

Chapter 2. Methodology

MetaWorks investigators used systematic review methods derived from the evolving science of review research.^{110, 111} These methods were generally applied according to standard operating procedures at MetaWorks and are displayed in Figure 1.

A Task Order, containing the original questions described above, was developed by AAN, submitted to AHRQ, then presented to MetaWorks. From this Task Order, MetaWorks researchers developed a Work Plan (see Appendix B), which was then reviewed by AHRQ, AAN, and the TEP. The work plan outlined the methods to be used for the literature search, study eligibility criteria, data elements for extraction, and methodological strategies to minimize bias and maximize precision during the process of data extraction and synthesis. After a preliminary review of the literature, a Topic Assessment and Refinement report was submitted to the AHRQ, AAN, and TEP, discussing the revised key questions and preliminary results of the literature searches (Appendix C).

Causal pathways relevant to the above questions were then developed (Appendix D). These pathways were not designed to function as clinical practice guidelines or algorithms for patient care decisions. They were constructed solely to guide the systematic review process for this project, and with the expectation that they might change as the project developed.

Literature Search

The published literature was searched from January 1, 1990 to December 31, 2000, using Medline, Current Contents®, and Cochrane Library databases. A manual search was performed of the bibliographies of all publications accepted for inclusion into the evidence base. In addition, the bibliographies of recent review articles were searched for potentially relevant citations. The retrieval cut-off date was February 1, 2001.

The Medline search included the following search strategies, with limits of publication dates 01/01/1990 to 12/31/2000, English language, Clinical Trial, and Human:

Diagnosis: (*Parkinson disease OR parkinson syndrome OR parkinsonism*) AND (*diagnosis OR medical errors OR accuracy OR sensitivity OR specificity*) OR (*diagnosis AND antiparkinsonian agents*).

Pharmacological Treatment: (*Parkinson disease OR parkinson syndrome OR parkinsonism*) AND (*treatment OR levodopa OR carbidopa OR amantadine OR anticholinergic OR selegiline OR deprenyl OR dopamine agonist OR bromocriptine OR pergolide or lisuride OR cabergoline OR pramipexole OR ropinirole OR tolcapone OR entacapone*).
(*Parkinson disease OR parkinson syndrome OR parkinsonism*) AND (*selegiline OR Vitamin E OR Vitamin C OR neuroprotective agents*).

The search cut-off date for pharmacologic studies was initially 1985, for the purpose of including studies of anticholinergic agents. However, no acceptable studies of anticholinergic agents were published between 1985 and 1990; therefore the search cut-off date was changed back to 1990, in accordance with the search cut-off date established for the other questions.

Surgical Treatment: *(Parkinson disease OR parkinson syndrome OR parkinsonism) AND (surgery OR pallidotomy OR brain tissue transplant OR deep brain stimulation)*

Psychiatric Treatment: *(Parkinson disease OR parkinson syndrome OR parkinsonism) AND (psychological OR psychotic OR mental disorder) AND (drug therapy OR drug interactions).*

Ancillary Treatment: *(Parkinson disease OR parkinson syndrome OR parkinsonism) AND rehabilitation.*

Genetics: *(Parkinson Disease OR parkinsonism OR Parkinson) AND genetics AND limit to review articles January 1, 1997-August 1, 2000.*

The search of the Current Contents CD-ROM database employed the same strategies. The Cochrane Controlled Trials Register search strategy was "Parkinson's Disease."

All citations and abstracts resulting from the above searches in Medline and Current Contents were downloaded and printed at MetaWorks.

To assist with the development of the evidence base, pertinent articles from the following Internet sites were reviewed:

American Parkinson Disease Association (<http://apdaparkinson.com>)

Medscape (<http://www.medscape.com>)

National Guidelines Clearinghouse (NGC; <http://www.guideline.gov>)

National Parkinson Foundation (<http://www.parkinson.org>)

Parkinson's Action Network (<http://www.parkinsonsaction.org>)

Parkinson's Disease Foundation (<http://www.parkinsons-foundation.org>)

United Parkinson Foundation (<http://www.aoa.dhhs.gov/aoa/dir/221.html>)

Clinical trials information (<http://www.parkinson-study-group.org>)

A list of potentially relevant studies was provided by the American Speech-Language-Hearing Association (ASHA). These citations were screened in the same manner as those identified by electronic searches.

