Review article

Nuclear Medicine Communications

Potential value of dual-time-point ^{18}F-FDG PET compared with initial single-time-point imaging in differentiating malignant from benign pulmonary nodules: a systematic review and meta-analysis

Yu-Yi Lin[a], Jin-Hua Chen[c], Hueisch-Jy Ding[e], Ji-An Liang[b,d], Jun-Jun Yeh[d,e,g,*] and Chia-Hung Kao[a,b,*]

We performed a meta-analysis to assess the potential value of dual-time-point (DTP) imaging as compared with initial single-time-point (STP) scanning with ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET in differentiating malignant from benign single pulmonary nodules. Data on the performance of DTP ^{18}F-FDG PET imaging in assessing lung nodules were extracted from articles of prospective or retrospective original research published between January 2001 and April 2010. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was used to assess the quality of study methodology. Heterogeneity in the results of the studies was assessed, and summary receiver operating characteristic (SROC) curves were constructed. Eleven studies comprising a total of 788 patients who underwent initial scanning, 778 of whom also underwent DTP imaging, were included in the final analysis. The quality of study methodology was judged to be moderate. Substantial heterogeneity in the results of the studies, with inconsistency (I^2) index values above 85%, reflected important differences in study methods and populations, including varying lesion sizes, ^{18}F-FDG avidity, uptake interval for delayed imaging, and threshold for positive result on DTP imaging. SROC curve analysis revealed a statistically nonsignificant trend toward higher sensitivity with DTP imaging, at moderate levels of specificity, when compared with initial STP scanning. The area under the curve (SE) values for DTP and initial STP imaging were 0.839 (0.079) and 0.757 (0.074), respectively. Although the results of our analysis do not support the routine use of DTP imaging with ^{18}F-FDG PET in the differential diagnosis of pulmonary nodules, this technique may provide additional information in selected cases with equivocal results from initial scanning. Further prospective research is required to better define the potential benefits of DTP ^{18}F-FDG PET imaging. *Nucl Med Commun* 00:000–000

Nuclear Medicine Communications 2012, **00**:000–000

Keywords: dual-time-point imaging, ^{18}F-fluorodeoxyglucose, meta-analysis, positron emission tomography, pulmonary nodules

[a]Department of Nuclear Medicine and PET Center, [b]Department of Radiation Oncology, [c]Biostatistics Center and Graduate Institute of Biostatistics, [d]Graduate Institute of Clinical Medicine Science and School of Medicine, College of Medicine, China Medical University, Taichung, [e]Department of Medical Imaging and Radiological Sciences, I-Shou University, Kaohsiung, [f]Chia-Yi Christian Hospital, Chia-Yi and [g]Chia Nan University of Pharmacy and Science and Tainan, Taiwan

Correspondence to Chia-Hung Kao, MD, Department of Nuclear Medicine and PET Center, China Medical University Hospital, No. 2, Yuh-Der Road, Taichung 404, Taiwan
Tel: +886 4 2205 2121 x7412; fax: +886 4 2233 6174;
e-mail: d10040@mail.cmuh.org.tw

*Jun-Jun Yeh and Chia-Hung Kao contributed equally to the writing of this article.

Received 30 April 2012 **Revised** 13 June 2012 **Accepted** 14 June 2012

Introduction

Solitary pulmonary nodules (SPNs) are common incidental findings on chest radiography or computed tomography (CT). These nodules can be benign or malignant, and differentiating between them has important treatment and prognostic implications. PET with ^{18}F-fluorodeoxyglucose (^{18}F-FDG) is widely used in the diagnosis of indeterminate SPNs on CT imaging. Traditionally, a standard uptake value (SUV) of 2.5 has been used as a threshold to help differentiate malignant from benign nodules [1,2]. Nevertheless, using ^{18}F-FDG PET to detect malignant pulmonary nodules has limitations. On the one hand, although many malignancies tend to be ^{18}F-FDG avid, exceptions include false-negative results for certain slow-growing tumors [3]. On the other hand, nonspecific ^{18}F-FDG uptake in benign processes, such as inflammation, may mimic that of malignant lesions on initial scanning, leading to false-positive results.

Various techniques in applying ^{18}F-FDG PET, including dual-time-point (DTP) or delayed imaging, have been reported to potentially improve diagnostic accuracy. Prior research showed continuous ^{18}F-FDG accumulation in malignant lesions, in contrast to stable or decreasing ^{18}F-FDG activity over time in benign processes [4]. Such differences in the rate of uptake on subsequent imaging might provide additional diagnostic value. A number of studies assessing the accuracy of DTP ^{18}F-FDG PET in differentiating malignant from benign pulmonary lesions have been published. These studies have reported a wide range of results, some of which appear conflicting. Furthermore, conclusions regarding accuracy were limited

0143-3636 © 2012 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/MNM.0b013e32835710d6

because of small study populations and the mostly retrospective and highly selective nature of patient inclusion.

Objective

We performed a meta-analysis to assess the potential value of DTP imaging as compared with initial single-time-point (STP) scanning with ^{18}F-FDG PET in differentiating malignant from benign single pulmonary nodules.

