

FDG-PET for diagnosing prosthetic joint infection: systematic review and metaanalysis

Thomas C. Kwee · Robert M. Kwee · Abass Alavi

Published online: 15 August 2008
© The Author(s) 2008

Abstract

Purpose The aim of this study was to systematically review and metaanalyze published data on the diagnostic performance of ^{18}F -fluoro-2-deoxyglucose positron emission tomography (FDG-PET) in detecting prosthetic hip or knee joint infection.

Methods A systematic search for relevant studies was performed of the PubMed/MEDLINE and Embase databases. Two reviewers independently assessed the methodological quality of each study. A metaanalysis of the reported sensitivity and specificity of each study was performed. Subgroup analyses were performed if results of individual studies were heterogeneous.

Results The inclusion criteria were met by 11 studies; there was a total sample size of 635 prostheses. Overall, the studies had good methodological quality. Pooled sensitivity and specificity of FDG-PET for the detection of prosthetic hip or knee joint infection were 82.1% (95%CI=68.0–90.8%) and 86.6% (95%CI=79.7–91.4%), respectively. Heterogeneity among the results of individual studies was present ($I^2=68.8\%$). Diagnostic performance was influ-

enced by type of joint prostheses (hip prostheses vs. knee prostheses) and type of reconstruction method used (filtered back vs. iterative) ($p=0.0164$ and $p=0.0235$, respectively). **Conclusion** In this metaanalysis, overall diagnostic performance of FDG-PET was moderate to high. Caution is warranted, however, because results of individual studies were heterogeneous and could not be fully explored. Future studies should further explore potential causes of heterogeneity and validate the use of FDG-PET for diagnosing prosthetic joint infection.

Keywords FDG-PET · Prosthesis · Arthroplasty · Hip · Knee · Infection

Introduction

Periprosthetic infection following total hip or knee arthroplasty is associated with significant morbidity and costs [1–3]. The infection rates following primary implantation and revision surgery are approximately 1% and 3% for hip prostheses and 2% and 5% for knee prostheses, respectively [4]. Differentiating prosthetic joint infection from aseptic loosening is of crucial importance for appropriate patient management; the treatment of an infected joint prosthesis generally involves both systemic antibiotics for an extended period and exchange arthroplasty in one or two stages, whereas aseptic loosening usually requires a single revision arthroplasty [1, 2]. Diagnosing prosthetic joint infection is difficult; clinical signs and symptoms, laboratory tests, radiography, and joint aspiration are insensitive, nonspecific, or both [5]. In addition, cross-sectional imaging modalities such as CT and MRI are hampered by artifacts produced by the prosthetic devices themselves [5]. Radio-nuclide imaging is less affected by metallic implants and

T. C. Kwee (✉)
Department of Radiology, University Medical Center Utrecht,
Heidelberglaan 100,
3584 CX Utrecht, The Netherlands
e-mail: thomaskwee@gmail.com

R. M. Kwee
Department of Radiology, University Medical Center Maastricht,
Maastricht, The Netherlands

A. Alavi
Division of Nuclear Medicine,
Hospital of the University of Pennsylvania,
Philadelphia, PA, USA

may be more useful [5]. Combined leukocyte–marrow scintigraphy has been reported to achieve a diagnostic accuracy of 90% or greater and is currently regarded as the imaging modality of choice for diagnosing prosthetic joint infection [5]. However, combined leukocyte–marrow scintigraphy is labor-intensive, time-consuming, not widely available, and potentially hazardous because of direct handling of blood products [5]. ¹⁸F-fluoro-2-deoxyglucose positron emission tomography (FDG-PET) enables visualization of hyperglycolytic inflammatory cells (leukocytes, macrophages, and other immunologically active cells) during infection; it may be an attractive alternative to combined leukocyte–marrow scintigraphy because it requires only one injection and scan and is more widely available [5]. Furthermore, treatment with antibiotics is not likely to affect the sensitivity of FDG-PET in delineating sites of infections because FDG does not rely on leukocyte migration, in contrast to combined leukocyte–marrow scintigraphy. However, controversial results have been reported on the diagnostic value of FDG-PET in detecting prosthetic joint infection and its utility is still under debate. The purpose of this study was, therefore, to systematically review and metaanalyze published data on the diagnostic performance of FDG-PET in detecting prosthetic hip or knee joint infection and to provide more insight into the causes of the controversial results in the literature.

Materials and methods

Search strategy

A computer-aided search of the PubMed/MEDLINE and Embase databases was conducted to find relevant published articles on the diagnostic performance of FDG-PET in detecting prosthetic hip or knee joint infection. The search strategy is presented in Table 1. No beginning date limit was used. The search was updated until 27 May 2008. To expand our search, bibliographies of articles which finally remained after the selection process were screened for potentially suitable references.

Study selection

Studies investigating the diagnostic performance of FDG-PET in detecting prosthetic hip or knee joint infection were eligible for inclusion. All reference standards used in the individual studies were accepted; however, when FDG-PET itself was part of the reference standard, the study was excluded. No language restriction was applied. Review articles, metaanalyses, abstracts, editorials or letters, case reports, guidelines for management, studies examining 15 or fewer patients with hip and/or knee prosthesis, studies performed in animals, and ex vivo studies were excluded. Studies that examined FDG with a gamma camera in coincidence mode were also excluded. Studies which provided insufficient data to construct a 2×2 contingency table to calculate sensitivity and specificity for detecting prosthetic hip or knee joint infection were excluded. When data were presented in more than one article, the article with the largest number of patients or the article with the most details was chosen.

Two researchers (T.C.K., R.M.K.) independently reviewed the titles and abstracts of the retrieved articles, applying the inclusion and exclusion criteria mentioned above. Articles were rejected if they were clearly ineligible. The same two researchers then independently reviewed the full-text version of the remaining articles to determine their eligibility for inclusion. Disagreements were resolved in a consensus meeting.