Exclusion Criteria

During Level I screening, all abstracts were downloaded, reviewed and evaluated for the following exclusion criteria:

- Reviews, meta-analyses (except those regarding diagnosis and genetics)
- Letters, case reports, editorials, and commentaries.
- Abstracts and unpublished study reports.
- Pharmacokinetic and pharmacodynamic studies.

- Animal or *in vitro* studies.
- Studies written in languages other than English.
- Studies published prior to 1990.
- Studies with < 10 patients.
- Cross-over studies.
- Studies where results for PD population cannot be separated from results from other populations.
- Studies not pertaining to diagnosis or treatment of PD.
- Treatment studies with < 24 weeks of treatment and followup.

Cross-over studies were excluded for several reasons. It is frequently difficult to extract information from the mid-point of the trial, before the cross-over occurs. The patient response in the second phase of a study of cross-over design may be impacted by treatment administered during the first phase. When patients drop out in the first phase, the patients entering the second phase may be different from the baseline population, introducing selection bias. Additionally, the number of parallel design RCTs that met the inclusion criteria comprised a large enough evidence base to justify the exclusion of studies of cross-over design, with all of their attendant difficulties in data extraction and interpretation.

Given that PD is a chronic condition, and that patients stay on medications for years, the most clinically relevant data comes from long-term trials. For this reason, treatment trials had to be greater than or equal to 24 weeks duration for acceptance. Furthermore, the most useful data for analysis concerning pharmacological treatment of PD are in randomized controlled trials (RCTs); therefore, only RCTs were accepted for studies pertaining to pharmacological treatment. For trials pertaining to surgical treatment, 24 weeks of followup were required; however, study designs other than RCTs were accepted, due to the scarcity of RCTs evaluating surgical procedures for PD. For trials pertaining to diagnosis, study duration and design were not restricted.

Full articles were retrieved for all abstracts passing Level I screening. The articles then underwent Level II screening, which consisted of evaluating the articles for the following inclusion criteria (See Appendix E):

Inclusion Criteria

Diagnosis:

- The following study designs: observational [prospective, retrospective, and cross sectional (XS)], or interventional [RCTs, non-randomized controlled trials (nRCTs), uncontrolled case series (UCSs), XS].

- Adult patients with potential diagnosis of PD.
- Studies addressing any diagnostic test to establish or support a diagnosis of PD.

Pharmacological Treatment:

- RCTs only.
- ≥ 24 weeks treatment and followup duration.
- Studies reporting at least one objective clinical outcome measure (efficacy or safety) on at least one of the following drugs or category of drugs:
 - L-dopa/Carbidopa (Sinemet)
 - L-dopa/Benserazide (Madopar)
 - Amantadine (Symmetrel)
 - Dopamine agonists: Bromocriptine (Parlodel), Pergolide (Permax), Ropinirole (Requip), Pramipexole (Mirapex), Andropinole, Cabergoline (Dostinex), Apomorphine, Lisuride (Dopergin)
 - Monoamine oxidase B (MAO-B) inhibitors: Selegiline (Deprenyl), Rasagiline (TVP-1012), Lazabemide
 - Catechol-O-methyltransferase (COMT) inhibitors: Tolcapone (Tasmar), Entacapone (Comtan)
 - Anticholinergic agents: Trihexylphenidyl (Artane), Benztropine (Cogentin), Procyclidine
- Studies involving neuroprotection with selegiline, Vitamin E (tocopherol), or Vitamin C.

Surgical Treatment:

- The following study designs: interventional (RCTs, nRCTs, and UCSs).
- ≥ 24 weeks study and followup duration.
- Must report at least one objective clinical outcome measure.
- Studies addressing surgery in adult patients with PD including:
 - Ablative or destructive surgery (thalamotomy, pallidotomy),

- Stimulation surgery or Deep Brain Stimulation (DBS),
- Transplantation surgery.

