REVIEW ARTICLE

Pathology-validated PET image data sets and their role in PET segmentation

Assen S. Kirov · Louise M. Fanchon

Abstract Positron emission tomography (PET)/computed tomography has recently been finding broader application for the diagnosis, treatment and therapy assessment of malignant disease. Accurate definition of the tumor border is extremely important for the success of localized tumor therapies. PET promises to provide the metabolically active tumor volume and, at present, it is used for target definition in a variety of tumors. This process is, however, subject to uncertainties of different origin. Resolving these uncertainties is challenging, since validating PET images and segmentation contours against tumor pathology is experimentally difficult. In addition to accurate lesion contouring, this challenges validation of PET tracers and investigations of tumor functional heterogeneity. In this paper, we briefly review the present studies providing PET image data sets with pathology validation. We focus on the specimen handling techniques aimed at achieving higher geometrical accuracy of the pathology-derived "ground truth". We also summarize the main findings obtained for the PET segmentation techniques which have been tested with the help of these data sets. Finally, we provide a critical summary of the current state of the art in pathological validation of PET images and briefly discuss future possibilities in this direction.

Color figures online at http://link.springer.com/article/10.1007/s40336-014-0068-9

A. S. Kirov (✉) · L. M. Fanchon
Department of Medical Physics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA
e-mail: kirova@mskcc.org

L. M. Fanchon
INSERM, UMR1101, LaTIM, Brest, France

Introduction

The establishment of various positron emission tomography (PET) tracers as biomarkers in oncology depends on quantification of tracer uptake in PET images. Segmentation of tumors in PET images is one approach to PET quantification and it is needed in localized cancer therapies to define the lesion borders [1, 2] and assess the effects of different treatment approaches. The stakes are especially high for hypo-fractionated radiation therapy, in which lethal doses need to be delivered to lesions that are often next to critical organs. Also, the use of PET/computed tomography (CT) guidance in interventional radiology procedures including tumor ablation [3] is increasing, and this requires both accurate and quick lesion border determination. Unfortunately, the accuracy of PET quantification is currently low due to the low spatial resolution of PET scanners, other PET imaging artifacts, and lack of "ground truth" for clinical images.

Recent advances in radiation therapy technology have opened the way for high-precision delivery of very high doses of radiation to a previously defined tumor target. Since the technical challenges of radiation delivery and of the immobilization or tracking of the target can be adequately addressed [4, 5], the problem of accurately defining the tumor target remains a major limitation [6]. While at present PET, is the main imaging modality which allows defining the tumor based on its metabolic properties, it has poor resolution and is subject to several artifacts from biological, physical and technical origin [7]. These factors challenge the tumor segmentation process [8, 9]. Similarly, accurate definition of the tumor border is needed in image-

guided interventions [3]. An example of this is PET/CT-guided percutaneous ablation, in which the interventional radiologist aims to conform the ablation volume to the PET-avid area in fused PET/CT images [10, 11]. Therefore, experimental verification of lesion margins derived from PET is of great importance.

Delineation of tumors in PET images can be performed manually or automatically. Automatic segmentation reduces inter-observer variability [12, 13]. As a result, many PET auto-segmentation (PET-AS) algorithms have recently been developed and these are reviewed elsewhere [9, 14, 15]. Evaluation of these algorithms can be performed on phantom-based images, simulated images, or on clinical PET images with either manual delineation or pathology validation. Experimental phantom images play a very important role in the initial evaluation of segmentation algorithms and phantom configurations with different degree of complexity have been used. These include different-sized spherical or cylindrical objects with uniform activity concentration similar to the NEMA image quality phantom [16–18] as well more complex phantoms, which represent real tumor shapes [19, 20]. The phantoms containing simple object shapes which can be filled with activity provide an easier way of producing PET images with ground truth, but the images are not very realistic in terms of tumor shape and non-uniformity of uptake and they are subject to cold wall artifacts [14, 18]. Simulated phantom images overcome this latter limitation and make it possible to produce images of realistically shaped lesions with more realistic activity distributions [21], either through some of the advanced Monte Carlo tools dedicated to nuclear medicine [22–24], or through simpler analytical forward projection approaches. A detailed overview of the different experimental and simulated phantoms used for evaluating PET-AS approaches is provided in the upcoming first report of Task Group No. 211 (TG211) of the American Association of Physicists in Medicine (AAPM) [9].

The major issues with manual delineation as a surrogate of truth are its highly subjective nature and low reproducibility. Although histopathology-derived ground truth may contain geometrical uncertainties, e.g., due to specimen deformation or shrinkage, it provides for a more clinically adequate evaluation of segmentation, since it does not contain the physical and biological biases which may be present in the PET image. Unfortunately, the number of PET image data sets accompanied with histopathology information is very limited, due to the significant difficulties in performing these studies. The growing role of PET prompted the publication during the last decade of several such investigations in which the features of the PET image are compared against histopathology findings of the excised specimen. The present review includes studies on this topic published in peer-reviewed journals identified through PubMed searches. These investigations are reviewed here with respect to their importance for evaluating PET images and segmenting lesions from these images. Many of the reviewed studies compared gross tumor volumes (GTVs) also defined by other imaging modalities: CT, and magnetic resonance imaging (MRI), which are here mentioned only briefly, when considered to be related to the accuracy of the pathology data or the PET images. PET image data sets obtained with the help of experimental and numerical phantoms contribute to resolving the physical sources of uncertainty and their impact on segmentation accuracy, as discussed elsewhere [9], but they do not address the link between histopathology and PET tracer uptake and are therefore not discussed here. Since in all of the reviewed studies (except one which also uses an additional tracer) the PET radiotracer used is ^{18}F-fluorodeoxyglucose (FDG) throughout the text by PET is meant FDG PET, unless stated otherwise.

PET image data sets with pathological validation

Available data sets

The literature describes two types of PET data sets with pathological validation (PSPVs) used for evaluating PET segmentation. One type is obtained by measuring certain volumetric characteristics (e.g., maximal diameter) of the tumor from the fresh specimen and comparing the measurements to the GTV delineated in the PET image. The other data set type aims at having an accurate reconstruction of the pathological volume of the tumor in three dimensions and may require special handling (e.g. fixation) and slicing of the specimen, and (in some of the studies) registering this volume with the CT and the PET. The latter data sets have the advantage of providing more complete pathological characterization of the tumor but the approach is more complex. The validity of such data sets as "gold standard" relies on the accuracy of the whole procedure and on the image registration method. Different procedures have been developed to avoid, as much as possible, deformations during extraction and processing of the specimen. Since these procedures may vary depending on the tumor location (ex: lung or head and neck tumors), they are described separately below. A list of the currently available data sets grouped by body location is given in Table 1.

First, methods permitting a 3D assessment of the pathology volume are reviewed. These studies have common steps that are discussed below by tumor location:

Table 1 Summary of current publications defining PET image data sets with pathological validation (PSPV)

Body location	Reference	Number of cases per study with histopathology data
Head and neck	Daisne et al. [25]	9 patients
	Burri et al. [38]	18 patients, 27 tumors
		(18 primary tumors and 9 cervical nodal metastases)
	Caldas-Magalhaes et al. [26]	10 patients
	Schinagl et al. [40]	12 patients with head and neck squamous cell carcinoma and 28 metastases in cervical lymph node(s)
Lung	Stroom et al. [27]	5 patients
	Van Baardwijk et al. [12]	23 patients
	Dahele et al. [32]	12 patients
	Yu et al. [29]	15 patients
	Wu et al. [43]	31 patients
	Wanet et al. [33]	10 patients
	van Loon et al. [28]	34 patients
	Meng et al. [30]	39 patients
	Schaefer et al. [31]	15 patients
Esophagus	Zhong et al. [41]	36 patients
	Han et al. [42]	22 patients
Rectum	Roels et al. [35]	15 patients
	Buijsen et al. [44]	26 patients
Colon and sigmoid	Chen et al. [45]	35 patients with colon cancer, 42 patients with sigmoid cancer
Cervix	Zhang et al. [34]	10 patients
Multiple locations	Sridhar et al. [39]	52 patients: 22 head and neck, 22 lung, 8 colorectal

NSCLC non-small-cell lung cancer

specimen fixation (for all except one), registration of the specimen with the images, and deformation corrections for some of them (Table 2). Such are the PSPVs generated by Daisne et al. [25], Caldas-Magalhaes et al. [26], Stroom et al. [27], Van Loon et al. [28], Yu et al. [29], Meng et al. [30], Schaefer et al. [31], Dahele et al. [32], Wanet et al. [33], Zhang et al. [34], and Roels et al. [35]. After this, we provide an overview of the volumetric studies.

