

ORIGINAL ARTICLE

An applicable method for PET/CT image quality assessment and comparison among three PET/CT systems with similar physical performance in cancer patients

Jian Tian¹, Kun Zhou¹, Zhaoheng Xie¹, Baixuan Xu², Jiahe Tian², Yingyin Chen³, Xiaoyi Zhu¹, Qiushi Ren¹

¹Department of Biomedical Engineering, Peking University, No.5, Beijing 100871, China; ²The General Hospital of the People's Liberation Army, No.28, Fuxing, Beijing 100039, China; ³Guangdong Provincial Key Laboratory of Medical Image Processing, School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China.

Summary

Purpose: This research proposes a method with specific procedure guideline for clinical PET/CT image quality assessment according to physicians' behavior of image interpretation and explore the relationship between image quality and image systems with similar physical performance.

Methods: Clinical PET/CT were divided according to body location: brain, chest, abdomen and pelvic cavity. We explored the lesions and suspicious regions where radiologists concerned most through eye-tracker and behavior observation study to generate an assessment checklist. Fifty-five patients who were statistically consistent in age, weight and height were studied. Thirty-seven were scanned with an experimental scanner A and control systems B or C because their clinical pathways required PET/CT examinations at short intervals, the other 18 were scanned with scanners A

and C. The grade of every system's PET, CT and PET/CT image performance on the four parts was calculated by subtraction of mean value and variance between experimental and control systems.

Results: The scoring checklist was set for PET, CT and PET/CT images in four parts respectively, and a standard procedure guideline was formulated for assessment. Using assessment criteria, the statistical results objectively reflected certain systems' superiority on certain modalities and certain parts of the body.

Conclusion: Our criteria for clinical PET/CT image quality assessment and comparison were efficient.

Key words: PET/CT, clinical criterion, image quality assessment, equipment comparison

Introduction

Combined positron emission tomography and computed tomography (18F-FDG PET/CT) is currently widely used in the clinical diagnosis, staging and treatment response monitoring of cancer to provide functional and morphological imaging for a patient from a single scanning session [1]. The most significant purpose of PET/CT clinical images is to provide sufficient information, from which clinicians can estimate cancer stage and/or assess response to therapy [2]. Clinical PET/CT im-

ages with "high quality" are crucial. Image quality has two meanings, one is fidelity, which represents the similarity between reconstructed PET/CT images and corresponding gradients slices within the body; the other one is intelligibility, which illustrates the image's capacity to provide information to an interpreter.

Since 1991, National Electrical Manufacturers Association (NEMA) published a series of test specifications named NEMA NU2 in America,

Corresponding author: Qiushi Ren, PhD. Department of Biomedical Engineering, Peking University, No.5, Yiheyuan Rd, Beijing 100871, China.

Tel: +86 010 62767113, Email: renqsh@coe.pku.edu.cn

Received: 05/05/2019; Accepted: 11/06/2019

which is based on the industry's opinions, especially the PET/CT manufacturer's attitude. Generally, image quality is tested by performance evaluation through standard NEMA quality phantoms. Rausch et al [3] presents how to use it to simulate lung tissue and background with different ^{18}F -FDG density in six spheres; the better the key performance (contrast, standard deviation and residual error) obtained, the higher the image quality will be. It is true to some extent, however, based on previous evidence [4], that there seems to be big conflict between using PET/CT system as a diagnostic imaging tool and using it as a quantitative imaging biomarker. Engineers are always updating new technology and optimizing PET/CT system [5] to pursue higher quantitative physical performance that ignored clinical needs like lesion detectability. In clinical practice, what matters most is whether there is sufficient clinical information that can be interpreted by the physicians.

Messerli [6], Orlandini et al [7] and Sah et al [8] developed various reconstruction algorithms to improve PET/CT image quality on clinical data. When evaluating the result of their method, Messerli and Orlandini et al use standard uptake value (SUV) only when use PET/CT just as a quantitative imaging biomarker. Sah et al. realizes the necessity and importance of experienced nuclear medicine physicians, artifacts, image sharpness, noise, and lesion detectability analyzed using a four-point scale.

While this simple assessment method is reasonable, effective and applicable, there is no explanation or certification. On the other hand, with the rapid increase of PET/CT usage and cancer incidence in recent years, experienced nuclear medicine physicians have become valuable and rare worldwide. Considering the complexity of the *in vivo* state and human bias, this paper aims to find out an applicable PET/CT image quality evaluation method, which is easy to use for not only experienced but also junior physicians.

Through behavioral observation study with eye tracking system, we find out what are physicians truly looking at and what they are concerned about when reading clinical images. Combining with cancer epidemiology statistics, we verified and summed up the key points with words physicians prefer and understand, which is the PET/CT image quality checklist and procedure guideline. This PET/CT image quality assessment method can assist in evaluating new equipment [9,10], new algorithms [11], regulating the labeling of PET/CT devices, comparing different devices to make judicious decisions [12], and validating the claims of instrument promotion. Besides, using this PET/CT image quality assessment method, we performed a comparison among three PET/CT systems from different manufacturers with 55 clinical trials to verify the effectiveness and feasibility, as well as the relationship between physical performance and clinical image quality.