Materials and methods

Literature search and selection criteria

A search of the Medline database for articles published in English, between January 2001 and April 2010, was performed using keywords and text words to identify studies evaluating the performance of DTP ^{18}F-FDG PET in differentiating lung nodules. The following search algorithm was used: 'lung nodule' or 'pulmonary nodule'; 'PET' or 'PET-CT' or 'PET/CT' or 'positron emission tomography' or 'fluorodeoxyglucose' or 'FDG'; 'dual-time-point' or 'dual phase' or 'double phase' or 'delayed phase'. To be included, an article had to be a contribution of prospective or retrospective original research in differentiating malignant from benign lung lesions using DTP ^{18}F-FDG PET performed with a dedicated PET or integrated PET/CT scanner and confirmed by pathological examination and/or clinical follow-up. Studies using modified gamma cameras for coincidence detection were excluded because of more limited resolution and the consequential lower sensitivity when compared with dedicated PET scanners [5,6]. Studies were also excluded if patient populations consisted of fewer than 10 patients, or data were unavailable for deriving 2×2 tables. The reference lists of selected studies were manually searched for additional relevant articles. Overlapping patient populations were avoided by including only the latest or the largest study.

Data extraction and quality assessment

Data regarding author, year of publication, study design, patient characteristics, DTP ^{18}F-FDG PET imaging, reference test, and diagnostic performance were extracted. In some articles, data were acquired using stand-alone PET scanners, whereas in others data were obtained using integrated PET/CT scanners. The CT component of the PET/CT examination in the latter studies was not used in this analysis. The description of the index test was considered acceptable when the radiopharmaceutical, administered dose, fasting before tracer injection, and the initial and delayed uptake intervals were all documented. The technical specification and quality of PET or PET/CT procedures were evaluated using recommended guidelines [7,8]. The quality of study methodology was assessed using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool [9].

Statistical analysis

To facilitate comparison of accuracy estimates at the study level, efforts were made to minimize the heterogeneity in the study populations and in the diagnostic criteria used to define positive results on DTP ^{18}F-FDG PET imaging. Whenever possible, patients with multiple pulmonary nodules were excluded from the final analysis, because these patients might have complex underlying conditions with different clinical implications, as opposed to those with SPNs. To reduce the heterogeneity due to a threshold effect, we used the same diagnostic criteria to recalculate the accuracy measures for each study with individual patient data. The constant threshold criteria used to compare the accuracy of DTP and initial STP imaging were an increase in maximal SUV of 10% or more and an initial maximal SUV of 2.5 or more, respectively. These threshold levels were chosen on the basis of prior research [1,2,4,10]. When multiple 2×2 tables corresponding to different diagnostic criteria for either DTP or initial imaging were listed in lieu of individual patient data, only the set of aggregated data corresponding to the threshold level closest to the appropriate cutoff value mentioned above was retained in the final analysis.

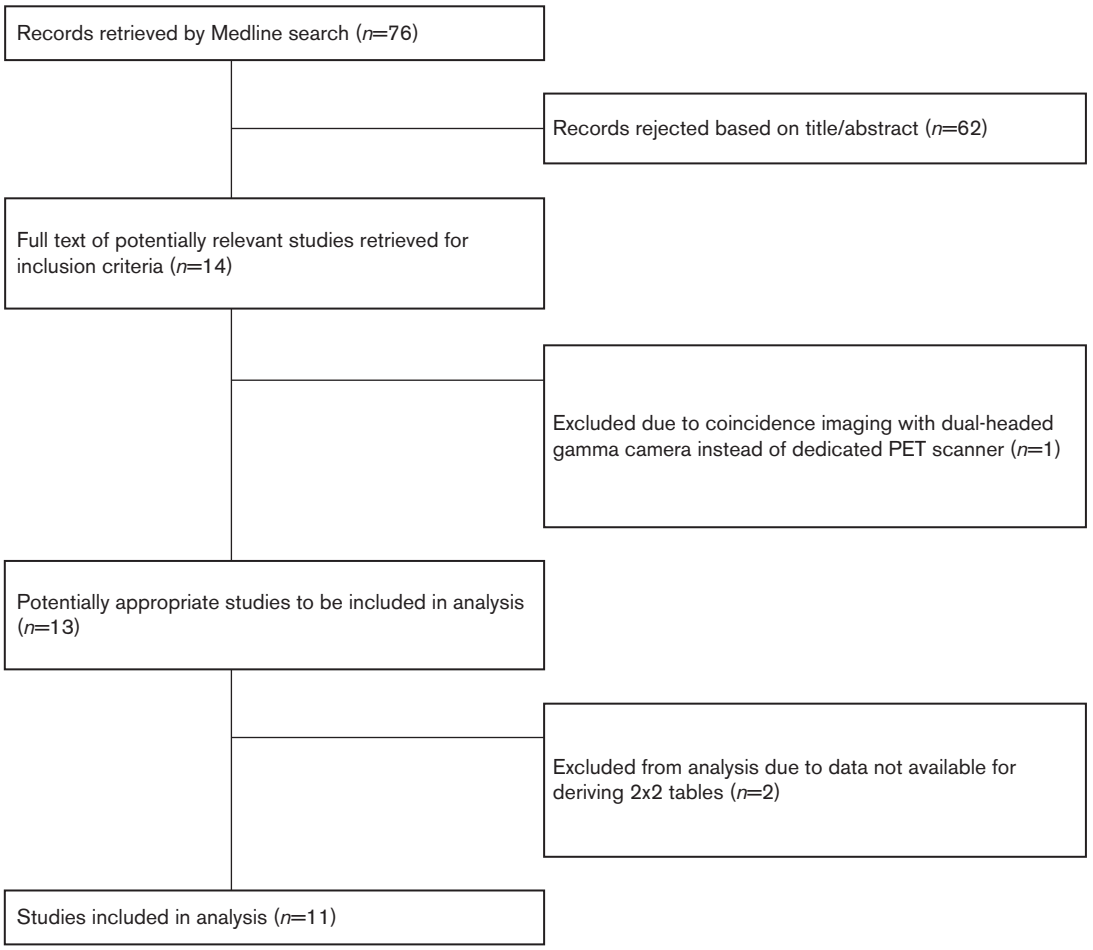
The heterogeneity in the results of the studies was assessed graphically with forest plots and statistically using the inconsistency (I^2) index, which describes the percentage of total variation across studies that is attributable to heterogeneity rather than to sampling error [11–16]. Summary receiver operating characteristic (SROC) curves were constructed to depict the relationship between sensitivity and specificity of the diagnostic tests across studies. The area under the curve (AUC) is a summary measure of the overall diagnostic accuracy. Whereas the full AUC and its SE were used to compare the overall accuracy of DTP with initial scanning, only the part of the curve corresponding to clinically relevant values of sensitivity and specificity was shown in the receiver operating characteristic space [17,18]. Statistical significance was defined at the conventional level of 0.05 in a two-tailed test.