Psychiatric Treatment:

- The following study designs: interventional (RCTs, nRCTs, and UCSs).
- ≥ 24 weeks study and followup duration.
- Studies addressing treatment of non-psychotic behavioral and psychological dysfunction in adult patients with PD.
- Studies addressing treatment of psychotic symptoms in adult patients with PD.
- Studies addressing use of antipsychotic medications in conjunction with antiparkinsonian agents.
- Studies addressing the use of atypical antipsychotic medications in management of adult patients with PD.
 - Clozapine (Clozaril)
 - Olanzapine (Zyprexa)
 - Quetiapine (Seroquel)

Ancillary treatment:

- The following study designs: interventional (RCTs, nRCTs, and UCSs).
- No minimum study duration.
- Studies reporting at least one of the following specific interventions:
 - Allied health interventions.
 - Occupational therapy (OT).
 - Physical therapy (PT).
 - Psychotherapy (counseling).
 - Speech therapy.

- Studies reporting at least one of the following specific outcomes:
 - Acute hospitalization.
 - Rehabilitation hospitalization.
 - Nursing home admission.
 - Work absenteeism.
 - Quality of Life (QoL).
 - Activities of Daily Life (ADL) assessment.

Genetics:

- Study design limited to recent review articles only.
- Adult patients undergoing genetic testing to support a diagnosis of PD.

General Considerations:

Studies pertaining to diagnosis were initially required to report sensitivity and specificity; however, as very few studies met this requirement, it was removed. Some of the peer reviewers commented that the inclusion criteria for surgical studies were less rigid than those for pharmacological studies. This disparity was due to the relative scarcity of RCTs pertaining to surgical treatment of PD.

For studies regarding ancillary treatment to be accepted, the initial requirement was that the study duration be at least 24 weeks. This resulted in acceptance of only six studies; therefore, this requirement was removed, and studies of < 24 weeks duration were also accepted.

Linked Studies

After the accepted studies were determined, linked studies were identified. These were studies in which the same patient population was reported in more than one publication. “Parent” studies were assigned, which contained primary data. “Child” studies contained supplemental information, such as followup data or additional analyses. Data elements were extracted from the parent studies, and supplemented by information presented in kin studies, when appropriate.

Rating the Evidence

All eligible studies were rated for both quality and level of evidence at the time of data extraction. Two established methods: 1) the Jadad method,¹¹² and 2) the Level of Evidence method¹¹³ were used (see Appendix B, Attachments B and C).

Data Extraction

Data Extraction Forms (DEFs) were designed in advance (see Appendix E), and pilot tested on a small sample of eligible studies. The pilot test allowed for necessary edits to the DEF to be made prior to implementation on all studies. Key data from each eligible study were extracted by a researcher recording data from original reports onto a DEF, and reviewed by a second researcher checking all DEF fields against the original report. Differences were resolved prior to data entry. In all cases, at least one physician reviewed each study. Dual review of all data served to reduce error and bias in the data extraction process. The data were then entered into MetaWorks' relational database of clinical studies, MetaHub™.

When trials consisting of several phases with different study designs were encountered, only data from the randomized phase was captured.

Key data elements sought for extraction from each study are listed in the Work Plan (Appendix B).

Database Development

Data were entered from the DEFs into a relational database of clinical trials. At the time each DEF was entered, 100 percent of the data entries were checked back against the original DEFs. In addition, a 20 percent random sample of data in the completed database was checked against the DEFs. Error rates in excess of 2 percent of QC-checked data would have triggered a 100 percent recheck of all data elements entered into the database.

Statistical Methods

The main goal of the statistical analysis was to estimate the difference in efficacy of various treatments for PD.

Summary Statistics

Data listings and summary data were prepared for study level characteristics, patient and treatment level characteristics, outcomes of interest, and safety data. When the database was complete, verified, and locked, data were entered into table shells. In general, study and patient characteristics and outcomes variables were summarized using standard descriptive statistics weighted by study sample size.

Diagnosis

Studies pertaining to diagnosis were synthesized with summary statistics only.

