This is an author version of the contribution published on:

Questa è la versione dell’autore dell’opera:


The definitive version is available at:

La versione definitiva è disponibile alla URL:

Title:

Abstract:
Objective: to systematically review and meta-analyze published data about the diagnostic performance of Fluorine-18 dihydroxyphenylalanine (18F-DOPA) positron emission tomography (PET) in patients with paraganglioma (PG).

Methods: A comprehensive computer literature search of studies published through June 30th, 2011 regarding 18F-DOPA PET or PET/computed tomography (PET/CT) in patients with PG was performed in PubMed/MEDLINE, Embase and Scopus databases. Pooled sensitivity and specificity of 18F-DOPA PET or PET/CT in patients with PG on a per patient- and on a per lesion-based analysis were calculated. The area under the ROC curve was calculated to measure the accuracy of 18F-DOPA PET or PET/CT in patients with PG.

Results: Eleven studies comprising 275 patients with suspected PG were included in this meta-analysis. The pooled sensitivity of 18F-DOPA PET and PET/CT in detecting PG was 91% (95% confidence interval [CI], 87-94%) on a per patient-based analysis and 79% (95% CI, 76-81%) on a per lesion-based analysis. The pooled specificity of 18F-DOPA PET and PET/CT in detecting PG was 95% (95% CI, 86-99%) on a per-patient based analysis and 95% (95% CI, 84-99%) on a per-lesion based analysis. The area under the ROC curve was 0.95 on a per patient- and 0.94 on a per lesion-based analysis. Heterogeneity between the studies was found.

Conclusions: In patients with suspected PG 18F-DOPA PET or PET/CT demonstrated high sensitivity and specificity. 18F-DOPA PET or PET/CT are accurate methods in this setting. Nevertheless, possible sources of false negative results should be kept in mind.
Keywords: positron emission tomography, PET/CT, Fluorine-18 dihydroxyphenylalanine, paraganglioma, pheochromocytoma.
Introduction

Paragangliomas (PG) are rare neuroendocrine tumours arising from chromaffin cells of the sympathetic and parasympathetic paraganglia located from the base of the skull to the urinary bladder. Catecholamine-secreting PG arising from the chromaffin cells of the adrenal medulla are referred to as pheochromocytomas, whereas sympathetic PG arising outside the adrenals are referred to as extra-adrenal PG [1]. A large tumour size (>5 cm) or capsular invasion may indicate malignancy; however, the only reliable evidence of malignancy is the presence of local tumour invasion and/or metastatic spread to distal sites, generally bone, liver, and lung or clearly identifiable lymph nodes [2,3].

Functional imaging methods are useful to provide accurate staging and extent of the disease in patients with PG. The information obtained by the combination of conventional and functional imaging methods may influence the management of these patients, especially in malignant and multifocal forms [4,5].

Recently, the use of positron emission tomography (PET) imaging in these tumours is growing rapidly and different positron emitters radiopharmaceuticals (with different uptake mechanisms) have been developed [4,5]. In particular, Fluorine-18 dihydroxyphenylalanine (18F-DOPA) has been proposed as a useful radiopharmaceutical for the imaging of catecholamine-secreting tumours [6]. DOPA enters the cells through the large amino acid transporter 2; then it is converted by aromatic amino acid decarboxylase to dopamine and transported into storage granules by vesicular monoamine transporter [7-9].

Several single-center studies have evaluated 18F-DOPA PET or PET/computed tomography (PET/CT) in patients with suspected PG, reporting different values of sensitivity and specificity; the purpose of our study is to systematically review and meta-analyze published data on the diagnostic performance of 18F-DOPA PET or PET/CT in patients with PG.

Methods
Search strategy

A comprehensive computer literature search of the PubMed/MEDLINE, Embase and Scopus databases was conducted to find relevant published articles on the diagnostic performance of $^{18}$F-DOPA PET and PET/CT in patients with PG. We used a search algorithm that was based on a combination of the terms: a) “DOPA” OR “dihydroxyphenylalanine” AND b) “PET” OR “positron emission tomography” c) “paraganglioma” OR “pheochromocytoma” OR “adrenal”. No beginning date limit was used; the search was updated until June 30th, 2011. No language restriction was used. To expand our search, references of the retrieved articles were also screened for additional studies.

Study selection

Studies or subsets in studies investigating the diagnostic performance of $^{18}$F-DOPA PET or PET/CT in patients with PG were eligible for inclusion. The exclusion criteria were: a) articles not within the field of interest of this review; b) review articles, editorials or letters, comments, conference proceedings; c) case reports or small case series; f) overlap in patient data (duplicate publication; in such cases the most complete article was included); g) insufficient data to reassess sensitivity or specificity from individual studies.