Results

The search strategy retrieved 76 articles, 62 of which were rejected on the basis of title or abstract. Another was excluded because of the use of coincidence imaging with a dual-headed gamma camera [19]. After review of the full text, two more articles were excluded because of unavailability of data for deriving the 2×2 tables [4,20]. The results of the literature search and study selection are summarized in Fig. 1.

Eleven studies comprising a total of 788 patients who underwent initial scanning, 778 of whom also underwent DTP imaging, were included in the final analysis [10,21–30]. The characteristics of these studies are summarized in Table 1. Only two of these studies were

Fig. 1



Results of the literature search and study selection.

Table 1 Characteristics of the included studies on DTP ^{18}F-FDG PET imaging in evaluating pulmonary lesions

References	Patients	^{18}F-FDG avidity criteria for inclusion	Uptake interval (min) after injection, mean (range)		Criteria for malignancy on DTP imaging	Reference standard
Matthies *et al.* [10]	36	–	69 (55–110)	122 (100–163)	Increase SUV_{mean} > 10%	HP; CFU 18–26 months
Demura *et al.* [21][a]	80	–	60	180	Increase SUV_{mean} > 0%	HP; CFU
Xiu *et al.* [22]	46	Initial SUV < 2.5	58 (50–110)	114 (99–163)	Increase SUV_{max} > 10%	HP; CFU > 24 months
Alkhawaldeh *et al.* [23][b]	265	–	60	100 (90–110)	Increase SUV ≥ 0%	HP; CFU
Lan *et al.* [24]	45	–	45–55	160 (150–180)	Increase SUV_{max} ≥ 10%	HP; CFU 6 months
Chen *et al.* [28]	27	Initial SUV_{mean} < 2.5	60	120	Increase SUV_{max} > 10%	HP only
Kim *et al.* [25]	30	SUV < 2.5	60	120	Change in SUV_{max} > – 2.3%	HP; CFU > 12 months
Laffon *et al.* [29]	38	Initial SUV_{max} > 2.5	78 (54–125)	159 (116–214)	Direction of change in SUV	HP; CFU
Suga *et al.* [26]	137	^{18}F-FDG-avid[c]	60 (56–65)	120 (110–131)	Increase SUV_{max} > 20%	HP; CFU 5–20 months
Cloran *et al.* [30]	113	Initial SUV_{max} < 2.5	64	121	Increase SUV_{max} > 10%	HP; CFU ≥ 24 months
Schillaci *et al.* [27][a,b]	30	–	50	90	Increase SUV_{max} ≥ 10%	HP; CFU > 10 months

CFU, clinical follow-up; DTP, dual-time-point; ^{18}F-FDG, ^{18}F-fluorodeoxyglucose; HP, histopathology; SUV, standard uptake value.
[a]Prospective studies.
[b]Alkhawaldeh *et al.* [23] and Schillaci *et al.* [27] included only patients with single pulmonary nodules.
[c]Suga *et al.* [26] included only patients with solitary ^{18}F-FDG-avid lesions (those with persistently higher uptake than that of the contralateral normal lung).

prospective in design; most were retrospective with small patient populations. Four of the 11 studies included only patients with pulmonary lesions of low ^{18}F-FDG avidity with initial SUV below 2.5. One study limited patient inclusion to only those with lung lesions with an initial SUV of 2.5 or more; another included only patients with ^{18}F-FDG-avid lung lesions – that is, those with persistently higher uptake than that of the contralateral normal

lung on both initial and delayed scans. Although delayed scanning occurred as early as 1.5 h after tracer injection in one study, and as late as 3 h after injection in another, it began at about 2 h after injection in six of the 11 studies. A 10% increase in mean or maximum SUV was chosen to be the threshold for positive results on DTP imaging in six of the 11 studies, whereas the others used values ranging from a decrease of 2.3% to an increase of 20% in SUV. Six of the 11 studies listed individual patient data, from which we recalculated accuracy measures using the constant threshold criteria mentioned above. With the exception of one retrospective study that included only patients with a final diagnosis based solely on the pathological results of surgical samples, most of the studies used a combination of histopathological examination and imaging or clinical follow-up of varying duration as the reference standard.