3D reconstruction of the pathological volume and image co-registration

Head and neck

Pathology processing Studies involving laryngectomy specimens have been published by Daisne et al. [25] and by Caldas-Magalhaes et al. [26]. There are both similarities and differences between the methods employed to preserve the shape of the excised larynx lesion. In both studies, the first step was to fix the specimen and to introduce rods as fiducial markers for aligning the slices of the specimen. However, the fixation procedures were different. Daisne et al. [25] fixed the specimen by placing it in a cast to which a gelatin solution was added, followed by

refrigeration intervals at decreasing temperatures (down to -80 °C). This procedure was previously developed and validated against another fixation procedure (formalin), using an animal model. Caldas-Magalhaes et al. [26] fixed the specimen with 10 % formaldehyde for extended period (at least 48 h) and embedded it in a solution of agarose at controlled temperature after which they cooled it at 5 °C until solidification. In both studies after fixation, the specimen was cut in few mm thick slices (between 1.7 and 2 mm [25] and around 3 mm [26]).

Daisne et al. [25] studied macroscopic slices of the specimen, whereas Caldas-Magalhaes et al. [26] investigated them at a microscopic level, for which they added an additional step. After removal of the agarose and decalcification, the macroscopic slices were embedded in paraffin and one 4-µm-thick section was obtained for each 3-mm-thick slice and stained with hematoxylin–eosin (H&E). The authors reported that shrinkage of the specimens occurred mostly during the last step of the pathology processing and that there was shrinkage of 12 ± 3 % between the microscopic and macroscopic sections. The extent of the deformations and shrinkage that occurred during surgery and formaldehyde fixation of the specimens was small (3 ± 1 % inside the cartilage skeleton).

Table 2 Summary of fixation and registration steps and deformation corrections applied in investigations performing detailed tumor shape and volume reconstruction from pathology sections

Reference/tumor type	Specimen fixation	Registration method	Slice thickness	Deformation corrections
Daisne et al. [25]/head and neck	Gelatin plus refrigeration sequence	Use wooden rods as fiducial markers RPC: semi-automated rigid method based on surface segmentation	1.7–2 mm	"Negligible," according to an animal model
Stroom et al. [27]/NSCLC	Formalin inflation of excised lobes, slicing and embedding in paraffin	RPC: Resliced CT to match orientation of macroscopic slices	5- to 10-mm macroscopic slices, 4-μm, microscopic slices	Deformation of the lung volume measurement: In 3 directions from ratios of the distances from the GTV edge to fiducial markers in the CT and pathology images
van Loon et al. [28]/NSCLC	Same as described by Stroom et al. [27]	Same as described by Stroom et al. [27]	Same as described by Stroom et al. [27]	Same as described by Stroom et al. [27]
Yu et al. [29]/NSCLC	Fixation of bisected tumor in 10 % formalin for at least 24 h	Orienting the specimen in the "in vivo geometry" and bisecting in the transverse plane	5- to 7-μm sections at 4-mm intervals	From digital photography of the bisected specimen with a ruler before and after fixation
Meng et al. [30]/NSCLC	Specimen fixed in 10 % formalin	Not mentioned	5- to 7-mm-thick slices at 4-mm intervals. Also whole-mount paraffin sections	Not mentioned
Schaefer et al. [31]/NSCLC	No fixation	Not mentioned	4–5 mm	Not considered
Caldas-Magalhaes et al. [26]/Laryngeal	Specimen fixed in 10 % formaldehyde for at least 48 h, and put in a box which was then filled with agarose. For histological analysis, the agarose was removed from the slices, which were then decalcified and embedded in paraffin	Three carbon rods were inserted in the specimen craniocaudal direction RPC: registration between the specimen reconstructed in 3D and the CT of the specimen after fixation; (CT$_{post}$) Registration of CT$_{post}$ with CT	For each 3-mm slice, one 4-μm section was obtained	Shrinking of the specimens mostly occurred during pathology processing to obtain H&E sections from thick sections
Wanet et al. [33]/NSCLC	After extraction, the lobes were inflated with liquid gelatin. Each specimen was placed in a box, which was then filled with gelatin and frozen	Wooden rods were placed in the box RPC: rigid	4-mm macroscopic slices	GTV was assumed to be non-deformable
Roels et al. [35]/rectal cancer	Each specimen was placed in a box, which was then filled with gelatin and kept at −20 °C for 2–3 days	Wooden rods were placed inside and around the specimen	2- to 3-mm-thick slices and microscopically thin cross-sections	Corrections for volume shrinkage after fixation by rigid registration of the microscopic to the macroscopic slices
Zhang et al. [34]/cervical cancer	Specimens were fixed with formalin for 24 h	Not mentioned	4-mm-thick slices and 4-μm sections	Volume shrinkage before and after formalin fixation of 65–97 % (mean 85 ± 10 %)

Radiology-to-pathology co-registration (RPC) procedures described are indicated in the third column

Daisne et al. [25] calculated that the loss of tissue during slicing of the specimen was close to one slice thickness per slice obtained. Caldas-Magalhaes et al. [26] measured a loss of 2 % for the whole specimen during slicing.

3D image registration The image registration method used by Daisne et al. [25] and developed in a previous study [36] consisted of sampling each volume to a common voxel size, and displaying simultaneously the axial, coronal

Axial	Coronal	Sagittal
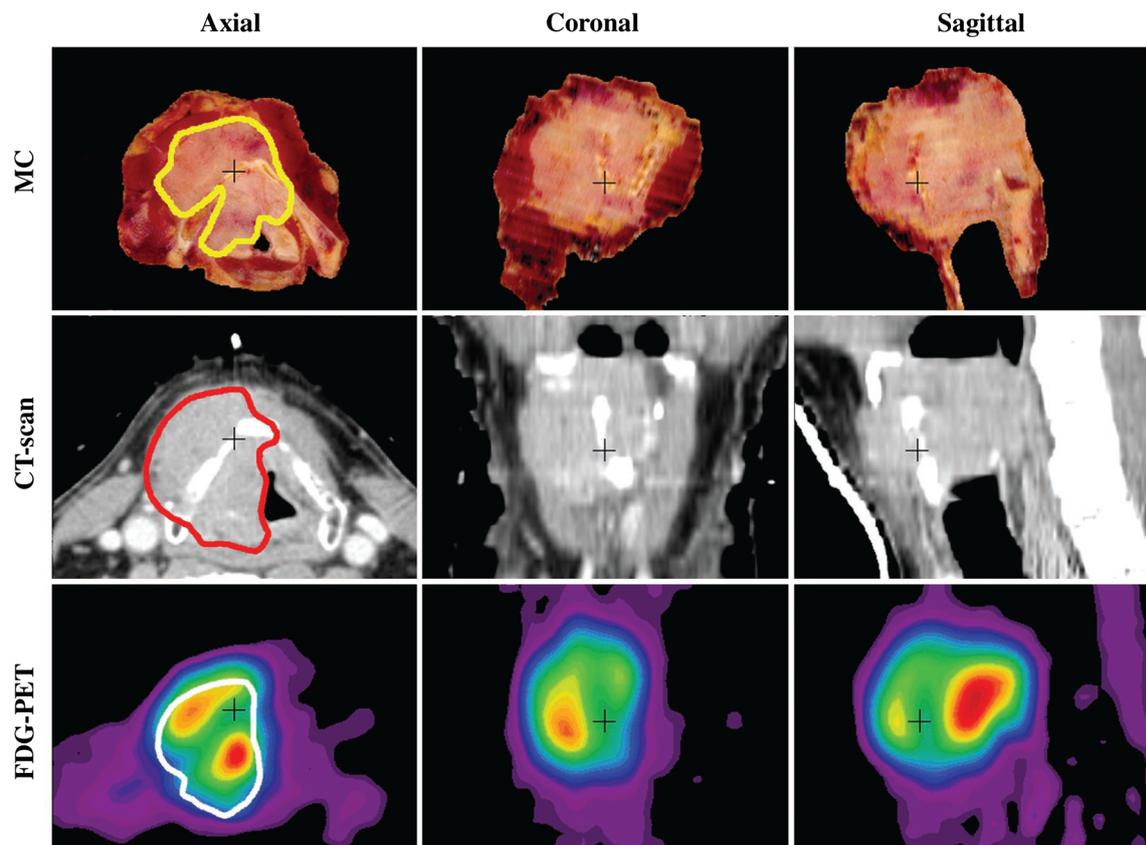
Fig. 1 Co-registered images of the macroscopic specimen (MC), computed tomography (CT) and FDG positron emission tomography PET (FDG PET) images in the transverse, coronal and sagittal planes published by Daisne et al. [25], with contours for each modality shown in the transverse plane. The "cross represents the same point in space for each of the modalities". Reprinted, with permission from Daisne et al [25]: Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology 233(1):93–100, RSNA, 2004 (color figure online)

and sagittal views. Thereafter, automatic segmentation of the CT scan was performed and the CT was overlaid and manually registered with the images from the other modalities. Mean co-registration precision between PET, MRI and CT images "as assessed from the Euclidian vectors was of the order of 1.1–2.4 mm". The accuracy of the pathology–radiology co-registration was not reported. Deformations during the whole process were considered negligible. Among the limitations of this approach, the authors highlighted difficulties related to image co-registration, which restricted the investigation to laryngeal carcinomas. The contours drawn on the macro-specimen, CT and PET images are illustrated in Fig. 1, which is here reproduced, with permission, from their original work [25].