Study quality

The methodological quality of the included studies was assessed in terms of the potential for bias (internal validity) and lack of generalizability (external validity). For this purpose, a checklist adapted from Kelly et al. [6] and Whiting et al. [7, 8] was used. The complete criteria list is presented in Table 2. Internal validity criteria and external validity scores were scored as positive (adequate methods) or negative (inadequate methods, potential bias). If insufficient information was provided on a specific item, a negative score was given. Two reviewers (T.C.K., R.M.K.)

Table 1 Search strategy and results as on 27 May 2008

No.	Search string	PubMed/MEDLINE	Embase
1	Fluorodeoxyglucose or 2-fluoro-2-deoxy-D-glucose or FDG or positron emission tomography or positron-emission tomography or PET	41,327	47,571
2	Arthroplasty or arthroplasties or arthroplastic or prosthesis or prostheses or prosthetic or endoprosthesis or endoprostheses or endoprosthetic	309,295	103,843
3	Infection or infectious or infected or septic or septicallly	1,078,831	922,906
4	No. 1 and no. 2 and no. 3	98	104

Table 2 Criteria list used to assess the methodological quality of the studies

	Criteria of validity	Positive score
Internal validity	Prospective study	Mentioned in publication
	Avoidance of withdrawal bias	<10% of patients who were examined by the index test did not undergo the reference test
	Avoidance of study examination bias	<10% of indeterminate or uninterpretable results
	Avoidance of diagnostic review bias	Blind interpretation of index test without knowledge of reference test
	Avoidance of test review bias	Blind interpretation of reference test without knowledge of index test
External validity	Avoidance of comparator review bias	Blinding index test to the other imaging modality, if more than one imaging modality was investigated
	Avoidance of spectrum bias	Only prostheses suspected of being infected were included (symptomatic prostheses only)
	Demographic information	Study location (country), age, and sex of patients reported
	Avoidance of selection bias	Consecutive series of patients or random selection of patients
	Standard execution of index test	Application of the same hardware and imaging protocol in all patients
	Avoidance of observer variability bias	Interpreter(s) of index test described

independently assigned the scores. Disagreements between the two researchers were discussed and resolved by consensus. Subtotals were calculated for internal (maximum six) and external (maximum five) validity separately. Total quality scores were expressed as a percentage of the maximum score.

Data analysis

Sensitivities and specificities of FDG-PET for the detection of prosthetic hip or knee joint infection (with corresponding 95% CIs) were calculated from the original numbers given in the included studies. Similarly, diagnostic odds ratios (DORs) of individual studies were calculated. The DOR is a single overall indicator of diagnostic performance and is, unlike sensitivity and specificity, independent of any threshold (cutoff) value [9]. In order to enable calculation of the DOR, a standard correction of adding 0.5 to all cells of the 2×2 contingency table was applied if the true-positive rate, false-positive rate, false-negative rate, or true-negative rate was zero. DORs of included original studies were displayed using forest plots.

Metaanalysis was performed using a bivariate random effects approach to pool the sensitivity and specificity [10]. This model assumes a bivariate normal distribution for the logit-transformed sensitivity and specificity values across studies, allowing for heterogeneity beyond chance due to clinical or methodological differences between studies. It incorporates and estimates the correlation that might exist between estimates of sensitivity and specificity within studies. A standard correction of adding 0.5 to all cells of the 2×2 contingency table was applied if the true-positive rate, false-positive rate, false-negative rate, or true-negative rate was zero. Estimates of the mean logit-transformed sensitivity and specificity were then obtained. Pooled

estimates of sensitivity and specificity with 95% CIs were calculated after antilogarithm transformation of these logit estimates. To improve visualization of the results, the 95% coverage region of the estimated bivariate distribution of the logit sensitivity and specificity was transformed back to receiver operating characteristic (ROC) axes [10]. Results of the included studies were also plotted in ROC space.

Heterogeneity among the results of individual studies was tested by subjecting the DORs of individual studies to the Higgins and Thompson test, calculating the I^2 statistic [11]. If the DOR is equal across studies, the only cause of heterogeneity is a difference in cutoff levels for prosthetic joint infection. If the DOR varies across studies, factors other than cutoff differences exist as well [9]. Heterogeneity was defined as $I^2 > 50\%$. Potential sources for heterogeneity were explored by subgroup analysis. Covariates analyzed were: study design (reported prospective study design vs. no or unreported prospective study design), way of patient recruitment (consecutive or random selection of patients vs. nonconsecutive, nonrandom selection, or unreported way of recruitment), patient spectrum (inclusion of only symptomatic prostheses vs. inclusion of both symptomatic and asymptomatic prostheses), type of joint prostheses (hip prostheses only vs. knee prostheses only), age of prostheses (only inclusion of prostheses older than 6 months vs. prostheses younger than 6 months were [also] included), reconstruction method (iterative reconstruction vs. filtered back projection), type of PET images reviewed (non-attenuation-corrected [NAC] images only or both NAC and attenuation-corrected [AC] images vs. AC images only), and way of image review (reported blinding to reference test vs. no or unreported blinding to reference test). Another important issue that requires subgroup analysis is the use of different criteria to diagnose prosthetic joint infection. Applied criteria for positivity can grossly be

divided into four groups: (a) FDG uptake in the periprosthetic soft tissue; (b) increased FDG uptake at the bone–prosthetic interface (BPI); (c) increased FDG uptake at the BPI, while emphasizing that FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis is not considered suggestive of infection (for hip prostheses only); and (d) other criteria. With regard to these criteria of positivity, subgroup analyses were performed as follows: (a) vs. (b, c, or d), (b or c) vs. (a or d), (b) vs. (a, c, or d), and (c) vs. (a, b, or d) (for studies or subsets in studies on hip prostheses only). Each of the predefined covariates was separately included in the bivariate model to compare the overall sensitivity and overall specificity between different strata, using a *z* test with the level of statistical difference set at 0.05.