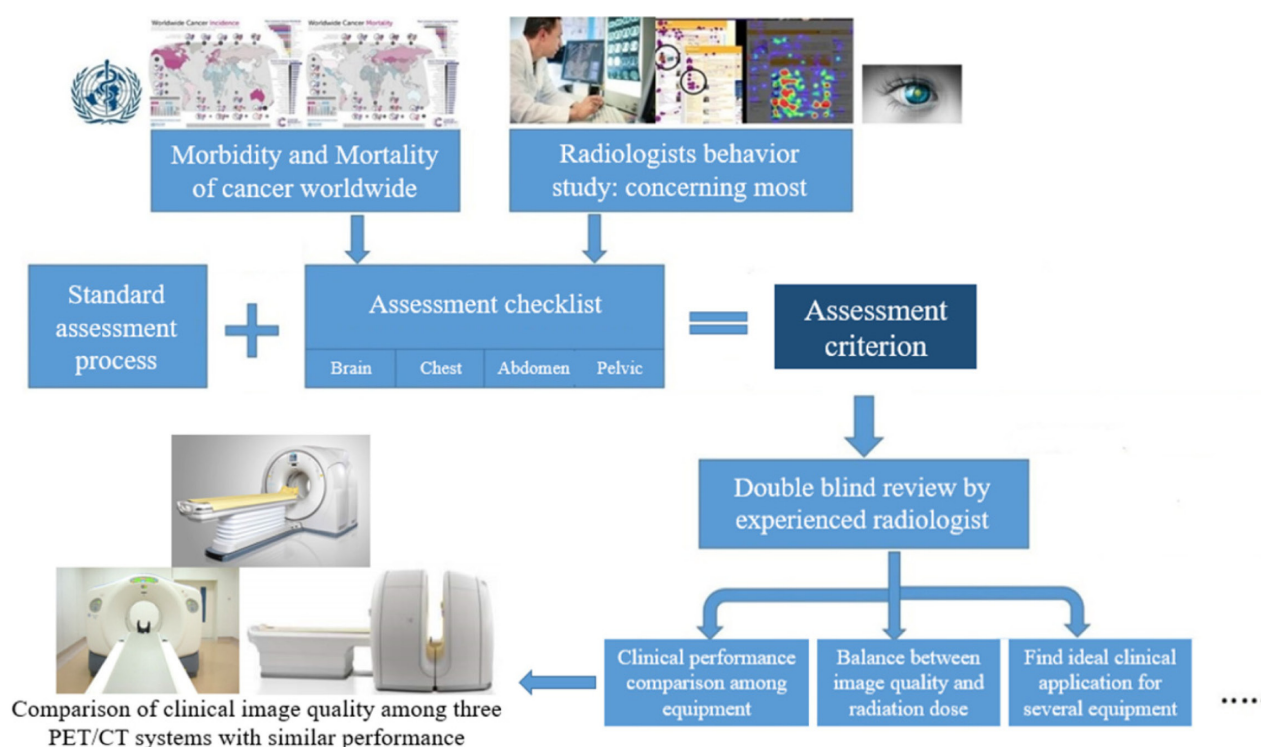


Figure 1. Flowchart of the method.

Methods

“Quality” contains what is truly necessary for assessing an image and what nuclear medicine physicians want to see most as well. In order to acquire an applicable image quality assessment method, we take epidemiological statistics and physicians’ behavior into consideration. Figure 1 shows the flowchart of our method. The assessment method is combined with a standard procedure guideline and checklist. The assessment checklist is divided into four parts: brain, chest, abdomen, and pelvis. We did experience nuclear medicine physicians’ behavior study at first to see what experts are concerning most. This includes too many details and some of them may be useless, then we use the data of cancer morbidity and mortality to select the key points which is the vital part physician are watching for clinical image reading.

After discussing with experienced physicians, we tried to describe the vital part with their words, which is our checklist. Finally, using this assessment method we invited qualified nuclear medicine physicians no matter how many years they have been working to do a double-blind review of clinical image quality among three PET/CT systems with similar performance to see whether the method is useful and to explore whether physical performance affects the image quality.

Setup for acquiring PET/CT image quality assessment checklist

Radiologist behavior observation study

We conducted a behavior observation study to determine what radiologists are looking at with a screen-based eye tracker (Tobii TX 300, Tobii Technology Co. Ltd.,