Caldas-Magalhaes et al. [26] detailed all the steps of the image co-registration process and reported the registration error associated with each step. The H&E slides were rigidly registered with the thick-slide photos, and a scaling factor was applied to these slides. The authors reported that the registration error between the pathology and the CT, MRI and PET images in the cartilage skeleton was on average 1.5, 3.0 and 3.3 mm, respectively. They found that the "GTV was a rigid and compact mass of tissue", and "that it maintained its shape during the procedure". They concluded that evaluating the GTV as delineated on the PET image with the GTV derived from the pathology specimen is feasible with an average overall accuracy below 3.5 mm inside the laryngeal skeleton. The delineation inaccuracies were larger than the inaccuracy of the registration error.

Non-small-cell lung cancer (NSCLC)

Pathology processing Lung tissue has a tendency to collapse. To compensate for deformations, in some studies involving NSCLC specimens, the lobe specimens were inflated using different materials and methods. In the method published on lung lobe processing, by Stroom et al. [27], the lobes were inflated with formalin. Inflation was stopped when the lobes attained a volume as close as possible to the lobe volume seen on the CT. Wanet et al. [33] used gelatin to inflate the excised lung lobes until they

were uniformly filled; and Dahele et al. [32] insufflated the specimens with 10 % formalin until the specimen was saturated and formalin was being exuded across the pleura.

However, in the three studies, significant deformations of the specimen from the in vivo status were found. Dahele et al. [32] reported that formalin was expelled from the lobes during cutting them into macroscopic sections, resulting in additional deformation which needs to be accounted for. To overcome this problem they embedded some of the specimens in agar before sectioning and tested several cutting methods. They reported that cutting with an electric rotor cutter improved the consistency of the sections and that embedding in agar was helpful in some of the cases.

Large deformations were observed between the CT and the macroscopic images of the specimen [27, 33], and were found to be anisotropic [27]. A similar observation was made by Siedschlag et al. [37] for a 10-mm-thick layer around the GTV "depending on circularity of the tumor and orientation of the specimen on the pathology table during processing." Gravity was expected to deform the specimens in a direction perpendicular to the table. Stroom et al. [27] mentioned that: "the volume of the well-inflated lung lobes on pathologic examination was still, on average, only 50 % of the lobe volume on CT."

For whole-mount sections, Stroom et al. [27] and Dahele et al. [32] embedded the macroscopic slices in paraffin blocks which were sliced into 4-μm sections and stained with H&E. They did not mention evaluating possible deformation between the macroscopic and the microscopic slices. Stroom et al. [27] found that the GTV was rigid enough not to deform, so the deformations were expected to affect mostly the microscopic disease extension (ME) measurements. In this work, the deformations of the GTV-surrounding tissue were measured using photos of the macroscopic specimen and the CT images (Fig. 2), and corrections were applied for the ME measurements. Wanet et al. [33] also assumed the GTV to be non-deformable. With some modifications, Stroom's method was then used in another study [28] to generate PSPV.

Other studies dealt with lung lobes, but tried a different technique, which did not involve inflating the specimen. For those studies, radiology–pathology image co-registration was not performed. Yu et al. [29] oriented the specimen to the in vivo geometry and bisected it in the transverse plane in the operating room. They took photographs of the specimen, both before and after fixation in 10 % formalin as well as after slicing the specimen with a microtome into 5- to 7-μ-thick slices to determine the volume correction. They reported a reduction to 82 ± 10 % of the original tumor volumes (range, 62–100 %) before and after fixation with formalin. Meng et al. [30], in a follow-up study, fixed the specimens in
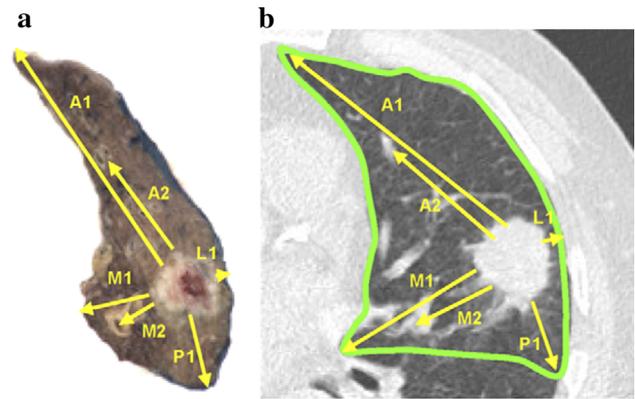
Fig. 2 From Stroom et al. [27]: "Example of procedure to determine deformations between pathology and computed tomography (CT) data. **a** Photograph of macroscopic slice with fiducial points indicated by *arrows*. **b** Aligned CT scan (after reslicing) with corresponding fiducial points indicated by *arrows*. By dividing arrow lengths at CT by corresponding lengths at pathologic examination, deformations were obtained." Reprinted, with permission from Stroom J, et al.: Feasibility of Pathology-Correlated Lung Imaging for Accurate Target Definition of Lung Tumors, International Journal of Radiation Oncology Biology Physics, 69:267–275, Elsevier, 2007 (color figure online)

formalin and subsequently sliced them to obtain whole-mount, H&E-stained slides after which they examined the ME. They did not correct for shrinkage as a result of fixation with formalin even though they had measured this in their previous study [29] and point that it may affect ME_{max}. Schaefer et al. [31] processed the specimens immediately after extraction, so formalin was not used and shrinkage was not considered. The specimens were sectioned into slices ranging from 4 to 5 mm in thickness and manual contouring of the macroscopic tumor extension area was performed for each slice. The accuracy of the technique was not reported.

3D image registration Stroom et al. [27] found that the CT-to-pathology deformation factors for their study were linear, anisotropic and ranged from 1.0 to 2.4 (average 1.8) over all three directions (Fig. 2). In this study, rigid corrections were applied to the pathology specimens to correlate them with pre-surgery scans. The authors also took the maximal ME for every patient and multiplied it by the deformation factors. Wanet et al. [33] rigidly registered the pathological volume with the CT and PET images. Dahele et al. [32] developed a method for 3D correlation of PET/CT images and whole-mount histopathology in NSCLC. They described qualitatively their experience in registering 3D PET/CT images with pathology and concluded that there "is no one definitive method for 3D volumetric" radiology–pathology correlation (RPC) in NSCLC and that using "large histopathology slides to whole-mount entire sections for digitization" allows rigid and manual

registration of histopathology reconstructions to CT and PET. They also pointed out that "timing between imaging and surgery and the use of respiratory-correlated PET and CT imaging" will become factors for robust RPC [32].

Rectal cancer

Pathology processing   Roels et al. [35] put each rectum specimen in a box immediately after extraction; wooden rods were placed inside and around the specimen for orientation and reconstruction purposes. The box was filled with a gelatin solution and stored at −20 °C for 2–3 days to freeze it. Slices with thickness of 2–3 mm were obtained and fixed in formaldehyde. Microscopically thin cross-sections of the tumor were then obtained and registered with the photos of the macroscopic specimen. Microscopic slices were corrected for the shrinkage that occurred during the fixation step and the GTV was delineated on these microscopic slides.

Cervical cancer

Pathology processing   Zhang et al. [34] aimed at determining the optimal SUV cutoff for FDG PET scans of patients with cervical cancer by matching the volume measured on the extracted specimen to GTV_{PET}. The pathology procedure includes fixing the extracted specimen with a 10 % formalin solution, cutting it into serial slices of 4 mm thickness and embedding them in paraffin. The macroscopic slices were then cut into 4-μm-thick histological sections and stained with H&E. They measured the tumor volume before and after formalin fixation and reported volume shrinkage of 65–97 % (mean 85 ± 10 %).

Volumetric measurements

In addition to the investigations mentioned above, there exist a large number of investigations in which the size or volume of the tumor was estimated from the surgical specimen without slicing it. The methods employed are discussed below.

Volume estimation

In a head and neck study, Burri et al. [38] estimated the pathological tumor volumes from the maximal 3D lengths of each tumor after resection and compared them with the volumes measured on the PET and CT images. The 3D diameters were also measured and used to calculate the pathological ellipsoid volume by Sridhar et al. [39] for head and neck, lung, and colorectal tumors. Schinagl et al. [40] measured the volume of lymph node metastases from

head and neck cancer using water immersion after removal of perinodal and fatty tissues.

Length measurement

The studies listed below compared one or more of the tumor dimensions of the surgical specimen from the patient with the corresponding lengths observed on CT, PET and in several cases also MRI scans. For esophageal squamous cell carcinoma, Zhong et al. [41] measured the gross tumor length on PET images and compared it with the tumor length as measured from the pathology specimen. The length of the esophagus was measured in vivo before removal. They corrected for the deformation of the surgical specimen by stretching it to the length as measured in vivo. The gross tumor length was then measured. Han et al. [42] followed this procedure for esophageal squamous cell carcinoma, but they added a fixation step of the specimen with 10 % formaldehyde. They then cut 0.5-cm-width tissue strips, and measured the longitudinal tumor length. They did not report correcting for shrinkage after fixation. The gross tumor lengths were compared to the lengths derived from FDG- and fluorothymidine- (FLT) PET images. To the best of our knowledge, this is the only PSPV investigation that also included a radiotracer other than FDG.