Statistical analyses were executed using Meta-DiSc statistical software version 1.4 (Unit of Clinical Biostatistics, Ramón y Cajal Hospital, Madrid, Spain) and SAS statistical software package version 9.1.3 (SAS Institute, Cary, NC, USA).

Results

Literature search

The computer-aided search revealed 98 articles from PubMed/MEDLINE and 104 articles from Embase (Table 1). Reviewing titles and abstracts from PubMed/MEDLINE revealed 20 articles potentially eligible for inclusion [12–31]. Reviewing titles and abstracts from Embase revealed 17 articles potentially eligible for inclusion, which were all already identified by the PubMed/MEDLINE search. Thus, 20 studies remained for possible inclusion and were retrieved in full-text version. After reviewing the full article, five articles [18, 21, 22, 25, 26] were excluded

because the same data were used in another article comprising a larger number of patients, one article was excluded because it did not investigate the diagnostic performance of FDG-PET in detecting prosthetic hip or knee joint infection [14], one article [16] was excluded because the same data were used in another article providing more study details, one article [29] was excluded because less than 15 patients with hip and/or knee prosthesis were investigated, and one article [31] was excluded because it appeared to be an abstract only. Screening references of the remaining articles resulted in one other potentially relevant article [32]. However, this article was excluded because less than 15 patients with hip and/or knee prostheses were investigated [32]. Thus, eventually 11 studies [12, 13, 15, 17, 19, 20, 23, 24, 27, 28, 30], comprising a total sample size of 635 prostheses, met all inclusion and exclusion criteria, and they were included in this systematic review. The characteristics of the included studies are presented in Tables 3, 4, and 5.

Methodological quality assessment

Methodological quality was assessed by 11 items. The scores for internal and external validity are presented in Table 6. The total score for combined internal and external validity, expressed as a fraction of the maximum score, ranged from 45% to 91% (median 82%).

Diagnostic performance

The results of the 11 included studies are presented in Table 7, their DORs are displayed in Fig. 1, and the corresponding ROC plot is displayed in Fig. 2. Stumpe et al. [20] provided two results (Table 5), but only the first result of their study was used for metaanalysis and assessment of heterogeneity. Sensitivity and specificity of

Table 3 Patient characteristics of included studies

Study and year	Country	No. of patients	Mean age in years (range)	Sex (M/F)	No. of prostheses	Age of prostheses
Chryssikos et al. [12], 2008	USA	113	59 (31–87)	54:59	127 (H)	12, 18, and 24 months
Garcia-Barrecheuren et al. [13], 2007	Spain	24	68 (37–81)	12:12	24 (H)	>6 months
Pill et al. [15], 2006	USA	89	NR (29–85)	NR	92 (H)	NR
Delank et al. [17], 2006	Germany	27	NR (45–82)	NR	36 (H+K)	0.8–19.4 years (<i>n</i> =27); NR (<i>n</i> =9)
Reinartz et al. [19], 2005	Germany	63	68 (43–88)	32:31	92 (H)	1–31 years
Stumpe et al. [20], 2004	Switzerland	35	69 (46–89)	23:12	35 (H)	12–260 months
Chacko et al. [23], 2003	USA	NR	NR	NR	53 (H)+36 (K)	NR
Vanquickenborne et al. [24], 2003	Belgium	17	NR (42–77)	8:9	17 (H)	2–163 months
Manthey et al. [27], 2002	Germany	23	70 (35–83)	9:14	14 (H)+14(K)	NR
Van Acker et al. [28], 2001	Belgium	21	66 (33–78)	8:13	21 (K)	7 months–9 years
Zhuang et al. [30], 2001	USA	62	NR (27–81)	NR	38 (H)+36 (K)	3 months–8 years

H hip prostheses, K knee prostheses, NR not reported

Table 4 FDG-PET parameters and image interpretation of included studies

Study and year	FDG dose	Time interval between FDG administration and scanning, acquisition time for emission scans	Reconstruction method	Review of AC and/or NAC images	Criteria for positivity	Interpreters
Chryssikos et al. [12], 2008	5.18 MBq/kg (maximum 370 MBq)	60 min, <30 min	NR	NR	Abnormally increased FDG uptake at the BPI (FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis was not considered suggestive of infection)	Three experienced observers
Garcia-Barrecheuren et al. [13], 2007	6.2 MBq/kg	≥40 min, 10 min per bed position	IR	AC	FDG uptake in areas of normality with intensity much higher than the synovial structures or adjacent soft tissues or uptake at the BPI with intensity greater than the synovial structures or adjacent soft tissues FDG uptake at the BPI with intensity much higher than the synovial structures or adjacent soft tissues or fistulous tract uptake	Two independent observers
Pill et al. [15], 2006	5.2 MBq/kg	60 min, NR	NR	NR	Abnormally increased FDG uptake at the BPI (FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis was not considered suggestive of infection)	Experienced observers
Delank et al. [17], 2006	370 MBq	60 min, NR	FBP	NR	Uptake of FDG in the periprosthetic soft tissue	Three PET investigators
Reinartz et al. [19], 2005	283±38 MBq	58±8 min, 12 min per bed position	IR	NR	Uptake of FDG in the periprosthetic soft tissue	Two experienced and board-certified nuclear medicine physicians
Stumpe et al. [20], 2004	300–400 MBq	30–40 min, NR	IR	AC and NAC	Diffusely increased FDG uptake of grade 3 or 4 (on a scale from 0 to 4) along the BPI both on AC and NAC images (focally increased FDG uptake of grade 1 or 2 was considered to indicate loosened total hip replacement)	Two independent board-certified experienced nuclear physicians
Chacko et al. [23], 2003	2.52 MBq/kg	60 min, NR	IR	NR	Abnormally increased FDG uptake at the BPI (FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis was not considered suggestive of infection)	Two readers
Vanquickenborne et al. [24], 2003	370 MBq	60 min, NR	IR	NR	FDG uptake of grade 2 or higher (on a scale from 0 to 3) with a pattern different from that observed in the control group	Two experienced nuclear medicine specialists
Manthey et al. [27], 2002	190–220 MBq (mean 200 MBq)	50 min, 7 min per bed position	FBP	NAC	Highly increased FDG uptake at the BPI	Two independent experienced readers
Van Acker et al. [28], 2001	3.7 MBq/kg	60 min, 10 min per bed position	IR	AC and NAC	Focal FDG uptake at the BPI on NAC images	Two nuclear medicine specialists
Zhuang et al. [30], 2001	4.22–4.56 MBq/kg	60 min, NR	IR	NR	Area of increased FDG uptake (compared with adjacent soft tissue) at the BPI (for hip prostheses: FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis was not considered suggestive of infection)	Two observers

