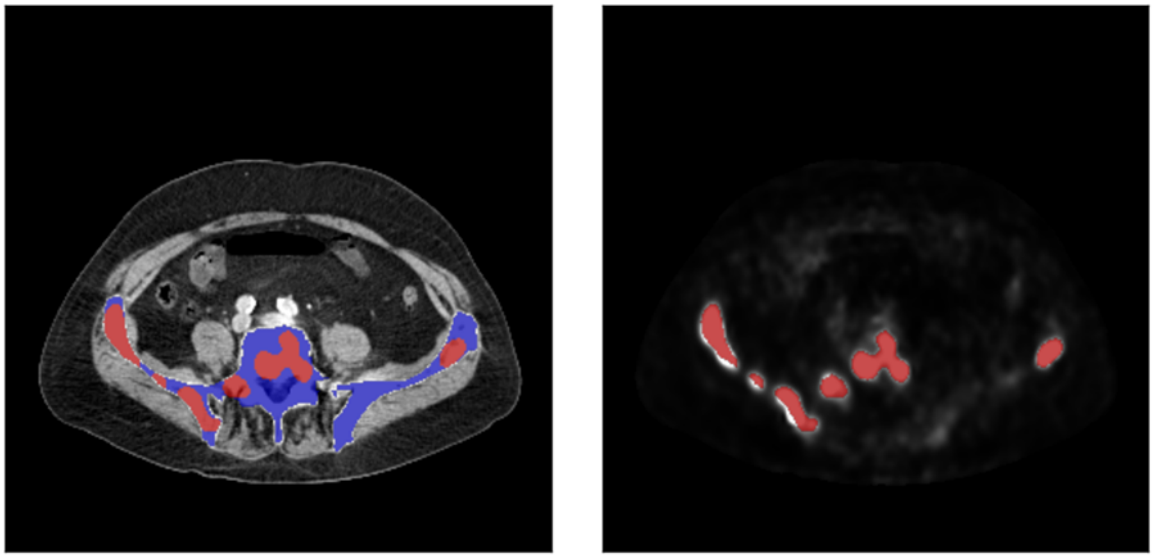

Figure 1: PET and CT image of one patient. The blue overlay shows the bone mask computed by the tool and the red overlay shows the tumour mask computed with a SUV-threshold of 3.

