Introduction
Disturbances of dopaminergic functioning are associated with numerous neurological disorders, whose diagnosis may be challenging, particularly in the early stages. Although in most cases the diagnosis of idiopathic Parkinson’s disease (PD) is straightforward in cases with classical presentation, delineating PD from other parkinsonisms may represent a diagnostic challenge.
Although the gold standard for the diagnosis of PD is post-mortem neuropathological examination [Rajput et al., 1991; Hughes et al., 1992; Hughes et al., 1993], neuropathological studies show that even at end-stage disease the clinical diagnostic accuracy for PD does not reach 100% [Hughes et al., 2001].
Drug-induced parkinsonism (DIP) frequently develops in patients treated with dopamine receptor blocking agents (DRBAs) [Gershanik, 1994] from few days to several years after medication exposure [Hausner et al., 1983]. After drug withdrawal, the majority of patients recover within few months, but, occasionally, motor improvement does not occur, raising the question of possible development of Parkinson’s disease (PD) [Fleming et al., 1970]. Clinically, DIP is characterized by symmetric symptoms, absence of tremor, occurrence of bucco-linguo-masticatory dyskinesias, and akathisia. However, asymmetry and tremor may also occur in DIP, making the differential diagnosis a challenge [Tolosa et al., 2003]. However, a reliable test to differentiate PD from other parkinsonian disorders would be of utmost importance because prognosis and management may differ considerably [Piccini et al., 2004], unnecessary medical examinations and therapies may be avoided.
Dopamine system imaging represents the most widely used diagnostic tool in differentiating PD from other parkinsonian disorders [Piccini et al., 2004]. The functional integrity of dopaminergic nigrostriatal pathway can be studied with single photon emission computed tomography (SPECT) imaging by using ligands of pre-synaptic dopamine transporter (DAT) such as Iodine-123-beta-CIT, Iodine-123-Ioflupane (FP-CIT), Iodine-123-IPT and 99mCT-TRODAT-1. A reduction of SPECT ligand binding to DAT correlates with the loss of presynaptic dopamine.
With regards to the differential diagnosis between PD and DIP, comparisons between dopamine system imaging and the ultimate gold standard (post-mortem neuropathological examination) do not exist; Furthermore, all studies adopt a surrogate gold standard in the form of a long-term clinical follow-up against which to assess sensitivity and specificity of imaging, which may severely affect the ability of SPECT to differentiate PD from DIP and other parkinsonisms [Vlaar et al., 2007]. The methodological heterogeneity across studies is also high, as different radiotracers and SPECT techniques are used, and sometimes different patient populations are investigated (relevant source of clinical heterogeneity). Finally, the generalizability of results (i.e. the external validity) is often limited, as many studies include only later-stage patients that are not representative for the diagnostic problem that one wants to address with a dopamine system imaging.
We therefore decided to systematically review the utility of dopamine system imaging using [¹²³I]FP-CIT SPECT in the differential diagnosis between PD and DIP. Our choice of focusing only on the use of [¹²³I]FP-CIT SPECT was justified in order to reduce methodological heterogeneity across studies and because this technique is more widely available for routine clinical use than PET or SPECT using other ligands.

Methods
Our aim will be to critically and systematically evaluate the literature to evaluate the diagnostic accuracy of [¹²³I]FP-CIT SPECT in the differential diagnosis between PD and DIP,
determining the sensitivity, specificity, positive and negative likelihood ratio (pLR, nLR) of this dopamine system imaging. We will include prospective and retrospective studies reporting data on $^{[123]}$I-FP-CIT SPECT performed in patients with DIP and PD. We will mainly focus on patients with unclear parkinsonism, evaluating the role of $^{[123]}$I-FP-CIT SPECT in differentiating between PD and DIP. However, we will consider also patients with a clinically defined diagnosis of PD and DIP, whose data will be analyzed separately. We will also perform an evaluation of studies comparing SPECT results in patients with DIP and healthy controls.