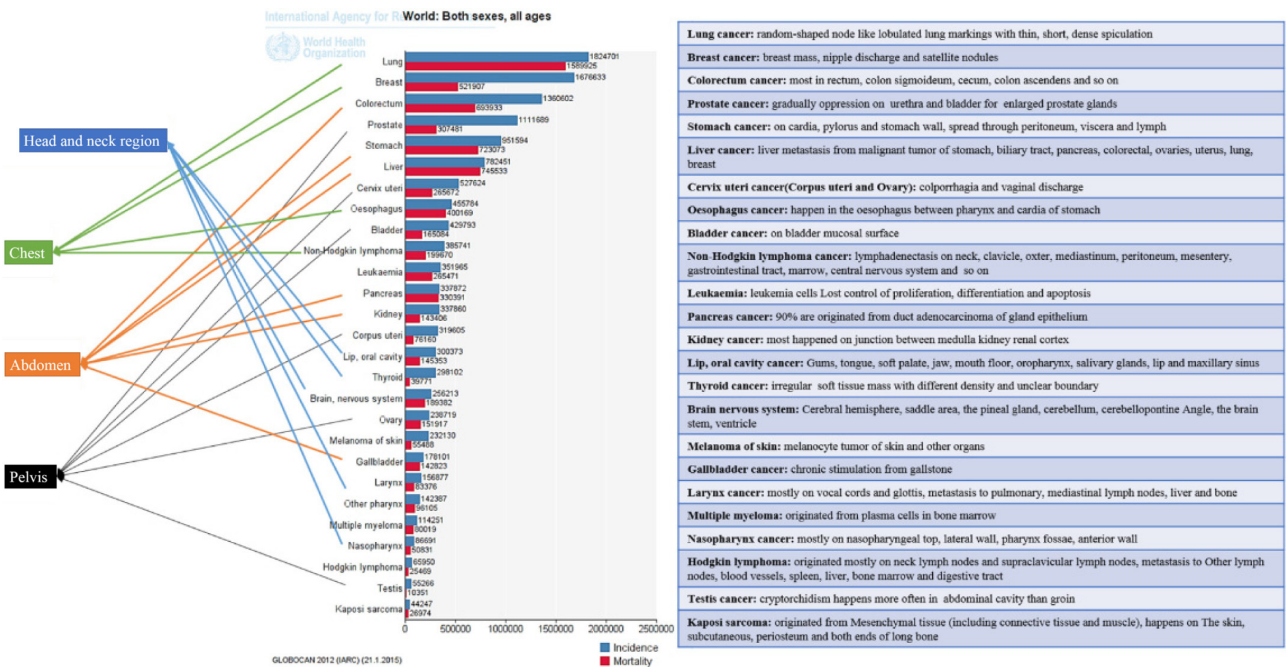


Figure 2. Division of the high incidence and mortality cancers into four parts to determine the necessary tissue for assessment.

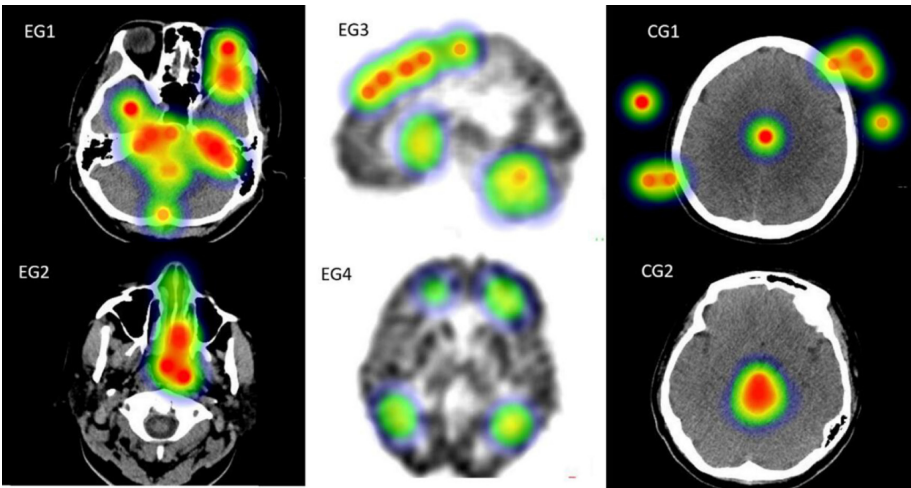


Figure 3. Typical heat maps of the eye tracker in the experimental and control groups.

