

Chapter 3. Results

In the following results, “k” refers to the number of studies, “t” refers to the number of treatment arms, and “n” refers to the number of PD patients.

Searches

The numbers of abstracts obtained from searches in Medline and Current Contents are displayed in Figure 2. The primary search in Medline (search window: 1990-2000) yielded 957 abstracts, the search in Current Contents (search window: 1990-2000) yielded 397, and the Cochrane Library search yielded 590, for a total of 1,944 citations. After 614 duplicates were identified, a total of 1,330 abstracts were downloaded into Reference Manager at MetaWorks. Another 174 potentially relevant citations were identified from manual bibliography checks. Thus, over the duration of this project, 1,504 abstracts identified from electronic searches and bibliography checks were screened against protocol-defined exclusion criteria. This does not include the searches for review articles and genetics, which yielded 377 and 149 citations, respectively. Of the 1,504 citations pertaining to diagnosis and treatment of PD, 791 were rejected during Level I screening of abstracts. The main reasons for Level I rejection were: studies not pertinent to PD, studies with less than ten patients or with study duration less than 24 weeks, studies not pertaining to diagnosis or treatment, and cross-over studies. Full-text papers of the remaining 713 studies were retrieved and screened at MetaWorks.

During Level II screening of full-text papers, 465 were rejected, resulting in 248 accepted studies. Evidence Table 1 summarizes the number of studies rejected during Level II screening or data extraction, organized by rejection reason. Comprehensive bibliographies for accepted studies may be found in Appendix H. Appendix I contains full bibliographies for rejected studies, organized by rejection reason.

The screening strategies were reviewed *a priori* with the TEP, TOO, and AAN representative. A few studies did not meet inclusion criteria, but the consulting PD experts believed they should be mentioned in this review. These studies are discussed in Appendix J, but were not extracted, entered into the database, or included in the statistical analyses.

Studies

Study Characteristics

Evidence Table 2 summarizes the main study-level characteristics of the 248 studies accepted for data extraction. These studies consisted of 180 parents and 68 kin studies. In the 180 parent studies, there were a total of 353 treatment groups, and 16,158 patients. Of the accepted studies, 59 parents (and 6 kins) pertained to diagnosis (t=141, n=3,369), 49 parents (and 36 kins) to pharmacological treatment (t=111, n=9,968), 42 parents (and 23 kins) to surgery (t=52, n=1,380), ten parents (and one kin) to psychiatric treatment (t=12, n=392), and 20 parents (and two kins) to ancillary treatment (t=37, n=1,049). The genetics section of this report was based on the contents of 16 papers which were reviewed, although not formally extracted.

Evidence Table 3 shows baseline treatment level characteristics of patients in all accepted studies. One study excluded patients < 60 years of age,¹¹⁹ but most studies did not focus specifically on elderly or young patients. The evidence base in both pharmacological and surgical studies is heavily weighted towards people under age 65.

Further details regarding treatment level variables will be described in the individual sections concerning diagnosis, pharmacological treatment, surgical treatment, psychiatric treatment, and ancillary treatment.

Diagnosis

There were 59 parent and six kin studies concerning diagnosis, consisting of 3,369 patients, 1,108 healthy controls, and 859 patients with other neurological diagnoses, such as secondary parkinsonism or essential tremor. The vast majority of studies were cross-sectional studies (k=46, n=2,055), six were retrospective observational studies (n=953), five were UCSs (n=278), and two were nRCTs (n=83). All were graded as level III evidence, and quality score could not be calculated because there were no RCTs.

Evidence Table 4 shows the number of studies, treatment arms, and patients who were evaluated by different diagnostic tests. The categories of diagnostic testing included: apomorphine or L-dopa challenge tests (k=5, t=6, n=229), autopsy studies (k=6, t=15, n=253), clinical or laboratory tests (k=10, t=26, n=1,412), color vision testing (k=2, t=3, n=35), MRI (k=3, t=8, n=140), olfactory testing (k=7, t=21, n=355), PD Test Battery (k=3, t=7, n=180), PET scans (k=8, t=21, n=185), SPECT scans (k=13, t=29, n=460), and other scans (k=2, t= 5, n=120).

An estimate of the diagnostic accuracy of any test should compare the test with a reference standard, which should be the best available method of assessing the presence or absence of the disease of interest.¹²⁰ The major difficulty in assessing diagnostic tests for PD is the lack of a validated reference standard for comparison.

Apomorphine and L-Dopa Challenge Tests

All five studies evaluating challenge tests with apomorphine or L-dopa were conducted in Europe: three in the United Kingdom^{121, 122, 123} and one each in Germany¹²⁴ and Italy.¹²⁵ Three studies were cross-sectional, and two were UCSs. They included 229 PD patients and 43 patients with other neurological diagnoses. Comparison of these studies is hampered by heterogeneity in many areas, including reference standard, study design, challenge test methodology, definitions of positive results, and outcomes reported.

In all five studies, patients were reported as having positive or negative apomorphine tests, and presence or absence of PD, based on long-term response to L-dopa. The sensitivities of the apomorphine tests ranged from 87 to 95 percent, and the specificities ranged from 75 to 95 percent. There are several problems, however, with comparing the sensitivity and specificity results from different studies. The major limitation is lack of a constant reference standard. In two studies, the reference standard was response to chronic L-dopa use. Apomorphine is a rapidly acting DA. Patients who respond to apomorphine would be expected to also respond to other DAs or to L-dopa, but it is not valid to equate L-dopa response with a diagnosis of PD, as other conditions may also respond to L-dopa. In three studies, the reference standard was clinical diagnosis of PD. The problem with using clinical diagnosis as the reference standard is that

autopsy studies have shown that clinical diagnosis may be wrong in up to 25 percent of cases.⁴² Until there is a valid reference standard, it will remain difficult to evaluate any diagnostic test for PD.

There are several other areas of heterogeneity among the apomorphine studies. The dosage of apomorphine varied, ranging from one to ten mg given subcutaneously, and the time interval between doses ranged from 30 to 120 minutes. In some studies, the apomorphine dose was adjusted for weight. Some studies used placebo, while others reported that the acute side effects associated with apomorphine would make blinding impractical. Criteria for positive results also varied from study to study, and included improvements of 15 or 25 percent in the tapping test, walking test, tremor scale, rigidity scale, or modified Webster score to define positive apomorphine test results. In several studies, some patients had equivocal results. The studies concluded, in general, that the apomorphine test might be predictive of response to L-dopa, but is not diagnostic of PD.

SPECT Scans

Thirteen studies, consisting of 460 PD patients, 191 healthy controls, and 64 patients with other neurological disorders, evaluated the use of SPECT scans in PD patients.^{126 - 138} Eleven of these studies were performed in Europe (eight in Germany (six by the same author), two in Italy, and one in the Netherlands), one in the US, and one in Japan. Ten studies were cross sectional, two were UCSs, and one was a nRCT. Five studies evaluated patients with early PD only,^{126 - 130} and eight evaluated patients with various stages of PD. Five of the studies compared SPECT results before and after administration of apomorphine,^{127, 128, 130, 131, 132} and one compared SPECT results to PET results.¹²⁹

Three studies evaluated presynaptic dopamine transporters, using the cocaine analogues [123]β-carboxymethoxy-tropane (β-CIT),¹³⁸ fluoropropyl-carbomethoxy-tropane (FP-CIT),¹³³ or iodopropenyl-carbomethoxy-tropane (IPT)¹²⁶ as radioligands. In all three studies, early PD patients had decreased presynaptic uptake compared with normal controls.

Nine studies evaluated postsynaptic dopamine receptors, using the ligand IBZM.^{127 - 132, 134, 135, 136} The results showed that normal or increased receptor binding of IBZM corresponded to positive response to L-dopa or apomorphine,¹²⁹⁻¹³² and decreased binding corresponded to negative response to L-dopa or apomorphine.^{127, 128, 130, 132, 136} There was, however, considerable overlap between patients with PD, atypical parkinsonism, and normal controls.^{127, 128, 132, 134}

One study using the ligand IBZM reported increased uptake in patients with early PD compared with normal controls, but decreased uptake in patients with more severe PD.¹³⁵ The authors theorized that in patients with advanced PD, there may be down-regulation of striatal post-synaptic dopamine receptors due to chronic exposure to exogenous L-dopa.

In one study, changes in global and regional cerebral blood flow (CBF) were measured.¹³⁵ Global CBF was lower in PD patients than in normal controls, and decreased with more advanced disease.

Combining the results of the SPECT scan studies is problematic because they varied in many ways. Some evaluated pre-synaptic function, while others evaluated post-synaptic function, and one evaluated cerebral blood flow. Some compared PD patients to normal controls, and some compared SPECT results to the results of apomorphine or L-dopa challenges. In many cases, there was overlap between the results of PD patients and controls. The great variation in

reporting results of SPECT scan studies precludes any conclusion regarding the utility of SPECT scans in diagnosis or management of PD.

PET Scans

Eight studies, including 185 patients, 144 healthy controls, and 52 patients with other neurological disorders, evaluated the use of PET scans in PD patients. Three studies were conducted in Finland,^{139, 140, 141} three in the United Kingdom,^{142, 143, 144} one in the United States,¹⁴⁵ and one in Japan.¹⁴⁶ All were cross-sectional studies.

As was true with the SPECT scan studies, the studies of PET scans reported their results in inconsistent fashion. Six studies reported that [¹⁸F] fluorodopa uptake was lower in the caudate and putamen nuclei of PD patients than in healthy controls.^{140, 142, 143, 144, 145, 146} One study reported that PET scans of patients with PD vs. atypical parkinsonism had different amounts of caudate ¹⁸F-Dopa uptake, blood flow, and glucose metabolism in the striatum,¹⁴⁶ while another reported that it was difficult to distinguish PD from atypical parkinsonian syndromes using PET scans.¹⁴² Two studies evaluated presence of striatal dopamine D2 receptors in PD patients. One reported that there were increased D₂ receptors in the putamen, but not the caudate, of patients with early PD.¹⁴¹ The other study reported that there were decreased D2 receptors outside the striatum in advanced, but not early, PD.¹³⁹ Further research is needed to evaluate the appropriate role of PET scans in diagnosis of PD.

Other Scans

Five studies presented results of imaging studies other than SPECT or PET scans. Three studies pertained to MRI scans,^{47, 147, 148} and one study each described results of proton nuclear magnetic resonance (NMR) spectroscopy¹⁴⁹ and transcranial color-coded real-time sonography (TCCS).¹⁵⁰ All studies were performed in Europe, although the NMR study was multicentric, and included sites in the United States. All studies were of cross sectional design.

The three studies of MRI scans in PD patients all reported their results differently. One study reported shortening of T₂ relaxation times in the substantia nigra, caudate, and putamen of PD patients compared with healthy controls, but there was some overlap between PD patients and controls, and the values did not correlate with disease severity.⁴⁷ One study showed significant differences in T₂ relaxation times between healthy controls and patients who had PD for greater than ten years, but not patients who had PD for shorter durations.¹⁴⁸ One study comparing MRI results in patients with PD, PSP, and MSA, showed hypointense and hyperintense signal changes in the putamen of nine of the 15 MSA patients, but none of the ten PSP or 65 PD patients, suggesting that this finding effectively rules out a diagnosis of PD.¹⁴⁷ These results suggest that MRI may be useful to rule out conditions other than PD, but are not useful in diagnosing PD.

