Decision Support for Treatment of Patients with Advanced Parkinson’s Disease

JERKER WESTIN
Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Rudbecklaboratoriet, Dag Hammarskjölds väg 20, Uppsala, Saturday, December 18, 2010 at 13:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in Swedish.

Abstract

The overall aim of this thesis was to develop, deploy and evaluate new IT-based methods for supporting treatment and assessment of treatment of advanced Parkinson’s disease. In this condition a number of different motor and non-motor symptoms occur in episodes of varying frequency, duration and severity. In order to determine outcome of treatment changes, repeated assessments are necessary. Hospitalization for observation is expensive and may not be representative for the situation at home. Paper home diaries have questionable reliability and storage and retrieval of results are problematic. Approaches for monitoring using wearable sensors are unable to address important non-motor symptoms.

A test battery system consisting of both self-assessments of symptoms and motor function tests was constructed for a touch screen mobile phone. Tests are performed on several occasions per day during test periods of one week. Data is transmitted over the mobile net to a central server where summaries in different symptom dimensions and an overall test score per patient and test period are calculated. There is a web application that graphically presents the results to treating clinical staff. As part of this work, a novel method for assessment of spiral drawing impairment useful during event-driven sampling was developed. To date, the system has been used by over 100 patients in 10 clinics in Sweden and Italy. Evidence is growing that the test battery is useful, reliable and valid for assessment of symptoms during advanced Parkinson’s disease.

Infusion of a levodopa/carbidopa gel into the small intestine has been shown to reduce variation in plasma drug levels and improve clinical response in this patient category. A pharmacokinetic-pharmacodynamic model of this intestinal gel infusion was constructed. Possibly this model can assist the process of individualization of dosage for this treatment through in numero simulations. Results from an exploratory data analysis indicate that severity measures during oral levodopa treatment may be factors to consider when deciding candidates for infusion treatment.

Keywords: Parkinson’s disease, telemedicine, motor test, tapping, spiral, self-assessment, home-assessment, outcome measure, electronic diary, patient-reported outcome, remote patient monitoring, levodopa, infusion, pharmacokinetic, pharmacodynamic

Jerker Westin, Department of Medical Sciences, Biomedical Informatics and Engineering, Akademiska sjukhuset, Uppsala University, SE-75185 Uppsala, Sweden.

© Jerker Westin 2010

ISSN 1651-6206
urn:nbn:se:uu:diva-132635 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-132635)
To the five million people living with Parkinson’s disease and those who care for them
List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


Reprints were made with permission from the respective publishers.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AADC</td>
<td>Amino acid decarboxylase</td>
</tr>
<tr>
<td>ADL</td>
<td>Activities of daily living</td>
</tr>
<tr>
<td>ADO</td>
<td>ActiveX data objects</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>Analysis of covariance</td>
</tr>
<tr>
<td>ASQ</td>
<td>After-scenario usability questionnaire</td>
</tr>
<tr>
<td>ASP</td>
<td>Active server pages</td>
</tr>
<tr>
<td>CGI</td>
<td>Clinical global impression</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>CPU</td>
<td>Central processing unit</td>
</tr>
<tr>
<td>CSUQ</td>
<td>Computer system usability questionnaire</td>
</tr>
<tr>
<td>DBS</td>
<td>Deep brain stimulation</td>
</tr>
<tr>
<td>DPSS</td>
<td>Data processing sub system</td>
</tr>
<tr>
<td>DSS</td>
<td>Decision support system</td>
</tr>
<tr>
<td>DWT</td>
<td>Discrete wavelet transform</td>
</tr>
<tr>
<td>EC50</td>
<td>Median effective concentration</td>
</tr>
<tr>
<td>EudraCT</td>
<td>European clinical trials database</td>
</tr>
<tr>
<td>FDA</td>
<td>US food and drug administration</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier transform</td>
</tr>
<tr>
<td>FOCE</td>
<td>First order conditional estimation method</td>
</tr>
<tr>
<td>HR-QOL</td>
<td>Health related quality of life</td>
</tr>
<tr>
<td>HTML</td>
<td>Hypertext markup language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext transfer protocol</td>
</tr>
<tr>
<td>ICC</td>
<td>Intra class correlation coefficient</td>
</tr>
<tr>
<td>IIV</td>
<td>Inter individual variability</td>
</tr>
<tr>
<td>IOV</td>
<td>Inter occasion variability</td>
</tr>
<tr>
<td>IT</td>
<td>Information technology</td>
</tr>
<tr>
<td>LNAA</td>
<td>Large neutral amino acid</td>
</tr>
<tr>
<td>MADRS</td>
<td>Montgomery Åsberg depression rating scale</td>
</tr>
<tr>
<td>MDS-UPDRS</td>
<td>Movement Disorder Society’s revised UPDRS</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini mental state examination</td>
</tr>
<tr>
<td>MRV</td>
<td>Movement rhythm variation</td>
</tr>
<tr>
<td>MSD</td>
<td>Mean squared deviation</td>
</tr>
<tr>
<td>MVC</td>
<td>Model view control</td>
</tr>
<tr>
<td>.NET</td>
<td>Microsoft .NET software framework</td>
</tr>
<tr>
<td>PC</td>
<td>Personal computer</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal digital assistant</td>
</tr>
<tr>
<td>PDQ-39</td>
<td>Parkinson’s disease questionnaire (39 items)</td>
</tr>
<tr>
<td>PEG</td>
<td>Percutaneous endoscopic gastrostomy</td>
</tr>
<tr>
<td>PKPD</td>
<td>Pharmacokinetic-pharmacodynamic</td>
</tr>
<tr>
<td>RAM</td>
<td>Random access memory</td>
</tr>
<tr>
<td>RDM</td>
<td>Remote device manager</td>
</tr>
<tr>
<td>ROM</td>
<td>Read only memory</td>
</tr>
<tr>
<td>RTM</td>
<td>Regression to the mean</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short form health survey (36 items)</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SDDV</td>
<td>SD of frequency-filtered drawing velocity</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>SMR</td>
<td>Standardized manual rating</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured query language</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum of squared residual errors</td>
</tr>
<tr>
<td>SSM</td>
<td>Sum of squares about the mean</td>
</tr>
<tr>
<td>SSRI</td>
<td>Selective serotonin reuptake inhibitor</td>
</tr>
<tr>
<td>TRS</td>
<td>Treatment response scale</td>
</tr>
<tr>
<td>UPDRS</td>
<td>Unified Parkinson’s disease rating scale</td>
</tr>
<tr>
<td>VB</td>
<td>Visual basic</td>
</tr>
<tr>
<td>WAST</td>
<td>Microsoft web application stress tool</td>
</tr>
<tr>
<td>WSTS</td>
<td>Wavelet spiral test score</td>
</tr>
<tr>
<td>WUI</td>
<td>Web user interface</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible markup language</td>
</tr>
</tbody>
</table>
Introduction

Parkinson’s disease

Parkinson’s disease (PD) is the second most common neurodegenerative disorder (after Alzheimer’s disease) and directly affects up to 5 million people worldwide. The prevalence of PD increases with age, with approximately 1% of those aged 60 years or older affected. Given the growing elderly population, the number of individuals with PD is expected to double by 2030 (Chen 2010).

PD is a slowly progressive disorder that affects a small area of dopamine producing cells in the mid-brain, vital for motor function control. Hoehn and Yahr (1967) described the progression of PD in five stages with increasing disability. The typical symptoms are slowness of movements, stiffness, instability and tremor, but also non-motor symptoms concerning e.g. cognition, mood, pain, sleep and sexual health are frequently occurring (Wolters et al. 2007). The cause of cell death is still unknown but one of the latest theories suggests that PD may be a prion disorder (Olanow and Prusiner, 2009). None of the currently available treatments can halt or even slow the progression of cell death. Standard symptomatic treatment at the initial stage of the disease is aimed at restoring depleted dopamine receptor stimulation. The most common therapy is levodopa with a stabilizer in tablet form. Levodopa is a precursor of dopamine; carbidopa and benserazide are stabilizers that prevent enzymatic breakdown of levodopa outside of the central nervous system. Forty years after its discovery (Carlsson, 2002), levodopa remains the most effective medication for PD. In fact, 70 to 80 percent of patients currently treated for PD receive levodopa therapy (Parkinson’s Disease Foundation’s webpage, www.pdf.org, last accessed Oct 2010).

In advanced stages of PD, patients treated with levodopa experience motor complications (or fluctuations) i.e. variation in level of motor function with swings within the order of hours or minutes. A number of different symptoms may occur with differing frequency and intensity. 50% of the patients have these problems after 5 years of receiving levodopa and almost 100% of the patients after 10 years (van Laar 2003). Both motor symptoms
and non-motor symptoms of advanced PD may fluctuate (Bayulchem and Lopez 2010).

PD medication targeting dopamine receptors must be individually tuned; underdosing does not relieve the symptoms and overdosing leads to side effects with uncontrolled movements (choreatic dyskinesias), sickness and hallucinations. The daily levodopa dosage requirement varies considerably and up to a 22-fold difference between individuals can be seen (Nyholm et al. 2010). The dosage may need to be adjusted daily taking account of food intake, physical exercise, time-of-day and mood. Since PD is progressive, there is a need to follow up treatments over time. The interval for appropriate dosage shifts upwards and gets narrower as the disease progresses (Mouradian et al. 1988).

Validation criteria for assessment methods

In 1987, Feinstein introduced the term “clinimetrics” to describe the field of research concerned with measurement in clinical medicine. Measures can be objective and directly measure some function of the body. They can also be subjective and assess perception of symptoms. The choice between objective and subjective measures should be based on the purpose of the measurement: Is the interest to measure the gait velocity, step-length etc or is it to assess if a patient experiences difficulty with walking? The selection of variables to measure depends on the phenotypes and concepts to be studied. The range of values to be measured is also important and different measurement instruments may be more or less suitable for early and late disease stages. Measurement instruments must provide meaningful and useful information, catching the targeted variables accurately enough for the intended purpose. Each measure (or item) of an instrument should be examined to assess whether or not it is redundant, understandable, variable, reproducible and if it displays floor or ceiling effects in the target population. The United States Food and Drug Administration (FDA) recently (2009) published guidelines for patient-reported outcome measures for use in medical product development. FDA recognizes that instrument development is an iterative process and that there is no single correct way to do it. There are however certain quality criteria that measurement instruments should meet (Marinus and van Hilten 2007):

Validity is the extent to which an instrument measures what it is supposed to measure. Content validity refers to if an instrument covers the important areas of the domain it aims to assess or not. It is a subjective criterion, which generally relies on literature reviews and expert opinions. When there exists a gold standard (or criterion) measure, the term criterion validity describes to
what extent results from an instrument agree with the gold standard. Construct (or concept) validity concerns how well an instrument relates to theoretical concepts (such as a symptom or group of symptoms) that are intended to be measured and is used when there is no gold standard available. Convergent (or concurrent) validity is a way to demonstrate construct validity by analyzing the extent to which an instrument relates to other instruments that measure the same concept. A good method for assessing convergent validity is by Spearman’s correlation coefficient. Known-groups validity is another way to demonstrate construct validity by examining if known phenotypic differences are reflected in the results from an instrument. Methods for assessment include significance tests for difference in median or mean and analysis of variance.

Reliability is the extent to which a set of measurements from an instrument is consistent, or whether or not measurement errors are small enough. Test-retest reliability refers to the stability of the instrument over time and is best assessed using the intraclass correlation coefficient (ICC). For rating scales, consistency between observers (inter-rater reliability) is another important reliability criterion. Internal consistency is a measure of the extent to which items are measuring the same concept and is assessed by Cronbach’s α coefficient. In addition to validity and reliability, a third quality criterion regards responsiveness, or the ability to detect clinically significant changes.

According to FDA (2009), the method for scoring items in an instrument should also be justified. Equally weighted scores for each item may be appropriate when the responses to items are independent. If however some items are dependent, their collected information is less than between independent items. These weighting concerns also apply when combining sub-scores into a single general score. Empirically determined patient preference ratings are one possible approach to weighting items. Qualitative research or statistical techniques should be used to justify the method chosen to combine items into scores.

Assessment methods in PD

In 2002, Ramaker et al. made a systematic evaluation of clinical rating scales for assessment of impairment or disability in PD. Regardless of the scale, there was a conspicuous lack of consistency concerning inter-rater reliability of many symptoms’ ratings.

The Unified PD Rating Scale, UPDRS (Fahn et al. 1987) is the most commonly used rating scale in clinical trials (Mitchell et al. 2000). According to Ramaker et al. (2002), the UPDRS generally showed moderate
to good reliability and validity. It consists of 4 parts aiming at measuring different aspects; Part 1 measures mentation, behavior, and mood, Part 2 activities of daily living (ADL), Part 4 clinical motor abilities and Part 4 complications of therapy. Parts 1, 2 and 4 are evaluated by an interview as patients remember their situation whereas Part 3 scoring is based on how the evaluator finds motor performance at the time of the patient examination. Each item of the scale is generally graded 0 to 4. A total of 199 points are possible in Parts 1 to 4. 199 represents the worst (total disability), 0 represents no disability. Visser et al. (2006) studied responsiveness of the UPDRS items and found that hand functions were the most responsive ADL items with ‘handwriting’ showing the largest response. In the Motor section, bradykinesia items showed the largest response, especially the item ‘finger taps’. A revision of the UPDRS was recently made by the Movement Disorder Society’s task force (Goetz et al. 2008).

