

Research Paper

Optimal Modality for Detecting Distant Metastasis in Primary Nasopharyngeal Carcinoma during Initial Staging: A Systemic Review and Meta-analysis of 1774 Patients

Cheng Xu^{1*}, Yuan Zhang^{1*}, Liang Peng^{1*}, Xu Liu¹, Wen-Fei Li¹, Ying Sun¹, Xu Zhang², Xiao-Ping Lin², Qing Liu³, Jun Ma¹✉

1. Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Canton, China;
2. Department of Nuclear Medicine, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Canton, China;
3. Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Canton, China.

* Cheng Xu, Yuan Zhang, and Liang Peng contributed equally to this work.

✉ Corresponding author: Jun Ma, Department of Radiation Oncology, Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, 651 Dongfeng Road East, Guangzhou 510060, People's Republic of China. Telephone: +86-20-87343469 Fax: +86-20-87343295 E-mail: majun2@mail.sysu.edu.cn

© Ivyspring International Publisher. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2016.11.15; Accepted: 2017.01.12; Published: 2017.04.10

Abstract

Purpose: To compare the diagnostic performance of two modalities commonly used for detecting distant metastasis in primary nasopharyngeal carcinoma (NPC): ¹⁸F-fluorodeoxyglucose (¹⁸F-FDG) positron emission tomography/computed tomography (PET/CT) and conventional work-ups (CWUs).

Methods: All topic-related studies were comprehensively searched and included. We determined sensitivities and specificities across studies, calculated negative and positive likelihood ratios (LR⁻ and LR⁺, respectively), and constructed summary receiver operating characteristic curves. Moreover, we compared the diagnostic performance of PET/CT and CWUs by analyzing studies that reported the results of these diagnostic methods on the same patients.

Results: The pooled sensitivity and specificity were 85.7% and 98.1% for PET/CT (1474 patients), and 38.0% and 97.6% for CWUs (1329 patients). In the head-to-head comparison of PET/CT and CWUs (1029 patients), PET/CT showed a significantly higher sensitivity (83.7% vs. 40.1%, $P < 0.001$) and lower LR⁻ (0.169 vs. 0.633, $P < 0.001$) than CWUs on a per-patient basis; no significant difference was observed in pooled specificity (97.7% vs. 97.8%, $P = 0.892$) or LR⁺ (36.416 vs. 16.845, $P = 0.149$). The superiority of PET/CT over CWUs was due mainly to the better diagnostic performance on bone metastasis. However, suboptimal sensitivity of PET/CT was reported in the aspect of detection of liver metastasis. Sensitivity analyses showed relatively poor sensitivity and LR⁻ of PET/CT compared to the original analysis.

Conclusions: PET/CT was superior to CWUs in detecting distant metastasis in primary NPC. However, the efficacy of PET/CT in detecting liver metastasis still requires further optimization.

Key words: Positron emission tomography/computed tomography, nasopharyngeal carcinoma, distant metastasis, diagnosis, meta-analysis.

Introduction

Nasopharyngeal carcinoma (NPC), one of the few head and neck malignancies prone to distant metastasis, is prevalent in Southern China, Southeast Asia, North Africa, the Middle East, and Alaska [1, 2].

In previous reports, 7.7%–20.3% patients had metastases at presentation, with the bones, lung, and liver as the most commonly affected sites [3-12]. Aggressive loco-regional radiotherapy with or

without chemotherapy is the recommended strategy for non-metastatic NPCs, and the long-term overall survival rate in these cases exceeds 80% [13, 14]. However, systemic palliative chemotherapy is the standard treatment for patients with metastasis, and the 1-year overall survival rate is only about 50% [1].

Distant metastasis is one of the most critical factors guiding treatment decisions in oncology and supposed to be diagnosed effectively [15]. Conventional imaging tests to detect metastasis, including ultrasonography, skeletal scintigraphy, and magnetic resonance imaging (MRI), require multiple tests and their accuracy is not much satisfactory [16, 17]. Integrated ^{18}F -fluorodeoxyglucose (^{18}F -FDG) positron emission tomography/computed tomography (PET/CT) can detect the increase in glucose metabolism in tumor cells and offer anatomical details at the same time. Moreover, it enables whole-body examination in a single test [18]. However, disadvantages of PET/CT, including exposure of patients to ionizing radiation, requirement for a cyclotron, and high costs, affect its wide application [19].

Conventional work-ups (CWUs; that is, skeletal scintigraphy, chest X-ray examination, and liver ultrasound) are widely used to detect metastasis in NPC, especially in developing countries, because of their low cost and accessibility [1, 20]. A large prospective study reported that CWUs presented equal effectiveness as PET/CT for patients who have both N0-1 classification and Epstein-Barr virus (EBV)-DNA less than 4000 copies/mL [9]. A previous meta-analysis on the diagnostic performance of PET/CT in newly diagnosed NPC merely enrolled a small sample size of 385 patients [21]. Moreover, it did not compare the diagnostic performance of PET/CT and CWUs.

Therefore, the optimal modality of choice on the diagnostic performance of detecting distant metastasis in primary NPC during initial staging still requires thorough investigation. We conducted this study to individually assess the overall value of ^{18}F -FDG PET/CT and CWUs, and to perform a head-to-head comparison between the two modalities.

Materials and Methods

Identification and eligibility of relevant studies

A prospective protocol was initially planned according to the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) [22]. We systemically searched in MEDLINE, EMBASE and Cochrane Library for eligible articles without restrictions to language or region;

additionally, we searched in the Chinese Biomedical Disc (CBMdisc) database for Chinese articles (last updated on October 3, 2016). The search algorithm was based on the following terms for all possible combinations: (1) fluorodeoxyglucose, FDG, positron emission tomography/computed tomography, or PET/CT; (2) conventional work-ups, CWUs; (3) nasopharyngeal carcinoma, nasopharyngeal cancer.

After the initial screening of the title and abstract of retrieved literatures, the full text of relevant articles was independently assessed by two investigators for inclusion (Y.Z. and L.P.) and any disagreements were resolved by consensus. The references in relevant articles were then manually screened for additional studies. The inclusion criteria were as follows: (a) ^{18}F -FDG PET/CT and CWUs were used for detecting whole-body distant metastasis in NPC; (b) Per-patient statistics including true positive, false positive, true negative, and false negative number were reported; (c) Among reports that pertained to overlapping patient cohorts, we retained the largest study to avoid duplication of information; (d) At least one of the following strategies was used as the reference standard: biopsy, imaging or clinical follow-up.

We excluded studies based on the following criteria: (a) Studies enrolled patients with residual/recurrent NPC; (b) Studies enrolled mixed patients with untreated and residual/recurrent disease, if relevant data regarding the untreated patients could not be obtained; (c) Studies enrolled patients with no evidence of distant metastasis on CWUs; (d) Case reports, conference abstracts and reviews.

Quality assessment and data extraction

Two investigators (Y.Z. and L.P.) independently evaluated the methodological quality of all included studies using the Quality Assessment tool for Diagnostic Accuracy Studies version-II (QUADAS-II) [23]. It consists of four key domains covering patient selection, index test, reference standard, and flow and timing. Each domain was assessed in terms of risk of bias and the first three were assessed in terms of concerns regarding applicability. If a study had \geq four items of low risk/concern, it was considered to be of high-quality. Moreover, signalling questions were used to help reach judgements on the domains of risk of bias.

In addition, for each report, we recorded the author names, year of publication, time range, country or region, sample size, technical parameters, interpreters, criteria defining positive PET/CT results, reference standards, follow-up time and so on. To assess the technical quality of PET/CT, we referred to the guidelines of the Society of Nuclear Medicine [24] and consulted two nuclear medicine physicians (X.Z.

and X.P.L.) experienced in PET/CT imaging.