The study on proton NMR spectroscopy showed no significant difference between patients and healthy controls, but subgroup analyses showed differences between elderly PD patients and controls and in treated vs. untreated patients.¹⁴⁹ There is insufficient evidence to determine the appropriate use of proton NMR spectroscopy to diagnose PD.

One study compared the results of TCCS in PD patients vs. normal controls.¹⁵⁰ The substantia nigra was undetectable by TCCS in 28 of 30 controls and 13 of 30 PD patients, because its echogenicity was identical to that of the adjacent brain tissue. Some PD patients with more severe disease had increased echogenicity of the substantia nigra, but the sensitivity of this

finding was only 40 percent, while the specificity was reported as 100 percent. There is insufficient evidence to determine the appropriate use of TCCS to diagnose PD.

Clinical Diagnosis

In current practice, the diagnosis of PD is made clinically. Two studies evaluated the accuracy of clinical diagnosis in PD. In one study, 402 patients who had been diagnosed with PD by general practitioners in North Wales were examined by PD specialists, using the UKPDS Brain Bank clinical diagnostic criteria outlined in Appendix A.¹⁵¹ Of the 402 patients, the diagnosis of parkinsonism was confirmed in 299 (213 with PD and 86 with possible PD or atypical parkinsonism), and 103 patients were found not to have parkinsonism (25.6 percent). The authors concluded that patients suspected of having parkinsonism should be referred early to a specialist for evaluation, given the apparent inaccuracy of clinical diagnosis by general practitioners.

There may be a measurable error rate for initial diagnosis by a specialist, as well. The 800 patients who had been involved in the Deprenyl and Tocopherol Antioxidative Therapy for Parkinson's Disease (DATATOP) study were observed for a mean of 6.0 ± 1.4 years (range 0.2 - 7.6 years).¹⁵² Of the 800 patients who had been diagnosed by experts as having PD, 65 (8.1 percent) subsequently received an alternative diagnosis.

Color vision discrimination in PD patients was evaluated in two studies. In one study, 16 patients with previously untreated PD and 16 age-matched controls were given color vision tests.¹⁵³ The PD patients had significantly worse color vision than the controls. In the second study, color vision was tested in 19 PD patients before and after treatment with L-dopa.¹⁵⁴ Nine control subjects were also tested twice, without L-dopa administration. Color discrimination improved significantly in PD patients after treatment with L-dopa, and did not change in the controls. Further testing is required to determine whether color vision testing is useful in diagnosing patients with PD.

In one study, visual evoked potentials (VEPs) were measured to compare visual impairment in 12 patients with PD, 12 with MSA, and 9 healthy controls.¹⁵⁵ The VEP patterns between PD and MSA were significantly different, leading the authors to speculate that VEPs might be useful in distinguishing PD from MSA.

Quantification of rigidity was reported in two studies. One study evaluated a computerized elbow device to quantify rigidity in 24 PD patients and 103 age-matched controls.¹⁵⁶ Basal (at rest) and activated (nontest arm performing flexion and extension exercises) rigidity values in both arms were significantly higher in PD patients than controls. In PD patients, activated rigidity values were higher than basal values, but the opposite was true in controls. In a similar study, basal and activated angular impulse scores, which reflect the relationship between change in force and time, were calculated in 20 PD patients and ten controls.¹⁵⁷ Angular impulse scores were significantly higher in PD patients than controls, and were higher with activation in patients, but not in controls. These studies suggest that objective measures of rigidity in PD patients may be useful in diagnosing and following disease progression in PD patients. Patients with atypical parkinsonism were not included in these studies.

Blood and CSF Tests

Three studies reported results of blood tests in PD patients. Peripheral blood lymphocyte (PBL) levels of dopamine were measured in 25 PD patients and 12 healthy controls.¹⁵⁸ PBL dopamine levels were significantly lower in untreated PD patients than in controls, and much higher in PD patients on L-dopa. The authors concluded that measuring PBL dopamine levels might be useful in diagnosing early PD.

A prospective double blind study of ³H-spiperone binding capacity to PBLs showed no significant differences in binding between patients with *de novo* PD, other parkinsonian syndromes, and healthy controls.¹⁵⁹

Plasma levels of pituitary and adrenal hormones were measured in 15 untreated PD patients and 12 healthy controls.¹⁶⁰ Integrated levels of adrenocorticotrophic hormone (ACTH), growth hormone (GH), and cortisol were significantly lower in PD patients than controls. Random prolactin (PRL) levels were nonsignificantly higher in PD patients, and nocturnal peak PRL levels were significantly higher in PD patients than controls. These results suggest abnormal pituitary function in PD patients, but do not provide useful tools for diagnostic testing in PD.

Two studies reported results of CSF analyses in PD patients. In one study, CSF somatostatin-like immunoreactivity (SLI) was measured in 15 patients with early PD, 8 with other forms of parkinsonism, and 26 controls.¹⁶¹ SLI was significantly higher in patients with PD than in controls. In patients with other forms of parkinsonism, SLI levels were higher, but the difference was not significant. When the PD patients were subgrouped according to degree of memory impairment, only patients with severe memory impairment had higher levels of SLI. Another study compared the CSF and plasma carnitine levels in 29 PD patients and 29 age-matched controls, and found no significant differences between patients and controls.¹⁶² There is insufficient evidence regarding the usefulness of testing CSF levels of somatostatin or carnitine in diagnosing PD.

Olfactory Tests

Seven studies, including 355 patients, 127 healthy controls, and 197 patients with other neurologic disorders, evaluated olfactory function in the diagnosis of PD. Two studies were performed in Japan,^{163, 164} two in the United States,^{165, 166} and one each in the United Kingdom,¹⁶⁷ France,¹⁶⁸ and Austria.¹⁶⁹ All were cross-sectional studies. All studies reported their results differently.

Three studies reported UPSIT results, but their reporting methods varied. One reported the number of PD patients with abnormal UPSIT scores (126 of 155 PD patients, or 81 percent).¹⁶⁷ Abnormal UPSIT scores were also reported in 11 of 72 patients with multiple sclerosis (15.3 percent), nine of 58 patients with motor neuron disease (15.5 percent), and eight of eight patients with AD (100 percent). The second study reported the bilateral UPSIT scores for 20 treated and 20 untreated PD patients compared with 20 controls.¹⁶⁵ Both treated and untreated patients had symmetrical significantly decreased olfactory function compared to controls. The third study¹⁶⁶ reported the mean UPSIT score for odor identification in 21 PD patients, 21 PSP patients, and 21 controls. The PD patients had significantly lower scores than the PSP patients or the controls.

The other studies of olfaction varied in their techniques of measurement. One study tested olfactory evoked potentials, and reported presence of olfactory dysfunction in ten of 20 PD patients, and zero of nine patients with AD.¹⁶⁴ One study reported the number of patients with

correct results in odor identification tests (OIT) and odor discrimination tests (ODT).¹⁶⁸ Thirty-seven of 80 PD patients had abnormal OIT results (46.3 percent), compared with five of 40 controls (12.5 percent). This difference was significant, but the percent of patients with abnormal ODT results did not differ significantly between PD patients and controls (28.0 vs. 16.4 percent). The authors concluded that PD patients have a defect in olfactory identification, but not in olfactory discrimination. Another study found that olfactory threshold and odor identification were significantly impaired in 21 PD patients compared with 19 controls, although there was no significant difference between the PD patients and 22 patients with AD.¹⁶⁹ One study compared odor detection threshold and recognition threshold, and found both to be significantly impaired in 18 PD patients compared with 10 controls.¹⁶³

Although all of the studies of olfactory function used different methods of measurement and reporting, they all were consistent in reporting that olfactory function is impaired in PD patients compared with healthy controls. There was not as much consistency in comparing results in patients with PD vs. atypical parkinsonism; therefore, there is insufficient evidence to support olfactory function testing to be used as a diagnostic tool at this time.

PD Test Battery

Three studies, all by the same author, evaluated the usefulness of a PD test battery in diagnosing PD.^{170, 171, 172} Two of these studies reported the performance of the PD test battery in PD patients,^{170, 171} and one reported its performance in first degree relatives of PD patients.¹⁷²

The PD test battery includes tests of motor function, olfaction, and depression. Results are combined in a logistic regression analysis into an equation that provides a "PD score," between 0 and 1.0. Scores ≤ 0.6 are considered to be suggestive of PD. The initial cross-sectional study, which evaluated 18 PD patients and 19 controls, showed a sensitivity of 94 percent and a specificity of 95 percent. The tests were then performed in a "validation group" of 103 PD patients and 122 controls, and the sensitivity and specificity decreased to 69 and 88 percent, respectively.¹⁷⁰

In a subsequent publication, the same authors performed the PD test battery on 205 patients with undiagnosed neurological conditions, and then followed the patients for at least one year, until they were diagnosed with a specific neurologic disease or determined to be neurologically normal.¹⁷¹ Fifty-nine of the 205 patients were subsequently diagnosed with PD, 106 with other neurologic diagnoses, and 40 were neurologically normal. Forty of the 59 PD patients had a PD test battery score consistent with the diagnosis of PD, and 37 of the 40 neurologically normal patients had a PD test battery score in the normal range. The authors reported that the test battery showed a sensitivity of 92 percent and a specificity of 68 percent, although they did not account for the 106 patients with other neurological diagnoses. These results are not adequate to consider the PD test battery to be the reference standard for PD diagnosis.

In one study, the PD test battery was administered to 78 asymptomatic first degree relatives of PD patients and 100 healthy controls.¹⁷² Eighteen of the 78 relatives and nine of the normal controls had abnormal PD test battery scores. These subjects would need to be followed over time to determine the predictive value of the test battery in asymptomatic people.

Autopsy

Six studies, including 253 patients with PD, 124 patients with other neurological disorders, and 76 controls without neurological disorders, reported results of autopsy data. Two of the studies were conducted in Canada,^{44, 173} two in the United States,^{174, 175} one in the United Kingdom,⁴² and one was multinational.¹⁷⁶

A commonly cited statistic is that up to 25 percent of patients with clinical diagnoses of PD are found to have different pathological diagnoses at autopsy. This stems from a study in which the brains from one hundred consecutive patients with clinically diagnosed PD were collected from sites within the United Kingdom, between 1987 and 1990.⁴² Neurologists had prospectively diagnosed the patients with PD. Autopsies of all 100 brains showed that 76 of the brains had a pathological diagnosis of PD, and 24 had been clinically misdiagnosed. The misdiagnoses included PSP, AD, MSA, vascular disease, isolated nigral atrophy without Lewy bodies, postencephalitic parkinsonism, and one case without abnormal findings. The authors then evaluated the sensitivity and specificity of various clinical manifestations of PD. If patients were required to have two of the three cardinal signs of PD for diagnosis (resting tremor, cogwheel rigidity, and bradykinesia), the sensitivity of clinical assessment was 99 percent, but the specificity was only eight percent. If they were required to have all three cardinal signs, the sensitivity decreased to 65 percent, but the specificity increased to 71 percent.