AC attenuation-corrected, BPI bone-prosthesis interface, FBP filtered back projection, FDG ¹⁸F-fluoro-2-deoxyglucose, H hip prostheses, IR iterative reconstruction, K knee prostheses, NAC nonattenuation-corrected, NR not reported, PET positron emission tomography

Table 5 Reference standards used in the individual studies

Study and year	Reference standard
Chryssikos et al. [12], 2008	Sepsis was confirmed if the patient met at least one of the following three criteria: An open wound or sinus in communication with the joint A systemic infection with pain in the hip and purulent fluid within the joint A positive result on at least three tests (ESR [≥ 25 mm/h], CRP [>0.9 mg/dL], joint aspiration, intraoperative frozen section, and intraoperative culture)
Garcia-Barrecheuren et al. [13], 2007	Prostheses were considered infected if: The same microorganism grew in at least two cultures obtained by aspiration and/or debridement The microorganism grew in a single culture with one of these three situations: Fistula with active drainage Purulent drainage at the time of debridement Evidence of acute or subacute purulent inflammation on biopsy of intraarticular tissue Poor presurgical evolution of the prosthesis within 1 year without a justified cause and definitive histopathological diagnosis of acute purulent inflammation postsurgery Poor postsurgical evolution of the prosthesis within 1 year without a justified cause and definitive histopathological diagnosis of acute purulent inflammation postsurgery
Pill et al. [15], 2006	Intraoperative histology and cultures
Delank et al. [17], 2006	Histological and microbiological examination of the interface tissue between bone and loosened prostheses, and intraoperative macroscopic findings
Reinartz et al. [19], 2005	Intraoperative findings, histological examination and microbiological cultures, or clinical follow-up (clinical assessments, plain radiography, and laboratory tests) 9–18 months. Arthroplasties which did not require revision or treatment with antibiotics during the follow-up period were considered to be uninfected
Stumpe et al. [20], 2004	In patients who underwent surgery, prostheses were considered infected if microorganisms were found in cultures or if local abscess formation or neutrophilic granulocytes were present. In the remaining patients, the diagnosis of infection was based on results of joint aspiration together with clinical follow-up ≥ 6 months. Prostheses were considered uninfected in patients with negative microbiologic results after joint aspiration; normal erythrocyte sedimentation rate, C-reactive protein level, and white blood cell count; and improvement in their clinical symptoms ≥ 6 months
Chacko et al. [23], 2003	Histopathological analysis of tissues obtained during revision arthroplasty or surgical findings or clinical follow-up ≥ 6 months
Vanquickenborne et al. [24], 2003	Cultures obtained during surgery or clinical follow-up ≥ 6 months
Manthey et al. [27], 2002	Operative findings or clinical follow-up up to 24 months
Van Acker et al. [28], 2001	Cultures obtained during surgery, clinical follow-up ≥ 6 months, or preoperative joint aspiration and culture
Zhuang et al. [30], 2001	Prostheses were considered infected if aspiration cultures grew organisms or if infection was verified at surgery. Prostheses were considered uninfected if an operative smear revealed no leukocytes and if intraoperative cultures obtained from suspected sites during surgery revealed no growth. Prostheses that did not require surgical exploration during clinical follow-up for 1 year were considered uninfected

Table 6 Quality assessment of included studies

Study and year	Criteria										Total scores		Percentage of maximum score	
	IV					EV					IV	EV		
	1	2	3	4	5	6	1	2	3	4				5
Chryssikos et al. [12], 2008	+	+	+	+	-	+	+	+	+	+	-	5	4	82
Garcia-Barrecheuren et al. [13], 2007	+	+	+	+	-	+	+	+	-	+	+	5	4	82
Pill et al. [15], 2006	+	+	+	+	-	+	+	-	-	+	+	5	3	73
Delank et al. [17], 2006	+	+	+	+	-	+	-	-	-	+	+	5	2	64
Reinartz et al. [19], 2005	+	+	+	+	-	+	-	+	+	+	+	5	4	82
Stumpe et al. [20], 2004	+	+	+	+	-	+	+	+	+	+	+	5	5	91
Chacko et al. [23], 2003	-	+	+	-	-	+	+	-	-	-	+	3	2	45
Vanquickenborne et al. [24], 2003	+	+	+	+	-	+	+	+	-	+	+	5	4	82
Manthey et al. [27], 2002	-	+	+	-	-	+	+	+	+	+	+	3	5	73
Van Acker et al. [28], 2001	+	+	+	+	-	-	+	+	+	+	+	4	5	82
Zhuang et al. [30], 2001	-	+	+	+	-	+	+	-	-	-	+	4	2	55

FDG-PET for the detection of prosthetic hip or knee joint infection ranged from 22.2% to 100% and from 61.5% to 100% with pooled estimates of 82.1% (95%CI=68.0–90.8%) and 86.6% (95%CI=79.7–91.4%), respectively. Heterogeneity among the DORs of individual studies was present ($I^2=68.8\%$). Overall specificity of FDG-PET in hip prostheses was significantly higher than that in knee prostheses (89.8% vs. 74.8%, $p=0.0164$). Overall specificity of studies using filtered back projection was significantly higher than that of studies using iterative reconstruction (98.3% vs. 82.3%, $p=0.0235$). No statistically significant differences were observed in sensitivities and/or specificities within the subgroups study design (reported prospec-