**Search methods for identification of studies**

**Electronic searches**
We will search the following databases:
- MEDLINE (accessed by Pubmed);
- the Cochrane Central Register of Controlled Trials (CENTRAL, The Cochrane Library);
- EMBASE.

We present the proposed search strategy for MEDLINE in Appendix. This strategy will be modified for use with the other databases. The electronic databases will be searched using the following medical subject headings (MeSH): “Parkinson Disease, Secondary”, “Tomography, Emission-Computed, Single-Photon”, as well as following free terms, combined in multiple search strategies with Boolean operators in order to find relevant articles: “drug-induced parkinsonism”, “SPECT” (see Appendix).

**Searching other resources**
We will contact experts in the field for information about any unpublished or ongoing studies. We will review the reference lists of articles retrieved by the electronic searches to check for other relevant reports not indexed in the electronic database. In order to provide a transparency of results as great as possible, and to allow readers to reproduce the methodology we adopted, and considering that in abstracts many methodological aspects are not declared and results are often synthesized, only in extenso papers and articles already published will be considered eligible for inclusion.

No language restrictions will be applied.

**Data collection and analysis**

**Selection of studies**
Two review authors (FB and AM) will independently screen all the titles and abstracts of publications identified by the searches to assess their eligibility. We will exclude publications that do not meet the criteria at this stage. Following screening, we will assess the full-text of potentially-eligible citations for inclusion. The review authors will reach consensus on the selection of trials and the final list of studies. We will discuss any disagreements and resolve them where possible. If we cannot reach consensus, we will consult a third member of the team (MT).

**Inclusion criteria**
We will include all cross-sectional studies of patients in which $^{[123]}$I-FP-CIT SPECT was tested as a means to differentiate between various parkinsonian syndromes, reporting data on
SPECT results in patients with PD and DIP. We will mainly focus on patients with unclear parkinsonism, evaluating the role of $[^{123}]$FP-CIT SPECT in differentiating between PD and DIP. However, we will consider also patients with a clinically defined diagnosis of PD and DIP, whose data will be analyzed separately. We will consider also studies reporting data on SPECT in patients with DIP and in healthy controls. No age, race or gender restrictions will be applied.

**Exclusion criteria**
The following exclusion criteria will be used:
1) studies not reporting data of SPECT in both patients with PD and with DIP or in patients with DIP and healthy controls;
2) studies not reporting information to allow the construction of the diagnostic 2 by 2 table with its four cells: true positives, false negatives, true negatives and false positives of both patients with PD and with DIP / patients with DIP and healthy controls;
3) studies using SPECT with ligands other than $[^{123}]$FP-CIT.

**Data extraction**
Two review authors (FB and AM) will independently extract data from the published reports where possible. We will use data extraction forms and resolve any disagreements by mutual agreement. We will record the rawest form of the data, when possible. In case of missing or incomplete data, we will contact the principal investigators of included trials and request additional information.

Following data will be extracted:
- author(s);
- year of publication;
- inclusion criteria;
- number of patients;
- age;
- gender;
- disease duration (years);
- Hohen-Yahr stage;
- reference diagnostic standard chosen;
- minimal duration of follow-up after SPECT (months);
- part of the striatum evaluated;
- cut-off point adopted to define SPECT abnormality;
- methods used to judge SPECT results (visual, quantitative);
- drug withdrawal before SPECT (hours);
- independent, blind comparison with clinical diagnosis.

A standardised data extraction form will be used and data on accuracy measure will be reported in a diagnostic 2 by 2 table to calculate sensitivity, specificity, and LRs.

**Methodological quality evaluation**
The methodological quality of each selected paper will be assessed independently by two reviewers (FB and AM).

The quality of the studies provisionally selected for inclusion in the review will be evaluated using the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool [Whiting et al., 2003], and summarized as outlined in Cochrane Handbook for Diagnostic Test Accuracy Reviews (Version 1.0.1 March 4th 2009) [Cochrane, 2009]. Revman 5 (version: 5.2.5), available online at: http://ims.cochrane.org/revman/download will be present summarized quality results.
Furthermore, we will evaluate whether studies explicitly reported a well-defined cut-off point to define SPECT abnormality.