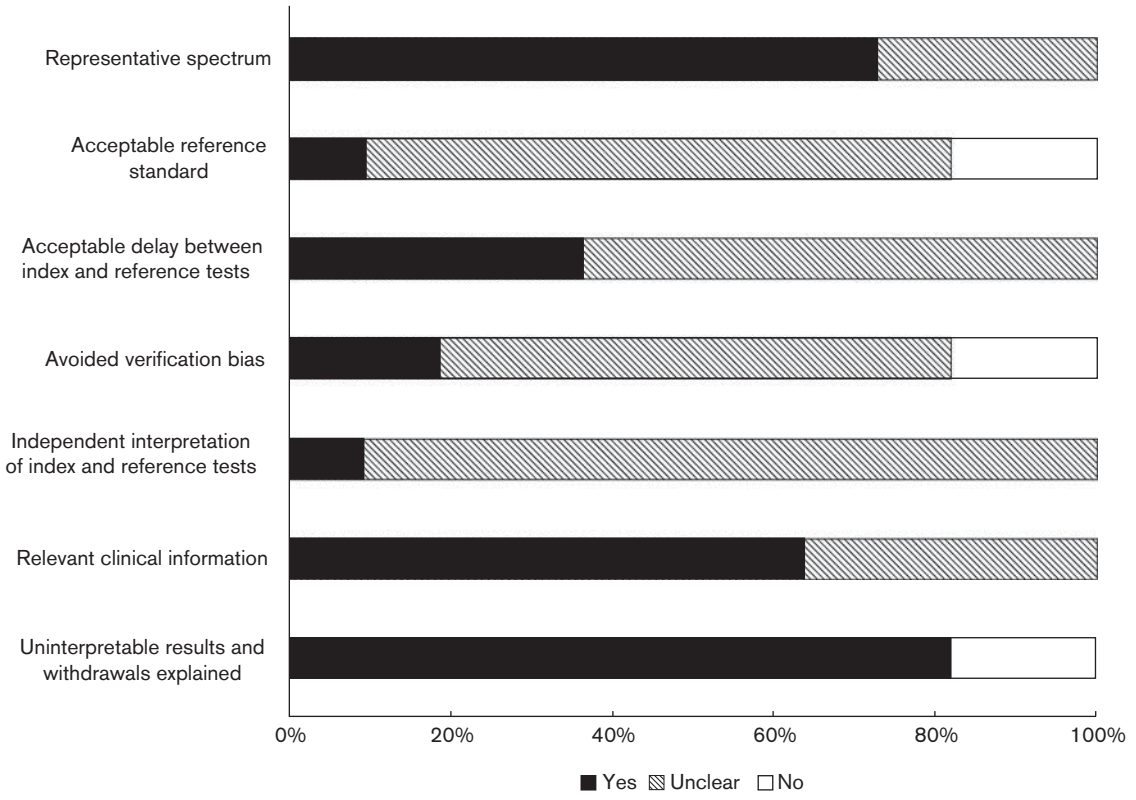
Overall, the quality of study methodology was judged to be moderate (Fig. 2). Inadequate reporting was found in multiple areas of the quality assessment. Despite the steps taken to reduce heterogeneity in the study populations and the diagnostic criteria on DTP imaging, there was substantial heterogeneity in the results of the studies, with inconsistency index values of 85 and 86% for sensitivity and specificity, respectively. The wide range of estimated accuracy with DTP imaging across studies (sensitivity 54–100% and specificity 14–93%) has been illustrated in forest plots (Fig. 3).

The SROC curve analysis revealed a statistically non-significant trend toward better overall diagnostic accuracy with DTP compared with initial STP imaging (Fig. 4). The AUC (SE) values for DTP and initial STP imaging were 0.839 (0.079) and 0.757 (0.074), respectively. Sensitivity was higher with DTP imaging at moderate levels of specificity. This potential advantage of DTP over initial STP scanning was diminished at higher levels of specificity.

Discussion

In a recently published meta-analysis on DTP ^{18}F-FDG PET in the diagnosis of lung nodules, Barger *et al.* [31] concluded that the additive value of DTP imaging was questionable. Our analysis was conducted without the knowledge of Barger's article and with a more focused aim of assessing the potential value of DTP versus initial STP ^{18}F-FDG imaging with dedicated PET scanners in evaluating patients with SPNs. Consequently, in contrast to Barger and colleagues, we excluded a study because of the use of modified gamma cameras for coincidence detection. Nevertheless, our final analysis included more

Fig. 2



Quality assessment of study methodology.