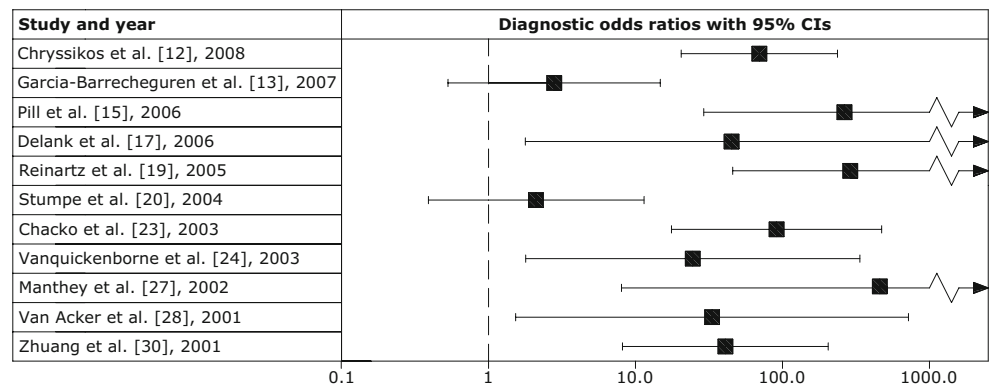
tive study design vs. no or unreported prospective study design), way of patient recruitment (consecutive or random selection of patients vs. nonconsecutive, nonrandom selection or unreported way of recruitment), patient spectrum (inclusion of only symptomatic prostheses vs. prostheses younger than 6 months were also included), age of prostheses (only inclusion of prostheses older than 6 months vs. prostheses younger than 6 months were [also] included), type of PET images reviewed (NAC images only or both NAC and AC images vs. AC images only), way of image review (reported blinding to reference test vs. no or unreported blinding to reference test), and criteria for positivity (four different comparisons) (Table 8).

Table 7 Results of included studies

Study and year	Sensitivity (%)		Specificity (%)	
	Value	95%CI	Value	95%CI
Chryssikos et al. [12], 2008	84.9	69.1–93.4	92.6	85.4–96.4
Garcia-Barrecheuren et al. [13], 2007	63.6	35.4–84.8	61.5	35.5–82.3
Pill et al. [15], 2006	95.2	77.3–99.2	93.0	84.6–97.0
Delank et al. [17], 2006	40.0	11.8–76.9	100	89.0–100
Reinartz et al. [19], 2005	93.9	80.4–98.3	94.9	86.1–98.3
Stumpe et al. [20], 2004	33.3 ^a 22.2 ^b	12.1–64.6 ^a 6.3–54.7 ^b	80.8 ^a 84.6 ^b	62.1–91.5 ^a 66.5–93.9 ^b
Chacko et al. [23], 2003	91.7	74.2–97.7	89.2	79.4–94.7
Vanquickenborne et al. [24], 2003	87.5	52.9–97.8	77.8	45.3–93.7
Manthey et al. [27], 2002	100	51.0–100	100	86.7–100
Van Acker et al. [28], 2001	100	61.0–100	73.3	48.1–89.1
Zhuang et al. [30], 2001	90.5	71.1–97.4	81.1	68.6–89.4
Pooled estimate	84.6	71.0–92.5	84.0	68.0–92.8

^a Reader 1^b Reader 2

Fig. 1 Forest plot with diagnostic odds ratios of included original studies (logarithmic scale)



Discussion

This systematic review and metaanalysis included 11 studies comprising a total sample size of 635 prostheses. Overall methodological quality of included studies was good. Metaanalytically, FDG-PET achieves moderate to high sensitivity and specificity in detecting prosthetic hip or knee joint infection. However, this result should be interpreted cautiously because significant heterogeneity was identified among the results of individual studies. Several causes may underlie this heterogeneity and explain the controversial results in the literature. Subgroup analysis revealed that overall specificity of FDG-PET in hip prostheses was significantly higher than that in knee prostheses, and overall specificity of studies using filtered back projection was (inexplicably) significantly higher than

that of studies using iterative reconstruction (Table 8). The lower specificity of FDG-PET in knee prostheses may be related to the relatively limited knowledge about the incidence and pattern of nonspecific FDG uptake around knee prostheses. Zhuang et al. [33] reported that increased FDG uptake around the femoral head and neck (possibly due to foreign body reaction to the material of the prosthetic joint) may persist for years following hip arthroplasty and can occur in both symptomatic and asymptomatic patients; it should not be interpreted as periprosthetic infection. Increased FDG uptake around the distal tip of the hip prosthesis is also nonspecific. However, FDG uptake along the interface between bone and hip prosthesis is virtually never seen in asymptomatic patients or in those with aseptic loosening and is, therefore, highly suggestive of infection [33]. Persistently increased nonspecific FDG uptake following knee arthroplasty has also been mentioned [33], but should be further investigated. More knowledge about the incidence and pattern of nonspecific FDG uptake around knee prostheses may improve the specificity of FDG-PET in detecting prosthetic knee joint infection. Despite the findings of Zhuang et al. [33], our subgroup analysis did not reveal any significantly higher sensitivity or specificity among studies which used FDG uptake at the BPI as criterion for positivity, while emphasizing that FDG uptake limited to the soft tissues or adjacent to the neck of the prosthesis was not considered suggestive of infection (Table 8). Metallic prosthetic material can cause artifacts on attenuation-corrected FDG-PET images and may also affect diagnostic performance. Goerres et al. [34] reported that the use of attenuation correction (both ⁶⁸Ge-based and CT-based) generates artifacts of apparently increased FDG concentration around metallic hip implants. The shape of the prosthesis, the absorption properties of the surrounding tissues, and the method of transmission scanning (⁶⁸Ge-based or CT-based) influence the appearance of such artifacts. It should be noted that all evidence regarding the diagnostic performance of FDG-PET in prosthetic joint infection has been acquired using stand-alone PET scanners, which use a