**Data synthesis**
Provided we think it clinically appropriate, and no important clinical and methodological heterogeneity is found, we plan to synthesise the study results in a meta-analysis.

Two comparisons will be performed:
1. DIP versus PD;
2. DIP versus healthy controls.

Sensitivity, specificity and the odds ratio will be calculated for each study separately, and the pooled diagnostic odds ratios (ORs) for all studies together. Considering the presence of methodological heterogeneity and threshold effect, due to differences in patients, SPECT machinery, radiotracers etc., we will use diagnostic ORs [Egger et al., 2005]. The diagnostic ORs express how much greater the odds of having the disease are for the people with a positive test result than for the people with a negative test result. It is a single measure of diagnostic test performance that combines both LRs.

The potential problems associated with sensitivities and specificities of 100% (i.e. studies with zeroes in one or more cells of the diagnostic 2 by 2 table) will be solved by adding 0.5 to all cells of the diagnostic 2 by 2 table [Cox, 1970; McGee, 2007]. Trials with a sensitivity of 100% and a specificity of 0% will not be excluded, however the pooled diagnostic ORs will also calculated without such studies (sensitivity analysis). The meta-analysis will be undertaken with the Review Manager software developed by the Cochrane Collaboration (5.1). Sensitivity, specificity, pLR and nLR with 95% CIs will be determined for each included study and for the summary estimate of pooled analysis using equations reported in Appendix [Simel et al., 1991; Newcombe, 1998; Devillé et al., 2002; Zamora et al., 2006a; McGee, 2007].

Sensitivity measures the proportion of positives that are correctly identified, whereas specificity measures the proportion of negatives that are correctly identified. The LR of a physical sign is defined as the proportion of patients with disease who have a certain finding divided by the proportion of subjects without disease who also have the same finding [McGee, 2007]. A pLR refers to the presence of the physical sign, whereas a nLR refers to the absence of that physical sign. The interpretation of LRs is straightforward: (1) values greater than 1 increase the probability of disease, and the greater the LR, the more compelling the argument for disease; (2) values between 0 and 1 decrease the probability of disease, and the closer the LR is to zero, the more the finding argues against the diagnose of disease; (3) values equal zero have no diagnostic values, as they do not change pre-test probability [McGee, 2007]. A pLR describes therefore how probability changes when the finding is present, whereas nLR describes how probability changes when the finding is absent.

OR of each study will be combined to obtain a summary estimate of value (and the corresponding 95% confidence intervals, CIs) using a random-effect model. Random-effects model, which considers both within-study and between-study variance to calculate a pooled ORs, will be used to summarize the ORs from the included studies. This model is considered more conservative than a fixed-effect, since it takes into account the variability between studies, thus leading to wider CIs [DerSimonian and Laird, 1986].

Accuracy measures for each study, pooled accuracy measures and diagnostic ORs will be obtained by performing a meta-analysis using Meta-DiSc software [Zamora et al., 2006a].

**Subgroup analysis**
We will separately analyze patients with unclear parkinsonism and patients with a clinically defined diagnosis of PD and DIP.
Assessment of heterogeneity
Visual inspection of the forest plots will be used to investigate the possibility of statistical heterogeneity. We will evaluate homogeneity among trial results using a standard Chi² test and the hypothesis of homogeneity will be rejected if the P value is less than 0.10. Assessment of statistical heterogeneity will be supplemented using the I² statistic which provides an estimate of the percentage of variability due to heterogeneity rather than a sampling error [Higgins et al., 2003].