The 39 item PD questionnaire (PDQ-39) (Jenkinson et al. 1997) is the most widely used disease-specific patient-reported rating scale in PD. This scale was analyzed by Hagell et al. (2007) and results provide general support for its acceptability and reliability. However it has limitations for assessing less severely affected patients. Hagell et al. (2008) also tested the most widely used general patient-reported rating scale, the SF-36 (Ware and Sherbourne 1992) with PD patients and found several limitations (meaningfulness and floor/ceiling effects). A number of less widely used rating scales for different aspects of PD (e.g. Marinus et al. 2004) are being developed under the framework of the SCOPA-PROPARK project (www.scopa-propark.eu, last accessed Oct 2010).

In presence of motor and non-motor fluctuations, single or a few observations or measurements, however extensive, cannot give full information on a patient’s condition. Weaver et al. (2005) highlighted the need for uniform, detailed reporting of comprehensive motor and non-motor outcomes at multiple time points. Isacson et al. (2008) also recommended repeated measurements for assessments during fluctuations. For the patients it can be difficult to remember their situation back in time and repeated observations by medical staff require hospitalization, which is expensive and may not be representative of the situation in the home environment.

Haaxma et al. (2008) proposed a timed motor test battery, which can reduce problems with inter-rater disagreements and time consumption for doing a full UPDRS motor examination. They found that their timed test battery was more sensitive to change than the UPDRS motor section. Nyholm et al. (2005) proposed a method, which allows for repeated clinical assessments of motor function: Video recordings were performed hourly on test days during tasks from four selected items of the UPDRS motor examination. These
tasks consisted of ‘finger tapping’, ‘alternating hand movements’, ‘rising from a chair’, and ‘walking’. In addition, dyskinesia was assessed according to the dyskinesia rating scale (Goetz et al. 1994). Investigators then rated overall motor performance based on the video recordings from -3 (severe PD symptoms) to +3 (severe choreatic dyskinesia) on a global treatment response scale (TRS), where 0 is ‘on’ without dyskinesia. ‘Off’ means an episode when PD symptoms are present and ‘on’ means relief from PD symptoms. An advantage of this method is that specially trained raters can evaluate many patients from different sites and that these raters can easily be blinded to which treatments the patients receive. The test days results were summarized as percentage “good on-time”; that is percentage ratings in the three best stages, -1, 0 and +1 (Nyholm et al. 2005).

Goetz et al. (2008) evaluated feasibility of home-videotaping in 10 patients in 30 minute intervals during two test days. All self-recorded video segments were clear with all agreed elements included. With the exception of one missed rating, patient-based self-ratings were performed at the right times. However, in spite of patient training concordance between physician’s and patients’ own on/off ratings was only 64%. To assess patients at home via self- (or caregiver-) recorded videos uploaded over internet, and then to rate the recordings centrally (e.g. on Nyholm’s TRS scale) has interesting prospects. Samii et al. (2006) found UPDRS scoring via video-conference equipment feasible, if good enough video quality could be obtained.

The most common method for assessing motor fluctuations today is paper home diaries, such as the one constructed by Hauser et al. (2000). Reimer et al. (2004) found that one full week, but not three days, of home diary data collection was adequate for capturing a sample of a longer time period. Paper home diaries are problematic since they may not be filled in at designated times (Stone et al. 2003). Another major problem with paper diaries, besides the reliability issue, is the inconvenient storage of, and access to, the responses. Both these problems are overcome by electronic diaries, which can show good compliance in PD (Nyholm et al. 2004, Nyholm et al. 2005, Lyons and Pahwa 2007).

Patients’ own assessment of function often does not correlate well with their actual performance in motor tests (Shulman et al. 2006). Therefore, more objective tests are warranted. However, non-motor symptoms of PD such as pain, depression and cognitive problems may affect quality of life at least as much as the motor symptoms do (Weintraub et al. 2008). Currently self-assessments or interviews are the only practical ways to assess the diversity of symptoms that may occur in order to determine if a treatment change was overall beneficial or not.
Most of the existing approaches to remote monitoring of PD symptoms have been based on the use of wearable sensors (e.g. Keijsers et al. 2003, Salarian et al. 2007, Giansanti et al. 2008, Giuffrida et al. 2009, Patel et al. 2009). Approaches to test motor function via appliances such as computer keyboards (Giovannoni et al. 1999, Hoffmann et al. 2008) and custom hardware testing devices (Goetz et al. 2009) have also been tried. These approaches often do not address the important issue of defining scores that combine different aspects of patient function in order to facilitate monitoring of the patients. To date, repeated assessments are usually summarized individually as percentages of responses at some levels (e.g. as percentage “good on-time” in a home diary or in the TRS) or as mean values of test results.

Decision support

According to van Bemmel (1997) medical informatics is a truly multidisciplinary and largely applied research field, which comprises theoretical and practical aspects of information processing and communication based on knowledge and experience derived from processes in medicine and health care.

In the diagnostic-therapeutic cycle, data are first generated by observation of patients. Information is then derived from the data by interpretation, leading to a diagnosis. Decision making regarding therapy thereafter is based on the diagnosis in combination with knowledge. By inductive reasoning with interpreted data, collected from many similar patients, new knowledge is obtained. This knowledge then can be used to support interpretation of other data and future decision making concerning therapy. Medical informatics supports medical sciences and all the phases of the diagnostic-therapeutic cycle by using methods and research on e.g. sensor technology, signal processing, measurement and test theory, statistics, data fusion, telecommunications, data warehousing, data mining, expert systems technology, mathematical modeling, control engineering, simulation, human computer interaction, and hardware and software engineering.

Shortliffe (1987) defines medical decision support systems (DSS) as computer systems designed to help health professionals make clinical decisions. Others (e.g. Wyatt and Spiegelhalter 1990) make narrower definitions so that the term medical DSS should only apply to active knowledge systems, which use patient data to generate case specific advice. Such system contains three main components: i) medical knowledge, for instance knowledge about a drug represented as a pharmacokinetic model; ii) patient data, for instance individual parameter values for the
pharmacokinetic model and iii) case-specific advice generation, for instance suggestion of drug dosage after applying the model to the data. In this work, Shortliffe’s broader definition, which does not necessarily require an inference engine and explicit advice generation, was applied. This view of medical DSS fully embraces IT-based methods and systems, which use medical knowledge, to support data collection and information processing and presentation in a manner useful for supporting clinical decision making.

Intestinal levodopa infusion

Levodopa in the presence of a stabilizer has a half-life of ca two hours and is absorbed in the small intestine. This mode of absorption combined with a short half-life causes problems when stomach emptying is irregular. Work driven by the neurology department at Uppsala University has shown that infusion of a levodopa/carbidopa gel (Duodopa®, Abbott Laboratories) into the duodenum can reduce fluctuations in both plasma levodopa levels (Nyholm et al. 2003) and clinical response (Nyholm et al. 2005) in patients with advanced PD. Administering infusion treatment is complex and requires special training of clinical staff and patients. Some patients can adjust their dosage well to their needs without assistance, whereas others may experience problems. For new candidates, the treatment is tried out using a nasal tube. Initial pump settings are calculated based on previous tablet dose and are adjusted on the basis of after-dose response during some days in the hospital. Typically, the pump is shut off at night. The day starts with a morning bolus dose to reach steady state and a continuous flow rate is supplied thereafter. In addition it is possible to take extra doses if needed depending on food intake, physical activity etc. An important issue in infusion treatment is how to individually adjust the continuous dose and morning bolus dose as accurately and fast as possible.

Pharmacokinetic-pharmacodynamic modeling

A core aspect of pharmaceutical development is to study the uptake, distribution, elimination and physiological effects of drugs. Mathematical modelling of these processes is known as pharmacokinetic-pharmacodynamic (PKPD) modelling. Typical models of drug absorption, distribution and elimination consist of systems of ordinary differential equations representing transport of molecules between “compartments” with some apparent volumes. The pharmacodynamic (or effect) models are often sigmoid functions relating the effect compartment drug concentration to the measured effect (e.g. Sheiner et al. 1979). An excellent interactive tutorial
was made by Barclay et al. at University of Otago, Christchurch, New Zealand and is available at www.icp.org.nz (last accessed Oct 2010).

In population PKPD modelling, data from several individuals are processed and population mean and individual variability parameters are estimated simultaneously. This is an efficient way of utilising data since data from individuals with very few observations can still contribute to the population model (Sheiner et al. 1992). In the non-linear mixed effects modelling software package NONMEM (Icon Development Solutions, Ellicott City, MD, USA), parameters are estimated using a maximum likelihood approach.

Levodopa in the presence of a stabilizer follows two-compartment pharmacokinetics (Chan et al. 2005) and its short-term effect on PD is often well described by sigmoid Emax models. As PD progresses, sigmoidicity increases (Adamiak et al. 2010, Contin et al. 2001, Troconiz et al. 1998, Harder et al. 1998). There is also a long-term effect of levodopa, which was included in a PKPD model by Chan et al. (2005).
Aims

The overall aim of this thesis work was to develop, deploy and evaluate new IT-based methods for supporting treatment and assessment of treatment of Parkinson’s disease with motor complications. The main underlying problem was the difficulty of following up and assessing effects of treatment in this patient category, where a number of different symptoms may occur with differing frequency and severity. Another problem was how levodopa/carbidopa intestinal gel infusion treatment can be optimized for an individual. Specifically, the aims of the different papers were:

- **Paper I:** To find patient variables related to degree of improvement with levodopa infusion and to identify and evaluate a predictive model.

- **Paper II:** To describe the design and preliminary evaluation of a new test battery for symptom home assessments during motor complications.

- **Paper III:** To describe and evaluate a new method for assessment of drawing impairment in traced Archimedean spirals on a touch screen.

- **Paper IV:** To describe a web application for feedback of test results from the test battery described in paper II.

- **Paper V:** To assess test-retest reliability, internal consistency, convergent validity, known groups’ validity, compliance and usability of the test battery described in papers II and IV.

- **Paper VI:** To identify and estimate parameters of a population pharmacokinetic-pharmacodynamic model for levodopa infusion.
Materials and methods

Paper I

Patients
Data from all patients who completed two previous studies according to the study protocols were analyzed. These were twelve patients (study 1, Nyholm et al. 2003) and eighteen patients (study 2, Nyholm et al. 2005) with advanced, clinically diagnosed PD, suffering from motor fluctuations and dyskinesia in spite of individually optimized treatment. Characteristics of the patients are presented in Table 1.

Study designs
Both studies had so-called “cross-over” designs, where half the patients (randomly selected) first received standard oral treatments (levodopa slow-release tablets in study 1 and any individual drug combination in study 2) and secondly the levodopa/carbidopa intestinal gel infusion (Duodopa). The other half first received infusion and they then switched to oral treatment. Between the treatment changes, there was a so-called “wash-out” period to avoid effects carried over from the first treatment. Doses for all treatments were individually adjusted during a “wash-in” period. At the end of each treatment period, patients’ conditions were assessed using the UPDRS. On some days, patients were video-recorded each hour or half-hour, performing certain movement tasks and motor performance was rated by neurologists on the TRS scale (see introduction section) from severe PD symptoms (-3) to severe involuntary movements (“choreatic dyskinesias”, +3). In study 2, the patients also self-assessed symptoms e.g. their ability to walk based on 5 categories from 1 (unable) to 5 (walks well) in an e-diary. Before starting the cross-over treatment periods, all patients were assessed at a “baseline” during the oral treatment.
Table 1. Baseline characteristics of the patients in study 1 (no. 101-112, Nyholm et al. 2003) and study 2 (no. 201-218, Nyholm et al. 2005)