For NSCLC lesions, van Baardwijk et al. [12] measured the maximal diameter (MD) by macroscopic examination of extracted lung tumors. Wu et al. [43] inflated and fixed the lung lobes for 12–24 h in 10 % neutral-buffered formalin. They measured the MD by macroscopic examination of sections of the specimens obtained at 3- to 5-mm intervals.

For rectal cancer, Buijsen et al. [44] measured the length of the tumor macroscopically with a ruler before slicing. Chen et al. [45] investigated maximum tumor diameters in colon and sigmoid cancer. Measurements on the specimen were performed after fixation in formalin and prior to slicing. They mentioned that they did not correct for possible shrinkage due to fixation in formalin.

Role of the data sets in tumor target determination

Since the focus of the present review is the role of pathology-determined tumor borders in the segmentation of PET images, comparison with CT- and MRI-determined tumor volumes is mentioned only where it relates to PET volumes. We group the results into three categories: comparisons of tumor volume sizes, evaluations of the accuracy of segmentation tools, and findings regarding the location of the tumor extensions with respect to the segmented volumes. The results obtained using the data sets in

each of these categories are briefly summarized below separately for each body location both from the reviewed articles as well as from subsequent investigations using the respective data sets (Table 3). To fully evaluate the value of the findings summarized below the reader should consider the limitations of the specimen handling and image registration procedures (where applicable) described in detail in the original articles.

As a general rule, the automatic PET segmentation tools including those discussed below should not be directly used in the clinic. Mistreatment could occur due to large variations between clinics and patients. Validation of the segmentation tools for each particular PET/CT scanner, scanning protocol, body location, disease type as well as careful review and editing of the tumor contours by an experienced physician for each patient are needed.

Head and neck lesions

Tumor volume comparisons

The main findings for laryngeal squamous cell carcinoma (LSCC) published by Daisne et al. [25] were that the GTV volumes were significantly smaller when determined from the surgical specimen than when determined from CT, MRI and FDG PET. At the same time, the macroscopic tumor extensions were not completely covered by any of the three imaging modalities. Caldas-Magalhaes et al. [26] reached similar conclusions, since they also found that the average GTVs determined from CT, MRI and PET (GTV_{CT}, GTV_{MRI} and GTV_{PET}) were all larger than the average GTV determined from pathology (GTV_{path}) and that GTV_{PET} was the closest to GTV_{path}, but that CT and MRI provided better tumor coverage. Burri et al. [38] also reported that the tumor volumes they measured on PET images are generally smaller than those measured on CT.

Evaluation of PET segmentation methods

To reach the above conclusion that average GTV_{PET} values are smaller than the average GTV_{path} values, Daisne et al. [25] used a signal-to-background ratio (SBR)-based algorithm [46] and attributed the observed discrepancies to potential inaccuracy of the automatic PET image delineation and the limited PET resolution. Caldas-Magalhaes et al. [26] reached a similar conclusion for manually drawn PET contours and pointed that in their study the segmentation inaccuracy was larger than the registration error.

Geets et al. [47] used seven cases of the Louvain LSCC laryngeal data set [25] to test the validity of a gradient-based segmentation method. They found that when applied on denoised and deblurred images this gradient-based

Table 3 Types of PET segmentation methods tested against histopathology for different tumor types or body locations

Tumor type or location	Types of PET segmentation methods tested against pathology
Head and neck	Manual [26]
	Fixed thresholds (SUV and % of maximum) [38, 39, 40]
	Adaptive threshold (signal-to-background ratio based) [25, 46, 50, 52]
	Gradient [39, 47]
	Fuzzy C-means [48, 50]
	Active contours [49, 50]
	Possibility theory [53]
	Multimodality using level sets [51]
Esophagus	Fixed threshold (SUV and % of maximum) [41, 42]
	Visual [41, 42]
Non-small-cell lung cancer (NSCLC)	Threshold [27–29, 33, 43]
	Adaptive threshold [12, 31, 33, 52]
	Gradient [47]
	Fuzzy C-means [48]
	Active contours [49]
	FLAB [54]
	Neural network [55]
	Multimodality using level sets [51]
Cervical cancer	Threshold [34]
Colon, rectal and sigmoid cancer	Threshold [39, 45]
	Adaptive threshold [35, 44]
	Gradient [35, 39]

method was more accurate than the SBR method also used above [46], although it did not totally cover the macroscopic tumor volume. Belhassen et al. [48] used the same set of seven cases [25] to compare the performance of three fuzzy C-means (FCM) clustering algorithms and found that incorporating à trous wavelet transform to improve accuracy for heterogeneous cases results in more accurate delineation. These authors also reported that all three techniques failed to fully encompass the macroscopic tumor volumes. Abdoli et al. [49] also used the Louvain LSCC data set [25] to compare a contourlet-based active contour PET-AS tool aimed at accounting for the noise and heterogeneity of PET images and found it to be superior to adaptive threshold and two FCM methods. Zaidi et al. [50] used the Louvain LSCC data set [25] to compare the performance of nine algorithms including five threshold methods, a level set method, a stochastic expectation–maximization method, fuzzy clustering-based segmentation (FCM) and a spatial wavelet-based FCM (FCM-SW) and found FCM-SW to be most accurate. Markel et al. [51] also used the Louvain LSCC data set [25] to evaluate a

multimodality segmentation tool using level sets and Jensen-Renyi divergence (JRD). They compared the results to those from Zaidi et al. [50], and found that the JRD approach was second to the FCM-SW method.

A possibility theory-based PET-AS tool, the 42 % threshold, and two adaptive threshold methods [46, 52] were tested and compared for the LSCC data set [25] by Dewalle-Vignon et al. [53]. The authors demonstrated the "validity" of their possibility theory approach, which was developed to account for the inherent uncertainty and accuracy in the PET images, with respect to the other methods tested, but remarked that the method "does not globally result in superior results to that of some adaptive thresholding."

SUV thresholds were tested for head and neck lesions by Burri et al. [38] and Schinagl et al. [40]. Burri et al. [38] determined the pathology volume from "the maximal tri-dimensional lengths of each tumor" and found that the default SUV threshold of their software and narrowing the SUV "window" by one standard deviation were most likely to underestimate the tumor volume, while a SUV of 2.5 was most likely to overestimate it, and that a threshold at "40 % or greater maximum" "appears to offer the best compromise between accuracy and reducing the risk of underestimating tumor extent." Schinagl et al. [40] compared several PET-AS tools (SUV = 2.5, two fixed threshold and two adaptive threshold) to volumes of lymph node metastases from head and neck cancer and found that the last four tools performed worse if the primary tumor was used as a reference. They did not see an advantage to adding PET for lymph node segmentation, but did recommend using a PET-AS tool for improving reproducibility and comparison between institutions for therapy planning and assessment.

Esophagus

Evaluation of PET segmentation methods

Zhong et al. [41] found "that the optimal PET method to estimate the length of gross tumor varies with tumor length and SUVmax; an SUV cutoff of 2.5 provided the closest estimation in this study," when compared to visual interpretation and 40 % of maximum SUV.

Han et al. [42] segmented their FLT PET images using visual delineation and several thresholds (SUV cutoffs of 1.3, 1.4, 1.5, and taking 20, 25 and 30 % of the SUV_{max}). For their FDG PET image segmentation they used: visual delineation, SUV 2.5, and 40 % of SUV_{max}. They used the same specimen stretching procedure as Zhong et al. [41] and found that an SUV cutoff of 1.4 for FLT PET and 2.5 for FDG PET gave GTV lengths closest to pathology.

Lung lesions

Tumor volume comparisons

Similarly to the investigations for head and neck tumors, for NSCLC, Schaefer et al. [31], reported that both CT and PET overestimated the pathological tumor volume and that the PET volume was closer to it. Interestingly, GTV_{path} was less than GTV_{PET} for all the patients included in that study. They also found significant differences between the PET and pathology volumes in the lower lobe, but not so for the upper lobe. Wanet et al. [33] also found that FDG PET provided an average volume that was closer to GTV_{path}, when compared to CT, but not for all patients. In four of the patients studied by Stroom et al. [27], GTV_{PET} values were 13, 7, 7, and 24 ml, while GTV_{path} values were 6, 4, 8, and 39 ml, respectively.

Evaluation of PET segmentation methods

Most of the original lung studies evaluated threshold and adaptive threshold methods against NSCLC pathology volumes. Schaefer et al. [31] found a correlation of an adaptive threshold algorithm (which uses the mean SUV above 70 % of SUV_{max} and background as parameters) with the pathology findings. Yu et al. [29] performed a search to identify a SUV that would result in the best match for GTV_{path}. They found that "The mean (±SD) %SUV and absolute SUV that produced the best agreement between GTV_{path} and GTV_{PET} were 31 ± 11 % and 3.0 ± 1.6, respectively." In addition, they found that "the optimal threshold was inversely correlated with GTV_{path} or tumor diameter."