Another study reviewed autopsy results in 59 patients who had been clinically diagnosed with parkinsonian syndromes over a 22-year period.⁴⁴ The initial clinical diagnosis was PD in 43 patients, and decreased to 41 patients in the final diagnosis. The diagnosis of PD was confirmed by autopsy in 28 of the 43 patients with an initial diagnosis of PD (65 percent) and 31 of the 41 patients with a final clinical diagnosis of PD (76 percent). The remaining pathological diagnoses included MSA, PSP, neurofibrillary tangle parkinsonism, drug-induced parkinsonism, substantia nigra cell loss without inclusions, and Jakob-Creutzfeldt's disease. The authors concluded that the clinical diagnosis of PD is more accurate in patients who have been affected for more than five years.

One study retrospectively reviewed clinical data on 34 patients with a pathologic diagnosis of PD, and 31 patients with a pathologic diagnosis of diffuse Lewy body disease (DLBD).¹⁷⁵ Significant differences between the two conditions included older mean age of onset and presence of myoclonus for DLBD patients, and presence of rest tremor and clinical response to L-dopa in PD patients.

A retrospective chart review for history of falling was performed in 77 patients with autopsy-confirmed diagnoses of parkinsonian disorders.¹⁷⁶ Only 11 of these patients had PD; the remainder had PSP, MSA, DLBD, or corticobasal degeneration (CBD). The frequency of recurrent falls was similar in all groups, but falls occurring at the onset of parkinsonian symptoms were common in PSP and absent in PD and DLBD. These results suggest that it may be important to include questions about falling when taking a history in patients with suspected PD, in order to rule out atypical parkinsonism.

One study sought to show the absence of resting tremor in a large proportion of PD patients; however, the opposite result was noted.¹⁷³ The authors reviewed clinical data from 22 years of observation in 47 patients with pathologically-confirmed parkinsonism, of which 30 had a pathological diagnosis of PD. All 30 patients had been noted to have resting tremor at some point during their disease (100 percent), while only six of the 17 patients with other forms of parkinsonism had a history of resting tremor (35 percent). This suggests that most patients with

PD have rest tremor at some point during the course of PD, while this may not be the case for patients with alternative diagnoses.

Interrater reliability for diagnosing PD was tested in a clinicopathologic study.¹⁷⁴ Six neurologists analyzed 105 clinical scenarios of patients with diagnoses blinded: PD (n=15), DLBD (n=14), or neither (n=76). Diagnoses had been confirmed by autopsy. The neurologists reviewed the clinical vignettes extracted from records of the patients' first and last clinic visits. For each patient, the neurologists gave an initial impression, based on clinical judgment after the first visit, without laboratory or neuroimaging data, and a final diagnosis, based on all information available at the last clinic visit. The median sensitivity and specificity of clinical diagnosis of PD at the first visit were 73.3 (range 53.5-80.0) and 85.6 (range 74.4-94.4) percent, respectively. At the last visit, these values increased to 80.0 (range 60.0-86.6) and 92.2 (range 82.2-96.7) percent. The median positive predictive value (PPV) increased from 45.9 (range 34.2 to 61.5) percent at the first visit to 64.0 (range 42.8-75.0) percent at the last visit. The median negative predictive value (NPV) was over 95 percent at both visits. The authors concluded that the low PPV and relatively high sensitivity suggest overdiagnosis of PD by neurologists.

In summary, the autopsy studies showed clinical diagnosis of PD to have a modest degree of accuracy, which may be improved by following patients over time. Aside from autopsy, there is insufficient evidence that any diagnostic tests have sufficient sensitivity or specificity to qualify as reference standards for the diagnosis of PD.

Pharmacological Treatment of PD

Forty-nine parent studies, composed of 111 treatment arms and 9,968 patients, were accepted for the pharmacological treatment section of this project (See Evidence Table 2). Twenty-six were graded as level I evidence, and 23 were Level II. None were Level III-V. The mean quality score was 3.6 (where 5 is best), and the median was 4, suggesting moderate to high validity of studies in this set.

The mean age of all patients in pharmacological treatment groups was 63.0 years (t=97, n=8,605, range 55.4-80.0 years). As shown in Evidence Table 3, gender was reported in 36 studies (t=82, n=7,774); 59.6 percent of patients were male and 40.4 percent were female. Race was reported in only two studies, and the vast majority of patients in these two studies were Caucasian.^{66, 177}

The mean disease duration of all patients in pharmacological treatment groups that reported this parameter was 5.0 years (t=84, n=6,369, range 0.6 –13.6 years). Disease duration was distributed as follows: 23 treatment arms reported a mean disease duration of < 2 years, 26 reported 2-5 years, 14 reported 5-10 years, and 21 reported ≥ 10 years.

Seven studies reported "on" UPDRS scores only,^{66, 178 - 183} two reported both "off" and "on" scores,^{184, 185} one reported the average of "off" and "on" scores.¹⁸⁶ The remainder did not specify whether their UPDRS scores were "off" or "on."

Studies were divided into early or advanced PD, based on the study authors' classification or disease characteristics reported in the studies. Thirty-two studies (t=74, n=7,405) that referred to patients as having "early" or "*de novo*" PD, or mean disease duration < 5 years were classified as early. Seventeen studies (t=37, n=2,563) that reported patients with "advanced" PD, mean disease duration > 5 years, or patients who suffered from fluctuations and dyskinesias due to long-term L-dopa treatment were classified as advanced. It must be recognized that this categorization has limitations. Disease duration is useful in classifying individual patients as

having early vs. advanced disease, but mean disease duration of < 5 years does not mean that all patients have short disease duration. Studies that reported mean disease duration did not always report the range of disease duration in the individual patients.

Motor fluctuations are typically seen in patients with advanced PD; however some studies in which the patients were identified as having "early" PD reported motor fluctuations and "off" times. Given these limitations, three of the 32 studies that were classified as early,^{187, 188, 189} and five of the 17 studies that were classified as advanced^{185, 190, 192, 193, 194} may have actually contained mixed populations of patients with early and advanced disease.

Numbers of treatment arms with each treatment combination are shown in Evidence Table 5. Determining the number of studies in which L-dopa was used was problematic, because some studies used L-dopa as a comparator drug, while others merely allowed investigators to give patients L-dopa as needed. This was further complicated by the fact that in the studies where L-dopa was discretionary, the number of patients who received L-dopa, and their dosage of L-dopa, was frequently not reported. In 41 treatment arms (n=3,927), L-dopa was the only anti-Parkinson drug prescribed. These include treatment arms that were labeled as placebo arms, but patients received L-dopa as needed.

In studies of patients with early PD, it was often difficult to ascertain whether or not patients had previously taken L-dopa. This is important to distinguish, because it may be assumed that patients who never received L-dopa have less severe PD symptoms than patients who had received L-dopa. In 12 studies, inclusion criteria required that patients had never received L-dopa prior to study entry.^{119, 180, 195 - 204} In 15 studies, some patients may have received L-dopa prior to study entrance, but it was not always possible to ascertain which patients had been on L-dopa.^{65, 66, 82, 178, 181, 188, 205 - 213} In five studies of patients with early PD,^{187, 189, 214, 215, 216} and all 17 studies of patients with advanced PD, all patients were on L-dopa prior to study enrollment.

As shown in Evidence Table 5, treatment with DAs alone was reported in six treatment arms (n=508), MAO-B inhibitors alone in five treatment arms (n=336), and placebo alone in four treatment arms (n=374). The small number of pure placebo arms is due to the fact that most placebo groups in the pharmacological literature are groups in which patients received both active drug and placebo.

L-dopa was combined with DAs in 33 treatment arms (n=2,935), with MAO-B inhibitors in seven treatment arms (n=700), and with COMT inhibitors in eight treatment arms (n=639). Seven treatment arms contained other combinations, including L-dopa/DA/MAO-B inhibitor (t=2, n=68), DA/MAO-B inhibitor (t=1, n=10), α -dihydroergocryptine (α -DHEC; t=1, n=62), α -DHEC and L-dopa (t=1, n=10), L-dopa/vitamin E (t=1, n=202), and L-dopa/vitamin E/MAO-B inhibitor (t=1, n=197). No studies of amantadine or anticholinergic medications met the criteria for inclusion into this systematic review.

Rejected pharmacological studies of interest that were not accepted for this review, but were deemed to be of interest by the TEP, are discussed in Appendix J. These studies were rejected mainly due to publication date prior to 1990, inadequate study duration, or cross-over design, and include studies of anticholinergic medications, pramipexole, pergolide, tolcapone, selegiline, and GM1 ganglioside.

Pharmacological Treatment: Early PD

The 32 studies that focused on patients with early PD consisted of 74 treatment groups and 7,405 patients. There were 58.4 percent males and 41.6 percent females (t=59, n=5,670). Mean

age, weighted by sample size, was 62.6 years, and ranged from 55.4 to 80 years (t=68, n=6,969). Mean disease duration, weighted by sample size, was 2.3 years, and ranged from 0.6 to 4.2 years (t=53, n=4,015).

The studies that were classified as describing patients with early PD included:

- ten studies of bromocriptine (with or without L-dopa) vs. L-dopa^{188, 195, 199, 204, 205, 209, 214, 215, 216}
- one study comparing two different doses of bromocriptine¹¹⁹
- two studies comparing different formulations of L-dopa^{197, 198}
- four studies of selegeline plus L-dopa vs. L-dopa alone^{181, 207, 210}
- one study of bromocriptine vs. selegiline plus L-dopa vs. L-dopa alone vs. selegiline plus bromocriptine²⁰⁸
- two studies of selegeline vs. placebo^{200, 202}
- two studies of ropinerole plus L-dopa vs. L-dopa alone^{65, 212}
- one study of ropinerole plus L-dopa vs. bromocriptine plus L-dopa²⁰⁶
- one study of pramipexole vs. placebo²¹³
- one study of pramipexole plus L-dopa vs. L-dopa⁶⁶
- one study of cabergoline plus L-dopa vs. L-dopa alone²⁰³
- one study of lisuride plus selegeline vs. lisuride alone²⁰¹
- one study of pergolide plus L-dopa vs. L-dopa alone¹⁸⁰
- one study of lisuride plus L-dopa vs. L-dopa alone¹⁷⁸
- one study of L-dopa alone vs. tocopherol plus L-dopa vs. selegeline plus L-dopa vs. selegeline, tocopherol, and L-dopa⁸²
- one study comparing four doses of lazabemide vs. placebo²¹¹
- one study of L-dopa vs two doses of tolcapone plus L-dopa¹⁸⁹
- one study of α -dihydroergocryptine (ADHEC) vs. placebo¹⁹⁶
- one study of ADHEC plus L-dopa vs. L-dopa¹⁸⁷

Pharmacological Treatment: Advanced PD

The 17 studies that focused on patients with advanced PD consisted of 37 treatment groups and 2,563 patients. There were 62.7 percent males and 37.3 percent females (t=23, n=2,104). Mean age, weighted by sample size, was 64.0 years, and ranged from 56.0 to 75.8 years (t=29, n=2,232). Mean disease duration, weighted by sample size, was 9.6 years, and ranged from 5.5 to 13.6 years (t=31, n=2,354).