<table>
<thead>
<tr>
<th>Patient no</th>
<th>Sex</th>
<th>Age (years)</th>
<th>Age at PD onset</th>
<th>Duration of levodopa therapy</th>
<th>Baseline Hoehn &amp; Yahr at worst</th>
<th>Total UPDRS 1+2+3+4</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>M</td>
<td>49</td>
<td>35</td>
<td>13</td>
<td>4</td>
<td>63</td>
</tr>
<tr>
<td>102</td>
<td>M</td>
<td>56</td>
<td>46</td>
<td>7</td>
<td>5</td>
<td>59</td>
</tr>
<tr>
<td>103</td>
<td>M</td>
<td>69</td>
<td>40</td>
<td>26</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>104</td>
<td>F</td>
<td>69</td>
<td>40</td>
<td>23</td>
<td>5</td>
<td>53</td>
</tr>
<tr>
<td>105</td>
<td>M</td>
<td>76</td>
<td>55</td>
<td>21</td>
<td>4</td>
<td>57</td>
</tr>
<tr>
<td>106</td>
<td>M</td>
<td>67</td>
<td>44</td>
<td>16</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>107</td>
<td>M</td>
<td>39</td>
<td>31</td>
<td>4</td>
<td>3</td>
<td>19</td>
</tr>
<tr>
<td>108</td>
<td>M</td>
<td>67</td>
<td>59</td>
<td>6</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>109</td>
<td>M</td>
<td>48</td>
<td>37</td>
<td>8</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>110</td>
<td>M</td>
<td>63</td>
<td>54</td>
<td>7</td>
<td>4</td>
<td>27</td>
</tr>
<tr>
<td>111</td>
<td>M</td>
<td>61</td>
<td>47</td>
<td>14</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>112</td>
<td>F</td>
<td>70</td>
<td>50</td>
<td>18</td>
<td>4</td>
<td>42</td>
</tr>
<tr>
<td>201</td>
<td>F</td>
<td>53</td>
<td>46</td>
<td>8</td>
<td>5</td>
<td>54</td>
</tr>
<tr>
<td>202</td>
<td>M</td>
<td>65</td>
<td>43</td>
<td>21</td>
<td>4</td>
<td>70</td>
</tr>
<tr>
<td>203</td>
<td>M</td>
<td>70</td>
<td>56</td>
<td>14</td>
<td>4</td>
<td>57</td>
</tr>
<tr>
<td>204</td>
<td>M</td>
<td>51</td>
<td>38</td>
<td>12</td>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>205</td>
<td>M</td>
<td>51</td>
<td>44</td>
<td>8</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>206</td>
<td>M</td>
<td>63</td>
<td>53</td>
<td>12</td>
<td>3</td>
<td>62</td>
</tr>
<tr>
<td>207</td>
<td>M</td>
<td>65</td>
<td>50</td>
<td>13</td>
<td>3</td>
<td>44</td>
</tr>
<tr>
<td>208</td>
<td>M</td>
<td>59</td>
<td>49</td>
<td>10</td>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>209</td>
<td>M</td>
<td>63</td>
<td>46</td>
<td>19</td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td>210</td>
<td>M</td>
<td>67</td>
<td>58</td>
<td>10</td>
<td>5</td>
<td>63</td>
</tr>
<tr>
<td>211</td>
<td>M</td>
<td>68</td>
<td>58</td>
<td>10</td>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>212</td>
<td>M</td>
<td>58</td>
<td>43</td>
<td>13</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>213</td>
<td>M</td>
<td>71</td>
<td>56</td>
<td>14</td>
<td>4</td>
<td>68</td>
</tr>
<tr>
<td>214</td>
<td>F</td>
<td>60</td>
<td>39</td>
<td>18</td>
<td>5</td>
<td>78</td>
</tr>
<tr>
<td>215</td>
<td>F</td>
<td>69</td>
<td>59</td>
<td>12</td>
<td>4</td>
<td>82</td>
</tr>
<tr>
<td>216</td>
<td>F</td>
<td>75</td>
<td>62</td>
<td>14</td>
<td>2</td>
<td>69</td>
</tr>
<tr>
<td>217</td>
<td>F</td>
<td>71</td>
<td>66</td>
<td>6</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>218</td>
<td>M</td>
<td>79</td>
<td>67</td>
<td>13</td>
<td>3</td>
<td>68</td>
</tr>
</tbody>
</table>

Mean Age at PD onset: 63.1 years
Mean Duration of levodopa therapy: 13.0 years
Mean Baseline Hoehn & Yahr at worst: 3.9
Mean Total UPDRS 1+2+3+4: 52.1

SD Age at PD onset: 9.2 years
SD Duration of levodopa therapy: 5.4 years
SD Baseline Hoehn & Yahr at worst: 0.8
SD Total UPDRS 1+2+3+4: 16.7

Median Age at PD onset: 65 years
Median Duration of levodopa therapy: 48 years
Median Baseline Hoehn & Yahr at worst: 13 years
Median Total UPDRS 1+2+3+4: 53.5

Min Age at PD onset: 39 years
Min Duration of levodopa therapy: 4 years
Min Baseline Hoehn & Yahr at worst: 2 years
Min Total UPDRS 1+2+3+4: 19

Max Age at PD onset: 79 years
Max Duration of levodopa therapy: 67 years
Max Baseline Hoehn & Yahr at worst: 26 years
Max Total UPDRS 1+2+3+4: 82
Outcome measures

Measures of disease severity were defined as UPDRS parts 2 (activities of daily living, ADL) and 3 (motor examination) and total UPDRS, as mean of self-assessments of mobility and satisfaction (only study 2), percentage functional on-time (proportion of TRS ratings from -1 to +1) and as mean squared deviation, MSD from the best condition (zero) of ratings on the TRS

\[
MSD = \frac{\sum_{i=1}^{n} (TRS_i)^2}{n}
\]

where n is the total number of ratings for the treatment period. Improvement measures (absolute, relative and ranked) were based on change in severity measures. If a high score in a severity measure represented bad function, improvement was defined as reduction in severity. On the other hand, if a high score represented good function, improvement was defined as increase in that measure.

Data analysis

Pearson correlation coefficients between measures of improvement and a number of patient-related numerical baseline variables, such as age, disease duration and disease severity were calculated. For control purposes, correlations between oral minus baseline treatment periods “improvement” and corresponding baseline severity measures also were calculated. For total UPDRS, this “baseline-to-baseline” difference (oral minus baseline) was used as a covariate in addition to baseline severity in an analysis of covariance (ANCOVA) model to estimate the effect of regression to the mean (RTM) (Barnet et al. 2005). The purpose of ANCOVA is to compare two or more linear regression lines. It is a way of comparing the Y variable among groups while statistically controlling for variation in Y caused by variation in the X variable (Mcdonald 2009).

Modeling

Plotting the various improvement measures vs their corresponding baseline severity measures generally revealed good linearity in the relations as judged by visual inspection. Therefore a linear model seemed appropriate for this range of data. Total UPDRS was selected since it was available at baseline in both studies and it represents a wide spectrum of PD symptom severity that can be assessed in clinical practice. Using total UPDRS data from study 2, a linear regression model was estimated. The obtained regression equation was
then applied to data from study 1 to predict improvement in that study. To assess quality of the prediction, goodness-of-fit ($r^2$) was calculated according to

$$r^2 = 1 - \frac{SSE}{SSM}$$

where

$$SSE = \sum_{i=1}^{n} (I_{actual_i} - I_{predicted_i})^2$$

and

$$SSM = \sum_{i=1}^{n} (I_{mean} - I_{actual_i})^2$$

where $I_{actual}$ is the observed improvement in total UPDRS and $I_{predicted}$ is the result after applying the linear equation. $I_{mean}$ is the mean of the actual improvements.

**Paper II**

**Description of the test battery**

A test battery for assessing symptoms in patients with advanced PD was developed for a personal digital assistant (PDA) in a telemedicine setting. The tests consist of both self-assessments of symptoms and motor function tests (tapping and drawing) on a touch screen. Assessments are performed several times per day in the patient’s home during test periods of one week in duration. Results are uploaded to a central server and processed into symptom dimensions and an ‘overall score’ reflecting the global condition of a patient during a test period. The system is illustrated in Figure 1.
Figure 1. The patient answers questions and performs tests on several occasions per day during test periods of one week’s length. Raw data is transmitted over the mobile net to a central server where summaries in different symptom dimensions and an overall test score per patient and test period are calculated and stored. There is a web application that graphically presents feedback on the test results to the treating clinical staff to assist decision making concerning treatment.

The test battery was implemented on a Qtek 2020i Pocket PC device with 3.5” touch screen with 240×320 dots resolution, running a Java virtual machine. The hand computer system software was separated in two parts, one to manage underlying hardware functions that are written in C++, and another to manage the test battery application, written in Java. The application uses a standard model view control (MVC) design pattern to keep the different areas of the system separated and standardized. The software was incorporated into the permanent read only memory (ROM) of the device. This ensures that if battery power runs out completely, the software installs itself automatically after the battery has been recharged. Data files are stored on a flash memory card and a communication protocol based on transfer of XML files enables transfer of patient data from the
PDAs to a central server and synchronization of the clock time. The server software and web application are described under the heading Paper IV.

The user interface of the PDA application consists of views and navigation between them. There is one set of views for clinical staff and another set of views for patients. The clinical staff log in to assign units to patients, to label test periods and to demonstrate the patient interface. The patients answer questions and perform motor tests. The question- and test sequence starts after “Start diary” is pressed during an open time slot. An audible signal notifies when it is time to enter; entry is only possible during a limited time. Each question text is written in a readable font on the upper part of the screen and answer alternatives are given below. All questions and tests except the last one have a button with text: “Next” at the bottom. The last screen in the sequence has the text: “Done”. Development of the user interface was done by iterating prototypes with input from selected pilot patients and clinicians.

The questions and their answer alternatives are given in Table 2a. Questions 1-6 relate to the previous 4h, or ‘this morning’, depending on the actual time-of-day. Question 7 relates to ‘right now’ and allows seven steps from −3 (very off) to +3 (very dyskinetic). Question 2 gives three answers in percent, whereas questions 1, 3, 4, 5 and 6 are of verbal descriptive scale type between 1 (worst) and 5 (best). Descriptions and illustrations of the user interface of each motor test are given in Table 2b. The tapping tests (Table 2b, #8-11) display square tapping areas (“buttons”) with a side of approximately 15mm, where at least one button will be active (red color in the illustrations). All buttons become inactive after 20 s after the first button was pressed. Results of each tapping test are assessed as speed and accuracy (proportion correct taps). The spiral test (Table 2b, #12-14) displays a picture of an Archimedean spiral, which the patient traces with the stylus from the center and out. Pixel-coordinates and times for the drawing are stored as each screen pixel is touched (event driven sampling). The spiral test is repeated 3 times per time slot. Results of the spiral test presented in paper II are standard deviation of frequency-filtered drawing velocity (SDDV) according to the method by Liu et al. (2005). Another method is described in paper III.
Table 2a. The self-assessment questions and corresponding answer alternatives

<table>
<thead>
<tr>
<th>Item#</th>
<th>Text</th>
<th>Answer alternatives</th>
</tr>
</thead>
</table>
| 1     | Have you had difficulty to walk (ca 100 meters) during the last four hours/this morning? | - Unable to walk  
- Difficult  
- With effort  
- Fairly well  
- Walks well |
| 2     | Mark the proportion of time you have been off, on and dyskinetic, during the last four hours/this morning. | Tap to place two lines in the figure and move them to adjust percentages.  
|       |                                                                 | Off: 30%  
On: 40%  
Dyskinetic: 30% |
| 3     | How much off have you been at worst during the last four hours/this morning? | - Extremely off  
- Quite a bit  
- Moderately  
- Slightly  
- Not at all off |
| 4     | How much dyskinetic have you been at worst during the last four hours/this morning? | - Extremely dyskinetic  
- Quite a bit  
- Moderately  
- Slightly  
- Not at all dyskinetic |
| 5     | Have you had painful cramps (dystonia) during the last four hours/this morning? | - All the time  
- Most of the time  
- Half the time  
- A little of the time  
- No cramps at all |
| 6     | Have you been satisfied with your functioning during the last four hours/this morning? | - Not at all satisfied  
- Slightly  
- Moderately  
- Quite a bit  
- Completely satisfied |
| 7     | How is your condition right now?                                      | - Very dyskinetic  
- Moderately dyskinetic  
- A little dyskinetic  
- On  
- A little off  
- Moderately off  
- Very off |
Table 2b. Motor tests definitions

<table>
<thead>
<tr>
<th>Test #</th>
<th>Test description</th>
<th>Test area</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Alternately tapping of two fields (un-cued) using right hand. Two active buttons (red) are displayed. The buttons both become inactive (black) 20 s after the first button was pressed.</td>
<td><img src="#" alt="Red" /> <img src="#" alt="Red" /></td>
</tr>
<tr>
<td>9</td>
<td>Same as 8 but using left hand.</td>
<td><img src="#" alt="Red" /> <img src="#" alt="Red" /></td>
</tr>
<tr>
<td>10</td>
<td>Two buttons, where one is active (red) and the second is inactive are displayed. The buttons alternately become active with increasing speed. Alternations start with 2 seconds per tap and speed increases with constant acceleration to 0.5 seconds per tap after 20 seconds. A sound is heard and a small filled circle is briefly (timeout) shown after a tap. Sounds on correct and missed taps are different. Dominant hand is used.</td>
<td><img src="#" alt="Red" /> <img src="#" alt="Red" /></td>
</tr>
<tr>
<td>11</td>
<td>Tapping an active field that changes location when tapped. Four buttons, where one is active (red) and the other three are inactive, are displayed. When an active button is tapped, one (randomly selected) of the other three buttons becomes active. Cueing is supplied as in test 10 above. Dominant hand is used.</td>
<td><img src="#" alt="Black" /> <img src="#" alt="Red" /> <img src="#" alt="Black" /> <img src="#" alt="Black" /></td>
</tr>
<tr>
<td>12-14</td>
<td>Tracing a pre-drawn Archimedean spiral from the centre and out using dominant hand without supporting it. The test is repeated three times and patients are instructed that it should be completed in approximately 10 seconds per drawing.</td>
<td><img src="#" alt="Spiral" /></td>
</tr>
</tbody>
</table>

Design considerations

Selection of self-assessment questions was based on questions from e-diaries used in two previous studies (Nyholm et al. 2004, Nyholm et al. 2005). Strongly correlated questions, unclear questions, questions that gave low variance in answers and questions giving answers that were near the maximum or minimum alternatives in these studies were removed or modified. Motor test designs were inspired by other tests and test batteries, such as those described by Giovannoni et al. (1999), Taylor Tavares et al. (2005), Williams et al. (2005) and Liu et al. (2005). The timed test battery by Haaxma et al. (2008) assesses (a) walking, (b) writing, (c) single and double-handed pegboard performance, (d) finger tapping, and (e) rapid alternating forearm movements. Literature searches were performed and user opinions were collected from clinicians and patients and items were added or
modified accordingly. When applicable, the methodology followed that described in the FDA’s Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (2009). The ambition was to capture a few symptom dimensions that were important to most sufferers from advanced PD. It was necessary to minimize the time and effort for the patients to secure feasible testing when performed repeatedly on a touch screen device at home.