Wanet et al. [33] evaluated gradient-based, adaptive threshold and fixed threshold PET-AS methods and found that a gradient-based method outperformed threshold-based techniques and also that there was "no statistical difference between the different imaging modalities and delineation methods" by performing volume matching analysis using the Dice similarity coefficient. Abdoli et al. [49] also used nine patients from the Louvain lung case data [33] to evaluate their active contour PET-AS tool and found it to be superior to the other methods they tested, as they also found for the laryngeal cases above.

The MAASTRO NSCLC data set [12] was used by several research groups. Van Baardwijk et al. [12] evaluated an automatic SBR-based PET-AS method and showed it to result in good correlation with pathology measurements and in reduction of the inter-observer variability. This data set was also used by Hatt et al. [54] to study the impact of tumor size and heterogeneity on the delineated volume. They found that the Fuzzy Locally Adaptive Bayesian (FLAB) algorithm (designed to account for

image uncertainty due to noise as well as image blurring due to limited resolution) gave results closer to pathology than the 50 % of the maximum PET intensity threshold, T_{50} [43] and an adaptive threshold method [52]. They also found that for more heterogeneous tumors the threshold-based techniques more strongly underestimated the tumor volumes and suggested that such methods should not be used for large heterogeneous NSCLC ^{18}F-FDG PET images.

The same data set [12] was also used by Belhassen et al. [48] to test the three FCM clustering algorithms, which they also tested against the laryngeal lesions (above). They found that the wavelet transform-enhanced FCM resulted in a smaller mean error of the maximal diameter estimation also for the NSCLC lesions. Markel et al. [51] also used the MAASTRO NSCLC data set [12] to evaluate their multimodality segmentation tool using level sets and JRD and found that JRD outperformed an SBR method when using only PET and noted further performance improvement when information from both PET and CT is used. Sharif et al. [55] used the MAASTRO data set [12] to evaluate an artificial neural network approach.

Wu et al. [43] contoured automatically GTVs on PET images at 20, 30, 40, 45, 50, and 55 % of the maximal intensity level. They found that GTV_{CT} correlated better with pathology than GTV_{PET} and that one of their CT window and level settings and a PET threshold of 50 % of the maximum level "had the best correlation with pathologic results."

Microscopic tumor extensions

Few of the studies reported ME findings. Stroom et al. [27] found that ME_{max}, defined as the maximum of the minimum distances from the GTV to each ME islet for each patient, varied between 0 and 9 mm before deformation correction (average 5 mm) with an average of 9 mm after the correction. A follow-up of this study published by van Loon et al. [28], using the same specimen processing and registration procedures, further examined ME for NSCLC and found an association of mean CT tumor density and GTV_{CT} with the presence of ME. Using a statistical model, they divided the patients into two groups with high and low probability of ME and found that the mean CT number and GTV_{CT} are significant predictors of ME presence. They also found that GTV_{PET} (automatically delineated using a 42 % threshold of the maximum SUV) as well as GTV_{CT} accurately represent the Clinical Target Volume determined from pathology, CTV_{path}, for patients with low risk of ME, but that both GTV_{CT} and GTV_{PET} underestimate CTV_{path} for patients with high risk of ME, on average by 19.2 and 26.7 mm, respectively. Meng et al. [30] determined the maximal ME from all islets for each patient

without considering direction. They found that ME_{max} was significantly correlated with SUV_{max} and the metabolic tumor volume (MTV). To cover 95 % of ME, they suggested margins varying between 1.93 and 9.60 mm depending on SUV_{max}.

Colon, rectal and sigmoid cancer

In rectal cancer, Roels et al. [35] compared the closeness of GTV_{PET} (obtained with adaptive threshold and gradient-based segmentation methods) and GTV_{MRI} to GTV_{path}. They found that GTV_{PET} obtained with the gradient-based segmentation was closer to GTV_{path} than GTV_{MR} or GTV_{PET} obtained with the adaptive threshold method. They also reported a spatial discordance between MRI- and PET-based tumor volumes of approximately 50 %, which could be in part related to rectal filling with MRI contrast.

Buijsen et al. [44] found that rectal tumor lengths determined by a SBR-based PET-AS method show the strongest correlation with lengths measured on pathology, compared with tumor lengths determined from the CT and MRI images. Chen et al. [45] tested segmentation thresholds at 20, 30, 40 and 50 % of SUV_{max} and found that a 30 % threshold of the PET maximum uptake provides an adequate tumor length and width for tumors in the colon and in the sigmoid.

Sridhar et al. [39] tested several threshold and a gradient segmentation methods for segmenting head and neck, lung and colorectal tumors and found the gradient method to have "superior correlation and reliability with the estimated ellipsoid pathologic volume."

Cervical cancer

Zhang et al. [34] searched for optimal segmentation thresholds and found that for their 10 cervical cancer patients the optimal percent and absolute SUV thresholds were 40.50 ± 3.16 % and 7.45 ± 1.10, respectively. They also found that the optimal percent SUV threshold was inversely correlated with GTV_{path} and tumor diameter and that the SUV threshold was positively correlated with SUV_{max}.

Discussion and conclusions

Summary and critical analysis of the literature

Using histopathology results of excised lesions to validate PET images is challenging. Therefore, efforts in this direction, including the papers summarized in this review, provide indispensable data toward solving the dilemma of how PET images should be used to define the tumor

volume. While all these investigations contribute toward finding a solution the problem, the most valuable are those that manage to provide an estimate of the 3D shape of the lesion based on pathology, since in addition to providing information about the tumor volume or diameter they may also locate the border of the lesion in the PET image. This was achieved through fixation of the specimen through freezing [25], inflation [27, 28] and/or placement in formalin [29] followed by corrections for tissue retraction and/or deformation. Despite these meticulous efforts to preserve the shape of the lesion after excision, fixation and slicing, the accuracy of the respective corrections for shrinkage and deformation and their effect on the validation accuracy is investigated only in a few studies [25–29, 35]. H&E staining was used in the studies in which microscopic histopathology analysis was performed.

Since the volumetric studies require less processing of the specimen, they present the possibility of having a larger number of patients and thus better statistics. These studies, however, do not provide sufficient information for strict evaluation of the segmentation methods, since as reported by Daisne et al. [25], even if similar in size, the GTV from the PET image may not overlap with the pathology volume.

Practically all the investigations reviewed in this paper used ^{18}F-loaded FDG except one (Han et al. [42]), which investigated both FDG and FLT PET. Also, most of the studies considered only GTV_{path} [25, 33, 34, 38, 56], but a few also evaluated the PET-derived GTV against both the GTV_{path} and the CTV_{path} [27, 28]. While the additional comparison with the CTV_{path} further complicates the investigation, it is very valuable in providing the CTV tumor margin, which in addition to being disease- and location- dependent may also be anisotropic.

The summarized studies also differ in how, and how much, GTV_{path} was used for evaluating various PET segmentation approaches. An upcoming review which lists the segmentation tools evaluated against PSPVs will be presented in the first TG211 report [9]. The majority of the pathology-validated PET image data sets reviewed here were originally presented by their authors in conjunction with some segmentation contours, although evaluating the contouring method may have not been their primary goal. The segmentation tools evaluated against the pathology in the original publications were mostly simple threshold or adaptive threshold methods. More advanced segmentation methods, which promise to be able to handle realistic tumors with irregular shape and non-uniform activity, have been evaluated in later publications against some of the PSPVs reviewed here [47–51, 53–55]. Important conclusions have been reached for these more advanced methods, as pointed out in the previous section. At the same time, the TG211 report [9] points about PSPVs that "several sources

of error in the production of these data sets should be acknowledged: (1) deformation of the surgical specimen after excision, (2) time difference between the PET scan and the specimen excision, (3) imperfect delineation of metabolic boundaries in digitized histopathology, and (4) imperfect co-registration between histopathology and PET image spaces."

As pointed out above, a few of the laryngeal and NSCLC investigations made the interesting observation that the deformation of the excised lesion can be neglected. Due to insufficiency of the data provided it is difficult to assess the meaning and accuracy of this statement, especially when observing the difference in lesion shape between the macro-specimens and the CT in Figs. 1 and 2. Probably what was meant was that the deformations of the lesions were much smaller than those of the surrounding soft tissues. As pointed out by many of the investigators, deformations both during fixation and slicing are possible.

Possible directions for improvement are to increase registration accuracy and reduce the time between patient scan and lesion excision. Applying correction factors for changes in the specimen during the fixation process [57] is necessary, although for some (e.g., laryngeal, cortical bone) specimens these changes may be small or negligible. Providing an estimate of the accuracy of the deformation correction factors is also desirable to verify that the level of accuracy is sufficient for evaluating PET segmentation methods. Free-breathing, non-gated PET scans were acquired in most studies except one [33], where the PET scan was gated. In some cases, the time between PET and surgery was long enough (up to 3 weeks) to expect tumor changes.