The same investigators extracted data from eligible studies independently by using a standardized data extraction form, and any disagreements were resolved by consensus. The investigators were not blinded to information regarding the journal name, authors, or affiliations, since this was unnecessary [25]. For each study, we recorded the number of true positive, false positive, true negative, and false negative findings of each modality in diagnosing distant metastasis. We used all available information, including findings per patient and per site (bones, lung, and liver).

Statistical analysis

We calculated *kappa* coefficient to evaluate the agreement between investigators regarding their answers for signalling questions; $P < 0.05$ indicated a good inter-rater reliability [26]. We explored the threshold effect inducing heterogeneity using Spearman correlation coefficient; $P < 0.05$ suggested presence of threshold effect [27]. In order to detect the heterogeneity due to sources other than threshold effect, we examined sensitivity, specificity, and negative and positive likelihood ratios (LR⁻ and LR⁺, respectively) using Cochran-Q and Chi-square (χ^2) test with the significance set at $P < 0.05$. I-square (I^2) statistic was also calculated to measure heterogeneities; I^2 -value $> 25\%$, 50% or 75% was considered to have mild, moderate or substantial heterogeneity [28].

We constructed summary receiver operating characteristic (SROC) curves to quantitatively summarize study results. A symmetric SROC curve was performed in the DerSimonian-Laird model when regression coefficient-*b*, the difference between the slope of the fitted regression line and zero, was non-significant ($P > 0.05$) [27]. We calculated pooled sensitivity, specificity, LR⁻, LR⁺, area under the curve (AUC), and Q* index (the maximum joint sensitivity and specificity) for each modality. The necessary precondition to demonstrate forest plots of those outcomes is that no substantial heterogeneity exists [27]. LR⁻ is defined as the ratio of (1 - sensitivity) over specificity, whereas LR⁺ is defined as the ratio of sensitivity over (1 - specificity). Although there is no absolute cutoff, a good diagnostic test would generally have LR⁺ > 5.0 and LR⁻ < 0.2 . The maximum joint sensitivity and specificity was defined as the point on the SROC curve that is intersected by a diagonal line that runs from the top left corner to the bottom right corner of the ROC diagram; this could be calculated by using the formula $Q^* = (1 + e^{-A/2})^{-1}$, where A is the summary log odds ratio (sensitivity/1 - specificity). Z test was used to compare all the pooled

outcomes between modalities.

The methodological quality summary of included studies was performed using Review Manager 5.3 (Cochrane Collaboration, Oxford, UK). Potential publication bias was investigated using Deeks' funnel plot by Stata software 12.0 (StataCorp, College Station, TX, USA); $P < 0.10$ indicated a high likelihood of publication bias [29]. All statistical analyses and figures were produced using Meta-Disc version 1.4 (Ramón Cajal Hospital, Madrid, Spain) [30].

Results

Study identification and description

After the rounds of selection presented in Figure 1, ten studies enrolling 1774 patients were included in our study. Table 1 presents the characteristics of the eligible studies. All studies were performed in Asia and enrolled patients with any stages, of which, six studies were published in English [3-7, 9] and four were in Chinese with English abstracts [8, 10-12]. Five studies were prospective [4-7, 9] while another five studies were retrospective [3, 8, 10-12]. One study compared CWUs with standalone ¹⁸F-FDG PET [4]; the remaining nine studies compared PET/CT with different diagnostic modalities, such as CWUs, CT, whole-body MRI, and CT+skeletal scintigraphy [3, 5-12].

The technical parameters of nine studies employing PET/CT are shown in Table 2. Generally, these studies followed most of the guidelines for performing PET/CT imaging [24]. Four studies implemented a comparison of PET/CT and CWUs on both patient and site basis [5, 6, 9, 10]. All studies used the visual interpretation to define positive PET/CT results; four of them used semi-quantitative criteria [3, 5, 7, 9]. Moreover, eight studies completed all diagnostic procedures within 7-14 days [3, 5-11]. Seven studies excluded participants with hyperglycemia [5-11]; another seven studies evaluated diagnostic findings of different modalities using double-blind method [3, 5-10].

Assessment of study quality

Two investigators showed a good agreement in their assessment of signalling questions, with a *kappa* coefficient of 0.823 ($P < 0.001$) (Supplementary Table S1). Generally, for patients with negative findings on PET/CT and CWUs, close clinical/imaging follow-up was used as the reference standard. If distant metastases were suspected on PET/CT or CWUs, biopsy was pursued whenever possible. If biopsy was not feasible or yielded negative results, clinical/imaging follow-up was performed. This procedure seems reasonable and appropriate in

practice but might subject to differential verification bias. Therefore, for studies regarding both clinical/imaging follow-up and biopsy as reference standard, the answer was “no” for question 2-3 of the flow and timing domain (i.e., did all patients receive the same reference standard). Moreover, no study reported whether or not the clinical/imaging follow-up was interpreted without knowledge of PET/CT or CWUs. Thus, both investigators answered “unclear” for question 2 of the reference standard domain (i.e., were the reference standard results blind to the results of the index test).

The methodological quality summary of all included studies is shown in the Supplementary Figure S1. The general quality was fair in terms of risk of bias. All studies had low risk in the index test domain; three studies had low risk in the patient selection domain [5, 7, 9]. Considering unclear blind method in reference standard and differential verification bias among patients, all studies were evaluated as “unclear” in the remaining domains. Owing to the stringent inclusion criteria and consistent population characteristics, the overall quality was good in terms of applicability. Only two

study was assessed as “high” in the patient selection domain because of the limitation of sample size [3, 11]. Thus, seven of ten studies were regarded as high-quality [4-7, 9, 10, 12], of which, three studies had five items that were evaluated as low risk/concern [5, 7, 9].

Diagnostic value of ^{18}F -FDG PET/CT

Nine studies were included in the meta-analysis of ^{18}F -FDG PET/CT (1474 patients) [3, 5-12]. There was no threshold effect inducing heterogeneity with the Spearman correlation coefficient of -0.261 ($P = 0.497$). Moreover, non-significant heterogeneity was reported in pooled LR- ($P = 0.632$; $I^2 = 0\%$), LR+ ($P = 0.582$; $I^2 = 0\%$); mild heterogeneity was reported in pooled specificity ($P = 0.188$; $I^2 = 28.9\%$) and sensitivity ($P = 0.061$; $I^2 = 46.3\%$).

The pooled sensitivity and specificity were 85.7% [95% confidence interval (CI), 80.1%-90.2%] and 98.1% (95% CI, 97.2%-98.8%), respectively (Figure 2A-2B). Likelihood ratio syntheses yielded an overall LR- of 0.180 (95% CI, 0.131-0.248) and LR+ of 35.182 (95% CI, 23.902-51.786) (Figure 2C-2D). SROC curve showed an AUC of 0.9812 and Q^* index of 0.9394.