Item 1 assesses the patient’s perception of ability to walk. Gait problems are both frequently occurring and important in advanced PD, as is reflected by its inclusion in all the major rating scales: UPDRS (Fahn et al. 1987, item #15), MDS-UPDRS (Goetz et al. 2008, items #2.12 and #2.13), PDQ-39 Jenkinson et al. 1997, item #5) and SF-36 (Ware and Sherbourne 1992, item #3i). It was presumed possible for most patients to determine their own walking ability well enough for self-assessment of this item. The alternative would have been to analyze gait patterns from wearable sensors and this would lead to increased cost and complexity, both in data collection and processing.

Item 2 assesses both perceived off-time and time with dyskinesia. These are typical outcomes from on/off home diaries (e.g. Hauser et al. 2000) and are included in UPDRS (items #32 and #39) and MDS-UPDRS (items #4.1 and #4.3). One aspect that was taken into account was that in the previous e-diary (Nyholm et al. 2005) patients often answered that they had been “off” most of the time in one question yet in response to another question they answered that they had been “dyskinetic” most of the time. In order to avoid such contradictions and to force patients to rate the predominant motor disability, this combined item was constructed.

Item 3 assesses the perceived magnitude of off-symptoms at worst which is included in items #2.10 and #4.4 of MDS-UPDRS. This self-assessed off-severity will have a different meaning from patient to patient, depending on individual symptom patterns during off-periods. Similarly, item 4 assesses the perceived magnitude of dyskinetic-symptoms at worst which is included in items #33 of UPDRS and #4.2 of MDS-UPDRS.

Item 5 assesses the amount of time the patient experiences painful cramps. This item is also found in item #4.4 of MDS-UPDRS and item #37 of PDQ-39.

Item 6 assesses the extent to which the patient has been satisfied with his or her overall functioning. This relates to items #1, #9a, #9d, #9e and #9h of the SF-36. The main purpose of this item was to account for individual differences in symptom patterns and preferences. Probably this item will to a
large extent measure mood/depressive symptoms and ability to manage daily activities. Hauser et al. (2009) used “subject-reported clinical global impression” (CGI) in the same spirit.

Item 7 asks the patient to self-assess his or her condition on the TRS scale (Nyholm et al. 2005) just before starting to do motor tests. It probably contains the information of items 3, 4 and 6 but in the temporal context of “right now”. The main purpose of this item was to relate the motor test results to self-assessments for assisting in development of combined test scores.

Items 8 and 9 assess un-cued alternating tapping ability. Alternating finger tapping speed is widely used as a measure of bradykinesia (slowness of movements) in PD. Not much learning seems to take place in PD patients after more than one day’s exercise with this test (Nutt et al. 2000). It is included in item #23 of UPDRS and #3.4 of MDS-UPDRS. Giovannoni et al. (1999) used alternating finger tapping in their computer keyboard tapping tests and this was used by Martinez-Martin et al. (2005) as well. It is well known that movements in PD can be facilitated by external cues (Freeman et al. 1993) and that cued and un-cued tapping involve different brain functions (Majsak et al. 1998).

Item 10 assesses the patient’s ability to follow a rhythm. Dysrhythmia is a feature frequently associated with motor disturbance in PD (Yahalom et al. 2004). Kao et al. (2003) found that movement rhythm variation (MRV) was a good method to provide quantitative data for assessment of hand dexterity. MRV improved in stroke patients along with recovery and improved in PD patients after levodopa treatment. Stegemöller et al. (2009) examined performance of unconstrained syncopated index finger flexion movements. Syncopated movements were paced by an acoustic tone that increased in frequency. The principal finding was a marked impairment in the ability of patients with PD to perform syncopated movements when without medication. Van den Berg et al. (2000) examined rhythmic forearm movements at a range of pacing frequencies in PD patients and controls. The PD group displayed marked coordination problems over and above the known clinical motor symptoms of PD. They suggested that these parkinsonian coordination problems are due to cortical dysfunction, which develop during later stages of PD.

Item 11 assesses dexterity and reaction time. Reaction time tests are frequently used in neurocognitive test batteries such as “neuromarker” (Williams et al. 2005) and dexterity is assessed using peg board tests (e.g. Haaxma et al. 2008). Jancovic et al. (1999) found that both these functions improved after neurosurgery in PD patients.
Item 12 assesses the patient’s ability to trace a pre-drawn Archimedean spiral. Writing and drawing tests are examples of a simple approach which investigators have used for a long time in the assessment of fine motor changes in movement disorders (Kraus and Hoffmann 2010). The most prominent example of this approach is spiral drawing. Spiral drawings are widely used for assessment of tremor in PD and Bain and Findley (1993) used spiral drawings for defining a rating scale for tremor severity. Spiral tests on PC tablets have been used for quantification of involuntary movements in PD such as dyskinesias (Liu et al. 2005) and tremor (Aly et al. 2007, Rudzinska et al. 2007). Even slowness and stiffness have been assessed from spiral drawings on PC tablets, using different calculated spiral “indices” (Saunders-Pullman et al. 2008) and spiral drawing completion time (Banaszkiewicz et al. 2009).

In order to determine the minimum necessary sampling frequency and sampling duration for reliable assessment during fluctuations, the following analysis was performed: If a population mean is estimated using the sample mean from \( n \) observations from a distribution with variance \( \sigma^2 \) and \( n \) is large enough, the central limit theorem can be applied to obtain an approximate 95% confidence interval for the population mean

\[
\left( \bar{x} - B, \bar{x} + B \right)
\]

where

\[
B = \frac{2\sigma}{\sqrt{n}}
\]

If the sampling error is required to be smaller than the bound \( B \), then

\[
n > \frac{4\sigma^2}{B^2}
\]

In case of the e-diary in the previous study by Nyholm et al. (2005) (study 2), variance \( \sigma^2 \) was rarely > 1.0 for any question answer. Typical mean values for responses were around 4. If one is willing to accept an error of up to 10% of a typical mean, this leads to

\[
n > \frac{4 \cdot 1.0^2}{(4 \cdot 0.1)^2} = \frac{4}{0.16} = 25
\]

The median number of missed entries per week in the study was two. Therefore, a reasonable number of samples to require should be at least 27. This yields a suitable assessment scheme of 4 samples per day over 7 days.
A further analysis was performed to assess the variance in mean values comparing all two week’s mean with only the seven first day’s mean and only the three first day’s mean. It turned out that the variance of differences between seven days and two weeks was small (around 0.04) and between three days and two weeks, it was larger (around 0.2). Pearson correlations between all two week’s mean and first seven day’s mean were strong (typically around 0.95), compared to 0.80 for the first three days. Similar to these results, one full week of using on/off diaries was found adequate by Reimer et al. (2004) whereas three days were insufficient.

**Responsiveness evaluation**

An evaluation with two pilot patients was performed before and after receiving new treatments for advanced PD. These patients were also involved in the user interface design. Both patients were scheduled to receive the treatment changes and no interventions were made in relation to this study.

One patient was a 45-year-old woman who first had oral treatment in combination with deep-brain stimulation (DBS). Later she received levodopa/carbidopa intestinal gel infusion (Duodopa), replacing the oral treatment, but keeping the DBS. The other patient was a 65-year-old man who first received oral treatment and later received DBS and reduced his oral dose. The test battery was used four times per day during test periods of approximately one week, both before and after the treatment change. Treatment differences were evaluated using the two-sided Mann-Whitney non-parametric statistical test. Due to the phase of development, some questions and tests were changed between the test periods and therefore comparable data were not available for some items.

**Usability and compliance evaluation**

The test battery was used by 65 patients with advanced PD in nine clinics in Sweden in an open longitudinal study (DAPHNE, Duodopa in Advanced Parkinson’s: Health Outcomes & Net Economic Impact, EudraCT no. 2005-002654-21). The patients entered diary responses and performed motor tests four times per day during one to ten periods of seven day’s length. On inclusion, the patients were either receiving treatment with levodopa/carbidopa intestinal gel infusion (n=36), or they were candidates for receiving this treatment (n=29). Baseline characteristics are shown in Table 3. During most of these test periods, UPDRS ratings were performed in afternoons at the start of the week.
A questionnaire regarding practical experiences in using the test battery were collected from the nine study nurses. This was followed-up one year later by telephone interviews. To assess the internal consistency of the test battery, Cronbach’s $\alpha$ for the test variables was calculated. In order to study the learning effect, tapping results from the first three days in a test period were compared to results from the remaining four days in the period. Compliance with the test battery was calculated as proportion actually completed test occasions per expected test occasions.

Table 3. Clinical features at baseline (mean ± one SD) in the DAPHNE study

<table>
<thead>
<tr>
<th>PATIENTS (gender)**</th>
<th>Oral levodopa*</th>
<th>Levodopa infusion</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>37 (23m; 14f)</td>
<td>40 (27m; 13f)</td>
<td>77 (50m; 27f)</td>
<td></td>
</tr>
<tr>
<td>AGE (years)</td>
<td>63.6 ± 7.6</td>
<td>65.9 ± 6.9</td>
<td>64.8 ± 7.3</td>
</tr>
<tr>
<td>DURATION OF LEVODOPA (years)</td>
<td>10.8 ± 4.5</td>
<td>15.7 ± 5.9</td>
<td>13.3 ± 5.7</td>
</tr>
<tr>
<td>DURATION OF DUODENAL LEVODOPA (years)</td>
<td>2.2 ± 1.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOEHN &amp; YAHR AT BEST</td>
<td>2.2 ± 1.1</td>
<td>2.5 ± 0.7</td>
<td>2.4 ± 0.9</td>
</tr>
<tr>
<td>HOEHN &amp; YAHR AT WORST</td>
<td>3.8 ± 1.0</td>
<td>3.9 ± 0.8</td>
<td>3.8 ± 0.9</td>
</tr>
<tr>
<td>TOTAL UPDRS AT VISIT (AFTERNOONS)</td>
<td>50.1± 15.3</td>
<td>46.9± 16.7</td>
<td>48.6 ± 16.1</td>
</tr>
</tbody>
</table>

* 27 of these 37 patients later started levodopa infusion treatment.
** 12 of these 77 patients did not use the test battery (7 were not Swedish-speaking and 5 were thought unable to handle the device or unwilling to use it)

Paper III

Data collection

The patients included in the DAPHNE clinical trial (Table 3), entered self-assessments, performed tapping tests and traced Archimedean spirals four times per day, during several test periods of seven days’ length. On each test occasion, three spirals were drawn and a self-assessment (Item 7, table 2a) was entered in one of seven categories of motor function: “very off”, “moderately off”, “a little off”, “on”, “a little dyskinetic”, “moderately dyskinetic” and “very dyskinetic”. On test occasions, the patients were instructed to place the hand computer on a table and to be seated in a chair. Using an ergonomic pen stylus and using the hand they normally write with, they were instructed to trace a pre-drawn Archimedean spiral, shown on the
screen, as accurately as possible, supporting neither hand nor arm. In addition to the spiral test, 20-seconds-long tapping tests were performed on each test occasion. There were three types of tapping tests (Table 2b). For the tapping tests, mean speed and proportion of correct taps were calculated.

Manual rating of spiral drawing impairment

A web interface was constructed to display the spiral drawings from the spiral test and to allow users (PD specialists) to rate observed drawing impairment. Along with the image, the drawing completion time in seconds, and the patient’s self-assessment of motor condition at the time of the particular test occasion were displayed. Given ratings were stored in a database. The web interface was organized in four “tracks”: “preliminary rating”, “training”, “standardized rating”, and “rater agreement”. In these tracks, users could display spirals, rate them and change previously given ratings.