Due to these potential sources of error, as well as the difficulty in accumulating more PSPVs, many of the PET-AS methods have also been tested against expert delineation or images from simulated and experimental phantoms as described in several reviews [9, 14, 15]. These reviews also list other very promising and advanced methods, which, to our knowledge, have not yet been tested against histopathology-based ground truth. This cannot be considered as a disadvantage of such PET-AS methods, bearing in mind the more accurate registration of the ground truth with PET images for experimental and numerical phantoms. Despite this, given the many factors that can affect a clinical image and may not be exactly represented in the simulations (e.g., biological uncertainty, image noise, etc.); testing the segmentation methods on clinical images with some type of pathology-based ground truth is highly desirable. As pointed out in the upcoming TG211 report, PET segmentation should ideally be evaluated against a combination of phantom and clinical images with reliable ground truth in a standardized way.
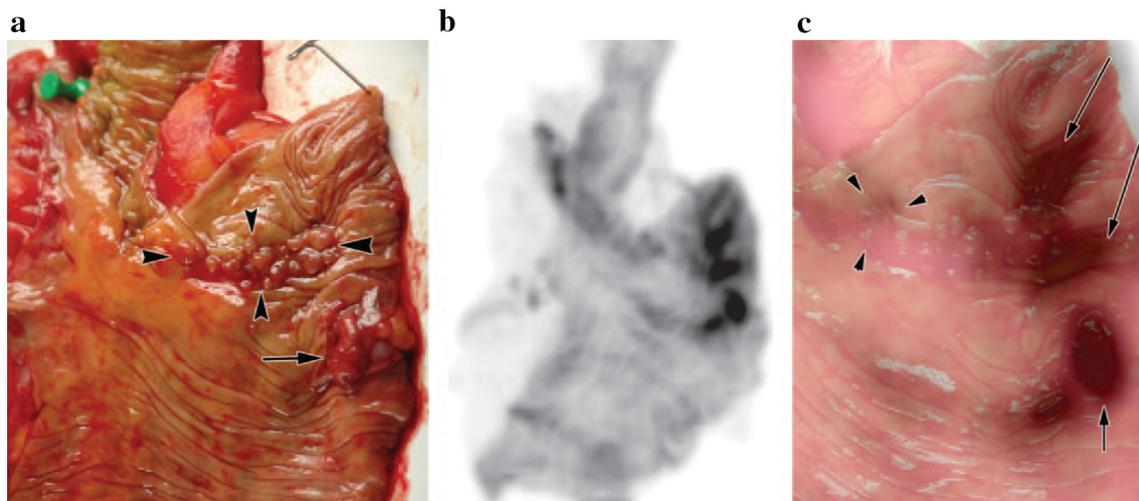
Fig. 3 From Gollub et al. [62]: "Cecal polyps (carpet lesions) in colon specimen resected at right hemicolectomy in 75-year-old woman. **a** Magnified gross specimen shows tubullovillous adenoma (carpet lesion) (*arrowheads*) and polyp (*arrow*). **b** PET image of gross specimen. **c** PET image fused with digital photograph of colon shows high FDG avidity of polyp (*short arrow*). FDG uptake is concentrated in folds of bunched-up mucosa near the carpet lesion (*long arrows*); some lesion activity may be present, but this is mostly bunched-up mucosa. An SUV of 2.9 is associated with a portion of the tubullovillous adenoma closer to the ileocecal valve (*arrowheads*)." Reprinted, with permission from Gollub et al: Feasibility of ex vivo FDG PET of the colon. Radiology 252(1):232–239, RSNA, 2009 (color figure online)

When considering the results from evaluation of various PET segmentation methods, it is very important to consider the PET scanners and protocols used in the different studies, since they may significantly affect the PET image and therefore the segmentation results. These differences between the scanner, protocol and procedures used by different institutions should be investigated and the accuracy of the segmentation algorithm should be tested by each user and the method adapted to his/her particular setting before using that algorithm for radiotherapy planning [8]. In general, validation of PET contours against the pathology-defined ground truth aims to resolve modifications of the PET image due to both physical artifacts and biological phenomena. However, since the physical artifacts are scanner- and protocol- dependent and the biological phenomena are patient-dependent, translating the results from published pathology validations to different patients in different institutions remains a challenge. Current efforts to standardize imaging protocols [7, 58, 59], as well as the work of several task groups (e.g., AAPM TG 174), will reduce differences between PET images due to physical factors, but this does not address patient-specific biological variations.

Despite the significant contribution of the investigations contributing PSPV, their number remains small and insufficient due to the substantial experimental burden and difficulties in producing pathology-based definitions of lesions and in registration of the pathology-derived ground truth with the PET image. The problem is compounded by the fact that variation of tumor type, stage and location in the body often results in large variations in the level and heterogeneity of PET tracer uptake in the tumor and in the surrounding healthy tissues. In addition, the recently observed heterogeneity of genetic mutations [60] introduces practically infinite degrees of freedom for the tumor genetic identity, which may also manifest in different metabolic representation. Therefore, continuing these efforts may be strongly affected by confirmation of the hypothesis that cancers of the same type have a common metabolic representation. More data of the kind summarized in this review, but with the addition of the extra dimension of tumor genetic mutations, will need to be accumulated to address this hypothesis [61].

Future directions

Approaches other than those summarized in this review (pre-excision PET/CT scan followed by pathological evaluation of the excised tumor) may contribute to resolve the above hypothesis. They include the possibility of carrying out post-excision PET, in other words a high-resolution micro-PET scan of an excised lesion containing a PET tracer as performed by Gollub et al. [62] (Fig. 3). This could allow for the correlation of pathology of excised lesions with ex vivo PET at higher resolution and could be helpful in providing further data on microscopic extensions and CTV definition. A recent investigation using ex vivo PET claimed that using such an approach is promising for evaluating segmentation techniques and provided images on the Internet to facilitate evaluation of segmentation

techniques [63]. It should be kept in mind however, that histopathological validation is needed, and that there may be differences between the PET image of an excised lesion and the clinical image of the same lesion due to physical artifacts, differences in the background activity and deformation of the lesion, which can all affect the segmentation process.

In a recently published study, Axente et al. [64] proposed another alternative for generating pathological data sets for PET segmentation validation. They tested their approach in a small animal model. It consisted of injecting a mouse with ^{14}C-FDG, which was sacrificed 80 min post injection. The tumor was then extracted and sliced. An autoradiography of the slices was acquired to image the activity distribution in the tumor, and a 3D reconstruction of the radiotracer distribution was performed. A PET scan was simulated based on the tracer uptake distribution. This method appears very promising to improve the accuracy of the pathology-based ground truth in the PET images, since the registration error was found to be very low.

Another opportunity for accumulating such data is to correlate the histopathology of the biopsy specimen obtained under PET/CT-guided biopsies [65] with the PET image. Such investigations would have the advantage of high spatial accuracy due to the visibility of the biopsy needle in the PET/CT image. In addition, performing autoradiography of the biopsy specimen provides an opportunity to determine the tracer distribution with higher spatial resolution than PET [66]. The data which can be obtained by such studies are limited to the point of biopsy needle insertion. This, however, may be partly compensated for by the large number of biopsy procedures performed and their routine use in oncology. Correlations of the specimen histopathology with the PET/CT image obtained during a biopsy procedure in the operating room would be spatially more accurate than current investigations. However, even for specimens extracted under CT guidance, correlations with patients' PET scans prior to the biopsy might also provide useful data, albeit with less spatial accuracy.

If the hypothesis described above is resolved and sufficient PET-histopathology correlations are accumulated for different tumor types, this may allow for more reliable definition of the lesion border from the PET image for localized therapies.