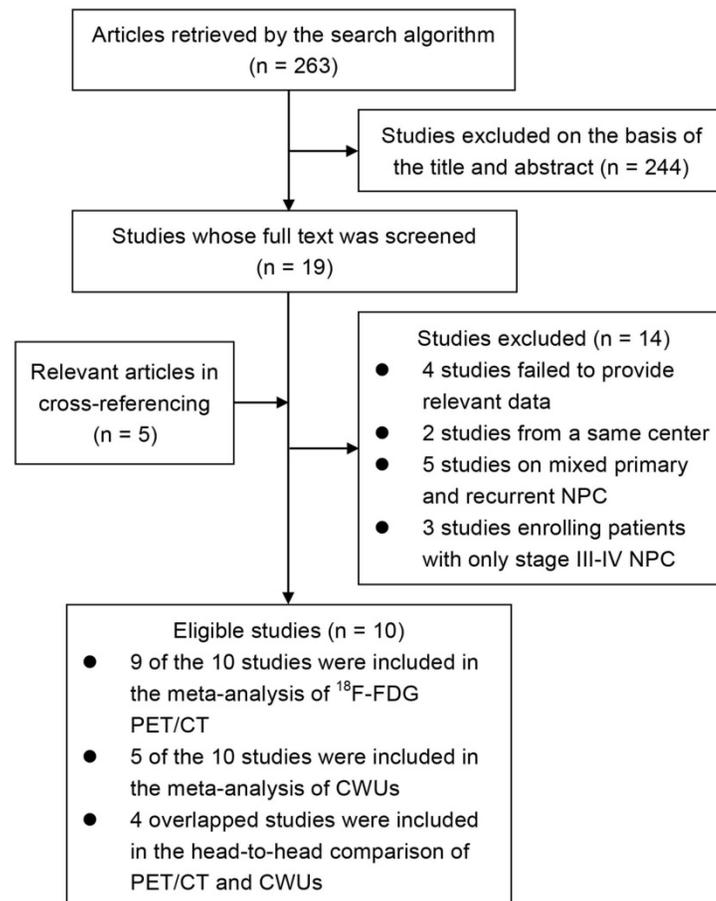


Figure 1. Flow chart describing the identification, inclusion and exclusion of studies.

Table 1. Characteristics of ten included studies on the diagnostic performance of ¹⁸F-FDG PET/CT and CWUs

First author / year	Time range	Country or region	Language	Type of design	Patients no. (M/F)	Mean age	Stage (AJCC)	Histologic type (WHO)	Reference standard	Follow-up time (mo.)	Metastasis prevalence
¹⁸F-FDG PET/CT versus CWUs											
Chua/2009 [5]	2005-2006	Singapore	English	Prospective	78 (60/18)	50.00	T1-4; N0-3	II-III	Biopsy, imaging and clinical follow-up	≥ 6	7.7%
Ng/2009 [6]	NR	Taiwan	English	Prospective	111 (84/27)	48.90	T1-4; N0-3	NR	Biopsy, imaging and clinical follow-up	≥ 12	14.4%
Zhang/2011 [10]	2004-2007	Mainland, China	Chinese	Retrospective	257 (201/56)	45.00	T1-4; N0-3	I-III	Biopsy, imaging and clinical follow-up	≥ 36	15.2%
Lin/2012 ^a [8]	2004-2009	Mainland, China	Chinese	Retrospective	216 (168/48)	45.00	T1-4; N0-3	I-III	Biopsy, imaging and clinical follow-up	NR	14.8%
Tang/2013 [9]	2007-2011	Mainland, China	English	Prospective	583 (474/109)	46.00	T1-4; N0-3	II-III	Biopsy, imaging and clinical follow-up	≥ 12	14.8%
¹⁸F-FDG PET/CT versus other modalities^b											
Chen/2006 [3]	2002-2004	Taiwan	English	Retrospective	20 ^c (14/6)	46.30	T1-4; N0-3	NR	Biopsy, imaging and clinical follow-up	≥ 6	10.0%
Ng/2009 [7]	2006-2007	Taiwan	English	Prospective	150 (111/39)	48.17	T1-4; N0-3	NR	Biopsy, imaging and clinical follow-up	≥ 12	12.0%
Wang/2007 [11]	2002-2005	Mainland, China & Australia	Chinese	Retrospective	18 ^d (NR)	52.00	I-IVb	NR	Biopsy, imaging and clinical follow-up	Median: 17	11.1%
Lin/2009 [12]	2004-2008	Mainland, China	Chinese	Retrospective	41 (25/16)	52.30	T1-4; N0-3	II-III	Biopsy	NR	4.9%
CWUs versus ¹⁸F-FDG PET alone											
Liu/2007 [4]	2002-2005	Taiwan	English	Prospective	300 (210/90)	50.00	T1-4; N0-3	II-III	Biopsy, imaging and clinical follow-up	≥ 12	20.3%

Abbreviations: ¹⁸F-FDG: ¹⁸F-fluorodeoxyglucose; PET: positron emission tomography; CT: computed tomography; CWUs: conventional work-ups; AJCC: American Joint Committee on Cancer; WHO: World Health Organization; T: primary tumor stage; N: node stage; M: male; F: female; no.: number; mo.: months; NR: not reported.

^a The comparison was performed between two matched groups which separately adopted ¹⁸F-FDG PET/CT and CWUs.

^b The study by Chua et al. and Tang et al. also compared ¹⁸F-FDG PET/CT with other modalities; we omitted them to avoid repetition.

^c This study enrolled 70 patients; only the 20 newly diagnosed patients were analysed.

^d This study enrolled 43 patients; only the 18 patients diagnosed during the initial staging were analysed.

Table 2. Parameters of ¹⁸F-FDG PET/CT from nine included studies

First author / year	Type of scanner (corporation)	FDG dose	Time btw FDG injection and scanning (min)	CT slice thickness (mm)	Acquisition mode	Reconstruction method	Criteria defining positive PET/CT result	Interpreters	Design of comparison
Chua/2009 [5]	PET/CT (Siemens)	370 MBq	60	NR	NR	NR	Semi-quantitative (three-point scale)	1NMP	CWUs; PET alone; CT of thorax & abdomen+SS
Ng/2009 [6]	PET/CT (GE)	370 MBq	50-70	3.00	2D	iterative	NR	1R+2NMP	CWUs
Zhang/2011 [10]	PET/CT (GE)	296-440 MBq	45-60	4.25	NR	iterative	NR	1R+1NMP	CWUs
Lin/2012 [8]	PET/CT (NR)	296-440 MBq	45	4.25	NR	NR	NR	NR	CWUs ^a
Tang/2013 [9]	PET/CT (GE)	5.55 MBq/kg	45-60	NR	3D	iterative	Semi-quantitative (three-point scale)	3NMP	CWUs; PET/CT+CWUs
Chen/2006 [3]	PET/CT (GE)	370 MBq	50-70	4.80	3D	iterative	Semi-quantitative (five-point scale)	1R+1NMP	PET alone; CT
Ng/2009 [7]	PET/CT (GE)	370 MBq	50-70	3.00	2D	iterative	Semi-quantitative (five-point scale)	1R+1NMP	Whole-body MRI
Wang/2007 [11]	PET/CT (GE)	270-370 MBq	40-60	NR	3D	iterative	NR	R+NMP ^b	CT+MRI
Lin/2009 [12]	PET/CT (Siemens)	550 MBq	45-60	4.25	NR	NR	NR	2R+2NMP	MRI

Abbreviations: ¹⁸F-FDG: ¹⁸F-fluorodeoxyglucose; PET: positron emission tomography; CT: computed tomography; CWUs: conventional work-ups; SS: skeletal scintigraphy; MRI: magnetic resonance imaging; btw: between; NA: not applicable; NR: not reported; 3D: three-dimensional; 2D: two-dimensional; R: radiologist; NMP: nuclear medicine physician.

^a The comparison was performed between two matched groups which separately adopted ¹⁸F-FDG PET/CT and CWUs.

^b No detailed number of interpreters was reported.