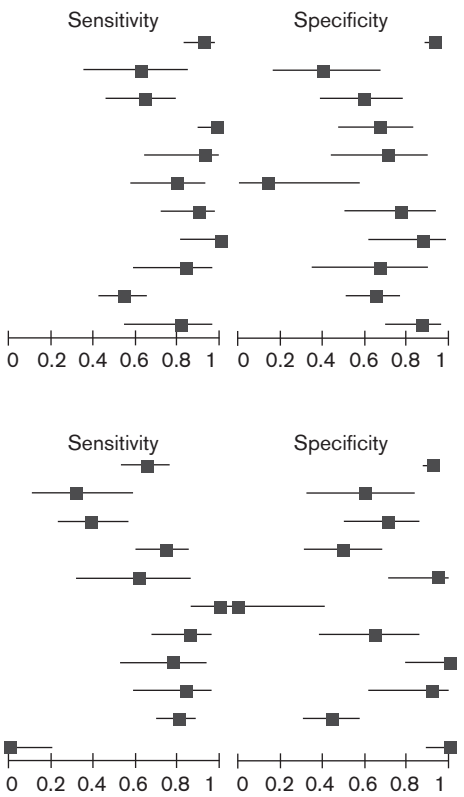
Fig. 3

DTP

References	TP	FP	FN	TN	Sensitivity	Specificity
Alkhawaldeh *et al.* [23]	57	14	5	179	0.92 [0.82, 0.97]	0.93 [0.88, 0.96]
Chen *et al.* [28]	10	9	6	6	0.63 [0.35, 0.85]	0.40 [0.16, 0.68]
Cloran *et al.* [30]	23	11	13	16	0.64 [0.46, 0.79]	0.59 [0.39, 0.78]
Demura *et al.* [21]	49	10	1	20	0.98 [0.89, 1.00]	0.67 [0.47, 0.83]
Kim *et al.* [25]	12	5	1	12	0.92 [0.64, 1.00]	0.71 [0.44, 0.90]
Laffon *et al.* [29]	19	6	5	1	0.79 [0.58, 0.93]	0.14 [0.00, 0.58]
Lan *et al.* [24]	25	4	3	13	0.89 [0.72, 0.98]	0.76 [0.50, 0.93]
Matthies *et al.* [10]	18	2	0	14	1.00 [0.81, 1.00]	0.88 [0.62, 0.98]
Schillaci *et al.* [27]	15	4	3	8	0.83 [0.59, 0.96]	0.67 [0.35, 0.90]
Suga *et al.* [26]	41	20	35	37	0.54 [0.42, 0.65]	0.65 [0.51, 0.77]
Xiu *et al.* [22]	13	4	3	26	0.81 [0.54, 0.96]	0.87 [0.69, 0.96]

STP

References	TP	FP	FN	TN	Sensitivity	Specificity
Alkhawaldeh *et al.* [23]	47	15	25	178	0.65 [0.53, 0.76]	0.92 [0.88, 0.96]
Chen *et al.* [28]	5	6	11	9	0.31 [0.11, 0.59]	0.60 [0.32, 0.84]
Cloran *et al.* [30]	14	8	22	19	0.39 [0.23, 0.57]	0.70 [0.50, 0.86]
Demura *et al.* [21]	37	15	13	15	0.74 [0.60, 0.85]	0.50 [0.31, 0.69]
Kim *et al.* [25]	8	1	5	16	0.62 [0.32, 0.86]	0.94 [0.71, 1.00]
Laffon *et al.* [29]	24	7	0	0	1.00 [0.86, 1.00]	0.00 [0.00, 0.41]
Lan *et al.* [24]	24	6	4	11	0.86 [0.67, 0.96]	0.65 [0.38, 0.86]
Matthies *et al.* [10]	14	0	4	16	0.78 [0.52, 0.94]	1.00 [0.79, 1.00]
Schillaci *et al.* [27]	15	1	3	11	0.83 [0.59, 0.96]	0.92 [0.62, 1.00]
Suga *et al.* [26]	61	32	15	25	0.80 [0.70, 0.89]	0.44 [0.31, 0.58]
Xiu *et al.* [22]	0	0	16	30	0.00 [0.00, 0.21]	1.00 [0.88, 1.00]



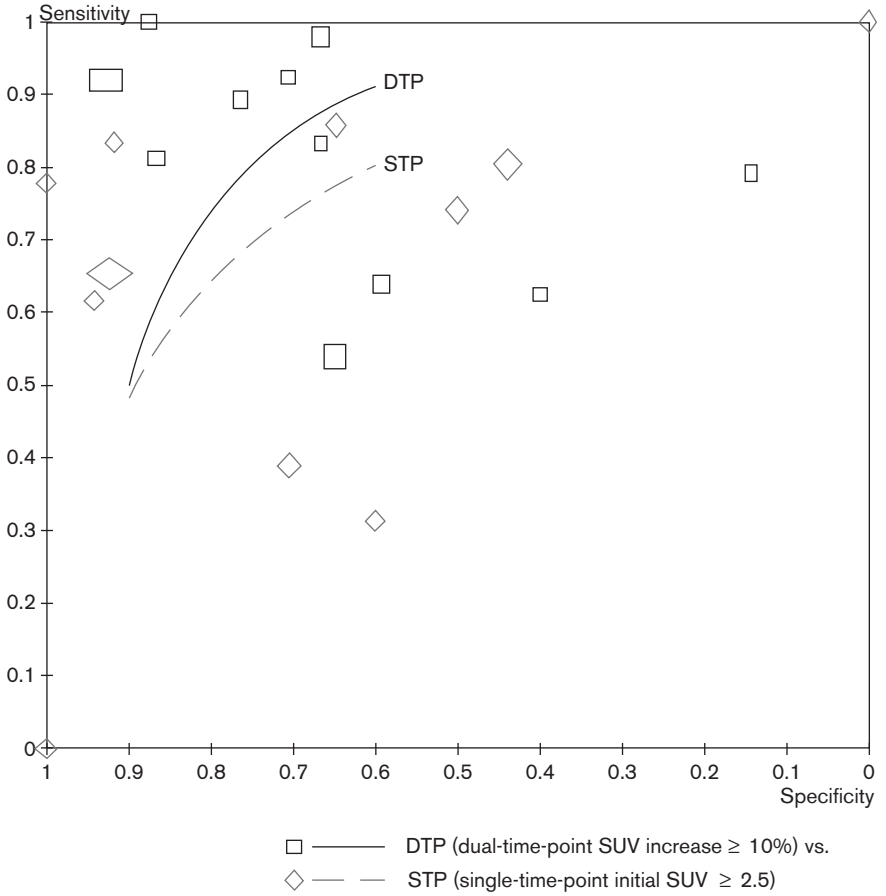
Forest plots, DTP (dual-time-point SUV increase $\geq$ 10%) vs. STP (single-time-point initial SUV $\geq$ 2.5). FN, false negative; FP, false positive; SUV, standard uptake value; TN, true negative; TP, true positive.

articles, albeit a smaller total number of patients, because we also excluded patients with multiple pulmonary nodules.