Three researchers (GT, FC and PC) 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 three 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.

Data extraction

For each included study, information was collected concerning basic study (authors, journal, year of publication, country of origin), patient characteristics (mean age, sex, number of patients with PG, number of patients with genetic mutations), technical aspects (device used, radiopharmaceutical injected dose, time between $^{18}$F-DOPA injection and image acquisition, carbidopa pretreatment, image analysis, applied reference standard). For each study the number of true positive, false
positive, true negative and false negative findings for $^{18}$F-DOPA PET or PET/CT in diagnosis of PG was recorded on a per patient- and on a per lesion-based analysis.

**Quality assessment**

Three independent reviewers (CdW, MRG and FDN) evaluated the methodology of the selected studies using the Quality Assessment Tool for Diagnostic Accuracy Studies (QUADAS) [10]. This 14-items tool is composed by five items related to verification bias (items 3, 4, 5, 6, 7) three items related to review bias (items 10, 11, 12), two items relating to generalizability and context and spectrum bias (items 1 and 2) and four to reporting (items 8, 9, 13, 14). Reviewers, who were blinded to the purposes of the meta-analysis, recorded a score of “1” for “yes” and “0” for “no” and “unclear” for each of the 14 items; all the disagreements were resolved by a consensus. Furthermore, studies were given an A, B, C, or D rating according to the quality assessment by Van den Bruel et al. [Van den Bruel A, Thompson MJ, Haj-Hassan T, Stevens R, Moll H, Lakhanpaul M, Mant D. Diagnostic value of laboratory tests in identifying serious infections in febrile children: systematic review. BMJ 2011;342:d3082]. Studies fulfilling all QUADAS items were rated as A. Studies with no or unclear total verification with the reference standard or with interpretation of the index test unblinded to the results of the reference standard were rated as D, while studies without an independent reference standard, with interpretation of the reference standard unblinded to the results of the index test or with an unduly long period between index and reference tests as C. All other studies were rated B.

**Statistical analysis**

Sensitivity and specificity of $^{18}$F-DOPA PET or PET/CT in patients with PG was obtained from individual studies on a per patient- and on a per lesion-based analysis. The reference standard was considered a combination of histologic examinations and clinical/imaging follow-up data. A random effect model is used for statistical pooling of the data in the case of heterogeneity between the studies ($p<0.1$); if there was no heterogeneity between the studies a fixed effect model for statistical pooling of the data is used. Pooled data are presented with 95% confidence intervals.
A I-square statistic was also performed to test for heterogeneity between studies. The area under the ROC curve was calculated to measure the accuracy of $^{18}$F-DOPA PET or PET/CT in patients with PG. Statistical analyses were performed using Meta-DiSc statistical software version 1.4 (Unit of Clinical Biostatistics, Ramón y Cajal Hospital, Madrid, Spain) [11].

**Results**

**Literature search**

The comprehensive computer literature search from PubMed/MEDLINE, Embase and Scopus databases revealed 63 articles. Reviewing titles and abstracts, 42 articles were excluded: 25 because not in the field of interest of this review, 17 as reviews or editorials. Twenty-one articles were selected and retrieved in full-text version; no additional study was found screening the references of these articles. From these 21 articles potentially eligible for inclusion, after reviewing the full-text article, six articles were excluded as case reports or small case series [12-17], one due to data overlap [18], and three due to insufficient data to calculate sensitivity or specificity of $^{18}$F-DOPA PET [19-21]. Finally, 11 studies, comprising a total sample size of 275 patients with suspected PG met all inclusion and exclusion criteria, and they were included in our meta-analysis [22-32] (Fig. 1). The characteristics of the included studies are presented in Tables 1-2.

**Quality assessment**

Table 3 shows the results of the quality assessment. Studies scored between 7 and 13 with a median score of 9 (interquartile range: 2). Most of the studies (6 studies: 54.5%), including those with larger study populations, scored between 8 and 9; 4 studies (36.4%) scored 10 or more whereas only 1 [31] achieved a score of 7. In this last study it was unclear if the index test and the reference standard were interpreted in blinded conditions (items 10 and 11) and if all uninterpretable/intermediate results were reported (item 13). Furthermore, patients did not receive the same reference standard regardless of the index test result and reference standard was not
independent from the index test (items 6 and 7). Moreover, methodology of the reference standard was not likely to correctly classify the target condition (item 3) and was not described in sufficient detail (item 9). Anyway, also in the study by Hoegerle 2002 [22], which achieved the highest score, the reference standard was not likely to correctly classify PG (item 3).