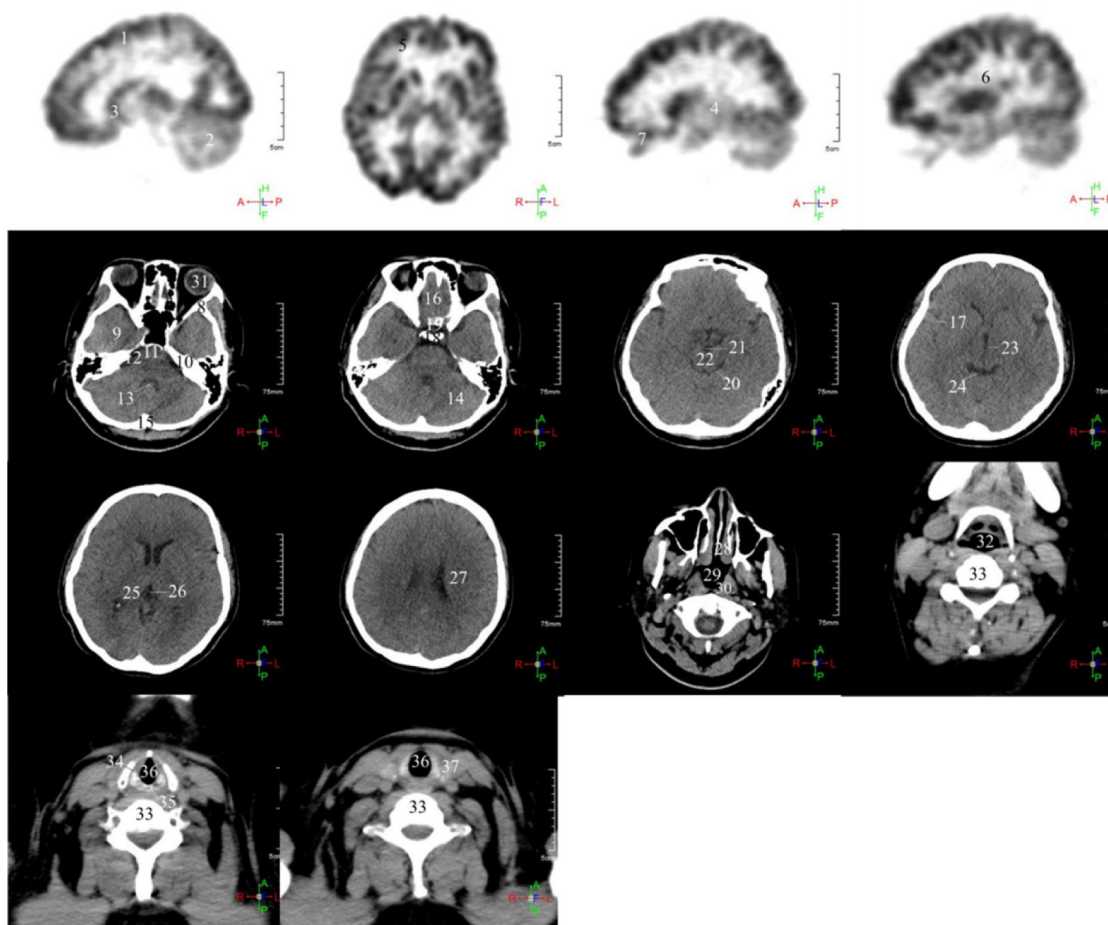


Figure 4. The key structures in the head and neck region. 1-cerebral sulci and gyri, 2-cerebellar sulci and gyri, 3-basal ganglia, 4-thalamus, 5-boundary between grey matter and white matter, 6-the blank area of paracele, 7-oculomotor muscle, 8-lesser wing of sphenoid bone, 9-temporal lobe, 10-petrous apex of temporal bone, 11-pons, 12- cerebellopontine angle cistern, 13-fourth ventricle, 14- cerebellar hemispheres, 15- internal occipital protuberance, 16- frontal lobe, 17-sella turcica, 18- dorsum sellae, 19- middle cerebral artery, 20- ambient cistern, 21-interpeduncular cistern, 22-mid-brain, 23-third ventricle, 24-quadrigeminal cistern, 25- thalamus, 26-pineal body, 27-lateral ventricle, 28- turbinate, 29-pharyngonasal cavity, 30-pharyngeal recess, 31-eyeball, 32-oropharyngeal cavity, 33-cervical vertebral body, 34-vocal cords, 35-arytenoid cartilage, 36- esophagus, 37-thyroid.

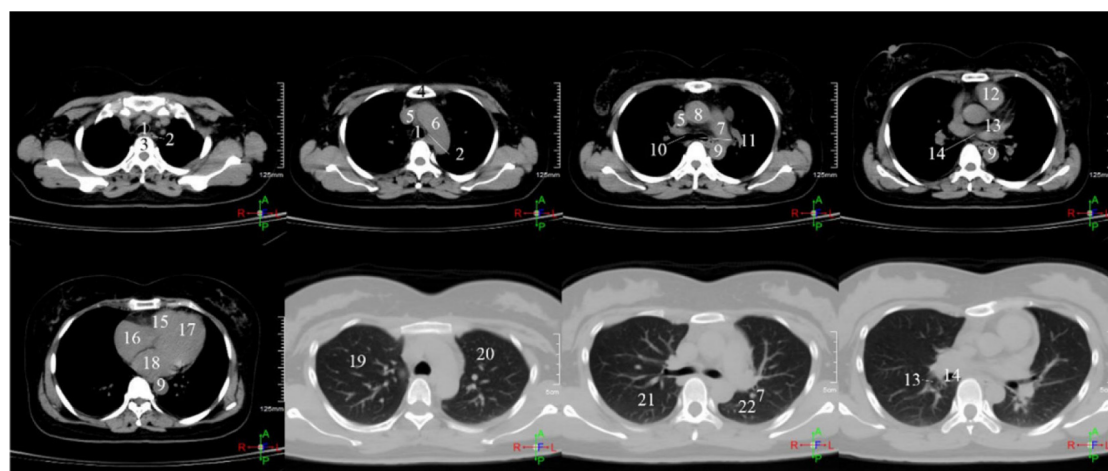


Figure 5. The key structures in the chest. 1-weasand, 2-esophagus, 3-thoracic vertebra, 4-sternum, 5-precava, 6-aortic arch, 7-left pulmonary artery, 8-ascending aorta, 9-descending aorta, 10-right main bronchus, 11-left main bronchus, 12-main pulmonary artery, 13-right pulmonary artery, 14-intermediate bronchus, 15-right ventricle, 16-right atrium, 17-left ventricle, 18-left atrium, 19-upper and middle lobe of right lung, 20-upper lobe of left lung, 21-lower lobe of right lung, 22-inferior lobe of left lung.

Stockholm, Sweden). The eye tracker records the exact location and time of eye movements. We selected 350 PET, 350 CT and 166 PET/CT fusion images that were displayed on a screen one by one. All images were scanned from a healthy people. We invited five experienced nuclear medicine physicians with more than 10 years working experience to take part in as the experimental group and five laypeople as the control group. All the people involved in the experiment were required to read all the image sets just as their usual work by themselves without any prior knowledge separately. As

shown in Figure 2, observing their behavior, we can tell that physicians with medical knowledge background are distinguished with normal people, physicians in the experimental group (EG1-4) have special concerns when reading cross-sectional clinical images, even for a long time, they still have strong preferences and focus on some specific regions. However, for the control group (CG1-2), it is easy to see that ordinary people do not have too many ideas about what to see and what should be seen and they are easy to lose their attention. The time they spend on reading a slice differs a lot, the longest