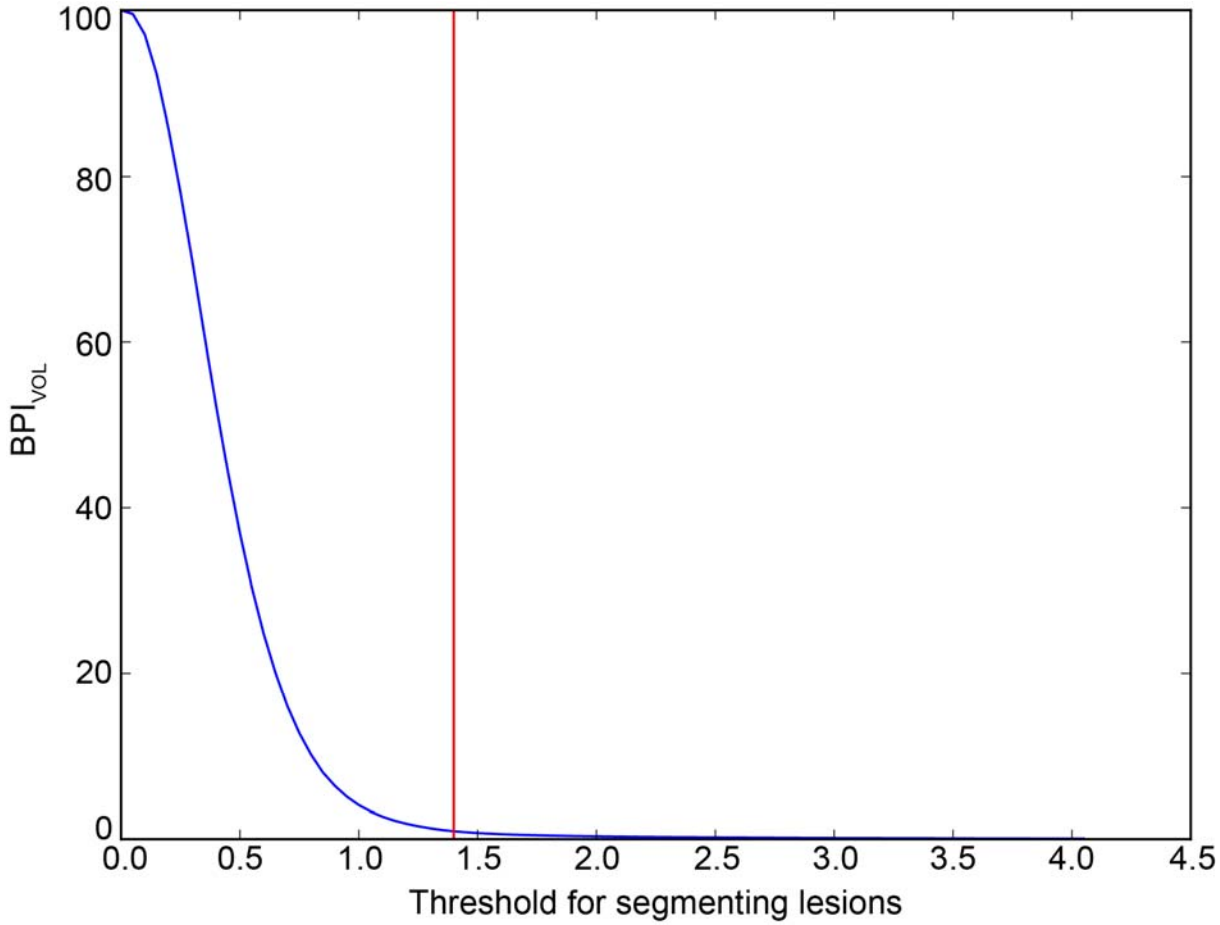

Figure 2: BPI$_{VOL}$ values obtained with different thresholds for one example of negative training patient. The red line indicates BPI$_{VOL}$ under 1 are obtained for thresholds superior to 1.4.

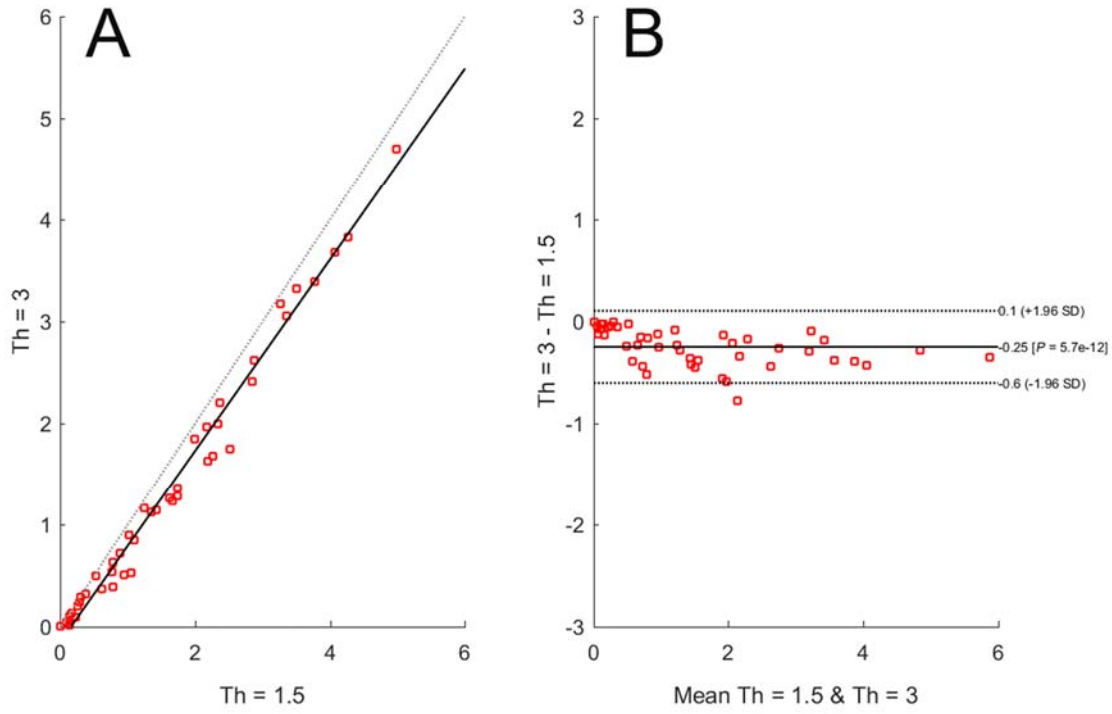

Figure 3: Comparison between BPI_{SUV} obtained with cut-off values 1.5 and 3. High correlation between the results is shown both using a scatter diagram (A) and in the Bland–Altman plot on which the differences between two BPI_{SUV} are plotted against their average (B). As expected, it shows a mean difference of -0.25 BPI_{SUV} (95% confidence intervals, +0.1 and -0.6 BPI_{SUV}), indicating systematically lower BPI_{SUV} values for a cut-off value of 3.