The studies that were classified as describing patients with advanced PD included:

- one study comparing two doses of bromocriptine plus L-dopa¹⁹²
- two studies of cabergoline plus L-dopa vs. placebo plus L-dopa^{185, 190}
- one study of cabergoline plus L-dopa vs. bromocriptine plus L-dopa¹⁹¹
- two studies comparing different preparations of L-dopa^{194, 217}
- two studies comparing pramipexole plus L-dopa vs. placebo plus L-dopa^{182, 184}
- one study comparing pramipexole plus L-dopa vs. bromocriptine plus L-dopa vs. placebo plus L-dopa¹⁸⁶
- one study comparing lisuride vs apomorphine²¹⁸
- one study comparing lisuride plus L-dopa vs L-dopa alone¹⁹³
- one study comparing pergolide plus L-dopa vs. placebo plus L-dopa¹⁸³
- one study comparing ropinirole plus L-dopa vs. placebo plus L-dopa²¹⁹
- two studies comparing entacapone plus L-dopa vs. placebo plus L-dopa^{177, 220}
- two studies comparing tolcapone plus L-dopa vs. placebo plus L-dopa^{86, 179}

Pharmacological Treatment of PD: Meta-analyses

Meta-analysis of DAs (plus L-dopa) vs. L-dopa alone. As shown in Evidence Table 6, seventeen studies provided sufficient data on one of the PD rating scales to calculate standardized mean differences between the pre-test/post-test change scores for a DA + L-dopa group and the pre-test/post-test change scores for an L-dopa group.^{65, 66, 178, 180, 182, 183, 185, 188, 190, 193, 195, 203, 208, 209, 212, 215, 216} Six studies investigated the effect of a DA+L-dopa versus L-dopa alone, but did not report enough data to allow the calculation of a pre-post effect size.^{184, 199, 204, 205, 214, 219}

Three of the studies in the meta-analysis examined patients naïve to L-dopa before the trial (all were studies of patients with early disease),^{180, 195, 203} seven of the studies examined patients with a mix of previous exposure to L-dopa (all early disease studies),^{65, 66, 178, 188, 208, 209, 212} and seven of the studies examined patients that were all previously exposed to L-dopa (all but two were studies of patients with advanced disease).^{182, 183, 185, 190, 193, 215, 216} Twelve of the studies in the meta-analysis examined treatment arms of patients with mean disease duration \leq five years (i.e., early disease),^{65, 66, 178, 180, 188, 195, 203, 208, 209, 212, 215, 216} and five studies examined treatment arms of patients with mean disease duration greater than five years (i.e., advanced disease).^{182, 183, 185, 190, 193} Six studies investigated bromocriptine,^{188, 195, 208, 209, 215, 216} while no other agonist was investigated more than three times.

In 11 of the studies in the meta-analyses, L-dopa was mandatory (i.e., patients were randomized to receive L-dopa or L-dopa plus a DA),^{178, 182, 183, 185, 188, 190, 193, 195, 209, 215, 216} and in five studies, the L-dopa was discretionary (i.e., patients were randomized to receive L-dopa or a DA, but L-dopa could be added at the practitioner's discretion).^{65, 66, 180, 203, 208}

A meta-analysis of differences between change in PD scores was performed. Figure 3 shows point estimates and 95 percent confidence interval error bars for the individual studies.

A fixed-effects meta-analysis showed that the change-score effect sizes (CHESs) were heterogeneous after sampling error was taken into account ($Q_E = 87.95, p < 0.001$). A random-effects model led to slightly positive but only marginally statistically significant effect for treatment with a DA + L-dopa versus L-dopa alone ($\hat{\delta} = 0.16, SE(\hat{\delta}) = 0.09$), where $\hat{\delta}$ represents the estimate of the mean effect size and $SE(\hat{\delta})$ represents its standard error. The 95 percent confidence intervals for this estimate are presented in Figure 3.

Given the presence of considerable heterogeneity, examination of study characteristics was warranted; however, multivariate analysis was made difficult by the presence of collinearity between the numerous predictors and the low number of studies. For instance, there was a strong relationship between stage of disease (early vs. advanced) and previous exposure to L-dopa. There was also a strong (yet unexplained) correlation between stage of disease and whether the DA investigated was bromocriptine; there were no studies in which bromocriptine was used and patients had a long disease duration. To simplify interpretation, univariate analyses for each study characteristic were conducted.

There were two study characteristics that were statistically significant: time of evaluation ($p = 0.009$) and type of L-dopa delivery ($p < 0.001$) (discretionary vs. mandatory). Studies with a duration of greater than one year had significantly lower effect sizes than studies with a duration of less than one year; studies with discretionary L-dopa delivery had significantly lower effect sizes than studies in which L-dopa delivery was mandatory in the dopamine agonist arm.

These findings suggest two different mechanisms at work. The first suggests that the effect of treatment with a DA+L-dopa, relative to L-dopa alone, may decline over time. The latter suggests that treatment which mandates L-dopa as an adjunct to a DA controls PD symptoms better than treatment which merely allows doctors to give L-dopa when they think it might be needed.

While the mechanisms are different, their individual impacts cannot be measured in this meta-analysis; these two variables were highly correlated ($r = .60, p = .015$), making it very difficult to separate their respective influences. Among the studies in which L-dopa delivery was mandatory, there was a mix of short- and long-term study durations; however, when L-dopa delivery was discretionary, the only studies present were long-term ones.

Meta-analyses within each level of each key study characteristic (e.g., a meta-analysis of studies with short disease duration, a meta-analysis of studies with long disease duration, a meta-analysis of studies with *de novo* patients, etc.) are presented in Figure 3. Studies in which the DA used was bromocriptine had a higher average effect size than those in which another DA was investigated. Studies of patients with advanced PD had a higher average effect size than those with patients with early PD, and studies with non-*de novo* patients had a higher average effect size than studies with *de novo* patients or patients with a mixed background. However, none of these differences were statistically significant.

Two groups of sensitivity analyses were conducted. In the first, three design characteristics (whether LOCF measurements were used, whether the study was blinded, and whether the study effect was known to be “on” or whether it was merely assumed to be “on”) were investigated univariately. None were close to being statistically significant ($p > 0.30$ for each). In the second group of sensitivity analyses, re-analyses of the data were conducted to determine whether any particularly large or small effect sizes were having an unbalancing effect on the overall results. The first set of re-analyses deleted the largest effect size from one study,²¹⁵ and the second set deleted the smallest effect size from a different study.⁶⁵ All meta-regressions using the new subsets of studies found substantively what the initial meta-regressions found: a significant negative effect for treatment duration, suggesting that DAs work better in studies of short duration.

There were three studies in which head-to-head comparisons between bromocriptine and other DAs were performed. The comparator drugs were cabergoline,¹⁹¹ ropinirole,²⁰⁶ and pramipexole.¹⁸⁶ All studies used L-dopa as a supplementary treatment. Two studies were in patients with advanced disease^{186, 191} and one in patients with early disease.²⁰⁶ Effect size could only be calculated for two of the studies. The average effect size (positive indicating that bromocriptine performed better) was not significantly different from zero.

Conclusions of DA+L-dopa vs. L-dopa meta-analyses. There is no evidence that different DAs vary in treatment effects. Meta-analysis suggests that in early PD, treatment with DAs plus L-dopa may control PD symptoms better than treatment with L-dopa alone, but this was not a consistent finding. However, given the wide heterogeneity in this small group of studies in type of treatment, focus of treatment, duration of treatment, and patient characteristics, it would be very difficult to detect such effects if they indeed existed.

Meta-analysis of selegiline (plus L-dopa) vs. L-dopa alone. Evidence Table 7 shows the three studies that compared the effect of selegiline and L-dopa versus the effect of L-dopa alone in patients with early PD.^{181, 207, 208} All studies looked at patients with short disease duration. Figure 4 shows point estimates and 95 percent confidence interval error bars for the individual studies, as well as for the overall meta-analysis.

A fixed-effects meta-analysis showed that the CHESs were heterogeneous ($Q_E = 11.79$, $p = 0.003$). A random-effects model led to a moderate sized estimate of mean effect that was statistically insignificant ($\hat{\delta} = 0.47$, $SE(\hat{\delta}) = 0.25$). Due to the small numbers of studies involved, reliable examination of the impact of study characteristics on treatment efficacy and sensitivity analysis were not possible.

Conclusions of Selegiline+L-dopa vs. L-dopa meta-analyses. While there is some evidence to suggest that selegiline + L-dopa may work better than L-dopa alone in controlling symptoms of PD, the difference between the two therapies in efficacy in controlling PD symptoms was statistically insignificant. However, the power to detect a difference between these two therapies was very small given the low number of studies involved and the wide variation between them.

Studies not included in selegiline meta-analysis. One study investigated the efficacy of the MAO-B inhibitor lazabemide relative to the efficacy of a placebo (as opposed to an L-dopa treatment).²¹¹ The average lazabemide effect was 0.302 ($p<0.05$), suggesting that lazabemide performed somewhat better than placebo. Two studies examined the effect of selegiline versus the efficacy of placebo.^{200, 202} In these placebo-controlled trials, the primary efficacy outcome was time until L-dopa treatment was required. In both cases, patients in the placebo arms needed L-dopa sooner than patients in the selegiline treatment arms.

In the DATATOP study, 800 patients with early PD were randomized to receive selegiline, vitamin E, the combination, or placebo.⁸² The primary endpoint was the time when L-dopa was required. Selegiline significantly delayed the L-dopa requirement, while vitamin E showed no evidence of benefit. This study could not be included in the meta-analysis because UPDRS scores at uniform time points prior to starting L-dopa were not available in the published literature.

Meta-analysis of COMT-inhibitors (plus L-dopa) vs. L-dopa alone. Five studies compared the effect of COMT inhibitors with L-dopa versus the effect of L-dopa alone in patients with PD.^{86, 177, 179, 189, 220} Except for one study,¹⁸⁹ all were in the setting of advanced disease. As shown in Evidence Table 8, three of the studies provided a pair of effects (each for a different dose of tolcapone: 100mg and 200mg). The remaining two studies investigated entacapone as a treatment. A total of six of the eight study arms investigated patients with disease duration of > 10 years. Figure 5 shows point estimates and 95 percent confidence interval error bars for the individual studies, as well as for the overall meta-analysis.

The fixed-effects meta-analysis of CHESs were homogeneous ($Q_E = 5.32, p=0.62$). The estimate of the mean was 0.33 ($SE(\hat{\delta}) = 0.056$), which was statistically significantly greater than zero ($p<0.001$). Error bars for each meta-analysis are presented in Figure 5.