Pharmacologic Treatment

The medications with sufficient data for comparisons were L-dopa, DAs, selegiline, and COMT inhibitors. The primary efficacy outcomes of interest were the standardized mean changes from baseline to common followup time points, as evaluated on the UPDRS scales. If total UPDRS score was not reported, scores from UPDRS III (Motor) or II (ADLs) were used. Both "off" and "on" scores were captured when reported. Most studies, however, did not report whether scores were "off" or "on."

When UPDRS scores were not available, S&E ADL scores, Webster, Columbia University Rating Scale (CURS),¹¹⁴ or H&Y scores were used instead, in the above order of preference. This order was based on frequency of reporting the different scales. Validation studies show that scores on the various PD rating scales are generally very highly correlated. For example, one study found a correlation of 0.79 between total UPDRS score and H&Y score, a correlation of -0.88 between S&E and UPDRS, and a correlation of 0.76 between scores on the Webster scale and the total UPDRS score.⁵⁴

We had initially intended to include the Northwestern University Disability Scale (NUDS), in the above evaluations; however, inconsistency in reporting methods prevented pooling results from different studies that used this scale.^{54, 115} This did not result in exclusion of any studies; while 19 treatment arms reported NUDS scores, all of them reported results of at least one other scale as well.

Placebo treatment arms frequently allowed discretionary L-dopa administration. For the purposes of analysis, placebo arms with discretionary L-dopa were categorized as L-dopa arms, despite incomplete reporting of the actual number of patients treated, or their specific results. In the DA meta-analysis, a separate analysis was performed comparing studies in which patients received discretionary L-dopa to studies in which patients were randomized to L-dopa.

Surgery

The surgical procedures analyzed were pallidotomy, DBS, and fetal cell transplants. Studies of thalamotomy were synthesized with summary statistics only, due to the small number of studies.

The primary efficacy outcomes of interest for surgery studies were the standardized mean changes from baseline to outcome, as evaluated on the UPDRS scores. If total UPDRS score was not reported, scores from UPDRS III (Motor) or II (ADLs) were used. When UPDRS scores were not available, S&E ADL, Webster, CURS, or H&Y scores were used instead, in the above order of preference.

For both pharmacological and surgical studies, safety outcomes were reported with summary statistics.

Ancillary Treatments

Results of ancillary treatments were synthesized with summary statistics only, due to the small number of accepted studies and the variety of evaluative techniques presented in the different studies.

Meta-Analyses

Meta-analysis of efficacy outcomes of pharmacological studies was performed for all RCTs reporting outcome data on at least one of the PD rating scales mentioned above. Meta-analysis of the primary efficacy outcomes of the surgery studies was also performed for all studies reporting the necessary outcome data. The effect sizes calculated and meta-analyzed were standardized mean differences.¹¹⁶ Appendix F describes interpretation of the size of standardized mean differences.

Effect sizes for pharmacological studies. In the meta-analyses of pharmacological studies, the effect size represents the standardized difference between two groups on the change in patients' scores from the beginning of the treatment to the end of the treatment. Optimally, this effect size was calculated from baseline and outcome data for the two treatment groups, and (preferably) change-score standard deviations; however, in some cases calculations were possible from study *p*-values.

Unbiased change-score effect sizes were calculated using the standard formula:

$$d_{CHANGE} = \left(1 - \frac{3}{4N - 9}\right) \frac{(\bar{Y}_{T_OUTCOME} - \bar{Y}_{T_BASELINE}) - (\bar{Y}_{C_OUTCOME} - \bar{Y}_{C_BASELINE})}{\sqrt{s_{pooled_CHANGE}^2}},$$

where *N* is the total study sample size.¹¹⁶ The numerator represents the difference between the treatment and control groups on the amount of change each underwent from the beginning of the study. The denominator contains the pooled variance of the treatment and control change-scores. See Appendix F for a description of how this variance was estimated when the source studies reported baseline and outcome data but failed to report change-score standard deviations. The change-score effect sizes measured the standardized difference in change between a treatment and a control group, to see if one treatment led to more improvement (or less decline) in PD than the other. Effect sizes were scaled such that a positive effect size indicated that the treatment under investigation worked better than the control, where the control was always therapy with L-dopa alone.