In a training phase, two raters studied the examples in the clinical handbook for assessment of tremor in spiral drawings in ten categories (Bain and Findley, 1993). In analogy, the same ordinal scale was applied for rating of drawing impairment in spiral drawings, not necessarily caused by tremor, but as well by other causes, such as choreatic dyskinesia or bradykinesia. First a rating of the overall drawing impairment was done and secondly a probable cause for the given impairment was marked. The ten categories of the ordinal scale were divided into four levels to assist in the assessment; categories 1-3 were defined as mild, 4-6 moderate, 7-9 severe and 10 extremely severe. The general impression of the shape of the spirals determined the level of severity; homogenous and symmetrical spiral shapes were rated as mild drawing impairment, larger deviations from the pattern of spiral shape were rated as moderate impairment and spirals with large interruptions, skewed or incomplete shapes were rated as severe. Drawings without signs of a spiral shape were rated as extremely severe. Within the levels of mild, moderate and severe, there were three scale steps to allow further refinement of the rating. Small and/or few deviations from the expected lines rendered low points on the scale. Completion time could assist in determining the cause of impairment – bradykinesia causes long completion times (Banaszkiewicz et al. 2009) whereas on-phase choreatic dyskinesia likely is associated with shorter time.

To achieve a common basis for assessment, one rater first browsed through spiral drawings in the “preliminary rating” track, to collect and rate at least ten representative examples of each of the ten drawing impairment categories. The examples were chosen from test occasions where all three spirals had similar degree of drawing impairment. Both raters then observed
these preliminary ratings in the “training track” and used them as templates for rating of drawings in other “tracks”.

In the “standardized rating” track, the web application displayed all three spiral drawings from three randomly selected (by a computer program) test occasions per patient. The users were instructed to rate each test occasion based on the global impression of all three spirals, drawn on the same occasion. Although drawn on the same occasion, sometimes drawings in this group of three were not homogenous and could not be given a common rating. These groups were replaced by drawings from other occasions. A “standardized manual rating” score (SMR) was defined as the mean of the two raters’ assessments.

In the “rater agreement” track, single spiral drawings from three randomly selected test occasions (different ones than those in “standardized rating”) from each patient were displayed in random order. These ratings were performed blinded to patients’ identity and to the other user’s rating.

The ”wavelet spiral test score” method

A novel method, using the discrete wavelet transform (DWT) and principal component analysis (PCA), was developed to process the spiral drawings in order to generate a ‘wavelet spiral test score’ (WSTS) to represent the drawing impairments. In the wavelet transform, filters of different cut-off frequencies are used to analyze the signal at different scales. The signal is passed through a series of high-, and low-pass filters to analyze high and low frequencies, respectively. The DWT analyzes the signal by decomposing it into its ‘approximate’ and ‘detailed’ information, which is accomplished with the use of successive high-pass and low-pass filtering and sub-sampling operations (Vetterli and Kovacevic, 1995). This procedure is repeated for further decomposition of the low-pass filtered signals. ‘Details’ are obtained when the signal is passed through the half-band high-pass filter whereas ‘approximation’ is obtained when the signal is passed through the half-band low-pass filter. Starting from the ‘approximation’, and ‘detailed’ coefficients, the inverse discrete wavelet reconstructs the signal, inverting the decomposition step by inserting zeros and convolving the results with the reconstruction filters.

In this work, decomposition of the signal was performed on the radius (square root of the sum of the squares of x and y coordinates). The type of wavelet chosen for the analysis and the number of the levels of decomposition are design parameters. In the current work and after experimentation, the decomposition was done into three levels using the Daubechies (db10) wavelet family to obtain the coefficients (Daubechies,
Accordingly, 256 features were obtained after performing DWT on radius.

Since there were a large number of features extracted by DWT and the aim was to create one single score, principal component analysis (PCA) (Jolliffe, 1986) was performed after extracting features. The main advantage of PCA is that it reduces the number of dimensions without much loss of information. PCA returns a square coefficient matrix with the size of feature vectors in which each column contains coefficients for one principal component. The data is expressed in terms of eigenvectors and these vectors are sorted by eigenvalues from the highest to the lowest to get the components in the order of contribution to variance. The first principal component represents the direction in feature space corresponding to maximum variance.

PCA was first applied on a subset of data, preselected on the basis of the 10% worst and 10% best tapping results. Tapping results were assessed based on both speed and accuracy (by summation of normalized results). By doing this, the first principal component of radius was assumed to provide a direction in feature space well representing PD symptom severity. Having both the feature vectors from all the spirals (full dataset) and the obtained coefficient matrix from the subset of the 10% worst and 10% best tapping results, the PCA coefficients thereafter were applied to the full dataset by multiplying the DWT of radius features by the coefficient matrix. An illustration of the concept of subset selection before calculating a first principal component, in order to find a desired direction in a multidimensional feature space is shown in Figure 2. The 10% worst tapping results can be represented by subset A and the 10% best results, by subset B.

Post-processing (calibration) of the first principal component was done using logarithmic and linear transformations, to bring it to a roughly linear interval scale between 0 and 10, comparable to the manual rating scale. To define the linear transformation, the rated spirals from the “training track” were used. This was done to reduce systematic errors, such as over-prediction of low impairment or under-prediction of high impairment.
Data analysis

Test-retest reliability was assessed by Spearman rank correlations (r). Since there were three spirals per test occasion (A, B, C), the mean of all three possible correlations (AB, AC, BC) was taken. Between-rater agreement and between-methods comparisons were also assessed by Spearman rank correlations. In order to avoid the problem of multiple test occasions per individual, 200 random samples of single test occasions per patient were drawn and mean correlations in this sample was calculated. Bland-Altman analysis of difference (Bland and Altman, 1986) was used to estimate the prediction error of the WSTS versus the SMR. Correlations between the different spiral scores and tapping tests results were assessed on test occasion level, first taking mean values of the three spiral drawings, while correlations with UPDRS were assessed after taking mean values over all (approximately $3 \times 28$) spirals drawn during each test period.
Paper IV

System description

For each test occasion, raw test data is sent from the PDA hand unit over the mobile net to the so-called remote device manager (RDM, Nordforce Technology). The RDM is a commercial product responsible for collecting and storing the data from the hand unit. There is a communication protocol between these two systems which handles the transfer of the data. The data processing sub system DPSS is a stand-alone application implemented in VB.NET, which incorporates knowledge to analyze and interpret the raw test battery data. It parses, processes and stores the data into database tables and at the same time calculates and stores test scores on test period level. A connection with RDM is first established followed by receiving and parsing XML data from files. Once files are received, they can be directly interpreted by the DPSS which runs during a specified time interval, e.g. every single hour. The data messages consist of patient identification, test period, hand unit identification, starting and ending time of the test occasion and responses to the test-battery items. The main reason for processing data centrally instead of locally is the risk of losing raw data in the remote devices if it is not uploaded regularly. Access to raw data centrally also facilitates future research and method developments. The data collection in the hand units is designed to minimize upload bandwidth.

For computing the spiral scores according to the previously described WSTS method, a single piece of M-code was written in Matlab® (MathWorks Inc). The M-function was encrypted and wrapped into a C# interface by using the Matlab Builder for .NET to be accessed by the sub system. This pre-built interface requires the Matlab Compiler Runtime (MCR) to be installed in the running machine. An error handler was designed so the application can recover from possible run-time errors without terminating by rolling back all the information and saving the error information to a log file.

The web user interface (WUI) is a feedback system comprising a secure web server and a database with web-based access for medical staff. The main role of the WUI in the overall system is to present test results to the end-user clients. Three-tier architectures ensure a high-level of availability where the different application components can be easily replicated to increase the overall performance (Casteleyn et al. 2009). WUI supports this architecture in which a web browser sends HTML requests using HTTP to a web server, which in turn passes the request to a common gateway interface (CGI) application program. This application server then sends requests to a database server, which generates the query result set and sends it back in
order to be formatted and presented to the end-user clients as an HTML page. The web application is built in ASP.NET as a code-behind model and ADO.NET is used as a standard way to connect to a database. Structured query language (SQL) is employed as a query language to extract and manipulate the data stored in Microsoft SQL-server. Data security is assured on three levels: forms authentication level; web-server level; and database-server level. Depending on the user credentials and their access permissions, the application limits what resources or functionalities are accessible (e.g. it enables access only to data belonging to a specific clinic). Some of the user requirement specifications for the web application are: easy to understand and use, user-friendly design, easy navigation and fast response. They were drafted in collaboration with experts and different prototype versions were developed in several iterations. To enable rapid patient status assessment, the information in the web application is displayed and ordered using a top-down approach where the general overview of the patient performance per test period is given. After the user logs in and selects a patient, the main page displays graphs of the patient scores on a test period level, focusing on an overall test score, symptom dimensions and daily summaries.

Calculation of test period summary scores
The rationale for selection of summary calculation methods was to define scores taking into account i) the severity, ii) the frequency and iii) the importance of occurring symptoms. Mean values are the obvious choices to represent levels and standard deviations are obvious for representing overall variation. In order to determine how much of the level and how much of the variation to include, principal component analysis was chosen. PCA is theoretically the best linear dimension reduction technique, in terms of least mean squared error, and the first principal component is the linear combination of the original variables (or direction in feature space) giving maximum variance (Joliffe 1986). To have at least three variables to base PCA on, also mean squared deviation from “the best” (MSD) was computed. This measure combines both level and variation. It would be possible to compute other time series related features, such as autocorrelations, moving averages and frequency transformations. There was however not rationale for doing this since possible gains in performance would be counterbalanced by a loss in interpretability. Some of the test-battery items measure the same construct and the information content of a test period with the test battery can be described by six dimensions: walking (based on item 1), satisfied (item 6), dyskinesia (items 2 and 4), spiral (all spirals), tapping (all tapping results) and off (items 2 and 3). In the interface, these dimensions are presented as regular hexagons (Figure 3). The ordering of the symptom dimensions were decided in collaboration with domain experts. Different criteria are used for determining the adequate number of principal
components to be used in model development. A popular one is to select a cumulative percentage of total variation to which it is desired that the selected principal components should contribute more than 70%. In this work, the justification of retaining only the first principal component for each test-battery dimension was based on the fact that a single value representing them was needed. Weighting of the symptoms dimensions was done using the UPDRS, since it is the most used scale for assessing PD today (Movement Disorder Society Task Force on Rating Scales for Parkinson’s Disease, 2003). The overall score was defined as a linear combination of the first principal components, with numerical weights estimated by regression technique to fit simultaneous clinical ratings on the UPDRS scale. Linear transformations were then used to scale dimensions and predicted overall score to a scale from 0 to 1 representing worst to best score respectively. The dataset used for optimizing this statistical procedure was collected from the clinical trial DAPHNE.

Figure 3. Illustration of test battery symptom dimensions during two test periods, arranged in regular hexagons in the WUI. A small ‘diamond’ represents bad function and a large ‘diamond’ represents good function.
Usability and response time evaluations

The web application was demonstrated to fifteen nurses from the nine clinics in Sweden who had experience using the mobile test battery device in the clinical trial DAPHNE, but they had not previously seen the web application. The evaluation was performed as a presentation session for the nurses where they asked the presenter to display certain data from a particular patient in whom they had an interest. The presenter took notes about the reactions from the group. At least one patient per clinic was showed.

One year later, the IBM Computer System Usability Questionnaire (CSUQ) was administered to evaluate the users’ satisfaction with the web application. The intended users are treating clinical staff, such as nurses. The questionnaire was web-based and more than one person per clinic involved in the study could check and update the answers to the questionnaire. CSUQ consists of 19 items on a seven-point Likert (1932) rating scale ranging from 1, “strongly agree” to 7, “strongly disagree”. Four usability scores can be calculated by averaging responses to the CSUQ items: Overall Satisfaction Score, System Usefulness, Information Quality, and Interface Quality (Lewis 1995). Data about gender, age and previous experience with computer applications were noted. The users were asked to perform a series of tasks using the web application, such as 1) login, 2) select a patient, 3) check the patient’s performance by looking at the graphs of summarized scores and compare these results with own clinical observations, 4) select and check the other patients’ results repeating steps 2 and 3, 5) complete the survey, and 6) logout.

Microsoft Web Application Stress Tool (WAST) was used to stress the web server by realistically simulating a large number of users accessing the web application at the same time. The main measurements include response time and throughput (Meier et al. 2004). The tests were run from a separate machine, running at 100 Mbits/sec with varying load levels (concurrent connections) of 1, 10, 50, and 100. The web server, IIS 6.0, runs on Windows Server 2003 with a CPU speed of 2.93 GHz and 1 GB of RAM. A script was developed, capturing a test scenario with a browser. The test variability of the metrics was measured by repeating each test 10 times per load level of 1 minute run time.
Paper V

Patients

Thirty-five subjects with a clinical diagnosis of idiopathic PD in Milan, Italy, were included in the TEVAL study. The patients were divided into two groups: i) advanced patients, experiencing on/off fluctuations (F), and ii) less severe, clinically stable patients (S). Patients of group F were matched on age (± one year) and gender with patients of group S. There were two study dropouts (withdrawn consent) and for three patients, a matching subject could not be found. So there were thirty patients in the matched pairs, used for comparison between groups. Clinical features are presented in Table 4. All available patients’ data were used for assessing usability and compliance.

Table 4. Baseline characteristics for the age and gender matched patient groups (mean ± one SD) in the TEVAL study.