There was little variance on most of the study characteristics, making the investigation of the impact of study duration, severity of disease, and baseline exposure of patients to L-dopa difficult. It is likely that this minimal variation in study characteristics contributed to the reason that the effect sizes were so homogenous. There was no statistically significant difference between the effect sizes of the three 100 mg tolcapone arms and the three 200 mg tolcapone arms ($p>0.50$), and no statistically significant difference between the effect sizes of the two arms in which average disease duration was less than five years and the six arms in which average disease duration was greater than ten years ($p=0.25$).

The significantly positive mean CHES for the COMT inhibitors remained significantly positive after the largest change-score effect was deleted ($p<0.001$).

Conclusions of COMT+L-dopa vs. L-dopa meta-analyses. Unlike the meta-analyses of the DAs and selegiline, the studies in the meta-analysis of treatment with COMT were very homogeneous with regard to disease duration, treatment duration, and baseline exposure of

patients to L-dopa. It is likely that this led to the consistent moderately positive effect for COMT + L-dopa versus L-dopa. It can safely be concluded that in the short term (\leq seven months), patients with advanced PD who receive combination treatment with COMT and L-dopa can expect a reduction in PD symptoms substantively greater than similar patients who are treated with L-dopa alone. Most of the studies examined the COMT inhibitor tolcapone; however, there was no evidence that tolcapone treatment was better or worse than treatment with entacapone. Also, there was no evidence that treatment with 200mg of tolcapone alleviated symptoms more than treatment with 100mg of tolcapone.

Due to reports of hepatotoxicity associated with tolcapone, the three tolcapone studies were reviewed for mention of hepatotoxicity.^{86, 179, 189} Of the 451 patients taking tolcapone, 16 patients were reported to have transient elevations in LFTs (3.5 percent), leading to withdrawal of six patients from the three studies (1.3 percent). Of the 227 patients on the lower dose of tolcapone (100 mg per day), seven patients had reports of elevated LFTs (3.1 percent), and one patient withdrew from the study (0.4 percent). Of the 224 patients on the higher dose of tolcapone (200 mg per day), nine patients had reports of elevated LFTs (4.0 percent), and five patients withdrew from the studies (2.2 percent). All of the COMT studies had a duration of seven months or less; therefore, no conclusions may be made regarding long-term safety.

Pharmacological Treatment of PD: On-Off Time

"On" and "off" time were captured, when reported, but these results were not amenable to meta-analysis because they were reported in non-standardized ways. While motor fluctuation assessment is an important component of determining optimal treatment of PD, the variation in methods of reporting this parameter precluded pooling results from different studies.

Pharmacological Treatment of PD: L-Dopa Doses

When L-dopa is used in combination with another drug to treat PD, one measure of efficacy is the dose of L-dopa required by patients on combination therapy. Where both baseline and outcome mean L-dopa doses were reported, the L-dopa doses prior to and after treatment were compared. In studies comparing DAs plus L-dopa to L-dopa alone, the mean daily L-dopa dose in the DA/L-dopa arms decreased from 624.7 mg (range 250.0 - 1,305.8) to 488.5 mg (range 136.0 – 940.4), whereas in the L-dopa monotherapy arms, the mean L-dopa dose decreased minimally, from 608.8 (range 242.7-940.4) to 594.0 mg (range 306.0 – 889.0).

In studies comparing COMT inhibitors plus L-dopa to L-dopa alone, the mean daily L-dopa dose in the COMT inhibitor/L-dopa arms decreased from 621.6 (range 270.6 – 865.8) to 514.8 mg (range 249.8 – 658.7), whereas in the L-dopa monotherapy arms, the mean daily L-dopa dose increased from 669.5 (range 364.3 – 948.0) to 681.9 mg (range 410.9 – 963.5).

In studies comparing selegiline plus L-dopa to L-dopa alone, many of the patients were not on L-dopa prior to the study. Therefore, the mean baseline dose of L-dopa is not available, but after treatment, the mean daily L-dopa dose in the selegiline/L-dopa arms was 388.0 mg (range 356.0 – 424.0), while in the L-dopa monotherapy group, it was 478.5 mg (range 426.0 – 543.0). All of these results suggest that combination therapy has greater efficacy than L-dopa alone in lowering L-dopa doses.

Pharmacological Treatment of PD: Dyskinesia Scores

While dyskinesia scores are of interest, particularly in patients with advanced PD, they could not be analyzed because only two pharmacologic studies reported dyskinesia scores.^{191, 193} Other studies described dyskinesias in a more qualitative fashion, which did not allow for evaluation with meta-analytic methods.

Pharmacological Treatment of PD: Safety

Heterogeneity in methods of reporting safety outcomes leads to imprecision in summarizing data from multiple studies. Some studies reported safety outcomes in terms of numbers of patients, while others reported numbers of adverse events (AEs). These values are not interchangeable, as one patient may suffer more than one event. Other studies reported only the most common or the most severe events. Numbers of AEs are also affected by the aggressiveness of the methods by which the investigators identify events. If investigators specifically ask about a particular AE (i.e., active monitoring), they are more likely to discover it than if they wait for patients to volunteer the information (i.e., passive monitoring).

For the purposes of this summary data, only AEs reported in terms of numbers of patients (not events) have been captured, except when zero or one event was reported, in which case zero or one patient was substituted, respectively.

The number of deaths, withdrawals, and most common AEs, classified by body system, are listed in Evidence Table 9. While withdrawals occurred more commonly for issues of safety rather than lack of efficacy, no studies reported treatment-related deaths. The table lists the most common or clinically important AEs, but is by no means comprehensive.

Due to the frequency and clinical relevance of neurological and psychiatric AEs, these have been reported separately, along with the incidence of the most common neurological and psychiatric symptoms (Evidence Table 10). Overall, the most common neurological AEs reported were aggravation of PD, dyskinesias, and akinesia. Sleeping disorders were the most commonly reported psychiatric AEs reported. Gastrointestinal AEs were the most common non-neurological, non-psychiatric AEs reported. As L-dopa was given concomitantly in most groups, it is difficult to separate the L-dopa AEs from those caused by other drugs.

Dizziness was more common in advanced than early disease (incidence 22.5 vs. 17.4 percent), as was dyskinesia (incidence 35.3 vs. 17.7 percent) and PD aggravation (33.7 vs. 15.4 percent). Thus, while efficacy did not differ between early and advanced disease, AEs were much more frequent in treatment groups of patients with advanced disease.

Surgical Treatment of PD

There were 42 parent and 16 kin studies concerning surgical treatment, encompassing 52 treatment arms and 1,380 patients (See Evidence Table 2).

The vast majority of studies were UCSs (k=35, n=1,145), and the remainder were RCTs (k=4, t=8, n=105), nRCTs (k=2, t=7, n=117), and one case-control retrospective study (t=2, n=13). Thirty-eight of the studies were graded as level III evidence, and four were level II. Quality score could be calculated only for the four RCTs, and was a mean of 3, reflecting moderate quality.

As shown in Evidence Table 3, gender was reported in 32 of the 42 surgery studies (t=40, n=891). There were 573 males (64.3 percent) and 318 females (35.7 percent). Mean age was reported in 41 studies (t=51, n=1,336). The mean age, weighted by sample size, was 60.8 years, with a range of 46.5 to 73.3 years. In 41 treatment arms, mean age was reported to be < 65, and ten treatment arms reported a mean age of ≥ 65, suggesting that younger patients tended to be enrolled in surgery trials. Disease duration was reported in 34 studies (t=43, n=1,058). The mean disease duration was 12.8 years, and ranged from 4.8 to 17.5 years. This was not unexpected, given that surgery is generally not performed until patients have become intolerant to medical therapy. Mean age of disease onset was reported in six treatment arms; three reported mean age of disease onset ≥ 50 years (n=102), and three reported mean age of disease onset < 50 years (n=49).

Treatment level characteristics of surgical studies are summarized in Evidence Table 11. Pallidotomy was evaluated in 20 treatment arms (n=764), thalamotomy in five treatment arms (n=134), DBS in 16 treatment arms (n=288), and tissue transplant in nine treatment arms (n=165). No surgery was performed on patients in two treatment arms (n=29).

Thirteen studies reported dyskinesia scores (t=16, n=426). This is an important outcome to assess in surgical patients, because patients commonly undergo surgery to reduce medication complications, such as dyskinesias. As dyskinesia is a medication side effect, dyskinesia scale results are reported in the "on" state. Pooling of these scales across studies is problematic, because studies reported different variations of the scales, the scales were not always defined, and standard deviations were generally not reported. Nearly all treatment arms showed improvement in mean dyskinesia scores, particularly contralateral scores, after surgery.

Scores that were reported less frequently include Beck Depression Inventory (BDI; t=4, n=62), postural instability and gait disturbances (PIGD; t=3, n=102), Perdue Pegboard Test (t=2, n=46), and Webster Score (t=1, n=12).

Two studies included concurrent control groups of patients who did not undergo surgery (n=29).^{221, 222} In both of these control groups, baseline UPDRS scores were lower (better) than post-study scores, the opposite pattern from that seen in all surgical groups, suggesting that patients who did not undergo surgery deteriorated clinically.

Timing of post-surgical evaluation may affect results, and may give an indication of duration of postoperative benefit. Approximately half of the surgical treatment arms reported results at less than one year, and the other half reported results at greater than one year. No consistent pattern emerged in comparing earlier vs. later results.

"On-off" time was of interest, because surgical patients have advanced PD, where motor fluctuations are particularly problematic; however, it was only reported in ten studies, using inconsistent methods of measuring this phenomenon.

Pallidotomy

Results of pallidotomy were reported in 20 treatment arms (n=764).²²¹⁻²⁴⁰ There were 16 treatment arms in which patients underwent unilateral procedures (n=491) and four in which patients underwent a mixture of unilateral and bilateral procedures (n=273; 107 were bilateral procedures).^{222, 228, 230, 240} Of the four studies that included results on patients with bilateral pallidotomies, two did not distinguish the results of the patients with unilateral vs. bilateral procedures,^{228, 240} one reported that there was no difference in outcome between the patients with

unilateral and bilateral procedures,²³⁰ and one reported that the patients who underwent bilateral procedures had symptoms of greater severity prior to the pallidotomies, and did not improve significantly after the procedures.²²²

Mean L-dopa dose in the nine pallidotomy studies that reported both baseline and outcome doses did not change significantly. The mean dose at baseline was 923.8 mg (range 545 – 1,125), and at endpoint was 921.6 mg (range 627 – 1,174).

Five pallidotomy studies reported "on-off" time. One study reported the number of patients with shorter "off" periods,²²⁴ one study reported the mean "off" time,²²⁸ one reported the number of patients whose "off" time improved, worsened, or was unchanged,²³⁴ one reported the percent of hours "on" and "off,"²³¹ and one used a study-specific scale, which was not defined.²³⁰ These results could not be pooled in a meaningful way, but all of these studies reported overall improvement in this parameter.