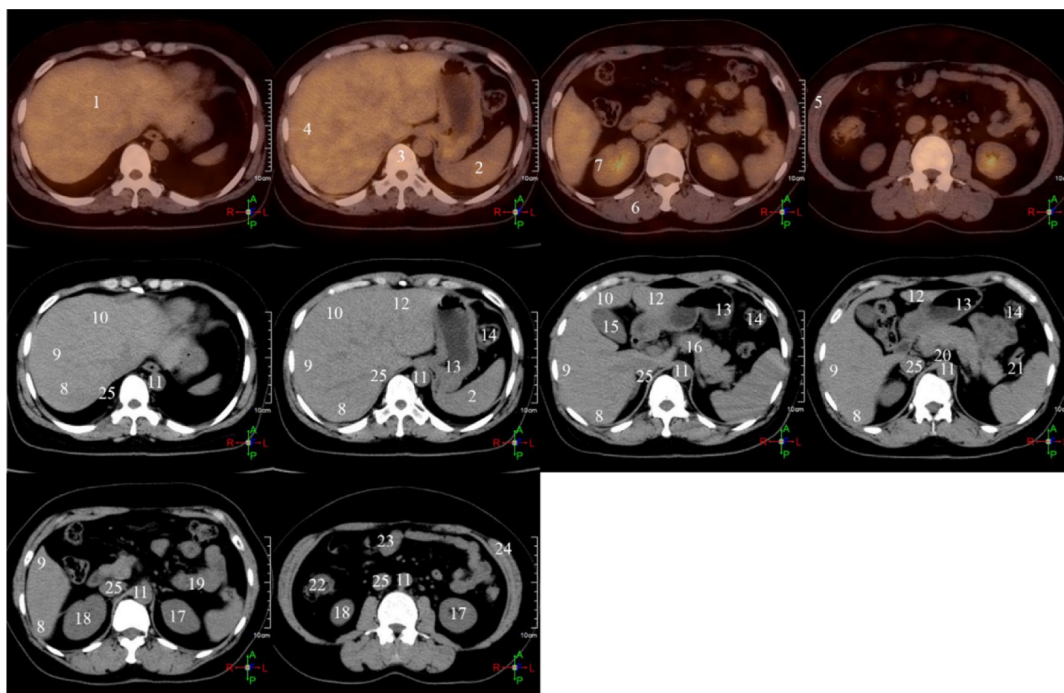


Figure 6. The key structures in the abdominal cavity. 1-liver, 2-spleen, 3-spine, 4-rib, 5-abdominal wall, 6-dorsal muscle, 7-renal cortex, 8-right posterior lobe of liver, 9-right anterior lobe of liver, 10-left internal lobe of liver, 11-abdominal aorta, 12-left external lobe of liver, 13-gastric wall, 14-splenic flexure of colon, 15-gall bladder, 16-pancreas, 17-left kidney, 18-right kidney, 19-small intestine, 20-celiac artery, 21-descending colon, 22-ascending colon, 23-transverse colon, 24-peritoneum, 25-inferior vena cava.

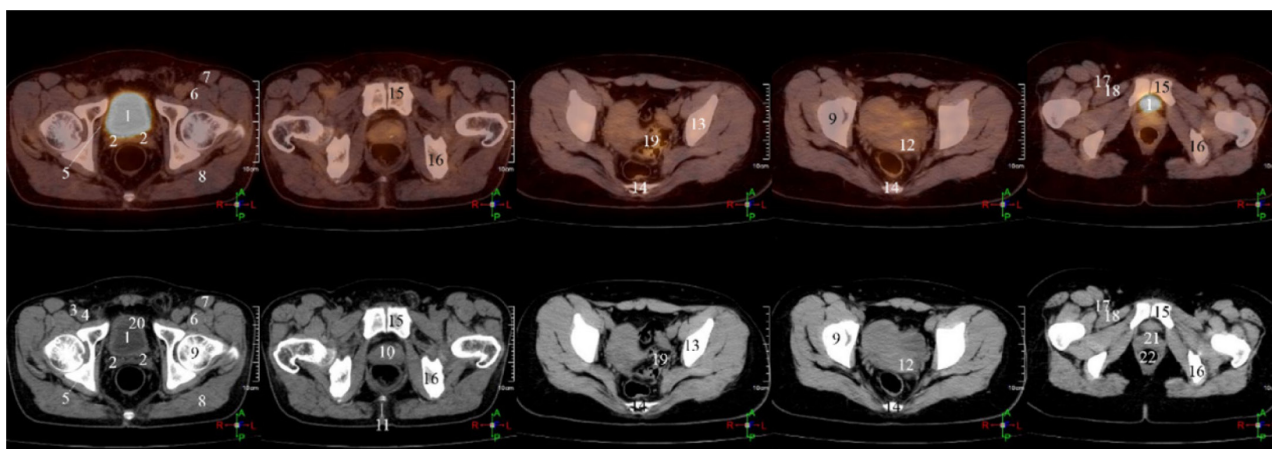


Figure 7. The key structures in the pelvic cavity. 1-bladder, 2-ureteral distal part, 3-femoral artery, 4-femoral vein, 5-musculus obturator internus, 6-iliopsoas, 7-sartorius, 8-gluteus maximus, 9-femoral head, 10-prostate gland, 11-coccyx, 12-uterus, 13-ilium, 14-sacrum, 15-phalanx, 16-ischium, 17-iliac artery, 18-rectus femoris, 19-small intestine, 20-bladder wall, 21-vagina, 22-anal canal.

and median time of EG group was 34.6 s and 8.4 s, while of CG group was 54.2 s and 1.3 s.

We divided the human body into four parts: head and neck region, chest, abdominal cavity and pelvis. There is an obvious anatomical boundary named diaphragm between the chest and abdominal cavity. The head and neck region mainly consists of bone tissues, while chest cavity is full of air, thus there is an automatic adjusted current value when scanning CT. The anatomic structure of abdominal and pelvic cavity is different with gender.

Then, we reviewed the heat map results and did an interview as soon as physicians finished their own eye-tracking test. Through communication, we distinguished the specific locations they were attending to recognize and why they paid so much attention to them. We summarized all the structures they mentioned, which are shown in Figures 3-6 which represent the parts most scrutinized or the structures must be clearly recognized in the head and neck region, chest, abdominal cavity and pelvic cavity.

Morbidity and mortality of cancer worldwide

There are 27 kinds of cancers with high incidence in both sexes and all ages all over the world. In order to divide these among the four human body sections, we made a deep analysis of every disease, such as the common locations of primary and metastatic lesions, their appearance on PET/CT images, and typical clinical manifestations.