<table>
<thead>
<tr>
<th>PATIENTS (n, gender)</th>
<th>Fluctuating (F)</th>
<th>Stable (S)</th>
<th>Combined group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 (13m; 2f)</td>
<td>15 (13m; 2f)</td>
<td>30 (26m; 4f)</td>
</tr>
<tr>
<td>AGE (years)</td>
<td>64.7 ± 6.9</td>
<td>64.6 ± 7.0</td>
<td>64.7 ± 6.8</td>
</tr>
<tr>
<td>AGE AT ONSET (years)</td>
<td>54.0 ± 8.7</td>
<td>56.5 ± 9.2</td>
<td>55.2 ± 8.9</td>
</tr>
<tr>
<td>DURATION OF DISEASE (years)</td>
<td>10.7 ± 4.4</td>
<td>8.1 ± 5.7</td>
<td>9.4 ± 5.2</td>
</tr>
<tr>
<td>L-DOPA EQUIVALENT DAILY DOSAGE (mg/day)</td>
<td>1024 ± 419</td>
<td>618 ± 416</td>
<td>814 ± 458</td>
</tr>
<tr>
<td>HOEHN &amp; YAHR ON</td>
<td>2.2 ± 0.4</td>
<td>1.8 ± 0.4</td>
<td>2.0 ± 0.4</td>
</tr>
<tr>
<td>HOEHN &amp; YAHR OFF</td>
<td>2.8 ± 0.9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Inclusion criteria for group F were: experiencing motor fluctuations, more than two hours per day in either off or moderate to severe dyskinesia. Inclusion criteria for group S were: clinically stable without motor fluctuations (UPDRS IV items 32-39 all = 0) but receiving treatment for PD and having a match in group F.

Exclusion criteria were the same for both groups: i) corrected score (for gender, age and education using the Italian table of correction (Measso et al. 1993)) of less than 27 in MMSE (Mini Mental State Examination (Folstein et al. 1975)); ii) presence of active psychiatric disorders; and iii) significantly impaired mobility for other reasons than idiopathic PD (i.e. severe joint disease); and iv) inability to handle the device for some other reason. There were no restrictions about prior and concomitant therapy; however medication regimes and doses were stable during the observation period. There were no exclusion criteria based on education level.
Study design

Inclusion and baseline assessments were performed on observation day 0. The patients then used the test battery at home for one week between day 1 and 8 (test period 1) and again after one week’s interval between day 14 and day 22 (test period 2). The Hauser home diary (Hauser et al. 2000) was filled in on days 3, 5, 7 in test period 1 and on days 16, 18 and 20 in period 2. Each test period ended with a clinic visit (days 8 and 22), including Hoehn and Yahr (1967) staging and UPDRS assessments in on-state (both groups) and in off-state (group F), if off occurred spontaneously during the time of the visit, including a few hours’ observation. The PDQ-39 scale for activities of daily living was also assessed at the end of the test periods. Medication times and dosages were kept constant throughout the study.

All patients were assessed at baseline by a neurologist and a neuropsychologist. The patient's date of birth, gender, height, weight, duration of disease, years on levodopa therapy, medical history, current PD treatment and other medications were noted. The following baseline scores were assessed: Hoehn and Yahr; MMSE, MADRS (Montgomery and Åsberg 1979); PDQ-39; and UPDRS. At baseline, patients received the handheld device (PDA) and were invited to perform a demo session, in order to get acquainted with it.

All texts in the test battery were translated into Italian. On test occasions, the patients were instructed to place the hand computer on a table and to be seated in a chair and to use an ergonomic pen stylus. The patients were instructed to complete tests each day in a test period, four times per day at sessions starting at specific hours (at 8, 12, 16, and 20 o clock) during both periods. These sessions were announced by a loud ringtone. After the first ringtone, the patient was instructed to perform the tests within one hour. During this time, if patient did not perform the tests, three other ringtones, (each 15 minutes) reminded the patient to complete the task. After this hour, the session was considered “lost” and entry was not possible until the next session.

The information content of the data from a test period with the test battery can be described by six summary measures (dimensions): ‘off’, ‘dyskinesia’, ‘walking’, ‘satisfaction’, ‘spiral’, and ‘tapping’ and an overall test score. All the equations for calculation of the summary measures were defined in the DAPHNE study and performance of the test battery’s overall score was assessed based on new data from the present TEVAL study.

At the end of the second test period, the patients were asked to fill out self-reported user satisfaction questionnaires at home. The IBM After-Scenario
Questionnaire (ASQ) is a three-item closed-ended ordinal questionnaire, based on three 7-point graphic Likert scales (1932). The items address three important components of user satisfaction with general computer systems usability: i) ease of task completion, ii) time to complete a task, and iii) adequacy of support information (on-line help, messages, and documentation) (Lewis 1995). The second part was a newly constructed questionnaire, consisting of five questions. Patients responded anonymously to the usability questionnaires. In practice, the patients were instructed to fill the forms anonymously (no name, no date, writing in capital letters) and drop them in a box in the office to collect them. The box was opened only after all questionnaires had been collected at the end of the study.

Data analysis
Compliance was assessed as the number of completed test occasions per test period, counting from the first test occasion and until the last expected occasion, approximately one week later. Compliance and results from the usability questionnaires were summarized using descriptive statistics. To assess the internal consistency of the test battery, Cronbach’s α for the six dimensions was calculated. Test-retest reliability was assessed using intra class correlation coefficients (ICC) of results from the two test periods, separately for the two patient groups and for the pooled data set (combined group). Convergent validity was assessed using Spearman rank correlations for the combined patient group. First, mean values of both test periods were taken for overall score and the other measures. Fisher’s Z-transformation statistic was used for testing if Spearman correlations were stronger than 0.5 (absolute values). In group F, the mean value of UPDRS and Hoehn and Yahr in on and off states was taken, whereas in group S only the on state existed. For seven of the patients in group F, UPDRS ratings were available in both on and off, whereas, the remaining eight patients in this group had only ratings in on. It was assumed that mean percent change was the same in these eight patients as in the other seven. The Mann-Whitney statistical test was used to test if the overall score’s median was different between the stable and fluctuating patient groups. In order to not introduce a bias, all weights for calculating the dimensions and overall scores were fixed to their previous values. In order to study potential learning effects in the motor tests, results were compared between the test periods.
Paper VI

Patients and study designs

The modeling involved pooling data from three studies and fixing levodopa disposition parameters to values found in literature (Chan et al. 2005). The disposition of the patients in all the studies is shown in Table 5. The first study involved levodopa plasma concentrations and motor function assessments from 12 infusion patients, studied on three occasions and was described previously by Nyholm et al. (2003). It turned out that data from this study alone were insufficient for PKPD model identification, since available levodopa plasma concentrations were generally stable. In order to successfully identify model parameters, doses giving both high and low concentrations and effects are required. A small additional study was therefore executed.

This second study, PEDAL, Parameter Estimation for Duodenal Administration of Levodopa, involved three patients studied on two occasions each. A ‘bolus’ dose (normal individual morning dose plus 50%) was given with the Duodopa pump after a washout during the night. Data collection continued until the clinical effect was back at baseline, or at most 4 hours. At this point, the patient’s normal infusion rate was started. Data were collected for another two hours. This procedure was performed on two non-consecutive days per patient. Blood samples and effect measurements were collected in 5 to 15 minutes intervals. The main effect variable was clinical assessment of motor function from video recordings on a treatment response scale (TRS) where -3 represents severe PD symptoms and +3 represents severe choreatic dyskinesia (Nyholm et al. 2005). Blood samples were immediately cold-centrifuged and plasma samples were frozen in -70°C until analyzed by HPLC as described by Nyholm et al. (2003).

The third study, Nyholm 2010, (Nyholm D, Johansson A, Lennernäs H, Aquilonius SM, Askmark H. Parkinsonism and dyskinesias in relation to plasma levodopa levels following different doses of duodenal levodopa/carbidopa infusion. 2010, unpublished data), involving five patients, was approved by the local ethics committee and the Swedish Medical Products Agency. Patients were studied during five 4-hour periods, in total 2½ days, using five different infusion rates. Doses well adjusted and judged as good as possible for an individual (optimized) were used on the first half-day (day 1), whereas on day 2, 120% of the optimized dose was given on the first half-day and 90% of the optimized dose was given the second half-day. On day 3, 80% of the optimized dose was given the first half-day and 110% was given the second half-day. All doses during test days
2 and 3 were blinded to patients. Sampling intervals were 20 to 30 minutes. Immediately after each blood sample, patients were video recorded while performing the standardized set of tasks for assessment on the TRS scale. Three raters independently performed the ratings in a randomized order and mean values were taken.

Table 5. Baseline characteristics for the patients in the reference study by Chan et al. 2005 and the patients in the three studies (Nyholm 2003, PEDAL and Nyholm 2010) used for model identification (mean ± one SD).

<table>
<thead>
<tr>
<th></th>
<th>Chan 2005</th>
<th>Nyholm 2003</th>
<th>PEDAL</th>
<th>Nyholm 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO. PATIENTS</td>
<td>20 (12m; 8f)</td>
<td>12 (10m; 2f)</td>
<td>3 (3m; 0f)</td>
<td>5 (3m; 2f)</td>
</tr>
<tr>
<td>(gender)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEIGHT (kg)</td>
<td>78.7 ± 12.4</td>
<td>66.7 ± 9.9</td>
<td>69.7 ± 14.2</td>
<td>62 ± 8.4</td>
</tr>
<tr>
<td>AGE (years)</td>
<td>59.8 ± 10.7</td>
<td>61.2 ± 11.0</td>
<td>62.3 ± 2.5</td>
<td>60.8 ± 6.1</td>
</tr>
<tr>
<td>AGE AT ONSET</td>
<td>56.1 ± 10.9</td>
<td>44.8 ± 8.6</td>
<td>44.3 ± 5.7</td>
<td>48.8 ± 7.9</td>
</tr>
<tr>
<td>(years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DURATION OF DISEASE (years)</td>
<td>3.8 ± 1.7 *</td>
<td>16.1 ± 8.0</td>
<td>18.0 ± 4.6</td>
<td>12.0 ± 2.9</td>
</tr>
<tr>
<td>HOEHN &amp; YAHR STAGE</td>
<td>2.5 ± 0.6 *</td>
<td>3.8 ± 0.6 **</td>
<td>2.7 ± 0.6 ***</td>
<td>4.4 ± 0.9 **</td>
</tr>
</tbody>
</table>

*From reference Nutt et al. 2002 ** Stage assessed at worst *** Stage assessed during on.

Modeling

The structural PKPD model was described by a set of ordinary differential equations. Various structural absorption models and effect (pharmacodynamic) models were evaluated using the value of the objective function in the NONMEM VI package (Icon Development Solutions, Ellicott City, MD, USA). This objective function is proportional to the log likelihood of observing the data given the model. The difference in the objective function between two nested models follows a chi-square distribution where a value change of four units or more between two hierarchical models differing in one parameter is statistically significant (p<0.05). Model identification and parameter estimation was done in two steps: Estimation of absorption parameters was based on dose and plasma levodopa concentration data, whereas pharmacodynamic parameters were estimated based on the concentration-time profile and effect data. Population mean parameter values, interindividual/interoccasion variability (IIV/IOV), residual variability and their standard errors (SE) were estimated using the first order conditional estimation (FOCE) method in NONMEM.
Results

Paper I

Correlations were found indicating that patients with more severe symptoms at baseline were most improved after infusion. It was found that a linear model relating baseline severity to improvement was adequate (Figure 4). The estimated regression equation was

\[ UPDRS_{\text{improvement}} = -14 + 0.60 \cdot UPDRS_{\text{baseline}} \]

Statistical tests estimating random effects (regression to the mean) showed no significant impact. Goodness-of-fit \( (r^2) \) of the prediction model based on baseline total UPDRS was 0.46. No correlations were found between any improvement measure and age, age at disease onset, disease duration or levodopa treatment duration.

Figure 4. Actual (circles) and predicted (squares) absolute improvement with infusion in total UPDRS score vs total UPDRS score at baseline for the patients in study 1. Predictions were based on a linear regression equation estimated from study 2 data.
Paper II

Treatment effect was detected in both pilot patients in many of the test variables. A large majority of the patients in the DAPHNE study were very compliant with using the test battery. Median compliance was 93% (26 out of 28 expected test occasions performed). Cronbach’s $\alpha$ was 0.81, suggesting good internal consistency. The learning effect between the three first days and the remaining four days in a test period was not detectable in any of the tapping tests. To the survey question: “How do you experience working with the diary?” two nurses responded ‘bad’, four responded ‘okay’ and three responded ‘good’. To the question: “How do you experience the patients feel about working with the diary?” two responded ‘bad’, five ‘okay’ and two responded ‘good’. According to the telephone interview, the nurses’ attitudes towards the diary one year later were generally the same as the year before.