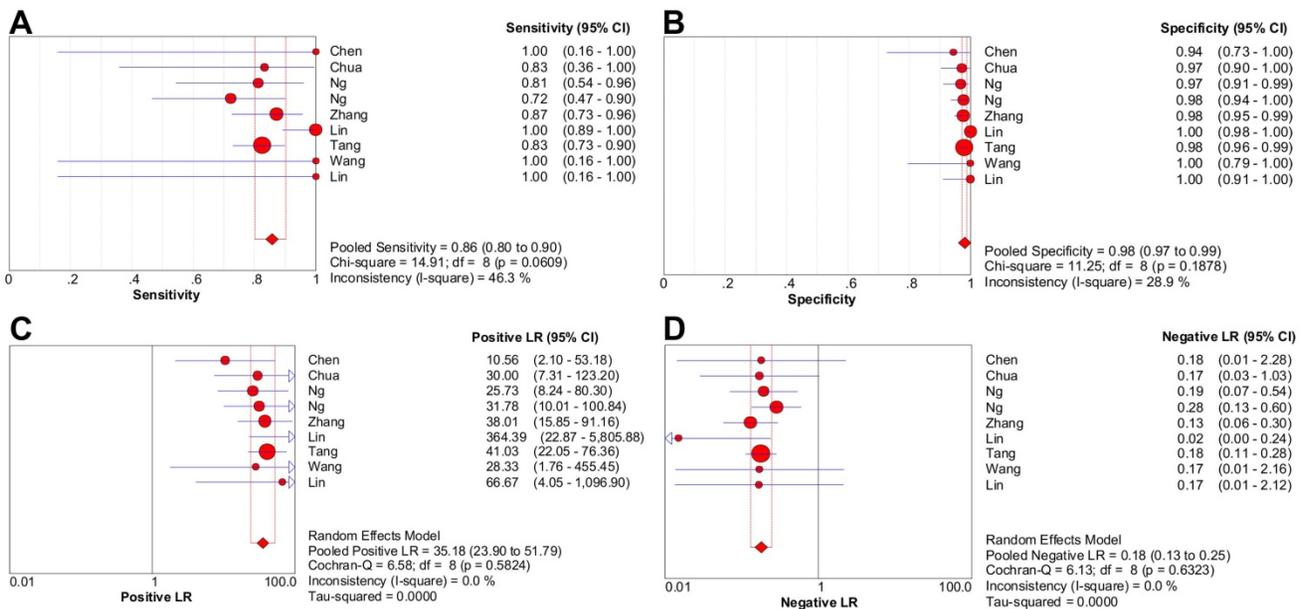


Figure 2. Forest plots of ^{18}F -FDG PET/CT on a per-patient basis in sensitivity (A), specificity (B), positive likelihood ratio (C), and negative likelihood ratio (D). CI: confidence interval; LR: likelihood ratio. Circles are the point estimates of studies with the 95% CIs indicated by horizontal bars. The size of the circles indicates the weight of each study. Diamonds are the summary estimates from the pooled studies with the 95% CIs indicated by horizontal bars. All pooled results were slightly different from the results reported in the text because of rounding.

Diagnostic value of CWUs

Considering the substantial heterogeneity detected in pooled sensitivity ($P < 0.001$; $I^2 = 87.6\%$) and LR- ($P < 0.001$; $I^2 = 80.2\%$), it was not appropriate to perform pooling analysis for all six studies [4-6, 8-10]. After discarding the only one study with high risk of bias [8], five studies (1329 patients) with moderate heterogeneity in pooled sensitivity, specificity, LR+ and LR- ($I^2 = 46.9\%$, 71.1%, 61.2%, 41.7%, respectively) were included in the meta-analysis of CWUs [4-6, 9, 10]. Moreover, no significant heterogeneity due to threshold effect existed since the Spearman correlation coefficient was 0.100 ($P = 0.873$).

The pooled sensitivity and specificity were 38.0% (95% CI, 31.4%–45.0%) and 97.6% (95% CI, 96.5%–98.4%), respectively (Figure 3A-3B). Likelihood ratio syntheses yielded an overall LR- of 0.653 (95% CI, 0.562–0.759) and LR+ of 14.748 (95% CI, 6.746–32.245) (Figure 3C-3D). SROC curve showed an AUC of 0.9040 and Q^* index of 0.8355.

Head-to-head comparison of ^{18}F -FDG PET/CT and CWUs

Four studies enrolling 1029 patients were included in the head-to-head comparison of diagnostic performance between PET/CT and CWUs (Table 3) [5, 6, 9, 10]. Spearman correlation coefficient was -0.400 ($P = 0.600$) in PET/CT and 0.200 ($P = 0.800$) in CWUs. All pooled outcomes of PET/CT had no significant heterogeneity (all I^2 -values = 0%).

Moderate heterogeneity was detected for CWUs in pooled sensitivity, LR-, and LR+ ($I^2 = 54.1\%$, 53.1%, 64.9%, respectively); marginally substantial heterogeneity was reported in pooled specificity ($I^2 = 76.5\%$).

The results showed that PET/CT had a significantly better pooled sensitivity (83.7% vs. 40.1%, $P < 0.001$) and LR- (0.169 vs. 0.633, $P < 0.001$) than CWUs. However, no significant difference was observed in pooled specificity (97.7% vs. 97.8%, $P = 0.892$) or LR+ (36.416 vs. 16.845, $P = 0.149$). Moreover, symmetric SROC curves for the diagnostic performance of PET/CT and CWUs were shown in Figure 4A-4B; DerSimonian-Laird model was used since the regression coefficient- b was -0.008 ($P = 0.9961$) in PET/CT and -0.613 ($P = 0.4797$) in CWUs. The AUC and Q^* index for PET/CT were greater than their counterparts for CWUs (0.9799 vs. 0.9137, $P = 0.040$; 0.9371 vs. 0.8462, $P = 0.014$, respectively).

When the comparisons were among specific sites, PET/CT had a higher pooled sensitivity and LR- than skeletal scintigraphy (89.8% vs. 42.0%, $P < 0.001$; 0.109 vs. 0.608, $P < 0.001$, respectively), chest X-ray examination (87.5% vs. 39.3%, $P < 0.001$; 0.133 vs. 0.670, $P < 0.001$, respectively) and liver ultrasound (72.7% vs. 35.7%, $P < 0.001$; 0.310 vs. 0.635, $P = 0.033$, respectively). However, the outcome also indicated relatively poor sensitivity and LR- of PET/CT in the aspect of detection of liver metastasis (72.7% and 0.310, respectively). Significantly improved results in AUC and Q^* index were only observed in the

subgroup of PET/CT versus skeletal scintigraphy ($P = 0.033, 0.009$, respectively). Thus, the superiority of PET/CT over CWUs was due mainly to the superb diagnostic performance on bone metastasis.

Moreover, the pooled outcomes of PET/CT had generally non-significant heterogeneity, while substantial heterogeneity was observed for chest X-ray examination and liver ultrasound (Table 3).