As shown in Figure 7, lip, oral cavity, thyroid, central nervous system, laryngeal, pharyngeal, and nasopharyngeal cancers are common in the head and neck region. Lung, breast, esophagus, non-Hodgkin lympho-

ma and Hodgkin lymphoma cancers tend to occur in the chest region. Colorectum, stomach, liver, pancreas and gallbladder cancers involve the abdominal region. Prostate, cervix uteri, bladder, corpus uteri, ovary, testis and kidney cancers involve the pelvic region. Leukemia, melanoma, multiple myeloma and Kaposi sarcoma always appear anywhere, so they are particularly attributed to systemic cancers. Take three diseases as examples: Nasopharyngeal cancer mostly occurs on top of the nasopharynx, followed by the lateral wall and pharyngeal recess, with metastatic spread possibly discovered around the sternomastoid muscle. For some patients with advanced-stage disease, it may affect lymph nodes in the axilla, mediastinum, peritoneum and groin; the lesion shows as incrassation and soft tissue masses on CT. Lung cancer is distributed mainly over the anterior segment of both superior lobes, posterior segment of left superior lobe and right middle lobe. Colorectal cancer is quite common successively in the rectum, sigmoid flexure, cecum, ascending colon, descending colon and transverse colon. Sarcoma originates from mesenchymal tissue (connective tissue and muscle), which comprises the skin, subcutaneous tissue, periosteum, and both ends of the long bones.

Assessment method: checklist and procedure guideline

With the epidemiological and cognitive data, we generated an assessment checklist for the four body regions and PET, CT, and or PET/CT fusion images with physician's description which is shown in the result 3.1. In addition to the assessment checklist, how to execute the assessment process is also quite important. Randomization of evaluation images, double blinding, and the

Table 1. Physical performance of three PET/CT systems

	Experiment system A	Control system B	Control system C
Manufacturer	AMIC	GE	Philips
Product Model	RAY-SCAN 64	Discovery VCT	Gemini TF
CT Detector Number	64	64	64
CT Diameter(mm)	750	700	700
CT Pattern Noise	$\leq 0.35\%$	$\leq 0.35\%$	$\leq 0.35\%$
CT Uniformity	$\pm 4\text{HU}$	$\pm 4\text{HU}$	$\pm 4\text{HU}$
CT Accuracy	Air: $-1000\text{HU} \pm 10\text{HU}$; Water: $0 \pm 4\text{HU}$	Air: $-1000\text{HU} \pm 10\text{HU}$; Water: $0 \pm 4\text{HU}$	Air: $-1000\text{HU} \pm 10\text{HU}$; Water: $0 \pm 4\text{HU}$
CT Spatial resolution	14 Lp/cm	14 Lp/cm	14 Lp/cm
CT Low contrast resolution	When aperture diameter $x \leq 0.900$ mm, more than 50% can be visible	When aperture diameter $x \leq 0.900$ mm, more than 50% can be visible	When aperture diameter $x \leq 0.900$ mm, more than 50% can be visible
CT Artifact	No artifact in water phantom	No artifact in water phantom	No artifact in water phantom
PET Diameter (mm)	700	700	700
PET Axial FOV (mm)	163	157	180
PET Transverse spatial resolution (mm)	4.9 (at 1 cm) 5.3 (at 10 cm)	5.0 (at 1 cm) 5.6 (at 10 cm)	4.7 (at 1 cm) 5.1 (at 10 cm)
PET Axial spatial resolution (mm)	5.1 (at 1 cm) 5.8 (at 10 cm)	5.6 (at 1 cm) 5.9 (at 10 cm)	4.7 (at 1 cm) 5.2 (at 10 cm)
PET sensitivity (cps/kBq)	7.1	6.2	7.2

qualification of reviewers should be carried out strictly.

When using this assessment method, we should follow the procedure guideline as below. First, scan and collect all clinical PET/CT image data from the participants, which are then divided in clinical images into the four parts: brain, chest, abdomen and pelvic cavity. The images in every part were numbered randomly. Based on the assessment checklist for PET, CT and PET/CT fusion images, three physicians who were blind to any experimental design would be invited to mark each image individually with 0 (unacceptable), 1 (barely recognized), 2 (acceptable) and 3 (clearly identified). The graders can do everything like zooming in, rotating and clicking on ailments to get more information. One statistician finally collects all the rating sheets.

Validation through comparison among three PET/CT systems

To verify the effectiveness and feasibility of our assessment method, we compared clinical image quality among three commercial PET/CT systems with similar performance. The physical performance of the three PET/CT systems are listed in Table 1. The key parameters for the CTs were identical, while the PET scanners varied slightly as to field of view (FOV), resolution, and sensitivity.

This experiment was approved by the Ethics Committee of the General Hospital of the People's Liberation Army and all volunteers involved in this experiment were clearly informed about the experiment details. Half of them were scanned under the experimental system A at first, and then were examined under the control system B or C within 1 week because they needed post-operative evaluation or re-staging which rely on their clinical pathway. The others were scanned in B or C first, followed by A. Finally, we selected those who had completed two scans and keep the patients' age, weight and height statistically consistent. 18F-FDG (46.25 MBq/kg) was used, and the dose of radiation was kept the same in the two groups.

After double-blind review, the scores of each system's performances of PET, CT and PET/CT on the four parts (brain, chest, abdomen and pelvic cavity) were concluded. The score difference between experimental and

control systems was calculated by subtraction of mean values. The tolerance was set at 2% of the maximum value of the control systems. When the tolerance was greater than the score difference, it should be added to positive variation, which means the score of the experimental system was higher than 98% of control system's scores. Wilcoxon's signed rank test was performed to determine whether there is significant difference in the paired-sample test. The test revealed significant difference, revealing the confidence of comparison.