Thalamotomy

Thalamotomy was described in five treatment arms (n=134),^{100, 222, 241, 242, 243} the vast majority of which were unilateral procedures. Most studies of thalamotomy were published prior to 1990; therefore, this database, with its search cut-off date of 1990, contains limited information regarding this procedure.

Very few treatment arms reported UPDRS, S&E, or H&Y scores in thalamotomy studies, and only one reported both preoperative and postoperative scores. The low number of studies that reported these parameters prevents any conclusion about the efficacy of thalamotomy to be drawn based on PD rating scales. All studies, however, reported overall improvement in tremor.

Deep Brain Stimulation

DBS was reported in 16 treatment arms (n=288), including stimulation of subthalamic nuclei (STN; t=8, n=135),^{244 - 250} globus pallidus (GPi, t=4, n=22),^{244, 246, 248} and thalamic nuclei (t=4, n=131).^{100, 251, 252, 253}

L-dopa doses decreased significantly after DBS: from a baseline daily mean of 1,018.8 mg (range 442 – 1,560 mg) to an endpoint mean of 455.2 mg (range 262 – 1,110 mg) per day. When the DBS studies were divided by nucleus location, the mean daily L-dopa dose decreased from 1,208.5 mg (range 729 – 1,560 mg) to 555.3 mg (range 262 – 850 mg) in the STN DBS groups (t=8), and increased from 863.0 mg (range 856 – 870 mg) to 1006.5 mg (range 903 – 1,110 mg) in the GPi DBS groups (t=2). Mean pre- and post-DBS L-dopa doses in the thalamic DBS groups were only reported in one study, and decreased slightly, from 649 to 610 mg per day.

Five DBS treatment arms reported mean dyskinesia scores, which improved after surgery in all cases.²⁴⁶⁻²⁵⁰ One STN study reported transient exacerbation of dyskinesias in the first postoperative weeks, which resolved with decreasing the dose of L-dopa and increasing the voltage of the stimulation.²⁴⁸

Tissue Transplants

Cell transplants were described in nine treatment arms (n=165), and included three groups of adrenal medulla transplants (n=91),^{254, 255, 256} five groups of human fetal brain cell transplants (n=62),^{257 - 260} and one group of porcine fetal brain cell transplants (n=12).²⁶¹ Results from the

studies of adrenal transplantation^{254, 255, 256} are not addressed in this report, as this procedure is no longer performed in PD patients, due to lack of efficacy and substantial morbidity.⁹⁷ Due to the small number of transplant studies, drawing conclusions regarding the efficacy of transplantation is problematic. One important study that was published too late to meet the inclusion criteria for this systematic review was an RCT comparing the outcomes of human embryonic tissue transplantation to sham surgery.¹⁰³ Some clinical improvement was noted in patients ≤ 60 years of age, but not in older patients. This study, which was notable for the development of late dystonias and dyskinesias in the active treatment arm, is discussed in detail in Appendix J.

Surgical Treatment: Meta-analyses

Pallidotomy. Fifteen pallidotomy treatment arms provided sufficient pre-post data on any of the PD rating scales to calculate pre-post standardized mean differences for “off” scores, and 12 treatment arms provided sufficient pre-post data to calculate standardized mean differences for “on” scores (Evidence Table 12). Figures 6 and 7 show point estimates and 95 percent confidence interval error bars for the individual study effect sizes for “off” and “on” scores, respectively.

Meta-analyses - “Off” effects. A fixed-effects meta-analysis showed that the “off” effect sizes were heterogeneous ($Q_E = 37.94, p < 0.001$). A random-effects model suggested that pallidotomy is effective in reducing “off” scores ($\hat{\delta} = 0.77, SE(\hat{\delta}) = 0.12$). Given the heterogeneity of effects, examination of study characteristics was warranted. Univariate meta-regressions were conducted which investigated three study characteristics: time since surgery (\leq one year vs. $>$ one year), average age of the participants, and average patient H&Y score at baseline. No predictors were statistically significant, although there was a marginally significant ($p = 0.095$) effect for time since surgery; the estimated effect size of pallidotomy on PD scale scores was 0.87 in studies with a duration of one year or less, and 0.36 in studies with a duration greater than one year. This suggests that pallidotomy may be effective mainly for the first year, but the number of long-term studies is too limited to make more than a tentative conclusion. The 95 percent confidence intervals for estimates of the mean effect size are in Figure 6.

Meta-analyses - “On” effects. A fixed-effects meta-analysis showed that the “on” effect sizes seemed homogeneous ($Q_E = 11.18, p = 0.43$). As mentioned previously, our rule was to use a random-effects model if the estimate of random-effects variation was greater than zero. In this case, the random-effect results are almost identical to those of the fixed-effects model. The random-effects model led to a small and statistically insignificant estimate of average effect for pallidotomy “on” scores ($\hat{\delta} = 0.13, SE(\hat{\delta}) = 0.08$). Univariate meta-regressions were conducted for three study characteristics: time since surgery (\leq one year vs. $>$ one year), average age of the participants, and average patient H&Y score at baseline. No study characteristics explained the significant amount of variation.

Sensitivity analyses were conducted to determine if any of the results seemed dependent on the inclusion of any one study. The only finding was that the effect of time since surgery was statistically significant if the largest effect size was deleted.²²⁹ The effect in this study is quite large (0.93), as opposed to the effects from the two other studies with surgery followups exceeding one year; the effect in one is 0.20,²²⁸ and the effect in the other is 0.10.²²³ The

followup times in these three studies were 16, 24, and 48 months, respectively. It might be that the effect of pallidotomy on “off” scores does not decline until after at least a year and a half (or more) has passed. This result is suggestive at best, but may be worthy of future within-study investigation. The 95 percent confidence intervals for estimates of the mean effect size are in Figure 7.

Deep Brain Stimulation. Fourteen DBS treatment arms provided sufficient pre-post data on any of the PD rating scales to calculate pre-post standardized mean differences for “off” scores, and eight treatment arms provided sufficient pre-post data to calculate standardized mean differences for “on” scores (Evidence Table 13). Figures 8 and 9 show point estimates and 95 percent confidence interval error bars for the individual studies for “off” and “on” scores, respectively.

Meta-analyses - “Off” effects. Given that DBS takes place at three different sites (the GPi, the STN, and the thalamus), three separate meta-analyses were conducted. The “off” scores of the four studies of DBS of the GPi were very homogeneous ($Q_E = 0.88, p > 0.50$). The fixed-effects estimated average effect of DBS-GPi on PD scale scores was significant and very large ($\hat{\delta} = 1.31$ ($SE(\hat{\delta}) = 0.33$)). The eight studies of DBS of the STN were quite heterogeneous ($Q_E = 55.68, p < 0.001$). A random-effects model led to a large and statistically significant estimate of average effect ($\hat{\delta} = 2.00$, $SE(\hat{\delta}) = 0.47$). Finally, the two studies of thalamic DBS seemed homogenous ($Q_E = 0.27, p > 0.50$), but the fixed-effects estimated average effect was near zero ($\hat{\delta} = -0.08$, $SE(\hat{\delta}) = 0.16$). This latter result is not surprising, as DBS of the thalamus is generally done for different reasons than DBS of the other two sites, i.e., it is not done to control severe PD, but to control tremor.⁹⁷

Three study characteristics (time since implantation of the DBS device, average age of patients, and average H&Y score at baseline) were explored as possible explanations for the heterogeneity in the STN effect sizes. Three univariate meta-regressions were conducted to test whether any of these characteristics explained significant variation; however, no characteristics did ($p > 0.05$). Sensitivity analyses were conducted to investigate whether any one study effect might be responsible for the excess variation; however, this was not the case.

Overall, meta-analyses of “off” effects showed that DBS led to significant improvement in “off” scores when performed on GPi or STN, and no significant change when performed on thalamic nuclei.

Meta-analyses - “On” effects. Two meta-analyses were conducted (there were no reports of “on” scores for thalamic DBS). The two studies of “on” effect sizes for DBS of the GPi were somewhat homogeneous ($Q_E = 1.56, p = 0.21$). A random-effects model led to a mean effect size near zero ($\hat{\delta} = 0.01$, $SE(\hat{\delta}) = 0.61$). The eight studies of DBS of the STN were heterogeneous ($Q_E = 14.96, p = 0.01$). A random-effects model led to a statistically significant estimate of average effect ($\hat{\delta} = 0.79$, $SE(\hat{\delta}) = 0.30$).

Three study characteristics (time since implantation of the DBS device, average age of patients, and average H&Y score at baseline) were explored as possible explanations for the heterogeneity in the STN effect sizes. Three univariate meta-regressions were conducted to test whether any of these characteristics explained significant variation; however, no characteristics did ($p > 0.05$). Sensitivity analyses were conducted to investigate whether any one study effect

might be responsible for the excess variation. The analyses showed that the “on” effect from one study (2.47) was causing most of the heterogeneity;²⁴⁵ a re-analysis without this effect size resulted in a meta-analysis showing little heterogeneity ($Q_E = 4.63, p=0.33$). The mean effect size ($\hat{\delta}=0.49$) was just short of being statistically significant ($p=0.06$). This suggests the possibility that DBS does not significantly impact “on” scores; a finding which would not be surprising, given that surgery is only performed on patients who are responsive to medication.

Tissue Transplant. Four tissue transplant treatment arms provided sufficient pre-post data on any of the PD rating scales to calculate pre-post standardized mean differences for “off” scores, and five treatment arms provided sufficient pre-post data to calculate standardized mean differences for “on” scores (Evidence Table 14). Figures 10 and 11 show point estimates and 95 percent confidence interval error bars for the individual studies for “off” and “on” scores, respectively.

Meta-analyses - “Off” scores. Meta-analysis showed that the “off” score effect sizes may be homogeneous ($Q_E = 3.18, p=0.36$). However, given the low number of studies in these meta-analyses, the power to detect heterogeneity was very low. To be conservative, a random-effects model was employed. Random-effects modeling shows a positive and statistically significant benefit for tissue transplants ($\hat{\delta}=0.88, SE(\hat{\delta}) = 0.21$). The 95 percent confidence intervals for these estimates are presented in Figure 10.

Meta-analyses - “On” scores. A fixed-effects meta-analysis showed that the “on” score effect sizes were heterogeneous after sampling error was taken into account ($Q_E = 9.36, p=0.05$). A random-effects model suggested a positive and statistically significant effect for tissue transplants ($\hat{\delta} = 1.09, SE(\hat{\delta}) = 0.34$), indicating that tissue transplants led to improvement in “on” scores. There were not enough studies to investigate whether variation in study characteristics might be responsible for the heterogeneity in the study effects (See Figure 11).

Conclusions of Surgery Meta-Analyses. Pallidotomy resulted in significant improvement in “off” scores and insignificant improvement in “on” scores. DBS of GPi and STN resulted in significant improvement in “off” scores, but no significant change in “on” scores. Thalamic DBS resulted in no significant change in “off” or “on” scores. Fetal cell transplantation resulted in significant improvement in both “off” and “on” scores.