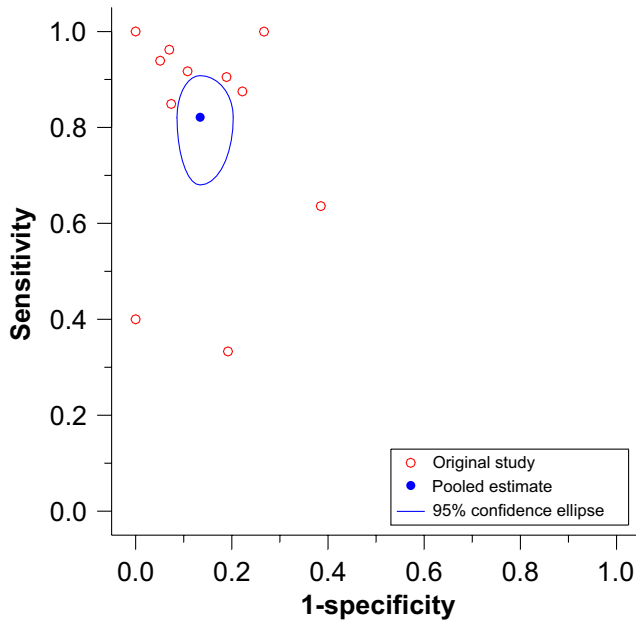


Fig. 2 ROC plot with pooled sensitivity and specificity (including 95% confidence ellipses) and results of included original studies for the detection of prosthetic hip or knee joint infection using FDG-PET

Table 8 Results of bivariate analysis with covariates

Study characteristic	No. of studies	Sensitivity		Specificity	
		Pooled value (%)	1 vs. 2 (<i>p</i> value)	Pooled value (%)	1 vs. 2 (<i>p</i> value)
1. Reported prospective study design	8	77.0	0.1629	86.3	0.8335
2. No or unreported prospective study design	3	91.7		87.7	
1. Consecutive or random selection of patients	5	77.6	0.6073	89.8	0.1970
2. Nonconsecutively, nonrandom selection, or unreported method of recruitment	6	84.0		81.8	
1. Inclusion of only symptomatic prostheses	9	83.4	0.7527	85.1	0.0955
2. Inclusion of both symptomatic and asymptomatic prostheses	2	78.2		95.2	
1. Hip prostheses only	9	82.6	0.3924	89.8	0.0164
2. Knee prostheses only	4	90.4		74.8	
1. Only inclusion of prostheses older than 6 months	4	70.2	0.3915	80.3	0.3103
2. Prostheses younger than 6 months were (also) included	4	84.2		89.2	
1. Iterative reconstruction method	7	82.0	0.4112	82.3	0.0235
2. Filtered back projection method	2	63.0		98.3	
1. Review of NAC images only or both NAC and AC images	3	73.7	0.8235	79.0	0.2021
2. Review of AC images only	1	63.6		61.5	
1. Reported blind interpretation of FDG-PET to reference test	9	79.6	0.2203	85.2	0.3461
2. No or unreported blind interpretation of FDG-PET to reference test	2	91.6		92.0	
1. Criterion (a) for positivity only	2	78.2	0.7527	95.2	0.0955
2. Criteria (b), (c), or (d) for positivity	9	83.4		85.1	
1. Criteria (b) or (c) for positivity	6	85.1	0.5209	87.5	0.7355
2. Criteria (a) or (d) for positivity	5	76.9		85.1	
1. Criterion (b) for positivity only	3	66.8	0.2650	83.0	0.5410
2. Criteria (a), (c), or (d) for positivity	8	85.0		87.7	
1. Criterion (c) for positivity only (for studies or subsets in studies on hip prostheses only)	4	89.7	0.2224	93.1	0.0887
2. Criteria (a), (b), or (d) for positivity (for studies or subsets in studies on hip prostheses only)	5	76.3		84.6	

Applied criteria for positivity divided into four groups: (a) FDG uptake in the periprosthetic soft tissue, (b) increased FDG uptake at the BPI, (c) increased FDG uptake at the BPI, while emphasizing that FDG uptake limited to the soft tissues adjacent to the neck of the prosthesis is not considered suggestive of infection (for hip prostheses only), (d) other criteria

AC attenuation-corrected, BPI bone–prosthesis interface, FDG ¹⁸F-fluoro-2-deoxyglucose, NAC nonattenuation-corrected, PET positron emission tomography

radionuclide source for attenuation correction. Combined PET/CT is replacing the stand-alone PET scanner in clinical practice, but may perform differently because it uses CT-based attenuation correction; this important issue should be further investigated. Goerres et al. [34] further reported that patient movement worsens attenuation artifacts, whereas attenuated-weighted iterative reconstruction appears to reduce the visibility of artifacts [34]. The presence of artifacts on attenuation-corrected images has also been observed in knee prostheses; in a phantom study, Van Acker et al. [28] showed that artifacts mimicking FDG uptake adjacent to a knee prosthesis can arise in attenuation-corrected images obtained with different methods of image reconstruction. In addition, Heiba et al. [35] reported the observation of an artifact within the joint space of total knee metallic prostheses in two patients on attenuation-