Although not statistically significant on the basis of full AUC comparison, our SROC curve analysis revealed a trend toward higher sensitivity with DTP imaging, particularly at moderate levels of specificity, suggesting a potential benefit in patients with suspicious SPNs and equivocal results on initial scanning. A single SUV, especially on initial imaging, may be inadequate for differential diagnosis because of potential 'confounding' factors such as lesion size, body habitus, plasma glucose level, uptake time interval, image reconstruction parameters, attenuation correction methods, and correction for partial volume effects [32–35]. DTP imaging provides an opportunity to determine a measure of the rate of uptake that may be less affected by factors other than the metabolic characteristics of suspected lesions. Prior research has demonstrated increasing uptake by malignant tumors over several hours after ^{18}F-FDG injection, as opposed to declining or stable uptake by surrounding normal tissue over time [36,37]. Our finding of potentially higher sensitivity with DTP imaging is consistent with the notion that increasing contrast between tumor and background activity improves the sensitivity of lesion detection.

Our review revealed substantial heterogeneity and apparently conflicting results in the original primary studies. Initial research showed that DTP imaging appeared to be useful in differential diagnosis [4]. The results of several subsequent clinical studies also indicated improved accuracy with DTP imaging in evaluating lung nodules [10,21–27]. Nevertheless, seemingly opposite conclusions were reached in a few studies [28–30]. The inconsistent results across the studies might reflect the fact that they could be measuring somewhat different aspects. Although small sample size could contribute to the variability due to chance, the major sources of heterogeneity could be attributed to the differences in study methods and populations, such as varying lesion size, malignancy risk, ^{18}F-FDG avidity, and uptake intervals. Some investigators attempted to alleviate reduced patient throughput by shortening the waiting time [20], whereas others opted for a longer uptake interval with the expectation of obtaining larger differences in uptake values [21].

Fig. 4



SROC curves, DTP (dual-time-point SUV increase ≥ 10%) vs. STP (single-time-point initial SUV ≥ 2.5). SROC, summary receiver operating characteristic; SUV, standard uptake value.

Different methods of image analysis and threshold measures, as well as the choice of reference standards, could also contribute to the heterogeneity.

A review of the individual studies may help illustrate how these differences might have impacted the accuracy estimates. Matthies *et al.* [10] found the sensitivity and specificity for initial and DTP imaging to be 80% (16/20, 95% confidence interval 0.56–0.94) and 94% (17/18, 0.73–1.00) and 100% (20/20, 0.83–1.00) and 89% (16/18, 0.65–0.99), respectively. Similarly, in a prospective study, Demura *et al.* [21] estimated the sensitivity and specificity for initial, delayed, and DTP imaging to be 74% (37/50, 0.60–0.85) and 50% (15/30, 0.31–0.69), 76% (38/50, 0.62–0.87) and 57% (17/30, 0.37–0.75), and 98% (49/50, 0.89–1.00) and 67% (20/30, 0.47–0.83), respectively, and concluded that DTP imaging was more accurate than STP scanning, except in patients with active granulomatous diseases.

Subsequently, four retrospective studies were published on the accuracy of DTP imaging for assessing lung lesions with initial SUV below 2.5. Xiu *et al.* [22] estimated the sensitivity and specificity to be 81% (13/16, 0.54–0.96) and 87% (26/30, 0.69–0.96), respectively. In comparison, Kim *et al.* [25] found a higher sensitivity of 92% (12/13, 0.64–1.00) and a lower specificity of 71% (12/17, 0.44–0.90). This trade-off could be attributed to both a threshold effect and the differences between patient populations, as Kim and colleagues used a lower threshold (a change in SUV_{max} of -2.3 vs. $+10\%$) in a population with larger lesion size (2.0 ± 0.8 vs. 1.1 ± 0.3 cm). In contrast, Chen *et al.* [28] found lesion-based sensitivity and specificity to be 63% (10/16, 0.35–0.85) and 40% (6/15, 0.16–0.68), respectively, and concluded that DTP imaging was not useful in populations with a high prevalence of granulomatous disease, such as tuberculosis. Chen and colleagues admitted to the possibility of selection bias, which could have contributed to the significantly lower specificity. Later, Cloran *et al.* [30] demonstrated lesion-based sensitivity of 63% (24/38, 0.46–0.78) and specificity of 59% (17/29, 0.39–0.76) and concluded that their results were consistent with and

might extend the applicability of Chen's findings, because the benign nodules in their study were mostly nongranulomatous infections.