The mean “off” effect size for pallidotomy (0.77) is lower than the mean effect size for DBS of the GPi (1.31). This implies that DBS of the GPi may be better than pallidotomy in controlling PD symptoms in the “off” state. Without head-to-head RCTs, however, any conclusions are tentative at best. It is also worthwhile to note that there are other possible benefits to surgery, such as reduction of dyskinesias and motor fluctuations, that could not be investigated in these meta-analyses. Finally, while current results of tissue transplantation are promising, too few studies have been done on fetal brain surgery to make any more than tentative conclusions about its effectiveness, and the recent RCT comparing tissue transplant to sham surgery raised important questions regarding the long-term safety of the procedure.¹⁰³

Surgery: Safety

Due to missing data and heterogeneity in methods of reporting AEs, summarization of surgical safety data suffers from the same limitations as summarization of pharmacological safety data. For the purposes of this summary data, only AEs reported in terms of numbers of patients (not events) have been captured, except when zero or one event was reported, in which case zero or one patient was substituted, respectively.

Evidence Table 15 lists the most common or clinically important AEs, but is by no means comprehensive. Transient AEs were not captured, as it was believed that decisions regarding the safety of surgery would be based mainly on long-term outcomes, not on transient perioperative complications.

Eighteen treatment arms reported the occurrence or absence of treatment-related deaths, which were uncommon except in tissue transplant studies, in which an 8.2 percent incidence of treatment-related deaths was reported. All of these deaths, however, were in adrenal transplant groups, and this procedure is no longer performed.

Reported AEs were primarily neurological or psychiatric, and included speech disorders (t=12, n=362, incidence = 6.1 percent); motor abnormalities (t=12, n=450, incidence = 5.6 percent); visual disturbances (t=8, n=320, incidence = 3.4 percent); depression (t=4, n=143, incidence = 6.3 percent); confusion, hallucinations, or psychosis (t=11, n=379, incidence = 4.2 percent); and dementia or impaired intellect (t=4, n=99, incidence = 5.1 percent). Some studies reported cerebral hemorrhage (t=11, n=266, incidence = 6.0 percent) and cerebrovascular events (t=4, n=106, incidence = 6.6 percent). The highest incidence of neurological AEs was reported in thalamotomy treatment groups (15.3 percent), but as these numbers were based on only two studies and 59 patients, the clinical relevance is unclear.

Psychiatric Treatment

There were ten accepted studies and one kin concerning treatment of psychiatric disorders in patients with advanced PD. Six studies were UCSs (n=114), two were RCTs (n=57), and two were retrospective observational studies (n=221). Eight studies were graded as level III evidence, and two were level II. Quality score could only be calculated for two studies, and was four in one study²⁶² and two in the other.²⁶³

Six accepted studies evaluated the efficacy of the atypical antipsychotic medication clozapine in managing PD patients with psychosis.^{264 - 269} None were RCTs. Four were UCSs (n=93), lasting from 12 to 24 months.^{264 - 267} Over seventy-five percent of patients in these studies demonstrated improvement in their psychotic behavior on a daily dose of 6.25 to 150 mg clozapine. The main adverse events reported were sialorrhea (reported in zero to 59 percent of patients), sedation (reported in two to 53 percent of patients), and confusion (reported in zero to 82 percent of patients). There were no reported cases of agranulocytosis. In the two retrospective reviews, charts of 221 PD patients on clozapine for control of psychotic symptoms were reviewed.^{268, 269} Patients received clozapine for one to 76 months (mean duration 15.2 months). There was a decrease in the number of patients with agitation, delirium, delusions, dementia, depression, visual hallucinations, psychosis sundowning, insomnia, vivid dreams, daytime napping, disorientation, memory loss, and abulia, although none of these symptoms resolved completely in all patients. Forty-six of 221 patients (20.8 percent) withdrew from the drug due to

adverse events. The most common adverse events were somnolence, amnesia, delirium, sialorrhea, and orthostatic hypotension. Granulocytopenia was reported in six patients, but all resolved with discontinuation of the drug, and there were no cases of agranulocytosis.

The efficacy and safety of risperidone,²⁷⁰ quetiapine,²⁷¹ piracetam,²⁶³ and citalopram²⁶² were evaluated in one study each. In a UCS of ten patients with advanced PD, cognitive decline and psychiatric symptoms, patients were treated with low doses of the atypical antipsychotic drug risperidone for 16 to 48 weeks (mean 34.8 weeks).²⁷⁰ While there was improvement in psychiatric symptoms in most subjects, two patients discontinued risperidone due to worsening of parkinsonism, and two developed delirium. The small size and uncontrolled design of this study does not allow conclusions to be drawn regarding the efficacy and safety of risperidone.

Quetiapine, another atypical antipsychotic, was openly administered for 12 months to 11 PD patients with psychosis.²⁷¹ Only five of the 11 patients completed one year of treatment. Withdrawals were due to dizziness, falling, obstipation, cerebrovascular accident, and lack of efficacy. Four of the five patients who completed the trial had improvement in their psychotic symptoms, particularly visual hallucinations. The small trial size and high dropout rate make these results difficult to interpret.

Piracetam, a drug that is structurally similar to γ -aminobutyric acid, was investigated in an RCT of 20 patients who were randomized to piracetam or placebo for 24 weeks, for treatment of intellectual impairment.²⁶³ There were no significant effects on any motor or cognitive features of PD.

Thirty-seven PD patients suffering from major depression participated in an RCT comparing citalopram, a serotonin-specific reuptake inhibitor (SSRI), to placebo.²⁶² After six, ten, 14, 26, 39, and 52 weeks, the citalopram was well tolerated, but no more efficacious than placebo.

Ancillary Treatment

For the purposes of this report, ancillary treatment included interventions other than medication or surgery. Eight studies concerning ancillary PD treatments were initially accepted for inclusion into the database.^{31, 272 - 278} Twelve additional studies and two kins were identified that did not meet the initial criteria for acceptance because they were less than 24 weeks in duration.^{279 - 292} When the study duration requirement was dropped for this category, the 12 studies were accepted. The results of the studies of ancillary treatments are presented in Evidence Table 16.

Of the 20 studies ultimately accepted, the majority were RCTs (k=13, t=28, n=866). Three were single-blinded, and the others were not blinded. There were two nRCTs (t=4, n=73), two cross-sectional studies (t=2, n=20), and three UCSs (n=90). Studies were graded as level I (k=2), II (k=11), or III (k=7) evidence, where I is best. The mean quality score for the 13 RCTs was 1.5 of a possible 5, where 5 is best, reflecting low quality.

Physical Therapy (PT)

PT was evaluated in six studies,^{272, 280, 282, 286, 288, 290} and one study compared music therapy (MT) to PT.²⁸⁵ Speech therapy was evaluated in four studies,^{31, 279, 281, 287} and swallowing therapy in one study.²⁸⁴ Facial mobility training was evaluated in one study.²⁸³ A health management program for PD was evaluated in two studies,^{275, 276} nurse practitioner participation in patient

care was evaluated in two studies,^{273, 277} and intensive, multidisciplinary, inpatient rehabilitation programs were evaluated in two studies.^{278, 289}

In the only long-term study of PT, 40 PD patients were divided into two groups of twenty patients each.²⁷² One group received conventional physiotherapy, which consisted of active and passive mobilization exercises to enhance postural control, balance, walking, and range of motion. The second group underwent sensory-enhanced physiotherapy, which consisted of coupling tasks with visual or auditory reinforcements, such as colored squares on the floor, or tones associated with certain movements. Each group had three four-week cycles of physiotherapy, with three months in between cycles. Baseline scores for H&Y, walking, dressing, eating, feeding, and hygiene were comparable between the two groups. At each endpoint tested (one, four, and 12 months), patients in the enhanced physiotherapy group performed better than the conventional physiotherapy group, in all scores. The scores in both groups improved immediately after each month of therapy, but returned to baseline in the conventional group after three months of no therapy, while the enhanced physiotherapy group's scores remained improved compared to baseline. The authors concluded that coupling rehabilitation with sensory stimulation leads to learning and retention of motor strategies in PD patients. Limitations of this study include its small size and the lack of randomization and blinding.

The remaining PT studies were all of less than three months duration. In one randomized, single-blind crossover study, advanced PD patients underwent intensive rehabilitation for one hour, three times a week, for four consecutive weeks.²⁸⁰ At the end of the month, patients were instructed to continue the exercise program at home. The control group received no specific instructions, and underwent the same rehabilitation program six months later. In both groups, the UPDRS total, mental, ADL, and motor scores were significantly improved immediately after the one month of rehabilitation, but returned to baseline six months later, suggesting that the beneficial effects of PT are not sustained when patients resume their usual activities.

One study compared a group of 16 PD patients who were treated with PT and various antiparkinson medications, with a group of 17 patients who were treated with medications only.²⁸² The PT group received PT for one hour, three times a week, for four months. After four months, patients in the PT group showed greater improvements than the control group in clinical rating scales and motor performance tests. Similar degrees of improvement were seen in patients with different degrees of symptom severity. Limitations of this study include the lack of randomization, and lack of followup after the PT had been discontinued.

In one study, 15 PD patients were randomly assigned to two training groups in which they were trained to perform specific arm movements.²⁸⁶ The patients in one group received auditory rhythmic cues, which consisted of tones to guide the timing of their movements. A group of age-matched volunteers who underwent the same training served as the control group. Speed of aimed movements was tested immediately after training and one hour later. Movement time improved to a similar degree in all groups, and did not change significantly after one hour. The short duration of this trial does not permit conclusions to be drawn about possible long-term efficacy of this type of training in PD patients.

In another study, 51 patients with early or mid-stage PD were randomized to participate in a ten-week program of exercises to improve spinal flexibility and axial mobility, or receive usual care.²⁸⁸ The therapy consisted of 30 individual sessions with a physical therapist, each session lasting 45 minutes to one hour. The usual care group was "wait listed" for therapy, and invited to participate in the program after the study was completed. After ten weeks, participants in the

exercise regimen improved in all three primary outcome variables, which were functional axial rotation (in degrees), functional reach (in inches) and time to go from supine to standing (in seconds). The control group did not change significantly in functional axial rotation or functional reach, although their time for moving from supine to standing increased to a similar degree as did the active patients. Limitations in this study include short duration, use of surrogate outcomes which may not reflect meaningful clinical changes, and lack of followup to determine if improvement was maintained.

In an RCT of 37 PD patients with gait impairments, 15 patients were randomized to a three-week home-based rhythmic auditory stimulation (RAS) program, which consisted of walking 30 minutes each day on a flat surface, stairs, and stop-and-go exercises to music at different tempos.²⁹⁰ One control group was given the same exercises without RAS, and the second control group was given no training. After three weeks, gait velocity on flat and inclined surfaces, cadence, and stride length all increased in the RAS group, velocity and stride length increased to a lesser degree in the exercise alone group, and did not change markedly in the untrained group. Some EMG patterns improved as well, but the changes were small and not consistent across muscles. Interpretation of the results of this study is limited by its short duration.