As far as the items are concerned, a brief description of the most important weaknesses is described. Only 1 paper (9.1%) [23] scored “no” on item 1 because of the choice of the Authors to study patients with a proven mutation of the SDHD gene, thus not representative of the population who receives F-DOPA PET/CT in practice. All others study populations included patients with suspected or known PG. Anyway, selection criteria (item 2) were clearly described only in 6 studies (54.5%). Only 3 studies (27.3%) used hystologic verification in all patients as reference standard (item 3); among the remaining 8 studies, 6 used it only when feasible. In 3 cases (27.3%) it was not possible to be sure that patient did not develop new lesions during the time elapsed from index test to reference standard. In none of the items from 5 to 9 unclear results were observed, but the quality was particularly poor on items 6 and 9. Only in 5 studies (45.5%) patients received the same reference standard regardless of the index test result (item 6) and only in 3 studies (27.3%) the execution of the reference standard (which was composite in most cases) was described in sufficient detail (item 9). On items 10 and 11 (blindness of index test and reference test respectively) there was the higher uncertainty: for both items, in 4 studies (36.4%) it was unclear whether the tests were executed in blind conditions or not. As far as withdrawals are concerned, it should be noted that in most studies no withdrawals were observed (item 14).

Finally, none of the studies achieved an A rating while 2 (18.2%) received a B rating and 4 (36.4%) got a C; all the other studies received a D rating.

**Diagnostic performance**

The diagnostic performance results of $^{18}$F-DOPA PET and PET/CT in the 11 included studies are presented in Tables 4-5. The sensitivity of $^{18}$F-DOPA PET and PET/CT calculated on a per patient- and per lesion-based analysis ranged from 77% to 100% and from 54% to 100%, with pooled
estimates of 91% (95% CI: 87-94%) and 79% (95% CI: 76-81%), respectively (Figs. 2-3). The included studies were statistically heterogeneous in their estimates of sensitivity on a per patient- (I-square: 57.3%) and per lesion-based analysis (I-square: 96.8%).

The specificity of $^{18}$F-DOPA PET and PET/CT calculated on a per patient- and per lesion-based analysis ranged from 75% to 100% and from 67% to 100%, with pooled estimates of 95% (95% CI: 86-99%) and 95% (95% CI: 84-99%), respectively (Figs. 4-5). The included studies were statistically homogeneous in their estimates of specificity on a per patient- (I-square: 0%) and per lesion-based analysis (I-square: 0%).

The area under the ROC curve was 0.95 on a per patient- and 0.94 on a per lesion-based analysis (Fig. 6)

Discussion

To our knowledge, this meta-analysis is the first to evaluate the diagnostic performance of $^{18}$F-DOPA PET and PET/CT in patients with PG. Several single-center studies have used $^{18}$F-DOPA PET or PET/CT in patients with suspected PG reporting different values of sensitivity and specificity (Tables 4-5). However, many of these studies have limited power, analyzing only relatively small numbers of patients. In order to derive more robust estimates of diagnostic performance of $^{18}$F-DOPA PET and PET/CT in patients with PG we pooled published studies. A systematic review process was adopted in ascertaining studies, thereby avoiding selection bias.

All the studies included in the review were shown to be of moderate quality according to QUADAS. Anyway, this tool has some limits because it is not supposed to be meant as a scale. In fact, items do have different relevance in the assessment of the quality: a study achieving a very high score, being fulfilled almost all items, could still have a debatable quality if it does not meet one of the most important items, such as the use of the same reference standard in all the patients. This clearly arises from the rating of studies according to Van den Bruel [Van den Bruel A, Thompson MJ, Haj-Hassan T, Stevens R, Moll H, Lakhanpaul M, Mant D. Diagnostic value of
laboratory tests in identifying serious infections in febrile children: systematic review. BMJ 2011;342:d3082]. In fact, it is noteworthy to observe that none of the studies achieved an A rating being 9 (81.8%) the studies with the lowest rating C and D. Another drawback of the QUADAS is that it does not take into consideration the sample size which is responsible for the precision of the study and its validity too. On the other hand, it is important to remember that the low quality could be also due to the limitations in carrying out these kinds of studies in the real clinical setting, where it might be difficult to confirm the final diagnosis in all patients. Pooled results of our meta-analysis indicate that $^{18}$F-DOPA PET and PET/CT demonstrate high sensitivity (91%) and high specificity (95%) on a per patient-based analysis and good sensitivity (79%) and high specificity (95%) on a per lesion-based analysis. Furthermore, the area under the ROC curve (0.95 on a per patient- and 0.94 on a per lesion-based analysis) demonstrates that $^{18}$F-DOPA PET and PET/CT are accurate methods for diagnosis of PG.