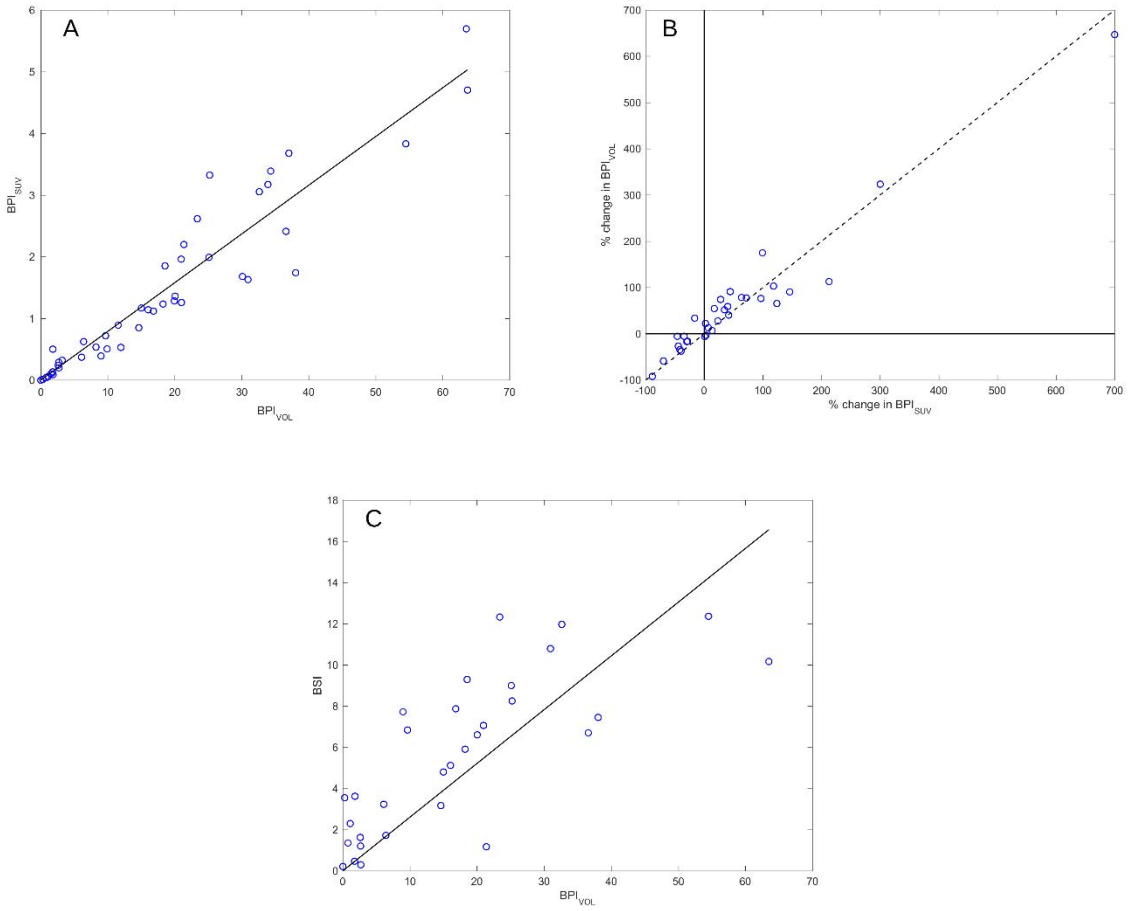

Figure 4: (A) BPI_{VOL} and BPI_{SUV} values. A statistically significant high correlation ($r=0.89$, $p<0.001$) was observed. Images from all 45 patients with bone metastases were used. (B) Change in BPI_{SUV} and BPI_{VOL} during therapy for 32 patients. A statistically significant high linear correlation ($r=0.97$, $p<0.001$) was observed. (C) BPI_{VOL} and BSI values for 31 patients before the beginning of the therapy. A moderate correlation ($r=0.76$, $p<0.001$) was observed.