corrected images. No uptake, however, was noted in the same location on the nonattenuation-corrected images [35]. Thus, verification of attenuation-corrected images against nonattenuation-corrected images may avoid false-positive results because of the abovementioned reasons. However, our subgroup analysis did not reveal any significantly lower sensitivity or specificity in the study which exclusively evaluated attenuation-corrected images (Table 8). In addition, no statistically significant differences in diagnostic performance were observed in the subgroup analyses according to study design, way of patient recruitment, patient spectrum, age of prostheses, and way of image review (Table 8). It should be noted, however, that results from our subgroup analysis may not be conclusive because of the relatively small number of included studies. Furthermore, it was not possible to perform subgroup

analyses according to FDG dose, time interval between FDG administration and scanning, acquisition time for emission scans, number and experience of interpreters, reference standard used, and way of interpreting the reference test because no (meaningful) stratifications could be made of the available data of included studies. A large multicenter study is required to further investigate potential sources of heterogeneity and validate the use of FDG-PET for diagnosing prosthetic joint infection. Another drawback of this metaanalysis is the use of different (imperfect) reference standards in the individual studies (Table 5), which may have lead to misclassification bias and may have affected the estimates of diagnostic performance of FDG-PET. However, because no perfect reference test exists yet for detecting prosthetic joint infection and all studies used a combination of reference standards (Table 5), we accepted this shortcoming.

Combined leukocyte–marrow scintigraphy is currently regarded as the imaging modality of choice for diagnosing prosthetic joint infection [5]. Two studies made a direct comparison between FDG-PET and combined leukocyte–marrow scintigraphy [15, 28]. Pill et al. [15] investigated 89 patients for revision of painful hip prosthesis. Of the 89 patients, 46 underwent both FDG-PET and combined leukocyte–marrow scintigraphy for a total of 51 hip prostheses. Although FDG-PET and combined leukocyte–marrow scintigraphy demonstrated comparable specificities (93% and 95.1%, respectively), FDG-PET exhibited a substantially higher sensitivity (95.2% and 50%, respectively) [15]. Van Acker et al. [28] investigated 21 patients with a painful knee arthroplasty. All patients underwent FDG-PET and 20 of 21 patients underwent combined leukocyte–marrow scintigraphy. Sensitivity and specificity of FDG-PET were 100% and 73%, respectively, and sensitivity and specificity of combined leukocyte–marrow scintigraphy were 100% and 93%, respectively [28]. Based on this small number of studies [15, 28], however, no definite conclusion can be drawn yet on the diagnostic performance of FDG-PET compared to that of combined leukocyte–marrow scintigraphy.

Antigranulocyte scintigraphy (AGS) with monoclonal antibodies or antibody fragments may be another attractive approach to detect prosthetic joint infection [36–38]. Unlike combined leukocyte–marrow scintigraphy, which requires time-consuming and potentially dangerous *in vitro* labeling of autologous leukocytes, AGS allows *in vivo* labeling of granulocytes in the inflamed tissue surrounding the prosthesis [36–38]. A recent metaanalysis on the diagnostic performance of AGS included 13 studies with a total sample size of 522 prostheses and reported independent random effects summary estimates of sensitivity and specificity of 83% and 80%, respectively [38]. Future studies are required to compare the diagnostic performance

of combined leukocyte–marrow scintigraphy, FDG-PET, and AGS and to assess which imaging modality is most cost-effective.

In conclusion, in this metaanalysis, overall diagnostic performance of FDG-PET was moderate to high. Caution is warranted, however, because results of individual studies were heterogeneous and could not be fully explored. Future studies should further explore causes of heterogeneity and validate the use of FDG-PET for diagnosing prosthetic joint infection.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Toms AD, Davidson D, Masri BA, Duncan CP. The management of peri-prosthetic infection in total joint arthroplasty. *J Bone Joint Surg Br* 2006;88:149–55.
2. Barberán J. Management of infections of osteoarticular prosthesis. *Clin Microbiol Infect* 2006;12(Suppl 3):93–101.
3. Bozic KJ, Ries MD. The impact of infection after total hip arthroplasty on hospital and surgeon resource utilization. *J Bone Joint Surg Am* 2005;87:1746–51.
4. Hanssen AD, Rand JA. Evaluation and treatment of infection at the site of a total hip or knee arthroplasty. *Instr Course Lect* 1999;48:111–22.
5. Love C, Tomas MB, Marwin SE, Pugliese PV, Palestro CJ. Role of nuclear medicine in diagnosis of the infected joint replacement. *Radiographics* 2001;21:1229–38.
6. Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, et al. The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol* 1997;70:1028–35.
7. 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.
8. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol* 2006;6:9.
9. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol* 2003;56:1129–35.
10. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol* 2005;58:982–90.
11. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
12. Chryssikos T, Parvizi J, Ghanem E, Newberg A, Zhuang H, Alavi A. FDG-PET imaging can diagnose periprosthetic infection of the hip. *Clin Orthop Relat Res* 2008;466:1338–42.
13. García-Barrecheguren E, Rodríguez Fraile M, Toledo Santana G, Valentí Nín JR, Richter Echevarría JA. FDG-PET: a new diagnostic approach in hip prosthetic replacement. *Rev Esp Med Nucl* 2007;26:208–20.