Validation materials

To make comparisons among three PET/CT systems with similar physical performance, as shown in Table 2, 37 patients (male: 21, female: 16, 56.5 ± 13.4 years old) were scanned with the experimental system A and control system B because their clinical pathway required frequent PET/CT examinations. The other 18 patients (male: 9, female: 9, 54.1 ± 13.6 years old) were scanned under the experimental system A and control system C. The patients' characteristics, such as age, weight and height, were confirmed to be in accordance with normal distribution. The experiment was approved by the Ethics Committee and all participants involved signed informed consent. P values of Kolmogorov-Smirnov test were 0.286, 0.462 and 0.461. There was no significant difference in gender, age, weight and height distribution. The Pearson χ^2 test of gender equals to 0.223 and p value was 0.637, the t-value of age equals to -1.012 and p value was 0.062, the t-value of weight equals to 1.910 and p value was 0.316, the t-value of height equals to -0.334 and p value was 0.740.

Results

Assessment checklist

As shown in Table 3, the assessment checklist was generated from Figure 2 and Figures 3-6. It included not only the most important structures needed to be seen but also the distinction among tissues with similar density. For PET/CT fusion images, images with good quality needed to be well

Table 2. Patient characteristics

Characteristics	A&B	A&C
Gender (male/female)	37 (21/16)	18 (9/9)
Age, years		
Mean \pm SD	56.5 \pm 13.4	49.2 \pm 13.1
Median (Min, Max)	58.4 (23.6,75.5)	52.4 (21.6,69.2)
Weight		
Mean \pm SD	66.1 \pm 10.0	69.5 \pm 14.3
Median (Min, Max)	65.0 (46.0,90.0)	65.0 (50.0,100.0)
Height		
Mean \pm SD	165.9 \pm 8.2	166.7 \pm 6.8
Median (Min, Max)	165.0 (150.0,180.0)	166.5 (152.0, 180.0)

Table 3. PET/CT image quality criteria checklist

		<i>Assessment checklist</i>	
Head and neck region	PET	1.	groove back of cerebrum and cerebellum
		2.	basal ganglia and thalamus
		3.	boundary between grey matter and white matter
		4.	the blank area of paracele
		5.	oculomotor muscle
		6.	pharyngeal
	CT	1.	subarachnoid space of sulcus and split brain
		2.	ventricle and cisterna
		3.	hypothalamus and brainstem
		4.	grey matter and white matter
		5.	tiny structure of bone
		6.	inner plate and outer plate of skull
		7.	soft tissues like eye, nose and scalp
		8.	maxillofacial muscles
		9.	lymphonodus and thyroid
		10.	vocal cords
		11.	laryngeal cartilage
		12.	cervical vertebra
	PET/CT	1.	distinguishable between PET and CT information
		2.	well-adjusted grey matter, white matter and ventricle
		3.	well-adjusted scalp outline
Chest		4.	well-adjusted anatomy and metabolic structure
	PET	1.	lung, mediastinum, soft tissue of chest wall
		2.	vascular bundle and lymphonodus of hilus pulmonis
		3.	bilateral nipples
		4.	mammary gland and ventricular muscle
	CT	1.	lung marking, trachea and primary bronchus
		2.	mediastinum, heart and aorta
		3.	esophagus, breast tissue
		4.	distinguishable among skin, fat and muscle
		5.	spine, rib and so on
		6.	tiny structure near centrum
		7.	cortex of bone, medullary cavity and soft tissue
	PET/CT	1.	well-adjusted anatomy and metabolic structure
		2.	distinguishable between PET and CT information
Abdomen	PET	1.	distinguishable among liver, spleen, bone and soft tissue
		2.	uniformity in liver
		3.	renal cortex
		4.	distinguishable between abdominal wall and dorsal muscle
	CT	1.	liver, biliary tract, spleen, muscle
		2.	retroperitoneal lesion and aorta
		3.	intestinal canal and mesentery
		4.	distinguishable between liver and intrahepatic bile duct
		5.	porta hepatis, pancreas, adrenal gland
		6.	each part in the stomach and intestine
		7.	renal cortex, collecting system and so on
	PET/CT	1.	well-adjusted anatomy and metabolic structure
		2.	distinguishable between PET and CT information
Pelvic	PET	1.	distinguishable between bladder and other soft tissue
		2.	bone and arthrosis
		3.	endometrium and adnexa uteri
		4.	muscle of buttock and leg
	CT	1.	adnexa uteri/ prostate and seminal vesicle
		2.	basin wall vessels
		3.	bladder wall
		4.	cortex of bone and articular cavity
		5.	inguinal canal
		6.	sacral canal and nerve
	PET/CT	1.	well-adjusted anatomy and metabolic structure
		2.	distinguishable between PET and CT information

Table 4. Comparisons (mean±SD) between experimental (A) and control systems (B, C) in the head and neck region, chest, abdomen and pelvis