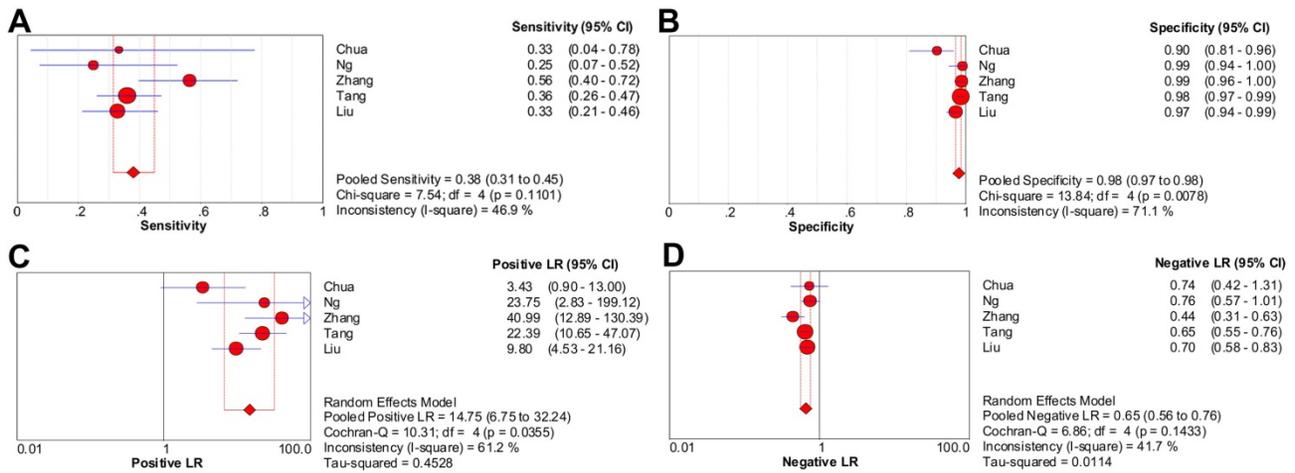


Figure 3. Forest plots of CWUs on a per-patient basis in sensitivity (A), specificity (B), positive likelihood ratio (C), and negative likelihood ratio (D). CI: confidence interval; LR: likelihood ratio. Circles are the point estimates of studies with the 95% CIs indicated by horizontal bars. The size of the circles indicates the weight of each study. Diamonds are the summary estimates from the pooled studies with the 95% CIs indicated by horizontal bars. All pooled results were slightly different from the results reported in the text because of rounding.

Table 3. Head-to-head comparison of the diagnostic performance between ¹⁸F-FDG PET/CT and CWUs on both patient and site basis

Test	Sensitivity (%) (95% CI)	<i>P</i>	Specificity (%) (95% CI)	<i>P</i>	LR- (95% CI)	<i>P</i>	LR+ (95% CI)	<i>P</i>	AUC	<i>P</i>	Q*	<i>P</i>
Patient based [5, 6, 9, 10]												
PET/CT	83.7 (76.7-89.3)	< 0.001	97.7 (96.5-98.6)	0.892	0.169 (0.117-0.244)	< 0.001	36.416 (23.459-56.528)	0.149	0.9799	0.040	0.9371	0.014
<i>I</i> ² (%)	0.0		0.0		0.0		0.0		NA		NA	
CWUs	40.1 (32.1-48.5)		97.8 (96.7-98.7)		0.633 (0.507-0.790)		16.845 (5.960-47.611)		0.9137		0.8462	
<i>I</i> ² (%)	54.1		76.5		53.1		64.9		NA		NA	
Site based [5, 6, 9, 10]												
Bones												
PET/CT	89.8 (81.5-95.2)	< 0.001	98.7 (97.8-99.3)	0.853	0.109 (0.060-0.199)	< 0.001	59.260 (33.998-103.291)	0.177	0.9907	0.033	0.9603	0.009
<i>I</i> ² (%)	0.0		37.4		0.0		0.0		NA		NA	
SS	42.0 (31.6-53.0)		98.8 (97.9-99.4)		0.608 (0.443-0.834)		31.348 (16.702-58.837)		0.9502		0.8907	
<i>I</i> ² (%)	64.1		0.0		53.6		0.0		NA		NA	
Lung												
PET/CT	87.5 (75.9-94.8)	< 0.001	99.0 (98.3-99.5)	0.117	0.133 (0.068-0.260)	< 0.001	85.828 (29.236-251.969)	0.799	0.9936	0.275	0.9680	0.170
<i>I</i> ² (%)	0.0		17.6		0.0		52.2		NA		NA	
CXR	39.3 (26.5-53.2)		99.6 (99.0-99.9)		0.670 (0.443-1.013)		62.462 (13.244-294.581)		0.9657		0.9130	
<i>I</i> ² (%)	81.2		48.7		75.9		39.5		NA		NA	
Liver												
PET/CT	72.7 (57.2-85.0)	< 0.001	100.0 (99.6-100.0)	0.381	0.310 (0.201-0.478)	0.033	250.740 (60.354-1043.400)	0.684	0.9934	0.483	0.9676	0.404
<i>I</i> ² (%)	0.0		0.0		0.0		0.0		NA		NA	
US	35.7 (21.6-52.0)		99.8 (99.2-100.0)		0.635 (0.424-0.952)		95.839 (8.146-1127.500)		0.9777		0.9329	
<i>I</i> ² (%)	55.8		79.8		31.2		63.9		NA		NA	

Abbreviations: ¹⁸F-FDG; ¹⁸F-fluorodeoxyglucose; PET: positron emission tomography; CWUs: conventional work-ups; CI: confidence interval; LR-: negative likelihood ratio; LR+: positive likelihood ratio; AUC: area under the curve; SS: skeletal scintigraphy; CXR: chest X-ray examination; US: ultrasound; NA: not applicable.

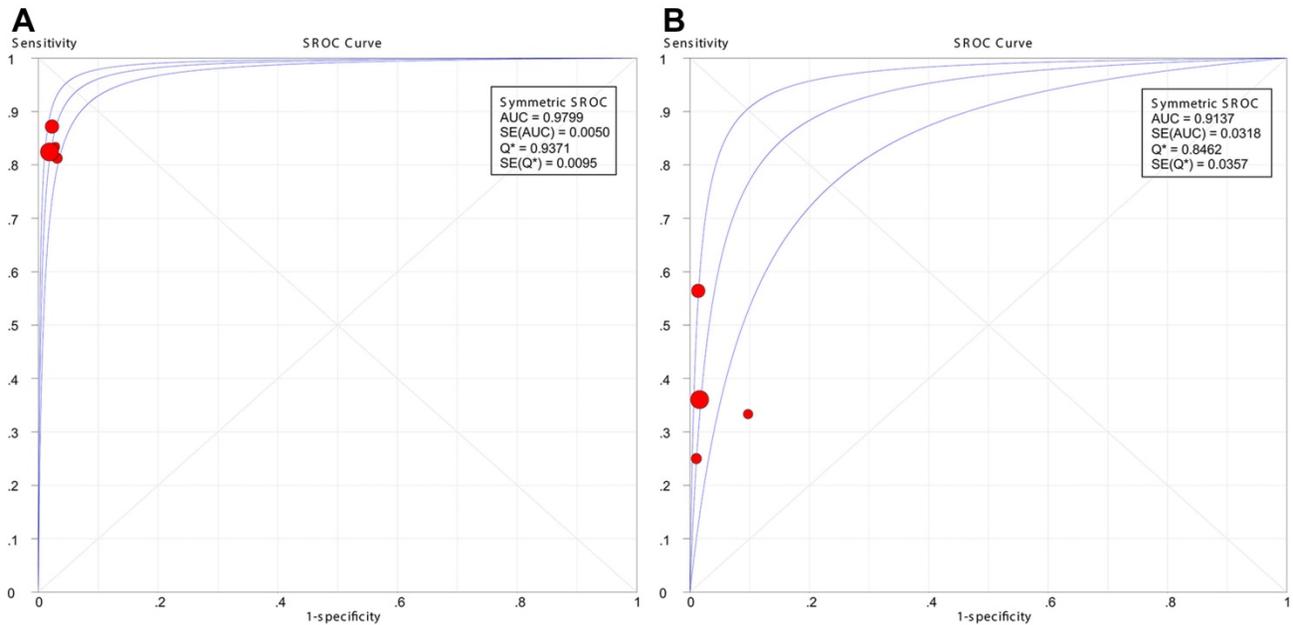


Figure 4. Summary receiver operating characteristic (SROC) curves for the diagnostic performance of ¹⁸F-FDG PET/CT (A) and CWUs (B). AUC: area under the curve; SE: standard error. The size of the circles indicates the weight of each study.