Nevertheless, possible sources of false negative results of these functional imaging methods should be kept in mind; they could be related to several factors such as the small size of the lesion, location of the tumor near organs with high physiologic $^{18}$F-DOPA uptake (such as the pancreas, biliary and urinary systems) or loss of $^{18}$F-DOPA uptake due to a tumour dedifferentiation. Genetic factors may also affect the $^{18}$F-DOPA uptake in PG; succinate dehydrogenase B (SDHB) gene mutations may result in extra-adrenal PG which show a sensitivity of $^{18}$F-DOPA PET lower than non-SDHB-related lesions [28]. It is possible that the high pooled sensitivity of $^{18}$F-DOPA PET and PET/CT observed in our analysis is related to the small number of patients with SDHB gene mutations enrolled in most of the studies (Table 1), except that of Timmers et al. [28].

The high specificity of $^{18}$F-DOPA PET and PET/CT can be explained by the fact that only neuroendocrine cells are able to take up, decarboxylate, and store amino acids and their amines. Few false positive $^{18}$F-DOPA PET findings are reported in the literature in patients with suspected PG. Kauhanen et al. described one patient with suspected PG recurrence and increased $^{18}$F-DOPA uptake in the right adrenal; histological verification showed a normal adrenal gland [25]. Timmers
et al. reported one patient with a gastrointestinal stromal tumor which was visualized by $^{18}$F-DOPA PET [28]; finally, Luster et al. described an adrenal mass with a mildly intense $^{18}$F-DOPA uptake, but clinical follow-up revealed no evidence of pheochromocytoma [29].

The included studies were statistically heterogeneous in their estimates of sensitivity. This heterogeneity is likely to arise through diversity in methodological aspects between different studies (Table 2 and 3). For example, some Authors used carbidopa pretreatment before $^{18}$F-DOPA PET examination; this drug, decreasing decarboxylation and subsequent renal clearance of DOPA may be used to increase the tumor to background uptake ratio; nevertheless, carbidopa pretreatment should have a low influence on the number of PG lesions depicted by $^{18}$F-DOPA PET, because these tumours usually have high $^{18}$F-DOPA uptake [5,18].

The baseline differences among the patients in the included studies (Table 1) may have contributed to the observed heterogeneity of the results too. However, such variability was accounted for in a random effect model.

A limitation of our analysis is the lack of the calculation of pooled sensitivity and specificity of $^{18}$F-DOPA PET or PET/CT in different forms of PG, for example adrenal vs extra-adrenal, sympathetic vs parasympathetic, functioning vs non-functioning, inherited vs sporadic, metastatic vs non-metastatic tumours; in fact, the frequent mixing of these forms of PG in the patient population of the included studies hampered the data extraction and the separate calculation of diagnostic performance of $^{18}$F-DOPA PET in such groups.

Nevertheless, according to the literature, $^{18}$F-DOPA PET and PET/CT seem to be accurate methods in both adrenal [22, 25, 26, 29] and extra-adrenal [23, 24, 30, 31], sympathetic [22, 25-27] and parasympathetic [23, 24, 30, 31], functioning [27] and non-functioning [23, 26, 31], metastatic and non-metastatic tumors [24, 28, 32]. Furthermore, $^{18}$F-DOPA PET and PET/CT seem to be accurate methods in both sporadic and inherited PG [27, 30], except in SDHB-related PG [28].

Finally, based on its high sensitivity and specificity, $^{18}$F-DOPA PET may be considered the first-line tracer in the diagnostic work-up of PG. Currently, the literature focusing on the use of $^{18}$F-
DOPA PET and PET/CT in PG remains limited; thus, further large multicenter studies will be necessary to substantiate the diagnostic accuracy of $^{18}$F-DOPA PET and PET/CT in patients with PG.

**Conclusions**

In patients with suspected PG $^{18}$F-DOPA PET and PET/CT demonstrated high sensitivity and specificity. $^{18}$F-DOPA PET and PET/CT are accurate methods in this setting. Nevertheless, possible sources of false negative results should be kept in mind.

**References**