Paper III

The test-retest reliability of the WSTS method was 0.77 whereas the test-retest reliability for the two raters was 0.71 and 0.70 respectively. Correlations (absolute values) between the WSTS and tapping results were low to moderate. The correlation between the WSTS and total UPDRS was 0.41 and WSTS had a lower correlation ($r = 0.38$) to the UPDRS sub section 3 (motor exam) than to sub section 2 (activities of daily living) ($r = 0.51$). WSTS had a correlation of 0.89 to the standardized manual rating (SMR). The 95% confidence interval for the prediction error for the WSTS was $\pm 1.5$ and the mean value (bias) was 0.39 scale units. The Bland-Altman plot of the prediction error is shown in Figure 5. Rater agreements were as follows: between raters, $r = 0.87$; between WSTS and mean rating, $r = 0.91$. 
Figure 5. Bland-Altman plot of the prediction error of the WSTS method with regard to the standardized manual rating (SMR, mean of manual ratings). The 95% confidence interval of the prediction error is ±1.5 scale units, and with a bias = 0.39. N= 182 test occasions from 62 patients. Prediction is based on the mean WSTS of the three spirals drawn at each occasion. A prediction error (y-axis) greater than zero constitutes an under-prediction and an error below zero constitutes an over-prediction.

Paper IV

The responses at the presentation session summarized in a qualitative manner were as follows: i) the web application is very useful, ii) the results during test periods showed agree with qualitative observations of the patient during that test period, for example, “one patient was in a bad condition in baseline, he improved after starting Duodopa, then he became worse again, 24-hour infusion started and the patient became better again; we can clearly follow this in the web application”, iii) comparisons between patients are possible (one patient is in a better/worse condition than another).

Responses to the CSUQ were obtained from seven of the nine clinics and the results were mixed. A majority of the clinics were quite satisfied with the usability although a sizeable minority were not. All evaluators were female and ages ranged from 38 to 61 years (mean value 49). Two out of seven asserted that they had much previous experience with computers, whereas four had some experience and one had little experience.
Concerning server performance, as the load level increased from 50 to 100, the number of requests that can be successfully served by the application per unit time starts to saturate. The average response time varied from page type to page type. For example, the login page introduced more delay compared to other pages; this was because it used the POST method to send data to the server. All average request times were in the range of 0.5 seconds indicating the users will not wait for too long for each page.

**Paper V**

Compliance with performing the tests (in N=70 expected test periods from 35 patients, including two drop-outs) was good: The median number of completed test occasions was 26 out of the 28 expected for perfect (100%) compliance. The mean value was 23 and six test periods (including drop-outs) out of seventy had fewer than 14 completed test occasions (50% compliance).

The median answer to all three questions in the IBM ASQ usability questionnaire was answer alternative 1 (I strongly agree). In the custom questionnaire, thirteen patients (37%) reported that they experienced a problem with the device. The majority of these problems were of technical nature, typical for a prototype device. Some patients found the tests being too boring whereas others thought it was difficult because the screen and displayed texts were too small. Thirty-three patients (94%) found the questions easy to understand and thirty-three also found the questions to be relevant. Four patients (11%) found the motor tests too difficult and thirteen patients (37%) found the motor tests too easy for them.

Internal consistency (Cronbach’s α) of the test battery dimensions was 0.85. The test-retest reliability (ICC, [95% confidence interval]) of the overall score was 0.88, [0.76, 0.94] in the combined group. The Spearman correlation (combined group, r, p-value for a relation stronger than 0.5) between overall score and: total UPDRS r= -0.64, p=0.11; and PDQ-39, r=-0.72, p=0.02. Median overall score of the test battery differed 18% between the two groups (p=0.0014). Plots of overall score and dimensions for the two patient groups and the two test periods are shown in Figure 6. Clearly, there seems to be a difference between the groups in all the dimensions. Learning effects are noticeable for tapping but not spiral drawing and take place in both groups.
Figure 6. Summary plots of test battery dimensions and overall score per patient group and per test period. Symbols represent mean values for the two patient groups and error bars represent one standard error of the mean (N=15). F is the fluctuating group and S is the stable group. For each dimension and overall score, high values represent good function.

Paper VI

An adequate model for absorption of the duodenal infusion of the levodopa/carbidopa gel is first order absorption (mean absorption time of 28.5 minutes) with bioavailability of 88% and a lag time of 2.9 minutes. The parameters were relatively well determined with standard errors of 4 to 43%. Individual variability (IIV/IOV) for bioavailability did not improve the model and is therefore not included. For absorption time IIV/IOV was 43% and it was large for the lag-time (140%).

The best pharmacodynamic model was of an effect compartment sigmoid Emax type with a steep sigmoidicity coefficient (11.6). The concentration giving half of the maximum effect, EC50 was 1.55 mg/L, but with an IIV/IOV of 64%. For the effect model, estimates of standard error could not be obtained for the final model due to difficulties with the numerical estimation routines. Figure 7 illustrates the model’s ability to predict actual plasma concentrations and motor ratings in the patients from PEDAL and the third study (Nyholm 2010). The proportional random error in the PK
predictions was 0.20 and the additive random error of the effect model predictions was 0.92.

Figure 7a. Actual and predicted (by population mean and individual models) levodopa concentrations for the 3 patients in the PEDAL study (panels 1011-1031) and the 5 patients in the Nyholm 2010 study (panels 2011-2053). Second last digit represents subject number and last digit is occasion (day) number.

Figure 7b. Actual and predicted motor ratings (by population mean and individual models). Same panel IDs as in Figure 7a.
Discussion

Data analysis findings

The first part of the research project concerned retrospective analysis of two previous studies with infusion (study 1, Nyholm et al. 2003, study 2, Nyholm et al. 2005). Findings from explorative data analysis of study 1 were indications that often too small extra doses had been given, since many times there was no visible increase in plasma concentration shortly after taking extra doses. Another finding was that in study 1, fluctuations while receiving continuous levodopa infusion were strongly correlated with years on levodopa therapy. This correlation was not found in study 2. However, in study 2 many patients were prescribed a selective serotonin reuptake inhibitor (SSRI) type antidepressant. If the patients with SSRI were removed from analysis, the correlation was present also in the second study, indicating a possible relation between SSRI and fluctuations and/or rate of progression. Since these are retrospective findings in two small studies, results should be verified in a prospective study. Nutt et al. (1997) found that a proportion of the variability of tapping speed during constant infusions could be explained by variation in plasma large neutral amino acid (LNAA) concentrations. Motor fluctuations were not associated with minor variations in levodopa concentrations. Fluctuations were more prominent in subjects who have taken larger daily doses of levodopa, implicating pharmacodynamic factors as well. Rampello et al. (2002) found that SSRI improved bradykinesia in PD and van de Vivjer et al. (2002) found that start of SSRI in PD was followed by increased use of dopaminergic drugs. Leo (1996) described sixteen cases of worsening PD symptoms after addition of SSRI. In conclusion, interaction between serotonin and dopamine systems in PD is an issue that deserves further attention.

Prediction of levodopa infusion outcome

The main finding described in paper I was that, the more severe PD symptoms the patients had during their oral treatment, the more improved they became with infusion. This finding was reproducible between two clinical studies for different measures of severity and improvement. The control analysis found that baseline minus oral values (‘improvement with
oral compared to baseline’) did not correlate with corresponding baseline severity values. Linearity in the relations between improvement measures and severity measures may not be preserved if patients with higher severity at baseline are included in the analysis. The increase in improvement will not be constant as severity increases towards total disability, and improvement could even decrease. Studies on more severely disabled patients are required to explore the validity of the linear model for this patient category.

Outcome measures revisited

In the total UPDRS score, highest weight is given to motor function impairments. According to the Movement Disorder Society Task Force on Rating Scales (2003), strengths of the UPDRS include its wide utilization, its application across the clinical spectrum of PD, its nearly comprehensive coverage of motor symptoms, and its clinimetric properties. However, non-motor symptoms such as pain, depression, cognitive problems and sleep problems may affect quality of life as much as the motor symptoms do. In a recent Chinese study aimed to identify the motor and non-motor factors associated with health related quality of life (HR-QOL) in early PD, the clinical factors that showed the highest predictive value for worse HR-QOL were non-motor symptoms, such as depression, sleep disorders, and fatigue. (Qin et al. 2009). The revised MDS-UPDRS now gives somewhat higher weight to non-motor symptoms (Goetz et al. 2008). The symptoms that individual patients have and which symptoms they rank as most problematic may vary between different stages of the disease and between individuals. Later in the disease, cognitive problems often have a large impact on HR-QOL (Weintraub et al. 2008, Miyasaki et al. 2006). Different cognitive tests have been successfully developed and deployed on a variety of mobile devices by Brian Tiplady (e.g. Tiplady et al. 2009). Possibly different symptoms should be assessed in different patient categories (e.g. early, intermediate and advanced disease).

Patients with PD fluctuations experience episodes of various symptoms. Symptomatic treatment of fluctuations aims to reduce i) the number of symptomatic episodes in a time period, ii) the durations of the episodes and iii) the severity of symptoms during the episodes. Not much information on weighting these aspects in PD can be found, but in epilepsy e.g. the Veterans Affairs cooperative study seizure rating scale combines frequency and severity in a weighted scoring system, summing all items in a total seizure score (Cramer 1993). The methodology proposed in this thesis, to take the first principal component of mean, SD and MSD for each symptom dimension, is a data-driven way to automatically obtain such weights based on information content. To account for individual differences in symptom
patterns and preferences, an individual’s own global self-assessment may be used.

This thesis emphasizes the need for repeated symptom assessments as well as the need for unification of multidimensional symptoms for screening patients and for outcome assessment of treatment changes. E.g. reducing medication to a dyskinetic patient may improve motor test scores but may bring him or her to a mental off; so self-assessed, the overall condition is worse. Work by others (Isacson et al. 2008, Weaver et al. 2005) support this view. Assessments must be done often enough to capture variations and reach acceptable measurement errors. Extensive assessment batteries are not feasible for repeated use and there is a trade-off between usability and ability to capture enough aspects of the symptoms. Many researchers (e.g. Haaxma et al. 2008, Joebges et al. 2003) agree that clinical rating scales are not the best choice for assessing motor function and that development of more quantitative and objective methods for measuring treatment effects in PD is highly warranted. Hagell et al. (2000) found that timed tests are valuable quantitative and objective measures in scientific as well as clinical assessments of PD. It is concluded that objective measurements of important motor symptoms should be included in unified assessment scores, but important non-motor symptoms (measured if possible, otherwise self-assessed or caregiver assessed) should not be forgotten. Unified assessment scores are necessary for assessment of overall disability but may not be sufficient for evaluating treatment outcomes. Some treatments are effective on some symptoms but not on other symptoms. If the symptom targeted has low weight in a unified score, responsiveness will be worse than if a sub-score for only the targeted symptom is used.

Continuous patient monitoring appears difficult. In the PEDAL study wrist-worn actigraphs were tested, but useful data was only collected during times when the patients performed the movement tasks of the TRS. If only data from those time points were retained, the dose-response of the increased morning dose could be followed, but at other times, readings were unrelated to the patient’s actual ability to move. Lloret et al. (2010) confirm the lack of specificity of simple wrist-worn actigraphy and suggest it may be suitable for dyskinesia assessment but not for on state and off state evaluation. Interpretation of movement patterns is difficult since they depend on context: Is it tremor or is the patient cutting his grass? Usability and perhaps integrity and cost aspects of applying wearable sensors are challenging and these approaches miss out completely on the non-motor problems.

Computer-assisted measurements of motor functions during controlled activities are often found feasible in literature (e.g. Goetz et al. 2009). Results from Popovic et al. (2008) indicate that hand trajectories from a
digitizing tablet are useful for assessing hand motor blocks and Gunzler et al. (2009) found that foot tapping may be a useful outcome measure for determination of dopaminergic medication effect. Stewart et al. (2008) found that lower extremity impairments had higher impact on HR-QOL compared to upper extremity impairments although both were important.

Many clinicians would agree to that a patient’s condition should preferably be assessed in his or her typical situation; that is typically at home. A major problem with paper diaries, besides the reliability issues (Stone et al. 2003), is the inconvenient storage of, and access to, assessment history which is available only in the raw format, as entered. To assess patients at home via self- (or caregiver-) recorded videos and then to rate the recordings centrally (Goetz et al. 2008, Samii et al. 2006, Nyholm et al. 2005) has interesting prospects. A drawback from using this technology in practice could be that manually rating videos takes much time from the clinician, but the method is feasible in studies. In the future, video-processing could potentially be used to automatically score symptoms in the videos. The use of e-diaries for capturing self-assessments is advocated. In a recent review article Buck et al. (2010) found that a single daily-diary instrument for assessing a broad spectrum of fluctuating non-motor symptoms would be helpful.

A novel test battery

A finding in study 2 where patients had both e-diary assessments and clinical ratings of motor function was that correlations between these measures were low. Significant improvement with levodopa infusion was seen in both diary answers and clinical ratings. Hence, patients and doctors agreed that there was improvement with infusion, but the magnitude of the perceived improvement according to the diary and improvement according to the rating of motor function differed. This led to the idea to combine e-diary questions with on-screen motor tests in a test battery to capture both aspects of patient condition.

An international patent application according to the patent cooperation treaty (PCT) was filed in February 2006 (WO/2006/088415, MOVEMENT DISORDER MONITORING) and received a positive patentability report. The invented test battery consists of a combined e-diary with PD-related questions and on-screen motor tests with a scheduler to restrict data collection and further comprising summary calculation methods. In patients with motor fluctuations, it should typically be used four times per day in the home environment, over test periods of one week. The aim of this test battery is to provide status information in order to evaluate treatment effects in clinical practice and research, follow up treatments and disease...
progression and to optimize treatment strategies. The ambition was to cover symptoms, which are important for most people suffering from PD with fluctuations, while still being feasible for the patients to perform independently at home.