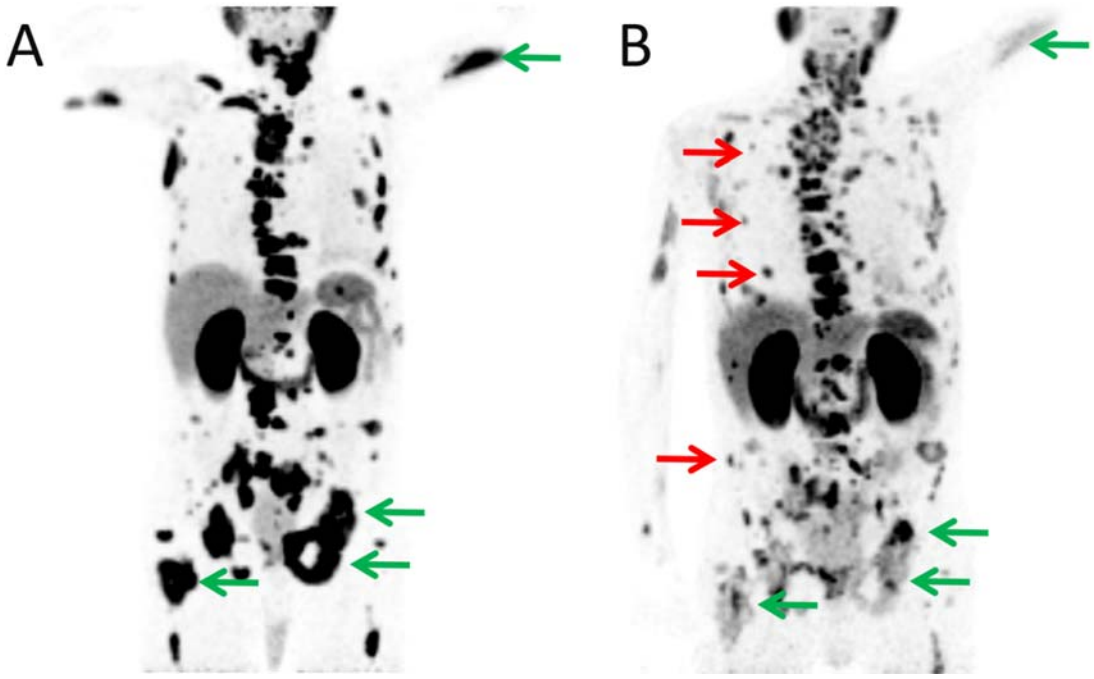

Figure 5: Maximum intensity projection of ^{68}Ga-PSMA-PET images of the same patient before (A) and after three cycles of Radium-223-dichloride therapy (B). A majority of the bone lesions show substantial response to therapy based on decreasing PSMA-expression (green arrows), but also a few new lesions (red arrows) appear in the follow-up scan. Despite an overal decrease in tumour load, based on PERCIST criteria the status of the patient was defined as progressive disease.

Tables

Table 1: Available data for the total of 60 patients.

	Test	Before therapy	After 3 cycles	After 6 cycles
Negative training patients	^{68}Ga-PSMA PET/CT	15	N/A*	N/A*
Patients undergoing Radium-223-dichloride therapy	^{68}Ga-PSMA PET/CT	45	32	20
	PSA value	43	33	21
	Bone scan	31	21	18

Legend: * not applicable,

Table 2: Comparison of expert reader and BPI and BSI classifications. A total of 32 patients for BPI and 19 patients for BSI were classified. Correlation is shown both for PERCIST using information from all target lesions and for bone involvement only (bone-PERCIST).

		BPI_{VOL}			BPI_{SUV}			BSI		
		Prog	Stab	Resp	Prog	Stab	Resp	Prog	Stab	Resp
PERCIST	Prog	17	2	6	17	3	5	8	4	4
	Stab	1	0	0	1	0	0	0	1	0
	Resp	2	0	4	0	1	5	0	0	2
Bone-PERCIST	Prog	15	2	2	15	2	2	7	2	2
	Stab	3	0	3	3	1	2	1	3	2
	Resp	2	0	5	0	1	6	0	0	2

Legend: Prog : progressive disease, Stab: stable disease, Resp: responsive to therapy.

SUPPLEMENTARY MATERIAL

SEMI-AUTOMATIC COMPUTATION METHOD FOR BPI_{VOL} AND BPI_{SUV}

After loading the PET and CT images, following steps were used for the computation of BPI_{VOL} and BPI_{SUV}:

Bed Removal in CT

The CT image was first thresholded at 15% of the range of its intensities (for an image with intensities between -1024 and 2500 Hounsfield Units, the threshold is -495.4 Hounsfield Units). From the resulting mask, the size of all connected components were computed. The biggest connected component was kept as the body mask. A further step of morphological hole closing was applied to each slice of the body mask to fill in the lungs and potential air in the abdomen. The bed, being outside of the body mask, was masked out of the image.