Sensitivity analysis
The diagnostic ORs in our study may be affected by differences in the individual cut-off points, hence introducing relevant methodological heterogeneity. We will therefore search for the presence of an (implicit) cut-off point effect between studies by calculating a Spearman correlation coefficient between sensitivity and specificity of all included studies. In case of strong negative correlation (i.e. ρ<0.6), indicative of strong cut-off effect [Moses et al., 1993], we will perform a sensitivity analysis adopting one common cut-off point for all studies. In this sensitivity analysis we will include only those studies adopting a cut-off point of 2 standard deviations (SD) below the binding rate of healthy controls as a definition for abnormal SPECT. We chose this approach to take into consideration the presence of a significant cut-off point effect in the included studies [Vlaar et al., 2007]. A further sensitivity analysis will be conducted excluding trials with a sensitivity of 100% and a specificity of 100% [Vlaar et al., 2007].

References


APPENDIX

Search strategy

("drug-induced parkinsonism" AND SPECT) OR ("Parkinson Disease, Secondary"[Mesh]) AND "Tomography, Emission-Computed, Single-Photon"[Mesh])

Equations used to calculate accuracy measures of SPECT

<table>
<thead>
<tr>
<th></th>
<th>Target disease (PD)</th>
<th>Absence of target disease (DIP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal SPECT</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Normal SPECT</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

$nc1 = a + nc2 = b + dz = 1.959964$

**Sensitivity**

$\text{Sensitivity} = \frac{a}{nc1}$

Lower limit = \(\exp\left(\ln\left(\frac{nc2 \times a}{nc1 \times b}\right) - z\sqrt{\left(\frac{c}{(a \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

Upper limit = \(\exp\left(\ln\left(\frac{nc2 \times a}{nc1 \times b}\right) + z\sqrt{\left(\frac{c}{(a \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

**Specificity**

$\text{Specificity} = \frac{d}{nc2}$

Lower limit = \(\exp\left(\ln\left(\frac{nc2 \times d}{nc1 \times c}\right) - z\sqrt{\left(\frac{a}{(c \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

Upper limit = \(\exp\left(\ln\left(\frac{nc2 \times d}{nc1 \times c}\right) + z\sqrt{\left(\frac{a}{(c \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

**Positive likelihood ratio**

$L+ = \frac{\text{Sensitivity}}{(1 - \text{Specificity})}$

Lower limit = \(\exp\left(\ln\left(\frac{nc2 \times a}{nc1 \times b}\right) - z\sqrt{\left(\frac{c}{(a \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

Upper limit = \(\exp\left(\ln\left(\frac{nc2 \times a}{nc1 \times b}\right) + z\sqrt{\left(\frac{c}{(a \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

**Negative likelihood ratio**

$L- = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$

Lower limit = \(\exp\left(\ln\left(\frac{nc2 \times c}{nc1 \times d}\right) - z\sqrt{\left(\frac{a}{(c \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

Upper limit = \(\exp\left(\ln\left(\frac{nc2 \times c}{nc1 \times d}\right) + z\sqrt{\left(\frac{a}{(c \times nc1)} + \frac{b}{(d \times nc2)}\right)}\right)\)

* When calculating LR, if any cell of the 2 x 2 table contained the value of zero, 0.5 was added to all cells, to avoid creating the unlikely LRs of 0 or infinity (McGee, 2007).

**Equation used to calculate diagnostic odds ratio (ORs)**

Diagnostic ORs = \(\frac{\text{Sensitivity}}{(1-\text{Sensitivity})}/(1-\frac{\text{Specificity}}{\text{specificity}})\)

**Statistical Formulae for pooling of proportions**

**Homogeneous sensitivity and/or specificity**

For the sensitivity:

\[
\text{Sensitivity}_{pooled} = \frac{\sum_{i=1}^{k} a_i}{\sum_{i=1}^{k} (a_i + c_i)}
\]

whereby $a = \text{true positives}$
\(c = \text{false negatives}\)
\(i = \text{study number}\)
\(k = \text{total number of studies}\)

with standard error:
\[
SE = \sqrt{\frac{p(1 - p)}{n}}
\]

whereby
\(p = \text{Sensitivity}_{\text{pooled}}\)
\(n = \sum_{i=1}^{k} (a_i + c_i)\)