Table 4. Summary of the diagnostic performance of ¹⁸F-FDG PET/CT and other modalities on a per-patient basis

Study	Sensitivity (%) (95% CI)	I ² (%)	Specificity (%) (95% CI)	I ² (%)	LR- (95% CI)	I ² (%)	LR+ (95% CI)	I ² (%)
Diagnostic performance of PET/CT from three meta-analyses								
Our study								
All related studies [3, 5-12]	85.7 (80.1-90.2)	46.3	98.1 (97.2-98.8)	28.9	0.180 (0.131-0.248)	0.0	35.182 (23.902-51.786)	0.0
Compared to CWUs [5, 6, 9, 10]	83.7 (76.7-89.3)	0.0	97.7 (96.5-98.6)	0.0	0.169 (0.117-0.244)	0.0	36.416 (23.459-56.528)	0.0
High-quality studies (≥ four items) [5-7, 9, 10, 12]	82.6 (76.0-88.1)	0.0	97.8 (96.7-98.6)	0.0	0.187 (0.135-0.285)	0.0	36.265 (24.150-54.459)	0.0
High-quality studies (five items) [5, 7, 9]	80.9 (72.3-87.8)	0.0	97.9 (96.5-98.8)	0.0	0.201 (0.137-0.295)	0.0	37.473 (22.500-62.412)	0.0
English publications [3, 5-7, 9]	81.3 (73.4-87.6)	0.0	97.7 (96.4-98.6)	0.0	0.200 (0.140-0.285)	0.0	32.091 (20.517-50.196)	0.0
Shen et al. [31] ^a	89.0 (84.0-93.0)	58.7	97.0 (96.0-98.0)	59.7	NR	NR	NR	NR
Vellayappan et al. [21]	87.0 (74.0-100.0)	0.0	98.0 (96.0-100.0)	0.0	NR	NR	NR	NR
Preliminary results on the diagnostic performance of other modalities								
CT of thorax and abdomen+SS [5]	66.7 (30.0-90.3)	-	91.7 (83.0-96.1)	-	NR	-	NR	-
Whole-body MRI [7]	77.8 (52.4-93.6)	-	98.5 (94.6-99.8)	-	NR	-	NR	-
PET/CT+CWUs [9]	83.7 (75.9-91.5)	-	97.0 (95.5-98.5)	-	NR	-	NR	-
PET alone+CWUs [4]	83.6 (NR)	-	93.7 (NR)	-	NR	-	NR	-

Abbreviations: ¹⁸F-FDG: ¹⁸F-fluorodeoxyglucose; PET: positron emission tomography; CT: computed tomography; CWUs: conventional work-ups; MRI: magnetic resonance imaging; SS: skeletal scintigraphy; CI: confidence interval; LR-: negative likelihood ratio; LR+: positive likelihood ratio; NR: not reported.

^a Only showed the pooled results of the subgroup of PET/CT which included patients with both newly diagnosed and recurrent disease.

Sensitivity analysis and preliminary diagnostic results of other modalities

The sensitivity analysis of PET/CT that individually included six studies with high-quality (i.e. ≥ four items of low risk/concern) [5-7, 9, 10, 12], three high-quality studies having five items of low risk/concern [5, 7, 9], and five studies published in English [3, 5-7, 9], yielded stable outcomes and excellent consistency among included studies; I²-values of all pooled results were 0% (Table 4). All pooled sensitivities were similar to each other and lower than the original analysis. The sensitivity

analyses of English publications and studies having five items of low risk/concern also showed relatively poor values in LR- of 0.200 and 0.201, respectively. Moreover, compared with the results from two previous meta-analyses, our study showed greater between-study consistency and obviously lower pooled sensitivity [21, 31]. Because of the limited number of topic-related studies, we only summarized the diagnostic performance of other modalities on a per-patient basis in Table 4. No significant difference in sensitivity or specificity was reported when each modality was compared with PET/CT in its own study [4, 5, 7, 9].

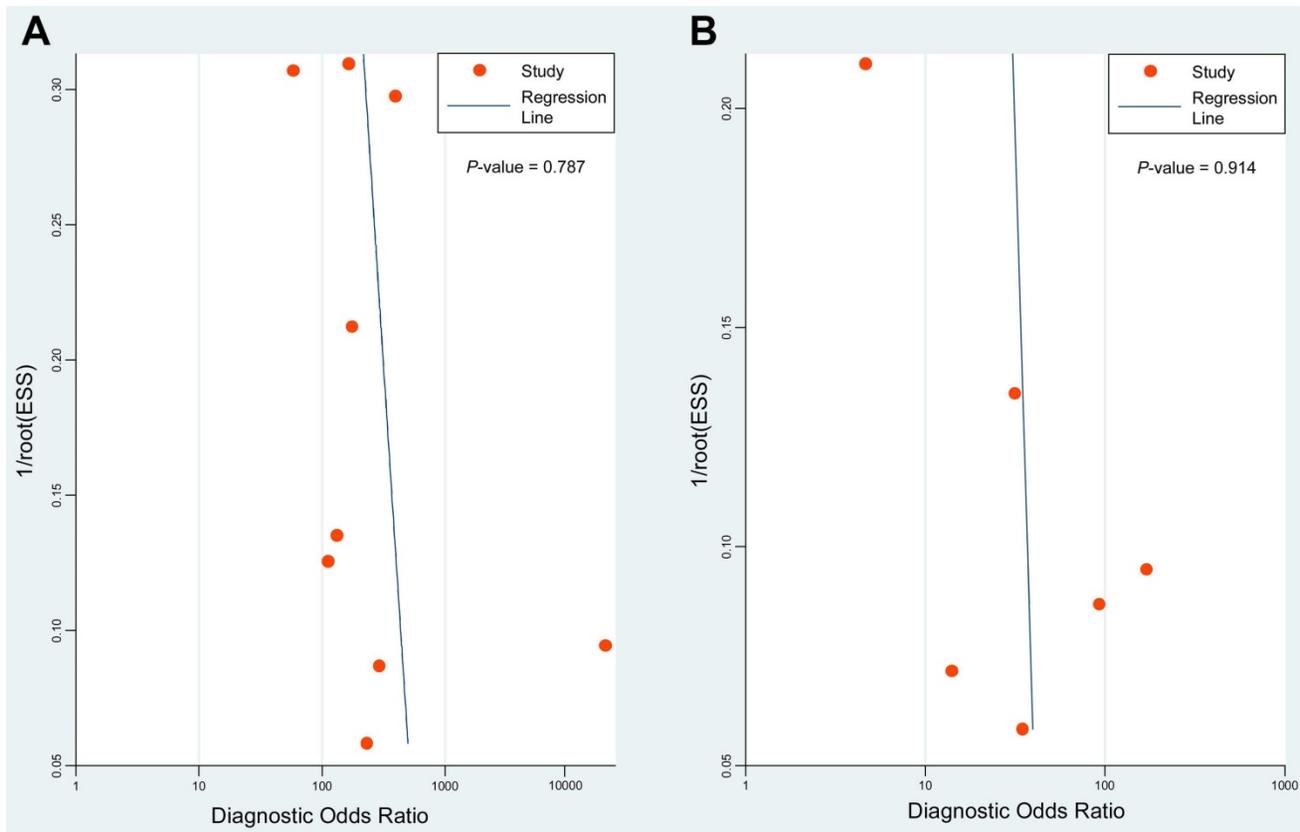


Figure 5. Deeks' funnel plots of ^{18}F -FDG PET/CT (A) and CWUs (B) to evaluate potential publication bias. ESS: effective sample size. $P = 0.787$ and 0.914 indicate symmetrical funnel shapes and suggest no publication bias.

Deeks' funnel plots were performed to assess potential publication bias. The results indicated symmetric funnel shapes and suggested no significant publication bias for studies using PET/CT ($P = 0.787$) or CWUs ($P = 0.914$) (Figure 5).