Bone Segmentation in CT

For the segmentation of the skeleton from CT, we used the first two steps of the method of Kang et al. *(10).*

Global thresholding. First, a low and a high threshold were computed by fitting a mixture of two Gaussian distributions $G_1(m_1,\sigma_1)$ and $G_2(m_2,\sigma_2)$ with respective means m_1 and m_2 and respective standard deviations σ_1 and σ_2 to the histogram of CT intensities (excluding the background). Without loss of generality, we assumed $m_1<m_2$. After fitting, the low threshold LT was computed as:

$$LT = \min(160, m_2+1.7\,\sigma_2)$$

The high threshold HT was computed as:

$$HT = LT+400$$

All pixels that had an intensity superior to HT were considered as bone. All pixels that had an intensity inferior to LT were considered as not bone. All pixels that had an intensity between LT and HT were considered as undetermined and were labelled in the next step. To avoid labelling endoprothesis as bones, all pixels with an intensity above 2000 were excluded from the bone mask.

Local thresholding. In a second step, the pixels that could not be labelled by global thresholding were considered. For each pixel, the local mean m_{loc} and local standard deviation SD_{loc} were computed within a 26-pixels neighbourhood. If the intensity of the pixel was superior to $m_{loc} - 0.8\ SD_{loc}$, it was labelled as bone, otherwise as not bone.

Morphological operations. To further correct the segmentation and include the bone marrow, morphological operations were applied. All connected components with a size inferior to 45 pixels were removed. The bone mask was dilated by one pixel, hole filling was applied to each slice and the bone mask was eroded by one pixel.

Lesions segmentation in PET

^{68}Ga-PSMA usually does not exhibit unspecific uptake in the bones and bone marrow. Therefore, regions of the skeleton with increased uptake can be considered as bone metastases. First, the SUV image was computed using the patient weight, injection time and acquisition time contained in the image header. Since the PET and the CT had been acquired in the same session on the same scanner, they were intrinsically aligned. The bone mask computed from the CT image was therefore applied to the SUV image to determine the location of bones. A user-chosen threshold was further applied to the bone locations to compute abnormal uptake. Normal uptake sites such as the kidneys are automatically excluded from the computation because they are not part of the bone mask.

BPI_{VOL} and BPI_{SUV} computation

After having computed the bone and the lesion masks, BPI_{VOL} and BPI_{SUV} were computed using their respective definitions, where *N* is the number of pixels *p* that belong to bone metastases:

$$BPI_{VOL} = 100 \times \frac{Bone\ metastases\ volume}{Skeleton\ volume} \qquad \text{(no unit)}$$

$$SUV_{mean} = \frac{1}{N} \sum_{\substack{p\ in\ bone \\ metastases}} SUV(p) \qquad \text{(unit: g/mL)}$$

$$BPI_{SUV} = BPI_{VOL} \times SUV_{mean}/100 \qquad \text{(unit: g/mL)}$$

SUV threshold choice

The SUV-threshold for lesion segmentation was chosen by the user. Rather than choosing it arbitrarily, we proposed a method to calculate it using a negative training patients cohort. For the cancer-negative patients, the value of BPI_{VOL} was the percentage of the skeleton that is falsely segmented as lesions.

Since the training cohort was chosen to be negative for cancer, BPI_{VOL} should have been 0 for these patients. However, because of the noise present in PET, in some pixels, the uptake was higher than normal background uptake. Depending on the chosen SUV-threshold, the thresholding of the PET image resulted in a non-empty set of bone lesions and BPI_{VOL} was computed as greater than 0.

We proposed that the user chose a percentage of false positives FP_{max} that he tolerated. For each patient of the cohort, BPI_{VOL} was computed using the method described above with thresholds ranging from 0 to 4 with steps of 0.05. A different BPI_{VOL} value was obtained for each SUV-threshold: with a SUV- threshold of 0, BPI_{VOL} was 100, and with a SUV-threshold of 4, for cancer-

negative patients, it was very close to 0. The recommended threshold t_{rec} for the patient was calculated as :

$$t_{rec} = \min\{t \in [0,4] | BPI_{VOL} \leq FP_{max}\}$$

The SUV-threshold for the testing cohort was then chosen as the maximum of all t_{rec} obtained for the training cohort.