		Grade difference		Tolerance value	Positive/negative variation		Z value		p value	
		A&B	A&C	A&B&C	A&B	A&C	A&B	A&C	A&B	A&C
Head and neck region	PET grade	0.659±0.532	0.181±0.452	0.360±0.001	14/8	15/3	-2.458	-2.188	0.014	0.029
	CT grade	0.000±0.000	-0.042±0.177	0.720±0.000	22/0	18/0	-4.690	-4.146	<0.001	<0.001
	PET/CT grade	0.023±0.074	0.056±0.202	0.240±0.000	20/2	15/3	-4.415	-2.974	<0.001	0.003
	Total grade	0.682±0.547	0.194±0.539	1.320±0.001	19/3	17/1	-3.596	-3.690	<0.001	<0.001
Chest	PET grade	0.09±0.39	0.13±0.39	0.24±0.00	23/4	12/6	-3.389	-1.432	0.001	0.152
	CT grade	-0.01±0.05	-0.00±0.00	0.42±0.00	27/0	18/0	-5.112	-4.243	<0.001	<0.001
	PET/CT grade	0.00±0.00	0.00±0.00	0.12±0.00	27/0	18/0	-5.196	-4.243	<0.001	<0.001
	Total grade	0.08±0.40	0.13±0.39	0.78±0.00	25/2	17/1	-4.649	-3.635	<0.001	<0.001
Abdomen	PET grade	0.04±0.20	-0.01±0.37	0.24±0.00	22/3	15/3	-3.285	-2.483	0.001	0.013
	CT grade	0.00±0.00	0.00±0.00	0.42±0.00	25/0	18/0	-5.000	-4.243	<0.001	<0.001
	PET/CT grade	0.00±0.00	0.00±0.00	0.12±0.00	25/0	18/0	-5.000	-4.243	<0.001	<0.001
	Total grade	0.04±0.20	-0.01±0.37	0.78±0.00	25/0	18/0	-4.662	-3.825	<0.001	<0.001
Pelvis	PET grade	0.14±1.13	0.08±0.81	0.19±0.02	19/3	17/1	-2.904	-3.113	0.004	0.002
	CT grade	0.00±0.00	0.00±0.00	0.36±0.00	22/0	18/0	-4.690	-4.243	<0.001	<0.001
	PET/CT grade	0.00±0.00	0.00±0.00	0.12±0.00	22/0	18/0	-4.690	-4.243	<0.001	<0.001
	Total grade	0.14±1.13	0.08±0.81	0.67±0.02	20/2	17/1	-2.938	-3.113	0.003	0.002

adjusted and distinguishable between PET and CT information.

Statistical result of comparison

The PET grade, CT grade, PET/CT grade and total grade of head, chest, abdomen and pelvis scanned by A and B and C are shown in Table 4. The grade difference means the difference in values between the experimental system and control system. A and B grade difference is the score of A subtract the score of B. The tolerance value was set to 2% of the maximum value of control systems B and C. Positive variation occurred when the tolerance value was higher than the grade difference, which means that the experimental system was better by 98% of the control system's score. The negative variation occurred when the tolerance value was lower than the grade difference, which means that the experimental system was inferior to 98% of the control system's score. Wilcoxon signed rank test was done on paired samples and the result is shown as Z values. Table 4 objectively reflects each system's superiority on certain modalities and certain parts of the body. The performance of the three systems was determined to be physically similar. However, system A was superior to B for the PET image of the whole body and the same was found for C, except for the PET image of the abdomen, where system C performed better than A. No significance difference of image quality on total grade was found, except for the pelvis where A was better

than B and C. For the PET/CT image, no significant difference was observed, while system A was better than C in the head and neck region.

Discussion

Focusing on lesions with high incidence and suspicious regions where the physicians have the most concern, our method fully demonstrates the value of the PET/CT image and stands for the clinicians' opinions. With the aid of double-blinding, random sampling and statistical analysis, the efficiency and applicability of image quality assessment criteria have been demonstrated. The results certify that there are significant differences among systems even with similar physical performance. It is necessary to assess a system considering a radiologists' opinion, as well as the clinical effect. In sum, our objective method for clinical PET/CT image quality assessment and comparison is efficient.

Acknowledgements

This work was supported by the National Key Instrumentation Development Project Foundation of China (2011YQ030114) and the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (81421004).

Conflict of interests

The authors declare no conflict of interests.

References

1. Townsend DW, Carney JP, Yap JT, Hall NC. PET/CT today and tomorrow. *J Nucl Med* 2004;45 (Suppl 1):4S-14S.
2. Martin CJ, Sutton DG, Sharp PF. Balancing patient dose and image quality. *Appl Radiat Isot* 1999;50:1-19.
3. Rausch I, Cal-Gonzalez J, Dapra D et al. Performance evaluation of the Biograph mCT Flow PET/CT system according to the NEMA NU2-2012 standard. *EJNMMI Phys* 2015;2:26.
4. Boellaard R. The engagement of FDG PET/CT image quality and harmonized quantification: from competitive to complementary. *Eur J Nucl Med Mol Imaging* 2016;43:1-4.
5. Mansor S, Pfaehler E, Heijtel D, Lodge MA, Boellaard R, Yaqub M. Impact of PET/CT system, reconstruction protocol, data analysis method, and repositioning on PET/CT precision: An experimental evaluation using an oncology and brain phantom. *Med Phys* 2017;44:6413-24.
6. Messerli M, Stolzmann P, Egger-Sigg M et al. Impact of a Bayesian penalized likelihood reconstruction algorithm on image quality in novel digital PET/CT: clinical implications for the assessment of lung tumors. *EJNMMI Phys* 2018;5:27.
7. Orlandini LC, Betti M, Fulcheri C, Dona M, Fisicaro D, Castagnoli A. Improvement in clinical evaluation of PET/CT images with high resolution algorithms. *Q J Nucl Med Mol Imaging* 2013;57:201-6.
8. Sah BR, Stolzmann P, Delso G et al. Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in 18F-FDG PET/CT studies. *Nucl Med Commun* 2017;38:57-66.
9. Siebert E, Bohner G, Dewey M et al. 320-slice CT neuroimaging: initial clinical experience and image quality evaluation. *Br J Radiol* 2009;82:561-70.
10. Reynes-Llompart G, Gamez-Cenzano C, Romero-Zayas I, Rodriguez-Bel L, Vercher-Conejero JL, Marti-Climent JM. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. *J Nucl Med* 2017;58:1155-61.
11. Vallot D, Caselles O, Chaltiel L et al. A clinical evaluation of the impact of the Bayesian penalized likelihood reconstruction algorithm on PET FDG metrics. *Nucl Med Commun* 2017;38:979-84.
12. Partovi S, Kohan A, Gaeta C et al. Image quality assessment of automatic three-segment MR attenuation correction vs. CT attenuation correction. *Am J Nucl Med Mol Imaging* 2013;3:291-9.