Discussion

If metastasis is missed during the initial staging, patients with NPC experience unnecessary morbidity and incur the costs of aggressive local-regional radiotherapy while losing the chance of receiving appropriate systemic chemotherapy at an early stage. Thus, the optimal modality to detect metastasis in primary NPC should have high sensitivity, whereby false-negative results can be minimized. The current meta-analysis demonstrated that ^{18}F -FDG PET/CT had higher sensitivity for metastasis detection compared to CWUs. Moreover, PET/CT allow simplification of the initial staging evaluation and reduce the anxiety of patients during testing, since it is single procedure as opposed to the multiple procedures involved in CWUs.

However, although PET/CT was superior to CWUs, the pooled sensitivity of PET/CT was not optimal, especially in the aspect of detection of liver metastasis. The sensitivity analysis also showed

generally lower values in pooled sensitivities than the original analysis. The following reasons may explain why. First, semi-quantitative evaluation by using standardized uptake value of PET/CT is very prone to variations by technical factors of the scanner system and biological factors of patient [32-34]. Although most of the included studies had restrictions on blood glucose level, time between FDG injection and scanning, and reconstruction method, there were still a lot of factors affecting the sensitivity of PET/CT, such as the partial volume effect, variable physiological uptake, and different acquisition parameters. Second, increasing evidence suggests that metastasis-initiating cells are cancer stem cells that may enter into a protracted period of dormancy before subsequent reactivation and proliferation [35]. Therefore, these metastasis-initiating cells might lack increased glucose metabolism and not FDG avid. In the future, technological improvements and standardization of PET/CT imaging procedures can help to further improve its sensitivity for detecting metastasis in primary NPC.

An important concern regarding PET/CT is the high cost, especially in developing countries. In an interesting study, Tang et al. tried to limit the use of PET/CT to patients with a higher-risk of metastasis at

presentation in order to reduce costs [9]. They divided patients into very low-risk, low-risk, and intermediate-risk groups for metastases based on node classification and pretreatment EBV-DNA levels. They demonstrated that the costs per true-positive case detected by PET/CT were \$47,458, \$14,188, and \$5,005 in the three groups, respectively. Therefore, they did not recommend routine PET/CT in the very low-risk and low-risk groups. However, clinicians should be aware that this study was not a cost-effectiveness study but a cost-minimization study, and the cost analyses did not include the consequences of the imaging tests and associated costs (e.g., the standard therapy for NPC if no distant metastases are detected, type of therapy if metastases are detected on staging, and extra imaging tests and biopsies required for both true- and false-positive results) or the difference in therapeutic effects (e.g., survival). Nonetheless, the selective PET/CT application approach based on appropriate stratification of NPC patients suggested by Tang et al. is appealing and warrants further investigation in future studies.

Our study differs from two previous meta-analyses in the respects shown in supplementary Table S2 [21, 31]. Although these meta-analyses reported that PET/CT had good diagnostic performance for metastasis in NPC, they showed relatively poor reliability, mainly because of differences in patient populations (e.g., patients with primary or residual/recurrent disease [36, 37], all diagnosed patients or patients with negative results on CWUs [38]) and the target condition (metastases to all sites or bones only [39]) across the studies included. More importantly, these meta-analyses did not compare the diagnostic performance of PET/CT and CWUs. Our study showed obviously lower values in pooled sensitivities and generally greater between-study consistency than the previous meta-analyses. Moreover, we included a large population of 1774 patients through a stringent inclusion and exclusion criteria, which helped to improve the overall reliability.

The present meta-analysis has several limitations that must be taken into account. First, previous studies have reported that the whole-body MRI with diffusion weighted imaging has equivalent diagnostic efficacy for detection of distant metastasis as PET/CT in other malignancies [19, 40]. However, limited number of topic-related studies on the diagnostic performance of whole-body MRI in NPC prevented us from further investigation. We summarized preliminary results on the diagnostic performance of other modalities, which indicated promising alternatives such as whole-body MRI and CT of

thorax and abdomen plus skeletal scintigraphy. However, those crude comparisons can not give an accurate conclusion. Second, the number of included studies was limited, especially for the head-to-head comparison. We excluded studies that enrolled patients with residual/recurrent disease, because previous treatment might influence the diagnostic accuracy of imaging modalities. It was reported that the sensitivity of PET/CT is significantly higher for recurrent NPC than primary disease [31]. Third, three included studies published in Chinese reported that the sensitivity and specificity of PET/CT were 100.0% and 100.0% [8, 11, 12]. Such results were too good to be believed. The small sample size was one of the reasons since two studies only involved less than fifty participants [11, 12]. Although we performed the sensitivity analysis of English publications and obtained stable outcomes, the influence of small sample size may still result into bias for the original analysis.

Conclusion

¹⁸F-FDG PET/CT were significantly better than CWUs in detecting metastasis in primary NPC during initial staging. However, the efficacy of PET/CT in detecting liver metastasis still needs to be optimized. In the future, studies performing cost-benefit analysis of PET/CT and exploring alternative modalities are needed.

Supplementary Material

Supplementary figures and tables.

<http://www.jcancer.org/v08p1238s1.pdf>

Abbreviations

NPC: nasopharyngeal carcinoma; ¹⁸F-FDG: ¹⁸F-fluorodeoxyglucose; PET: positron emission tomography; CT: computed tomography; CWUs: conventional work-ups; MRI: magnetic resonance imaging; EBV: Epstein-Barr virus; DNA: deoxyribonucleic acid; CBMdisc: Chinese Biomedical Disc; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analysis; QUADAS: Quality Assessment tool for Diagnostic Accuracy Studies; LR+: positive likelihood ratios; LR-: negative likelihood ratios; χ^2 : Chi-square; I^2 : I²-square; SROC: summary receiver operating characteristic; AUC: area under the curve.

Acknowledgements

This work was supported by Grants from National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (No. 2014BAI09B10), Science and Technology Project of Guangzhou City, China (No. 14570006), the Planned

Science and Technology Project of Guangdong Province, China (No. 2013B020400004), Health & Medical Collaborative Innovation Project of Guangzhou City, China (No. 201400000001), and National Natural Science Foundation of China (No. 81230056).

Competing Interests

The authors have declared that no competing interest exists.