The effect of MT on emotional well being and QoL was evaluated in a single-blinded RCT in which 32 PD patients were randomized to participate in sessions of MT or PT weekly for two months.²⁸⁵ At three months, MT patients demonstrated improvements in UPDRS ADL, motor, and bradykinesia scores, although rigidity scores were unchanged. PT patients demonstrated no significant change in UPDRS ADL, motor, or bradykinesia scores, but the rigidity score improved significantly. QoL was measured by a happiness measurement scale, and was improved in the MT group, but unchanged in PT patients. One limitation of this study is the validity of comparing these two very different therapies. The PT sessions consisted of group exercises, and involved minimal interaction among participants. The MT sessions were of longer duration than the PT sessions, and involved more active participation. Another limitation is that the final UPDRS and QoL measurements were taken only one month after completion of the programs; therefore, the durability of the improvements cannot be assessed.

Speech/Swallowing Therapy

Two studies, both by the same author, evaluated the effects on intensive speech treatment in PD patients.^{31, 287} The two trials evaluated a total of 80 patients. Forty-eight patients were treated with the Lee Silverman Voice Treatment (LSVT), an intensive speech therapy program in which high-effort loud phonation is emphasized, with the goal of improving respiratory, laryngeal, and articulatory functions during speech. Patients received four weekly one-hour sessions of LSVT. Thirty-two patients had the same number of placebo speech therapy sessions, in which they were trained to increase their respiratory muscle activity during inspiration and expiration. The study durations were one and twelve months. A variety of measures of auditory function were performed in the different studies. Both studies supported the efficacy of LSVT for improving vocal intensity and decreasing the impact of PD on communication. Respiratory treatments alone were not effective. In the 12-month study, the LSVT group improved or maintained vocal intensity above pretreatment levels 12 months after their training was completed, whereas the placebo group had statistically significant deterioration of vocal intensity levels from before treatment. However, the 12-month study contained only 22 PD patients, and the generalizability of these results is unclear.

The Lombard effect, which describes the phenomenon that most people will increase their voice intensity when speaking in the presence of masking noise, was tested in a cross-sectional study of ten patients.²⁷⁹ All patients had been judged to have low vocal intensity by a speech-language pathologist. They were instructed to read a paragraph aloud with "normal auditory feedback," then read it again while listening to white noise through headphones. All ten PD patients showed a marked increase in speech intensity while listening to white noise. Speaking rate and speech intelligibility did not improve consistently with the white noise, and in fact worsened in some cases. It is not possible to extrapolate the effects of a one-time exposure to white noise on long-term voice intensity of PD patients.

A longer-term study evaluated the effect of a one-month voice rehabilitation program on 20 moderate-severity PD patients with complaints regarding their oral communication skills.²⁸¹ H&Y stage 1 patients were excluded, because they generally do not have speech difficulties, and stage 5 patients were excluded because their severe motor impairment would make participation difficult. After the one-month program, patients had increased vocal intensity, and decreased complaints of weak, monotonous, and unintelligible speech. Twelve of the patients complained of dysphagia prior to the program, compared with zero complaints afterwards. While these results are promising, longer-term trials are needed to adequately assess the efficacy of this treatment.

The effect of swallowing training on PD patients with swallowing disorders was evaluated in ten PD patients and 12 healthy volunteers.²⁸⁴ Subjects underwent an initial evaluation which consisted of a modified barium swallow and electromyogram (EMG) to evaluate the time it took to initiate their swallowing reflex (premotor time, or PMT). Subjects were then given one session of swallowing training. PMTs were initially elongated in the PD patients, and decreased significantly after the training, while they were normal and unchanged in healthy controls. Studies of longer duration are needed to assess the clinical significance and durability of these results.

Other Therapies

One study evaluated the effects of orofacial physiotherapeutic treatment (OPT) on facial mobility of PD patients.²⁸³ OPT consisted of brushing and applying ice to muscles, blowing through a straw, and other exercises to stimulate the facial muscles. Eight patients were randomized to receive OPT twice a week for four weeks, and eight patients received no therapy. After four weeks, measurements of facial movement were significantly improved in the OPT group patients, but there were no significant differences in the measurements of the control group patients. Repeat measurements one month after treatment completion showed similar findings.

One study reported the results of transcranial magnetic stimulation (TMS), a procedure in which a magnetic coil was positioned over the motor cortices of ten PD patients, who then received 30 stimuli twice a day for ten days.²⁷⁴ Mean UPDRS scores improved by 20.9 to 33.3 percent in total, mentation, ADL, and motor scales, and the improvements persisted after six months of followup. No adverse events were reported. Given the small number of patients in this study, no definitive statements may be made regarding the efficacy of TMS; however, these preliminary results appear favorable.

Two studies evaluated the effectiveness of PROPATH, a patient education and health promotion program designed for PD patients.^{275, 276} PROPATH participants receive an introductory videotape and educational pamphlets that provide detailed advice on daily coping with physical, emotional, and psychological aspects of PD. Patients periodically complete detailed questionnaires in which they rate the severity of their symptoms and their ability to perform ADL. Both studies were unblinded RCTs, in which patients were randomized to PROPATH participation or usual care. Patients were followed for six months in one study (n=400), and 12 months in the other study (n=50). In the six-month study, medical utilization was lower in PROPATH patients, when measured by numbers of doctor visits, hospital days, or sick days, although only the change in number of doctor visits was statistically significant. The control group had no change in the numbers of doctor visits or hospital days, and a decrease in sick days. QoL scores improved in patient global assessment in the PROPATH group, but the change was not statistically significant. In the 12-month study, patient perception of general health and psychological well-being improved significantly in the PROPATH group and worsened in the control group. Patient satisfaction with care and health care utilization was not significantly different between the two groups. Thus, there are some inconsistencies between the results of the two PROPATH studies.

Two studies assessed the value of a nurse practitioner or PD nurse specialist in management of PD.^{273, 277} Both studies were unblinded RCTs, involving a total of 225 PD patients. In one study, patients were randomized to see a PD nurse specialist or a neurologist. Minimal differences were noted in a one-year followup of these patients. In another study, patients were randomized to receive home visits from a nurse practitioner or usual care. After six months, there was no significant difference in psychosocial functioning between the two groups.

In a single-blinded RCT, 12 patients with moderately advanced PD were randomized to participate in a four-week, inpatient, multidisciplinary rehabilitation program administered by physical, occupational, and speech therapists.²⁷⁸ The eight control patients received no rehabilitation. UPDRS total, Webster, and scales of functional independence all improved significantly in the active group after the four weeks of treatment, but did not change significantly in the control group. Five months after the program had been discontinued, the above scores were still improved from prior to the intervention, but worse than they were immediately following the rehabilitation program.

An uncontrolled study of a five to ten-day inpatient multidisciplinary rehabilitation program evaluated QoL, using the Nottingham Health Profile (NHP), in 58 PD patients before and one month after completion of the program.²⁸⁹ Patients showed significant improvement in total score, pain, emotional reactions, and physical mobility, but no significant change in energy, sleep, or social isolation. The authors did not report whether this improvement lasted for longer than one month.

In summary, the 20 studies of ancillary treatment in PD reviewed in this Evidence Report showed modest improvement in some parameters after treatment with PT or MT, and significant improvement in vocal intensity after LSVT. Studies of multidisciplinary rehabilitation programs, PROPATH, or nurse practitioner interventions yielded mixed efficacy results. Evaluation of literature pertaining to ancillary treatment of PD is hampered by poor quality studies.

Genetics

Due to the lack of prospective trials regarding genetic testing for PD, it was decided that the genetics review would be presented as a summary of recent articles on the topic, including review articles. Evidence for the existence of a genetic component for PD has been reported since the 1880s, when a neurologist described a family history of PD in up to 15 percent of his patients.^{293, 294} However, experts currently believe that most cases of PD are sporadic, and that family history does not appear to confer increased risk of developing PD.²⁹⁵

Early twin studies showed low concordance rates, and argued against a genetic etiology of PD. More recent studies have refuted this claim.²⁹⁵ In a study of nearly 200 twin pairs in which at least one twin had PD, there was increased concordance in monozygotic twins, but only in patients who were diagnosed with PD before age 50.²⁹⁶ The authors concluded that genetics do not appear to play a major role in PD with typical age of onset (age > 50), but may be more important for cases with younger age of onset (≤ 50).

Another twin study evaluated the [18-F] fluorodopa PET scans of 34 patients and their monozygotic or dizygotic twins who did not have PD.²⁹⁷ Ten of the 18 monozygotic twins and three of the 17 dizygotic twins had PET scans that showed decreased uptake of fluorodopa in the striatum, consistent with PD, although none of them had clinical evidence of PD. Subjects were followed for up to seven years after their initial evaluation. All asymptomatic monozygotic co-twins showed progressive loss of dopaminergic function over seven years, and four developed clinical PD, but none of the dizygotic twin pairs became clinically concordant. This study suggests that there is a substantial genetic component to the etiology of PD that has been unrecognized because standard clinical diagnostic criteria are insufficiently sensitive. It also shows that functional imaging modalities, such as [F-18] Fluorodopa PET, may be useful tools for future studies of genetic or environmental risk factors for PD.

In an epidemiologic study using a comprehensive genealogic computerized database of over 600,000 Iceland residents over the past 11 centuries, all relatives of PD patients were traced, to examine the evidence for a genetic component of PD risk.²⁹⁸ Risk ratios (RRs) were calculated for the relatives of all 772 PD patients, and also for the subgroup of PD patients with late-onset PD (n=560). Siblings of PD patients had the highest RRs (6.3 for all PD patients, 6.7 for patients with late-onset PD), followed by offspring (3.0 and 3.2, respectively), and nieces and nephews (2.3 and 2.7, respectively). The results of this study suggest that there may be a substantial genetic contribution (at least in this highly interrelated ethnic population), not only in patients with "young-onset" disease (as has been the finding in twin studies), but also in PD patients with typical age of onset.

Mutations associated with PD have been identified in several genes, and it appears that different mutations can produce the same parkinsonian phenotype. The *α -synuclein* gene on chromosome 4q21-23, and the *parkin* gene on chromosome 6q25-27 have been studied extensively.²⁹⁴ *α -synuclein* is a protein that has been identified as a major component of Lewy bodies and a part of the amyloid plaque in AD.²⁹⁵ A point mutation in the *α -synuclein* gene has been identified in some cases of autosomal dominant familial PD in families of Greek or Italian descent.²⁹⁴ Many mutations in the *parkin* gene are associated with early-onset, autosomal-recessive PD in Japanese and European families^{299 - 302} while other *parkin* mutations may be associated with a protective factor for sporadic (non-familial) PD.³⁰³

It is likely that other genes will be identified that are involved with familial PD, but the currently available evidence suggests that the vast majority of PD cases are not familial, and have no known associated genetic component.²⁹⁵ When more information is known about the specific genetic abnormalities in PD patients, specific intracellular genetic manipulation may become possible, with the goal of treating, curing, and even preventing PD.^{304, 305}