In the largest of these studies, Alkhawaldeh *et al.* [23] retrospectively evaluated the impact of partial volume correction and DTP imaging. On initial scanning without partial volume correction, they found a high specificity of 92% (178/193, 0.88–0.96) and a low sensitivity of 65% (47/72, 0.53–0.76), which was attributed to the significant proportion (34%) of lesions smaller than 1.5 cm. With partial volume correction, sensitivity increased significantly to 90% (65/72, 0.81–0.96) with a consequential drop in specificity to 80% (154/193, 0.73–0.85), whereas on DTP imaging a similar gain in sensitivity to 92% (57/62, 0.82–0.97) was not accompanied by a loss in specificity, which remained high at 93% (179/193, 0.88–0.96). Last but not the least, in a small prospective study comparing DTP PET imaging with contrast-enhanced CT within a PET-CT examination, Schillaci *et al.* [27] estimated the sensitivity and specificity for initial, delayed, and DTP PET imaging to be 78% (14/18, 0.52–0.94) and 92% (11/12, 0.62–1.00), 78% (14/18, 0.52–0.94) and 67% (8/12, 0.35–0.90), and 83% (15/18, 0.59–0.96) and 67% (8/12, 0.35–0.90), respectively. Contrast-enhanced CT demonstrated the lowest accuracy. They concluded that DTP PET imaging was the most sensitive, whereas initial PET imaging was the most specific.

In summary, most of the studies were retrospective and prone to bias. Quality assessment of study methodology was problematic because of poor reporting. A major limitation of our analysis was the varying time interval between initial and delayed scans across studies, which raised the question of appropriateness of our attempt to apply the same diagnostic criteria across the studies. The substantial heterogeneity in the results of the studies reflected important differences in study methods and populations. Although the heterogeneity made it impractical to obtain precise overall estimates of diagnostic accuracy, it did provide an opportunity to improve the understanding of how these estimates might vary in different settings. Our analysis revealed the potential for overall improved accuracy with DTP versus initial STP imaging, which might translate into higher sensitivity while maintaining a moderate level of specificity. Potentially higher sensitivity with DTP imaging may be beneficial to patients with malignant but curable SPNs without significant initial ^{18}F-FDG uptake, provided the consequences of false-positive results are acceptable. Clinicians contemplating the use of DTP imaging in specific scenarios should carefully weigh the additional costs and potential benefits. These decisions should be individualized on the basis of available information, such as morphological data from CT imaging, as most malignant ground-glass opacities may not exhibit high ^{18}F-FDG uptake, whereas the degree of ^{8}F-FDG avidity of benign inflammatory lesions depends on disease activity.

Conclusion

Although the results of our analysis do not support the routine use of DTP imaging with ^{18}F-FDG PET in the differential diagnosis of pulmonary nodules, this technique may provide additional information in selected cases with equivocal results from initial scanning. Further prospective research is required to better define the potential benefits of DTP ^{18}F-FDG PET imaging.

Acknowledgements

The authors thank Chih-Hsin Muo of China Medical University, College of Public Health, and China Medical University Hospital, Management Office for Health Data, for assistance with data management and statistical analysis. They express their appreciation for the grant support for this study (DMR-100-078 and DMR-100-110) provided by their hospital and the Taiwan Department of Health Clinical Trial and Research Center for Excellence (DOH101-TD-B-111-004) and the Taiwan Department of Health Cancer Research Center for Excellence (DOH101-TD-C-111-005).

Conflicts of interest

There are no conflicts of interest.

References

1 Patz EF Jr, Lowe VJ, Hoffman JM, Paine SS, Burrowes P, Coleman RE, Goodman PC. Focal pulmonary abnormalities: evaluation with F-18 fluorodeoxyglucose PET scanning. *Radiology* 1993; **188**:487–490.
2 Knight SB, Delbeke D, Stewart JR, Sandler MP. Evaluation of pulmonary lesions with FDG-PET. *Chest* 1996; **109**:982–988.
3 Pauwels EK, Ribeiro MJ, Stoot JH, McCready VR, Bourguignon M, Mazière B. FDG accumulation and tumor biology. *Nucl Med Biol* 1998; **25**: 317–322.
4 Zhuang H, Pourdehnad M, Lambright ES, Yamamoto AJ, Lanuti M, Li P, *et al.* Dual time point ^{18}F-FDG PET imaging for differentiating malignant from inflammatory processes. *J Nucl Med* 2001; **42**:1412–1417.
5 Zhang H, Tian M, Oriuchi N, Higuchi T, Tanada S, Endo K. Detection of lung cancer with positron coincidence gamma camera using fluorodeoxyglucose in comparison with dedicated PET. *Eur J Radiol* 2003; **47**:199–205.
6 Tatsumi M, Yutani K, Watanabe Y, Miyoshi S, Tomiyama N, Johkoh T, *et al.* Feasibility of fluorodeoxyglucose dual-head gamma camera coincidence imaging in the evaluation of lung cancer: comparison with FDG PET. *J Nucl Med* 1999; **40**:566–573.
7 Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD, Siegel BA, *et al.* Procedure guideline for tumor imaging with ^{18}F-FDG PET/CT 1.0. *J Nucl Med* 2006; **47**:885–895.
8 Schelbert HR, Hoh CK, Royal HD, Brown M, Dahlbom MN, Dehdashti F, Wahl RL. Procedure guideline for tumor imaging using fluorine-18-FDG. Society of Nuclear Medicine. *J Nucl Med* 1998; **39**:1302–1305.
9 Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003; **3**:25.
10 Matthies A, Hickeson M, Cuchiara A, Alavi A. Dual time point ^{18}F-FDG PET for the evaluation of pulmonary nodules. *J Nucl Med* 2002; **43**:871–875.
11 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**:557–560.
12 Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technol Assess* 2005; **9**:1–113, iii.