14. Stumpe KD, Romero J, Ziegler O, Kamel EM, von Schulthess GK, Strobel K, et al. The value of FDG-PET in patients with painful total knee arthroplasty. *Eur J Nucl Med Mol Imaging* 2006;33:1218–25.
15. Pill SG, Parvizi J, Tang PH, Garino JP, Nelson C, Zhuang H, et al. Comparison of fluorodeoxyglucose positron emission tomography and (111)indium-white blood cell imaging in the diagnosis of periprosthetic infection of the hip. *J Arthroplasty* 2006;21(Suppl 2):91–7.
16. Parvizi J, Ghanem E, Menashe S, Barrack RL, Bauer TW. Periprosthetic infection: what are the diagnostic challenges? *J Bone Joint Surg Am* 2006;88(Suppl 4):138–47.
17. Delank KS, Schmidt M, Michael JW, Dietlein M, Schicha H, Eysel P. The implications of 18F-FDG PET for the diagnosis of endoprosthetic loosening and infection in hip and knee arthroplasty: results from a prospective, blinded study. *BMC Musculoskeletal Disord* 2006;7:20.
18. Mumme T, Reinartz P, Alfer J, Müller-Rath R, Buell U, Wirtz DC. Diagnostic values of positron emission tomography versus triple-phase bone scan in hip arthroplasty loosening. *Arch Orthop Trauma Surg* 2005;125:322–9.
19. Reinartz P, Mumme T, Hermanns B, Cremerius U, Wirtz DC, Schaefer WM, et al. Buell U. Radionuclide imaging of the painful hip arthroplasty: positron-emission tomography versus triple-phase bone scanning. *J Bone Joint Surg Br* 2005;87:465–70.
20. Stumpe KD, Nötzli HP, Zanetti M, Kamel EM, Hany TF, Görres GW, et al. FDG PET for differentiation of infection and aseptic loosening in total hip replacements: comparison with conventional radiography and three-phase bone scintigraphy. *Radiology* 2004;231:333–41.
21. Cremerius U, Mumme T, Reinartz P, Wirtz D, Niethard FU, Büll U. Analysis of (18)F-FDG uptake patterns in PET for diagnosis of septic and aseptic loosening after total hip arthroplasty. *Nuklearmedizin* 2003;42:234–9.
22. Mumme T, Reinartz P, Cremerius U, Hermanns B, Müller-Rath R, Neuss M, et al. [F-18]-fluorodeoxyglucose (FDG) positron emission tomography (PET) as a diagnostic for hip endoprosthesis loosening. *Z Orthop Ihre Grenzgeb* 2003;141:540–6.
23. Chacko TK, Zhuang H, Nakhoda KZ, Moussavian B, Alavi A. Applications of fluorodeoxyglucose positron emission tomography in the diagnosis of infection. *Nucl Med Commun* 2003;24:615–24.
24. Vanquickenborne B, Maes A, Nuyts J, Van Acker F, Stuyck J, Mulier M, et al. The value of (18)FDG-PET for the detection of infected hip prosthesis. *Eur J Nucl Med Mol Imaging* 2003;30:705–15.
25. Kisielinski K, Cremerius U, Büll U, Hermanns B, Wirtz DC, Niethard FU. First experiences with fluorodeoxyglucose-positron-emission tomography (FDG-PET) in the evaluation of painful total knee and hip joint replacements. *Z Orthop Ihre Grenzgeb* 2003;141:153–9.
26. Chacko TK, Zhuang H, Stevenson K, Moussavian B, Alavi A. The importance of the location of fluorodeoxyglucose uptake in periprosthetic infection in painful hip prostheses. *Nucl Med Commun* 2002;23:851–5.
27. Manthey N, Reinhard P, Moog F, Knesewitsch P, Hahn K, Tatsch K. The use of [18 F]fluorodeoxyglucose positron emission tomography to differentiate between synovitis, loosening and infection of hip and knee prostheses. *Nucl Med Commun* 2002;23:645–53.
28. Van Acker F, Nuyts J, Maes A, Vanquickenborne B, Stuyck J, Bellemans J, et al. FDG-PET, 99mTc-HMPAO white blood cell SPET and bone scintigraphy in the evaluation of painful total knee arthroplasties. *Eur J Nucl Med* 2001;28:1496–504.
29. De Winter F, van de Wiele C, Vogelaers D, de Smet K, Verdonk R, Dierckx RA. Fluorine-18 fluorodeoxyglucose-positron emission tomography: a highly accurate imaging modality for the diagnosis of chronic musculoskeletal infections. *J Bone Joint Surg Am* 2001;83:651–60.
30. Zhuang H, Duarte PS, Pourdehnad M, Maes A, Van Acker F, Shnier D, et al. The promising role of 18F-FDG PET in detecting infected lower limb prosthesis implants. *J Nucl Med* 2001;42:44–8.
31. Love C, Pugliese PV, Afriyie MO, Tomas MB, Marwin SE, Palestro CJ. Utility of F-18 FDG imaging for diagnosing the infected joint replacement. *Clin Positron Imaging* 2000;3:159.
32. De Winter F, Van de Wiele C, Vandenberghe S, de Bondt P, de Clercq D, D'Asseler Y, et al. Coincidence camera FDG imaging for the diagnosis of chronic orthopedic infections: a feasibility study. *J Comput Assist Tomogr* 2001;25:184–9.
33. Zhuang H, Chacko TK, Hickeson M, Stevenson K, Feng Q, Ponzio F, et al. Persistent non-specific FDG uptake on PET imaging following hip arthroplasty. *Eur J Nucl Med Mol Imaging* 2002;29:1328–33.
34. Goerres GW, Ziegler SI, Burger C, Berthold T, Von Schulthess GK, Buck A. Artifacts at PET and PET/CT caused by metallic hip prosthetic material. *Radiology* 2003;226:577–84.
35. Heiba SI, Luo J, Sadek S, Macalental E, Cacavio A, Rosen G, et al. Attenuation-correction induced artifact in F-18 FDG PET imaging following total knee replacement. *Clin Positron Imaging* 2000;3:237–9.
36. Becker W, Goldenberg DM, Wolf F. The use of monoclonal antibodies and antibody fragments in the imaging of infectious lesions. *Semin Nucl Med* 1994;24:142–53.
37. Skehan SJ, White JF, Evans JW, Parry-Jones DR, Solanki CK, Ballinger JR, et al. Mechanism of accumulation of 99mTc-sulesomab in inflammation. *J Nucl Med* 2003;44:11–8.
38. Pakos EE, Trikalinos TA, Fotopoulos AD, Ioannidis JP. Prosthesis infection: diagnosis after total joint arthroplasty with antigranulocyte scintigraphy with 99mTc-labeled monoclonal antibodies—a meta-analysis. *Radiology* 2007;242:101–8.