References

- Chua ML, Wee JT, Hui EP, Chan AT. Nasopharyngeal carcinoma. *Lancet*. 2016; 387: 1012-24.
- Sham JS, Cheung YK, Chan FL, Choy D. Nasopharyngeal carcinoma: pattern of skeletal metastases. *Br J Radiol*. 1990; 63: 202-5.
- Chen YK, Su CT, Ding HJ, Chi KH, Liang JA, Shen YY, et al. Clinical usefulness of fused PET/CT compared with PET alone or CT alone in nasopharyngeal carcinoma patients. *Anticancer Res*. 2006; 26: 1471-7.
- Liu FY, Lin CY, Chang JT, Ng SH, Chin SC, Wang HM, et al. 18F-FDG PET can replace conventional work-up in primary M staging of nonkeratinizing nasopharyngeal carcinoma. *J Nucl Med*. 2007; 48: 1614-9.
- Chua ML, Ong SC, Wee JT, Ng DC, Gao F, Tan TW, et al. Comparison of 4 modalities for distant metastasis staging in endemic nasopharyngeal carcinoma. *Head Neck*. 2009; 31: 346-54.
- Ng SH, Chan SC, Yen TC, Chang JT, Liao CT, Ko SF, et al. Staging of untreated nasopharyngeal carcinoma with PET/CT: comparison with conventional imaging work-up. *Eur J Nucl Med Mol Imaging*. 2009; 36: 12-22.
- Ng SH, Chan SC, Yen TC, Chang JT, Liao CT, Ko SF, et al. Pretreatment evaluation of distant-site status in patients with nasopharyngeal carcinoma: accuracy of whole-body MRI at 3-Tesla and FDG-PET-CT. *Eur Radiol*. 2009; 19: 2965-76.
- Lin S, Li X, Wu H, Lu J, Liang B, Peng X, et al. Efficiency comparison between PET/CT and conventional work-up for evaluating distant metastasis of nasopharyngeal carcinoma. [in Chinese]. *J Clin Otorhinolaryngol Head Neck Surg*. 2012; 26: 529-32.
- Tang LQ, Chen QY, Fan W, Liu H, Zhang L, Guo L, et al. Prospective study of tailoring whole-body dual-modality [18F]fluorodeoxyglucose positron emission tomography/computed tomography with plasma Epstein-Barr virus DNA for detecting distant metastasis in endemic nasopharyngeal carcinoma at initial staging. *J Clin Oncol*. 2013; 31: 2861-9.
- Zhang GY, Wei WH, Li YZ, Xu T, Wu HB, Wang QS, et al. The role of PET-CT in diagnosing distant metastasis of nasopharyngeal carcinoma. [in Chinese]. *Cancer Res Clinic*. 2011; 23: 294-8.
- Wang GH, Lau EW, Shaker R, Binns DS, Hogg A, Drummond E, et al. Clinical application of (18)F-FDG PET/CT to staging and treatment effectiveness monitoring of nasopharyngeal carcinoma. [in Chinese]. *Chin J Cancer*. 2007; 26: 638-42.
- Lin QY, Zhao HG, Zhao JH, Lin CH. Comparison of diagnostic value between 18F-FDG PET/CT and MRI in nasopharyngeal carcinoma. [in Chinese]. *J Jilin Univ (Med Ed)*. 2009; 35: 1163-66.
- Lee N, Xia P, Quivey JM, Sultanem K, Poon I, Akazawa C, et al. Intensity-modulated radiotherapy in the treatment of nasopharyngeal carcinoma: an update of the UCSF experience. *Int J Radiat Oncol Biol Phys*. 2002; 53: 12-22.
- Tham IW, Hee SW, Yeo RM, Salleh PB, Lee J, Tan TW, et al. Treatment of nasopharyngeal carcinoma using intensity-modulated radiotherapy-the national cancer centre singapore experience. *Int J Radiat Oncol Biol Phys*. 2009; 75: 1481-6.
- Sonnenschein C, Soto AM. Cancer Metastases: So Close and So Far. *J Natl Cancer Inst*. 2015; 107: djv236.
- Groheux D, Hindie E, Delord M, Giacchetti S, Hamy AS, de Bazelaire C, et al. Prognostic impact of (18)FDG-PET-CT findings in clinical stage III and IIB breast cancer. *J Natl Cancer Inst*. 2012; 104: 1879-87.
- Apolo AB, Riches J, Schoder H, Akin O, Trout A, Milowsky MI, et al. Clinical value of fluorine-18 2-fluoro-2-deoxy-D-glucose positron emission tomography/computed tomography in bladder cancer. *J Clin Oncol*. 2010; 28: 3973-8.
- Lardinois D, Weder W, Hany TF, Kamel EM, Korom S, Seifert B, et al. Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography. *N Engl J Med*. 2003; 348: 2500-7.
- Gong J, Cao W, Zhang Z, Deng Y, Kang L, Zhu P, et al. Diagnostic efficacy of whole-body diffusion-weighted imaging in the detection of tumour recurrence and metastasis by comparison with 18F-2-fluoro-2-deoxy-D-glucose positron emission tomography or computed tomography in patients with gastrointestinal cancer. *Gastroenterol Rep (Oxf)*. 2015; 3: 128-35.
- Kumar MB, Lu JJ, Loh KS, Chong LM, Soo R, Goh BC, et al. Tailoring distant metastatic imaging for patients with clinically localized undifferentiated nasopharyngeal carcinoma. *Int J Radiat Oncol Biol Phys*. 2004; 58: 688-93.
- Vellayappan BA, Soon YY, Earnest A, Zhang Q, Koh WY, Tham IW, et al. Accuracy of (18)F-fluorodeoxyglucose-positron emission tomography/computed tomography in the staging of newly diagnosed nasopharyngeal carcinoma: a systematic review and meta-analysis. *Radiol Oncol*. 2014; 48: 331-8.
- Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009; [Epub ahead of print].
- Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011; 155: 529-36.
- Delbeke D, Coleman RE, Guiberteau MJ, Brown ML, Royal HD, Siegel BA, et al. Procedure guideline for tumor imaging with 18F-FDG PET/CT 1.0. *J Nucl Med*. 2006; 47: 885-95.
- Berlin JA. Does blinding of readers affect the results of meta-analyses? University of Pennsylvania Meta-analysis Blinding Study Group. *Lancet*. 1997; 350: 185-6.
- Kelsey JL, Evans AS, Thompson WD. *Methods in Observational Epidemiology* (2nd ed). New York, USA: Oxford University Press; 1996.
- Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006; 6: 31.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003; 327: 557-60.
- Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005; 58: 882-93.
- Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. *Med Decis Making*. 1993; 13: 313-21.
- Shen G, Zhang W, Jia Z, Li J, Wang Q, Deng H. Meta-analysis of diagnostic value of 18F-FDG PET or PET/CT for detecting lymph node and distant metastases in patients with nasopharyngeal carcinoma. *Br J Radiol*. 2014; 87: 20140296.
- Vriens D, Visser EP, de Geus-Oei LF, Oyen WJ. Methodological considerations in quantification of oncological FDG PET studies. *Eur J Nucl Med Mol Imaging*. 2010; 37: 1408-25.
- Boellaard R. Standards for PET image acquisition and quantitative data analysis. *J Nucl Med*. 2009; 50 (Suppl 1): S11-S20.
- Adams MC, Turkington TG, Wilson JM, Wong TZ. A systematic review of the factors affecting accuracy of SUV measurements. *AJR American journal of roentgenology*. 2010; 195: 310-20.
- Giancotti FG. Mechanisms governing metastatic dormancy and reactivation. *Cell*. 2013; 155: 750-64.
- Chang JT, Chan SC, Yen TC, Liao CT, Lin CY, Lin KJ, et al. Nasopharyngeal carcinoma staging by (18)F-fluorodeoxyglucose positron emission tomography. *Int J Radiat Oncol Biol Phys*. 2005; 62: 501-7.
- Comoretto M, Balestreri L, Borsatti E, Cimitan M, Franchin G, Lise M. Detection and restaging of residual and/or recurrent nasopharyngeal carcinoma after chemotherapy and radiation therapy: comparison of MR imaging and FDG PET/CT. *Radiology*. 2008; 249: 203-11.
- Yen TC, Chang JT, Ng SH, Chang YC, Chan SC, Lin KJ, et al. The value of 18F-FDG PET in the detection of stage M0 carcinoma of the nasopharynx. *J Nucl Med*. 2005; 46: 405-10.
- Liu FY, Chang JT, Wang HM, Liao CT, Kang CJ, Ng SH, et al. [18F]fluorodeoxyglucose positron emission tomography is more sensitive than skeletal scintigraphy for detecting bone metastasis in endemic nasopharyngeal carcinoma at initial staging. *J Clin Oncol*. 2006; 24: 599-604.
- Ohno Y, Koyama H, Onishi Y, Takenaka D, Nogami M, Yoshikawa T, et al. Non-small cell lung cancer: whole-body MR examination for M-stage assessment--utility for whole-body diffusion-weighted imaging compared with integrated FDG PET/CT. *Radiology*. 2008; 248: 643-54.