13 Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; **21**:1559–1573.
14 Lijmer JG, Bossuyt PM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002; **21**:1525–1537.
15 Glasziou PP, Sanders SL. Investigating causes of heterogeneity in systematic reviews. *Stat Med* 2002; **21**:1503–1511.
16 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; **21**:1539–1558.
17 Walter SD. Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data. *Stat Med* 2002; **21**:1237–1256.
18 Walter SD. The partial area under the summary ROC curve. *Stat Med* 2005; **24**:2025–2040.
19 Núñez R, Kalapparambath A, Varela J. Improvement in sensitivity with delayed imaging of pulmonary lesions with FDG-PET. *Rev Esp Med Nucl* 2007; **26**:196–207.
20 Conrad GR, Sinha P. Narrow time-window dual-point ^{18}F-FDG PET for the diagnosis of thoracic malignancy. *Nucl Med Commun* 2003; **24**: 1129–1137.
21 Demura Y, Tsuchida T, Ishizaki T, Mizuno S, Totani Y, Ameshima S, *et al*. ^{18}F-FDG accumulation with PET for differentiation between benign and malignant lesions in the thorax. *J Nucl Med* 2003; **44**:540–548.
22 Xiu Y, Bhutani C, Dhurairaj T, Yu JQ, Dadparvar S, Reddy S, *et al*. Dual-time point FDG PET imaging in the evaluation of pulmonary nodules with minimally increased metabolic activity. *Clin Nucl Med* 2007; **32**:101–105.
23 Alkhawaldeh K, Bural G, Kumar R, Alavi A. Impact of dual-time-point (18)F-FDG PET imaging and partial volume correction in the assessment of solitary pulmonary nodules. *Eur J Nucl Med Mol Imaging* 2008; **35**:246–252.
24 Lan XL, Zhang YX, Wu ZJ, Jia Q, Wei H, Gao ZR. The value of dual time point (18)F-FDG PET imaging for the differentiation between malignant and benign lesions. *Clin Radiol* 2008; **63**:756–764.
25 Kim IJ, Kim SJ, Kim YS, Lee TH, Jeong YJ. Characterization of pulmonary lesions with low F-18 FDG uptake using double phase F-18 FDG PET/CT: comparison of visual and quantitative analyses. *Neoplasma* 2009; **56**: 33–39.
26 Suga K, Kawakami Y, Hiyama A, Sugi K, Okabe K, Matsumoto T, *et al*. Dual-time point ^{18}F-FDG PET/CT scan for differentiation between ^{18}F-FDG-avid non-small cell lung cancer and benign lesions. *Ann Nucl Med* 2009; **23**:427–435.
27 Schillaci O, Travascio L, Bolacchi F, Calabria F, Bruni C, Cicciò C, *et al*. Accuracy of early and delayed FDG PET-CT and of contrast-enhanced CT in the evaluation of lung nodules: a preliminary study on 30 patients. *Radiol Med* 2009; **114**:890–906.
28 Chen CJ, Lee BF, Yao WJ, Cheng L, Wu PS, Chu CL, Chiu NT. Dual-phase ^{18}F-FDG PET in the diagnosis of pulmonary nodules with an initial standard uptake value less than 2.5. *Am J Roentgenol* 2008; **191**:475–479.
29 Laffon E, de Clermont H, Begueret H, Vernejoux JM, Thumerel M, Marthan R, Ducassou D. Assessment of dual-time-point ^{18}F-FDG-PET imaging for pulmonary lesions. *Nucl Med Commun* 2009; **30**:455–461.
30 Cloran FJ, Banks KP, Song WS, Kim Y, Bradley YC. Limitations of dual time point PET in the assessment of lung nodules with low FDG avidity. *Lung Cancer* 2010; **68**:66–71.
31 Barger RL Jr, Nandalur KR. Diagnostic performance of dual-time (18)F-FDG PET in the diagnosis of pulmonary nodules: a meta-analysis. *Acad Radiol* 2012; **19**:153–158.
32 Zasadny KR, Wahl RL. Standardized uptake values of normal tissues at PET with 2-[fluorine-18]-fluoro-2-deoxy-D-glucose: variations with body weight and a method for correction. *Radiology* 1993; **189**:847–850.
33 Kim CK, Gupta NC, Chandramouli B, Alavi A. Standardized uptake values of FDG: body surface area correction is preferable to body weight correction. *J Nucl Med* 1994; **35**:164–167.
34 Hickeson M, Yun M, Matthies A, Zhuang H, Adam LE, Lacorte L, Alavi A. Use of a corrected standardized uptake value based on the lesion size on CT permits accurate characterization of lung nodules on FDG-PET. *Eur J Nucl Med Mol Imaging* 2002; **29**:1639–1647.
35 Weber W, Young C, Abdel-Dayem HM, Sfakianakis G, Weir GJ, Swaney CM, *et al*. Assessment of pulmonary lesions with ^{18}F-fluorodeoxyglucose positron imaging using coincidence mode gamma cameras. *J Nucl Med* 1999; **40**:574–578.
36 Hamberg LM, Hunter GJ, Alpert NM, Choi NC, Babich JW, Fischman AJ. The dose uptake ratio as an index of glucose metabolism: useful parameter or oversimplification? *J Nucl Med* 1994; **35**:1308–1312.
37 Basu S, Kung J, Houseni M, Zhuang H, Tidmarsh GF, Alavi A. Temporal profile of fluorodeoxyglucose uptake in malignant lesions and normal organs over extended time periods in patients with lung carcinoma: implications for its utilization in assessing malignant lesions. *Q J Nucl Med Mol Imaging* 2009; **53**:9–19.