References

1. MacManus M, Nestle U, Rosenzweig KE, Carrio I, Messa C, Belohlavek O, Danna M, Inoue T, Deniaud-Alexandre E, Schipani S, Watanabe N, Dondi M, Jeremic B (2009) Use of PET and PET/CT for radiation therapy planning: IAEA expert report 2006-2007. Radiother Oncol 91(1):85–94

2. Nestle U, Weber W, Hentschel M, Grosu AL (2009) Biological imaging in radiation therapy: role of positron emission tomography. Phys Med Biol 54(1):R1–R25

3. Shyn PB (2013) Interventional positron emission tomography/computed tomography: state-of-the-art. Tech Vasc Interv Radiol 16(3):182–190

4. LoSasso T (2003) Quality assurance of IMRT. In: A practical guide to intensity-modulated radiation therapy. Medical Physics Publishing, Madison

5. Nehmeh SA, Erdi YE, Meirelles GSP, Squire O, Larson SM, Humm JL, Schoder H (2007) Deep-inspiration breath-hold PET/CT of the thorax. J Nucl Med 48(1):22–26

6. Njeh CF, Dong L, Orton CG (2013) Point/Counterpoint. IGRT has limited clinical value due to lack of accurate tumor delineation. Med Phys 40(4):040601

7. Boellaard R (2009) Standards for PET image acquisition and quantitative data analysis. J Nucl Med 50(Suppl 1):11S–20S

8. Kirov AS, Schmidtlein CR, Kang H, Lee N (2012) Rationale, instrumental accuracy, and challenges of PET quantification for tumor segmentation in radiation treatment planning, in Positron Emission Tomography-Current Clinical and Research Aspects. In: Hsieh C-H (ed). ISBN:978-953-307-824-3, InTech. http://www.intechopen.com/books/positron-emission-tomography-current-clinical-and-research-aspects/rationale-instrumental-accuracy-and-challenges-of-pet-quantification-for-tumor-segmentation-in-radia

9. Hatt M, Lee J, Schmidtlein CR, Naqa IE, Caldwell C, Bernardi ED, Lu W, Geets SDX, Gregoire V, Jeraj R, MacManus M, Mawlawi O, Nestle U, Pugachev A, Schöder H, Shepherd T, Spezi E, Visvikis D, Zaidi H, Kirov AS Report of AAPM TG211: Classification and evaluation strategies of auto-segmentation approaches for PET. Under review by the AAPM

10. Schoellnast H, Larson SM, Nehmeh SA, Carrasquillo JA, Thornton RH, Solomon SB (2011) Radiofrequency ablation of non-small-cell carcinoma of the lung under real-time FDG PET CT guidance. Cardiovasc Intervent Radiol 34(Suppl 2):S182–S185

11. Ryan ER, Sofocleous CT, Schoder H, Carrasquillo JA, Nehmeh S, Larson SM, Thornton R, Siegelbaum RH, Erinjeri JP, Solomon SB (2013) Split-dose technique for FDG PET/CT-guided percutaneous ablation: a method to facilitate lesion targeting and to provide immediate assessment of treatment effectiveness. Radiology 268(1):288–295

12. van Baardwijk A, Bosmans G, Boersma L, Buijsen J, Wanders S, Hochstenbag M, van Suylen RJ, Dekker A, Dehing-Oberije C, Houben R, Bentzen SM, van Kroonenburgh M, Lambin P, De Ruysscher D (2007) PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. Int J Radiat Oncol Biol Phys 68(3):771–778

13. Hatt M, Cheze Le Rest C, Albarghach N, Pradier O, Visvikis D (2011) PET functional volume delineation: a robustness and repeatability study. Eur J Nucl Med Mol Imaging 38(4):663–672

14. Lee JA (2010) Segmentation of positron emission tomography images: some recommendations for target delineation in radiation oncology. Radiother Oncol 96(3):302–307

15. Zaidi H, El Naqa I (2010) PET-guided delineation of radiation therapy treatment volumes: a survey of image segmentation techniques. Eur J Nucl Med Mol Imaging 37(11):2165–2187

16. Erdi YE, Mawlawi O, Larson SM, Imbriaco M, Yeung H, Finn R, Humm JL (1997) Segmentation of lung lesion volume by adaptive positron emission tomography image thresholding. Cancer 80(12 Suppl):2505–2509

17. NEMA NU 2-2001 (2001) Performance measurements of positron emission tomographs. National Electrical Manufacturers Association, Rosslyn, VA, USA

18. Drever L, Robinson DM, McEwan A, Roa W (2006) A local contrast based approach to threshold segmentation for PET target volume delineation. Med Phys 33(6):1583–1594

19. Shepherd T, Berthon B, Galavis P, Spezi E, Apte A, Lee J, Visvikis D, Hatt M, de Bernardi E, Das S, El Naqa I, Nestle U, Schmidtlein CR, Zaidi H, Kirov A (2012) Design of a benchmark platform for evaluating PET-based contouring accuracy in oncology applications. In: European Association for Nuclear Medicine Annual Congress, 27–31 Oct 2012, Milan, Italy, Eur J Nucl Med Mol Imaging, Vol 39, Suppl 2, p S264

20. Zito F, De Bernardi E, Soffientini C, Canzi C, Casati R, Gerundini P, Baselli G (2012) The use of zeolites to generate PET phantoms for the validation of quantification strategies in oncology. Med Phys 39(9):5353–5361

21. Le Maitre A, Segars WP, Marache S, Reilhac A, Hatt JM, Tomei S, Lartizien C, Visvikis D (2009) Incorporating patient-specific variability in the simulation of realistic whole-body 18F-FDG distributions for oncology applications. Proc IEEE 97(12):2026–2038

22. Harrison R, Gillispie S, Schmitz R, Lewellen T (2008) Modeling block detectors in SimSET. J Nucl Med 49(Suppl 1):410

23. Jan S, Santin G, Strul D, Staelens S, Assie K, Autret D, Avner S, Barbier R, Bardies M, Bloomfield PM, Brasse D, Breton V, Bruyndonckx P, Buvat I, Chatziioannou AF, Choi Y, Chung YH, Comtat C, Donnarieix D, Ferrer L, Glick SJ, Groiselle CJ, Guez D, Honore PF, Kerhoas-Cavata S, Kirov AS, Kohli V, Koole M, Krieguer M, van der Laan DJ, Lamare F, Largeron G, Lartizien C, Lazaro D, Maas MC, Maigne L, Mayet F, Melot F, Merheb C, Pennacchio E, Perez J, Pietrzyk U, Rannou FR, Rey M, Schaart DR, Schmidtlein CR, Simon L, Song TY, Vieira JM, Visvikis D, Van de Walle R, Wieers E, Morel C (2004) GATE: a simulation toolkit for PET and SPECT. Phys Med Biol 49(19):4543–4561

24. Jan S, Benoit D, Becheva E, Carlier T, Cassol F, Descourt P, Frisson T, Grevillot L, Guigues L, Maigne L, Morel C, Perrot Y, Rehfeld N, Sarrut D, Schaart DR, Stute S, Pietrzyk U, Visvikis D, Zahra N, Buvat I (2011) GATE V6: a major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy. Phys Med Biol 56(4):881–901

25. Daisne JF, Duprez T, Weynand B, Lonneux M, Hamoir M, Reychler H, Gregoire V (2004) Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. Radiology 233(1):93–100

26. Caldas-Magalhaes J, Kasperts N, Kooij N, van den Berg CA, Terhaard CH, Raaijmakers CP, Philippens ME (2012) Validation of imaging with pathology in laryngeal cancer: accuracy of the registration methodology. Int J Radiat Oncol Biol Phys 82(2):e289–e298

27. Stroom J, Blaauwgeers H, van Baardwijk A, Boersma L, Lebesque J, Theuws J, van Suylen RJ, Klomp H, Liesker K, van Pel R, Siedschlag C, Gilhuijs K (2007) Feasibility of pathology-correlated lung imaging for accurate target definition of lung tumors. Int J Radiat Oncol Biol Phys 69(1):267–275

28. van Loon J, Siedschlag C, Stroom J, Blauwgeers H, van Suylen RJ, Knegjens J, Rossi M, van Baardwijk A, Boersma L, Klomp H, Vogel W, Burgers S, Gilhuijs K (2012) Microscopic disease extension in three dimensions for non-small-cell lung cancer: development of a prediction model using pathology-validated positron emission tomography and computed tomography features. Int J Radiat Oncol Biol Phys 82(1):448–456

29. Yu J, Li X, Xing L, Mu D, Fu Z, Sun X, Sun X, Yang G, Zhang B, Sun X, Ling CC (2009) Comparison of tumor volumes as determined by pathologic examination and FDG-PET/CT images of non-small-cell lung cancer: a pilot study. Int J Radiat Oncol Biol Phys 75(5):1468–1474

30. Meng X, Sun X, Mu D, Xing L, Ma L, Zhang B, Zhao S, Yang G, Kong FM, Yu J (2012) Noninvasive evaluation of microscopic tumor extensions using standardized uptake value and metabolic tumor volume in non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 82(2):960–966

31. Schaefer A, Kim YJ, Kremp S, Mai S, Fleckenstein J, Bohnenberger H, Schafers HJ, Kuhnigk JM, Bohle RM, Rube C, Kirsch CM, Grgic A (2013) PET-based delineation of tumour volumes in lung cancer: comparison with pathological findings. Eur J Nucl Med Mol Imaging 40(8):1233–1244

32. Dahele M, Hwang D, Peressotti C, Sun L, Kusano M, Okhai S, Darling G, Yaffe M, Caldwell C, Mah K, Hornby J, Ehrlich L, Raphael S, Tsao M, Behzadi A, Weigensberg C, Ung YC (2008) Developing a methodology for three-dimensional correlation of PET-CT images and whole-mount histopathology in non-small-cell lung cancer. Curr Oncol 15(5):62–69

33. Wanet M, Lee JA, Weynand B, De Bast M, Poncelet A, Lacroix V, Coche E, Gregoire V, Geets X (2011) Gradient-based delineation of the primary GTV on FDG-PET in non-small cell lung cancer: a comparison with threshold-based approaches, CT and surgical specimens. Radiother Oncol 98(1):117–125