Effect sizes for surgery studies. The meta-analysis of surgery data examined pre-post surgery standardized mean differences in PD scores:

$$d_{PRE \rightarrow POST} = \left(1 - \frac{3}{4N - 9}\right) \frac{(\bar{Y}_{T_OUTCOME} - \bar{Y}_{T_BASELINE})}{\sqrt{s_{pooled_PRE \rightarrow POST}^2}}.$$

These effect sizes simply compare pre-test to post-test scores to determine if there was any improvement at all due to the surgery. This means that the only comparisons between the efficacies of different types of treatment (e.g., pallidotomy versus DBS) that can be made are indirect comparisons.

Meta-analyses of pharmacological and surgery studies. Based upon available data, three sets of meta-analyses were conducted for the pharmacological studies: one that compared the efficacy of DAs (with L-dopa allowed) to L-dopa given alone, one that compared the efficacy of selegiline (with L-dopa allowed) to L-dopa given alone, and one that compared the efficacy of COMT inhibitors (with L-dopa allowed) to L-dopa given alone.

There were insufficient studies to allow a meta-analysis of any pharmacological treatment against placebo, with no L-dopa involved.

Based upon available data, three sets of meta-analyses were calculated for the surgery studies: one that investigated whether pallidotomy was associated with improved "off" or "on" scores, one that investigated whether DBS was associated with improved "off" or "on" scores, and one that investigated whether fetal brain cell transplants were associated with improved "off" or "on" scores. There were insufficient studies to perform a similar meta-analysis on thalamotomy treatment arms.

After effect sizes and their expected variances were calculated for a given set of studies, a fixed-effects meta-analysis was conducted within each set.¹¹⁶ The chi-square homogeneity statistic (Q_E) was calculated for each meta-analysis to determine whether there was any variation in the study effects that could not be explained due to sampling error. Given the low number of studies in each meta-analysis, there was very low power to detect effect heterogeneity. Thus, a more conservative random-effects model was utilized to calculate the final estimates and confidence intervals when the estimate of random-effects variation (defined as τ or Δ^2) was greater than zero.¹¹⁷ The random-effects model accounts for treatment variation not explainable due to sampling error, and thus leads to wider confidence intervals for its parameters than the fixed-effects model. When data permitted, fixed-effects and/or random-effects meta-regressions (mixed-model meta-analyses that consider study characteristics as predictors of treatment effect) were examined as well.^{117, 118} Common study characteristics investigated were mean patient age, severity of disease at baseline, disease duration, and the time between initial treatment and post-test.

All meta-analyses and meta-regressions were performed using SPSS 10.1 and procedures written in SAS/IML 8.1.

Sensitivity Analysis. When the number of studies in the meta-analysis permitted, sensitivity analysis was performed to examine whether any design characteristics were associated with treatment effects. Characteristics (covariates) of interest included:

- 1) whether study data were intention-to-treat (ITT) or completers
- 2) adequacy of blinding, as reflected in Jadad score.¹¹²

The data were also inspected for "outliers" – study effects that were extreme enough, either in their value or in the value of their study characteristics, that they might by themselves "skew" the estimate of the mean effect, the estimate of effect size heterogeneity, or the relationship between a study characteristic and the study effects.

Role of Consultants

Eight people from both academic and community settings comprised the TEP (Appendix G). They all received copies of the Work Plan and its revision, causal pathways, topic refinement, study listings, and draft report. When TEP members provided feedback, MetaWorks investigators reviewed their comments, and applied them as deemed appropriate. Additionally, during the course of the project, monthly conference calls were instituted among MetaWorks, the topic nominator (AAN), the TOO, and the co-investigator from LDI. During these conference calls, project updates were provided and issues of concern were addressed.

Peer Review

A group of 19 peer reviewers (Appendix G) was assembled to review a draft version of this report. The panel was composed of neurologists, a neurosurgeon, an internist, two statisticians, a speech-language pathologist, and two PD patients. All reviewers were asked to complete peer review form relative to the content of the first draft of this report (Appendix G), and were also invited to provide additional written comments. Seven of the eight TEP members and 13 of the 19 peer reviewers provided feedback on the draft Evidence Report. All responses from the TEP and peer reviewers were reviewed and, where appropriate, incorporated into the final report.