34. Zhang Y, Hu J, Lu HJ, Li JP, Wang N, Li WW, Zhou YC, Liu JY, Wang SJ, Wang J, Li X, Ma WL, Wei LC, Shi M (2013) Determination of an optimal standardized uptake value of fluorodeoxyglucose for positron emission tomography imaging to assess pathological volumes of cervical cancer: a prospective study. PLoS ONE 8(11):e75159

35. Roels S, Slagmolen P, Nuyts J, Lee JA, Loeckx D, Maes F, Vandecaveye V, Stroobants S, Ectors N, Penninckx F, Haustermans K (2009) Biological image-guided radiotherapy in rectal cancer: challenges and pitfalls. Int J Radiat Oncol Biol Phys 75(3):782–790

36. Daisne JF, Sibomana M, Bol A, Cosnard G, Lonneux M, Gregoire V (2003) Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. Radiother Oncol 69(3):237–245

37. Siedschlag C, van Loon J, van Baardwijk A, Rossi MM, van Pel R, Blaauwgeers JL, van Suylen RJ, Boersma L, Stroom J, Gilhuijs KG (2009) Analysis of the relative deformation of lung lobes before and after surgery in patients with NSCLC. Phys Med Biol 54(18):5483–5492

38. Burri RJ, Rangaswamy B, Kostakoglu L, Hoch B, Genden EM, Som PM, Kao J (2008) Correlation of positron emission tomography standard uptake value and pathologic specimen size in cancer of the head and neck. Int J Radiat Oncol Biol Phys 71(3):682–688

39. Sridhar P, Mercier G, Tan J, Truong MT, Daly B, Subramaniam RM (2014) FDG PET metabolic tumor volume segmentation and pathologic volume of primary human solid tumors. AJR Am J Roentgenol 202(5):1114–1119

40. Schinagl DA, Span PN, van den Hoogen FJ, Merkx MA, Slootweg PJ, Oyen WJ, Kaanders JH (2013) Pathology-based validation of FDG PET segmentation tools for volume assessment of lymph node metastases from head and neck cancer. Eur J Nucl Med Mol Imaging 40(12):1828–1835

41. Zhong X, Yu J, Zhang B, Mu D, Zhang W, Li D, Han A, Song P, Li H, Yang G, Kong FM, Fu Z (2009) Using 18F-fluorodeoxyglucose positron emission tomography to estimate the length of gross tumor in patients with squamous cell carcinoma of the esophagus. Int J Radiat Oncol Biol Phys 73(1):136–141

42. Han D, Yu J, Yu Y, Zhang G, Zhong X, Lu J, Yin Y, Fu Z, Mu D, Zhang B, He W, Huo Z, Liu X, Kong L, Zhao S, Sun X (2010) Comparison of (18)F-fluorothymidine and (18)F-fluorodeoxyglucose PET/CT in delineating gross tumor volume by optimal threshold in patients with squamous cell carcinoma of thoracic esophagus. Int J Radiat Oncol Biol Phys 76(4):1235–1241

43. Wu K, Ung YC, Hornby J, Freeman M, Hwang D, Tsao MS, Dahele M, Darling G, Maziak DE, Tirona R, Mah K, Wong CS (2010) PET CT thresholds for radiotherapy target definition in non-small-cell lung cancer: how close are we to the pathologic findings? Int J Radiat Oncol Biol Phys 77(3):699–706

44. Buijsen J, van den Bogaard J, Janssen MH, Bakers FC, Engelsman S, Ollers M, Beets-Tan RG, Nap M, Beets GL, Lambin P, Lammering G (2011) FDG-PET provides the best correlation with the tumor specimen compared to MRI and CT in rectal cancer. Radiother Oncol 98(2):270–276

45. Chen SW, Chen WT, Wu YC, Yen KY, Hsieh TC, Lin TY, Kao CH (2013) Which FDG/PET parameters of the primary tumors in colon or sigmoid cancer provide the best correlation with the pathological findings? Eur J Radiol 82(9):e405–e410

46. Daisne JF, Sibomana M, Bol A, Doumont T, Lonneux M, Gregoire V (2003) Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: influence of reconstruction algorithms. Radiother Oncol 69(3):247–250

47. Geets X, Lee JA, Bol A, Lonneux M, Gregoire V (2007) A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur J Nucl Med Mol Imaging 34(9):1427–1438

48. Belhassen S, Zaidi H (2010) A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET. Med Phys 37(3):1309–1324

49. Abdoli M, Dierckx RA, Zaidi H (2013) Contourlet-based active contour model for PET image segmentation. Med Phys 40(8):082507

50. Zaidi H, Abdoli M, Fuentes CL, El Naqa IM (2012) Comparative methods for PET image segmentation in pharyngolaryngeal squamous cell carcinoma. Eur J Nucl Med Mol Imaging 39(5):881–891

51. Markel D, Zaidi H, El Naqa I (2013) Novel multimodality segmentation using level sets and Jensen-Renyi divergence. Med Phys 40(12):121908

52. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rube C, Kirsch CM (2005) Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer. J Nucl Med 46(8):1342–1348

53. Dewalle-Vignion AS, Betrouni N, Lopes R, Huglo D, Stute S, Vermandel M (2011) A new method for volume segmentation of pet images, based on possibility theory. IEEE Trans Med Imaging 30(2):409–423

54. Hatt M, Cheze-le Rest C, van Baardwijk A, Lambin P, Pradier O, Visvikis D (2011) Impact of tumor size and tracer uptake heterogeneity in (18)F-FDG PET and CT non-small cell lung cancer tumor delineation. J Nucl Med 52(11):1690–1697

55. Sharif MS, Abbod M, Amira A, Zaidi H (2010) Artificial neural network-based system for PET volume segmentation. Int J Biomed Imaging. 1–11, article id 105610

56. Yu W, Fu XL, Zhang YJ, Xiang JQ, Shen L, Jiang GL, Chang JY (2009) GTV spatial conformity between different delineation methods by 18FDG PET/CT and pathology in esophageal cancer. Radiother Oncol 93(3):441–446

57. Hsu PK, Huang HC, Hsieh CC, Hsu HS, Wu YC, Huang MH, Hsu WH (2007) Effect of formalin fixation on tumor size determination in stage I non-small cell lung cancer. Ann Thorac Surg 84(6):1825–1829

58. Das et al (2014) Upcoming report of APPM TG 174: Utilization of 18F-fluorodeoxyglucose positron emission tomography (FDG-PET) in radiation therapy, In preparation

59. Boellaard R, O'Doherty MJ, Weber WA, Mottaghy FM, Lonsdale MN, Stroobants SG, Oyen WJ, Kotzerke J, Hoekstra OS, Pruim J, Marsden PK, Tatsch K, Hoekstra CJ, Visser EP, Arends B, Verzijlbergen FJ, Zijlstra JM, Comans EF, Lammertsma AA, Paans AM, Willemsen AT, Beyer T, Bockisch A, Schaefer-Prokop C, Delbeke D, Baum RP, Chiti A, Krause BJ (2010) FDG PET and PET/CT: EANM procedure guidelines for tumour PET imaging: version 1.0. Eur J Nucl Med Mol Imaging 37(1):181–200

60. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med 366(10):883–892

61. Mitsudomi T, Suda K, Yatabe Y (2013) Surgery for NSCLC in the era of personalized medicine. Nat Rev Clin Oncol 10(4):235–244

62. Gollub MJ, Akhurst TJ, Williamson MJ, Shia J, Humm JL, Wong WD, Paty PB, Guillem JG, Weiser MR, Temple LK, Dauer LT, Jhanwar SC, Kronman RE, Montalvo CV, Miller AR, Larson SM, Margulis AR (2009) Feasibility of ex vivo FDG PET of the colon. Radiology 252(1):232–239

63. Prieto EH, Pardo JL, Peñuelas FJ, Richter I, Martí-Climent JA, Gómez-Fernández JM, García-Velloso M, Valero MJ, Garrastachu M (2014) Validation of segmentation techniques for positron emission tomography using ex vivo images of oncological surgical specimens. Rev Esp Med Nucl Imagen Mol 33:79–86

64. Axente M, He J, Bass CP, Sundaresan G, Zweit J, Williamson JF, Pugachev A (2014) An alternative approach to histopathological validation of PET imaging for radiation therapy image-guidance: a proof of concept. Radiother Oncol 110(2):309–316

65. Cerci JJ, Neto CCP, Krauzer C, Sakamoto DG, Vitola JV (2013) The impact of coaxial core biopsy guided by FDG PET/CT in oncological patients. Eur J Nucl Med Mol Imaging 40(1):98–103

66. Kirov AS, Fanchon L, Dogan S, Moreira AL, Apte A, Schmidtlein CR, Carlin SA, Schöder H, Solomon SB, Humm JL (2014) In situ histopathological correlation of FDG uptake by autoradiography of needle biopsy specimens obtained under PET-CT guidance, Society of Nuclear Medicine and Medical Imaging Annual Meeting, St. Louis, Missouri, June 7–11 2014, J Nucl Med 2014, 55 (Suppl 1):586