Content validity was established through literature surveys and expert consultations. Generally the test battery assesses patients’ perception of common symptoms and their ability to perform fine-motor tasks. The tapping tasks assess tapping speed, hand dexterity, reaction time and attention to visual stimuli. These are in turn affected by symptoms such as akinesia, bradykinesia, coordination deficits, tremor, dyskinesia and cognitive problems. The spiral drawing task assesses drawing impairment, which is most likely strongly related to handwriting ability. It is affected mainly by involuntary movements, such as tremor and dyskinesias, but also by other symptoms. Compliance with, and usability of, the test battery were good in 35 Italian and 65 Swedish patients and the test-retest reliability was good. Convergent validity with rating scales is adequate and known-groups validity and responsiveness has been demonstrated.

Hardware-related problems in the DAPHNE and TEVAL studies consisted mainly of broken charger contacts and touch screens and memory cards that “popped out”. Battery replacements were needed once per year. Usability results of the web interface with nurses were not optimal. Explanations for this are that help and explanation functions were not fully implemented and it was possible to reach too detailed results which were not easily interpreted. Clearly there is potential to improve the user interface and training of clinical staff may be necessary. One explanation for the surprisingly good patient compliance is that a support service monitored if patient data was coming in as expected and contacted the clinic if it did not appear.

An advantage of the test battery is that symptom summaries and history are easily accessible in a web interface. The use of an overall score may facilitate screening patients and helps avoiding sub-optimization of treatments. Since symptom profiles are so different in this patient group, an overall score can be beneficial for deciding if a treatment change leads to an improvement of a patient’s general condition or not. The test battery meets the requirements of adequate data capture, namely frequent, time-stamped assessments in home environment. In the context of home assessments making the checkups more efficient and less frequent, it may benefit patients living far away from the clinic.

An advisory board consisting of 14 senior neurologists evaluated the test battery system after presentation and demonstration of its functionality. The
neurologists were based in the following countries: USA 3, Germany 3, Italy 2, Spain 1, Netherlands 1, UK 1, Sweden 1, Denmark 1 and Finland 1. They responded to questions by pressing keypads. To a question about their overall impression of the system, eleven had a positive impression, one had a neutral impression and two had a negative impression. The most important benefits they could see were an increased ability to identify patients who are not doing well and facilitated follow-up and optimization of an individual’s treatment. The system was seen as most important for complicated patients and for regional patients. The general conclusion of the board was that the test battery system was recognized as a tool that will assist in management of patients.

Pharmacokinetic-pharmacodynamic properties of duodenal levodopa

This is the first attempt to identify a PKPD model for the levodopa/carbidopa intestinal gel formulation (Duodopa). The absorption time appears slightly faster and bioavailability is larger for infusion of the levodopa/carbidopa gel compared to oral levodopa, while pharmacodynamic parameters are similar to those found in other studies (Adamiak et al. 2010, Contin et al. 2001, Troconiz et al. 1998, Harder et al. 1998). On average 88% of infused dose will be absorbed and it will take 51.4 minutes until a momentary bolus dose (without any subsequent continuous dose) yields peak effect from an unmedicated condition in a typical patient. The effect is very sensitive to concentration changes around the EC50 in this patient group and there is a large individual variability in EC50 (64%), which corresponds to the large inter-individual variability in dosage (Nyholm et al. 2008).

This also is the first attempt to model both PD symptoms and peak-dose dyskinesia symptoms in the same scale (the TRS). As seen in Figure 7, the absorption and effect models were reasonably successful in fitting observed data. Clearly motor ratings are more difficult to predict than levodopa plasma concentrations. Effect parameters’ inter-individual variability, i.e. the therapeutic window, is large and therefore the specific individual’s prediction deviates substantially from the average individual’s prediction and the observed data.

This PKPD modeling is based on data from in total 20 patients, which may be considered a small number, but on the other hand these were intensively studied on several occasions. Furthermore, literature information was used to support the present investigation with respect to levodopa disposition. It is well known that other factors than levodopa concentration may influence PD motor symptoms. These factors include stress-level, food intake, time-of-
day, physical activity, intake of other pharmaceuticals affecting
dopaminergic or other receptors. None of these factors are included in the
present model and these will therefore appear as unexplained inter- or intra-
individual variability. The selected model had the best ability to explain
observed data among all the tested models. If a model is having too many
parameters, there is always a possibility that it will be biased towards the
data used during estimation. The selected “best” model has 27 parameters
and its ability to predict new patients’ dose-response should be tested with a
new data set.

Contributions
This thesis has had a wide focus in applied research in collaboration with
industry, sometimes bordering on product development. This is considered
as an advantage since it hopefully means that the results have a high
probability of actually coming to practical use and thus making a real impact
for sufferers of PD and the people caring for them. Hopefully it can also be
considered to have merits as a multidisciplinary research effort. The
particular problems addressed in this thesis however have lead to some
solutions, which may be more generally applied:

The method to take the first principal component of mean, SD and MSD for
each feature, is a data-driven general way to automatically obtain weights for
combining level and variation, based on information content. The WSTS
method for assessment of spiral drawing impairment is useful during event-
driven sampling, which reduces the amount of data to be transferred
compared to methods using constant sampling rates. This is beneficial for
scaled-up storage and narrow-band wireless transfer. The method of
unification by selecting subsets of data in order to define a desired direction
in feature space should also be generally applicable.

A look ahead
In summary, there are four steps in the research concerning the test battery;
first selection of when, what and how to measure to capture the relevant
symptoms; second comes validation that the measurements are accurate and
measure the right things; third comes providing access to information useful
for decision making and fourth comes analysis of the implications of use.
Evidence is growing that the test battery is useful, reliable and valid for
assessment of symptom status during motor fluctuations. However, this
evidence comes from clinical studies, where special support and
requirements were provided to the clinical staff. A very important
outstanding question is if compliance and usability are similar in daily clinical practice, outside of study settings. A further question is whether there are country-specific issues, posing different requirements for introduction into clinical practice. There is yet not evidence that using this test battery will improve treatment of patients during clinical practice or that a cost benefit analysis is favorable.

To analyze benefit of using the test battery in the everyday clinical practice, a “market acceptance test” is planned. One question addressed is if it can detect a need for treatment changes and to do a cost-benefit analysis. A further question is whether there are country-specific issues, posing different requirements for introduction into clinical practice. Development issues include ability to modify the test battery using a configuration editor (to facilitate tailoring the test battery to individual users, changing language etc.); porting to newer hardware and regulatory certification as a medical device.

An important issue in infusion treatments is how to individually adjust morning bolus dose and continuous infusion rate as accurately and as fast possible. In order to utilize the PKPD model for \textit{in numero} simulation experiments to assist dose titration, a model-based case-based method is being developed. The functionality is not yet evaluated. The method finds individualized doses to patients using dose-effect time series and a PKPD model and a case database and some algorithms. Simulations may also be helpful for training purposes to illustrate the expected outcome of certain dosages for typical patients, sensitive or tolerant patients.

Ongoing work includes an interface for dosage information for infusion treatments including an on-line decision support system based on dosage from pumps. Functions will include summary of treatment periods and alerts for unusual or unwanted dosage. Pump data is collected using a wireless interface and transferred via the test battery’s hand unit. There is an idea to use data-driven methods for simultaneous assessment of both impairment and cause in a combined score per test occasion with the test battery, similar to Nyholm’s TRS (from very off to very dyskinetic). Together with the model-based case-based method for dose finding, this can form a basis for an active decision support system for generating advice for individually optimized drug dosage.
Conclusions

A test battery system for symptom assessment during PD motor fluctuations was developed, deployed and evaluated. Its main contribution is a novel and reliable way to capture and easily access symptom information from patients’ normal home environment. As part of this work, a novel method for assessment of spiral drawing impairment useful during event-driven sampling was developed. Pharmacological properties of the levodopa/carbidopa intestinal gel were examined and results from exploratory data analysis indicate that severity measures during oral treatment (e.g. from the test battery) may be factors to consider when deciding candidates for infusion treatment. Specifically, the conclusions of the different papers are:

- **Paper I:** The more severe PD symptoms the patients had during their oral treatment, the more improved they became with infusion. This finding was reproducible between two clinical studies for different measures of severity and improvement. A numerical prediction model (linear regression) built using data from one study was reasonably successful in predicting outcome in the other study.

- **Paper II:** The test battery consists of both self-assessments of symptoms and motor function tests and was implemented on a touch screen mobile phone. It could detect treatment changes in two pilot patients, both in self-assessed responses, tapping results and spiral drawing. It had remarkably good patient compliance in a clinical study with 65 Swedish patients, considering the advanced disease stages in the group. Usability with nurses at nine clinics was acceptable considering the circumstances.

- **Paper III:** A method, using discrete wavelet transform and principal component analysis, was developed to process spiral drawings in order to generate a wavelet spiral test score, useful during event driven sampling. The scoring method could assess PD-related drawing impairments well comparable to trained raters. The reliability was acceptable and convergent validity with a standardized rating, UPDRS and tapping results was better than for two previous methods.
• **Paper IV:** A web-based system was developed, which processes various time series of motor test results and self-assessments during test periods into scores for different symptom dimensions and an overall test score, reflecting the global condition of the patient during the test period. The system is able to summarize and present information to clinical staff in a useful manner.

• **Paper V:** Compliance with, and usability of, the test battery were good in a majority of 35 Italian patients and test-retest reliability was good. Internal consistency and convergent validity with UPDRS and PDQ-39 rating scales was adequate. Known-groups validity was demonstrated since it was possible to use the overall score to separate fluctuating from stable patients.

• **Paper VI:** Absorption of the levodopa/carbidopa gel can be adequately described with a first order absorption with a bioavailability and a lag-time. The absorption time appears slightly faster and bioavailability is larger for infusion of the levodopa/carbidopa gel compared to oral levodopa. The sigmoidicity is high and there is a large individual variability in effect parameters. The residual error magnitudes indicate that the models developed provide predictions of a relevant quality.
Beslutsstöd för behandling av patienter med avancerad Parkinsons sjukdom

Uppgiften för avhandlingsarbetet har varit att utveckla, tillämpa och utvärdera IT-baserade metoder till stöd för att behandla och utvärdera behandling av Parkinsons sjukdom i komplikationsfas med dosglapp och överrörlighet. Detta är ett tillstånd som förekommer hos fler än hälften av alla som haft en Parkinsondiagnos i mer än fem år. En behandling som har visat sig fungera bra för dessa patienter är infusion av en levodopa/karbidopa gel, direkt till tunntarmen med en bärbar pump. En bakomliggande problemställning har varit svårigheten att följa upp och bedöma effekt av behandlingar för denna patientkategori, där ett antal olika symptom kan förekomma med varierande frekvens och intensitet. En annan problemställning har varit hur infusionsbehandling kan optimeras.


Arbete II beskriver design och första utvärdering av ett testbatteri-system för en handdatortelefon för hemmabruk. Testbatteriet består av schemalagda frågor till patienten och motoriktester (”tapping” och spiralritning) direkt på pekskärmen. Skattningar genomförs några gånger om dagen under testperioder om ca en vecka. Testresultaten skickas till en centraldator och behandlas där med statistiska metoder. En beräkningsmetod för

I arbete VI togs en simuleringsmodell (populationsfarmakokinetsisk-farmakodynamisk) fram, som kan bli till hjälp för att ställa in rätt dos av infusionsbehandling för olika personer i olika stadien av sjukdomen. Modellen ger en ökad insikt i farmakologin bl.a. vad gäller upptaget över tarmväggen och effekten vid olika sjukdomstillstånd och doseringsnivåer. En framtida målsättning är att simuleringsexperiment ska kunna användas i utbildningssammanhang. En klinisk studie designades och genomfördes särskilt för att samla data för denna modellering.

Arbetet har alltså resulterat i ett testbatteri-system som underlättar uppföljning av behandling för patienter med Parkinsons sjukdom och en modell som relaterar dosering till förväntad plasmakoncentration och effekt vid infusionsbehandling med levodopa/karbidopa gel. Pågående arbete innefattar hur en sådan modell kan användas i utbildningssyfte samt för att anpassa dosering hos enskilda individer.
## Terminology

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akinesia</td>
<td>Inability to initiate movement</td>
</tr>
<tr>
<td>Bradykinesia</td>
<td>Slowness of movement</td>
</tr>
<tr>
<td>Carbidopa</td>
<td>Inhibitor of AADC (metabolizes levodopa)</td>
</tr>
<tr>
<td>Chorea</td>
<td>Involuntary dancing-like movement</td>
</tr>
<tr>
<td>Dexterity</td>
<td>Skill of performing tasks with hands</td>
</tr>
<tr>
<td>Dopaminergic</td>
<td>Related to or activated by dopamine</td>
</tr>
<tr>
<td>Duodopa</td>
<td>Levodopa/carbidopa intestinal gel</td>
</tr>
<tr>
<td>Dyskinesia</td>
<td>Distortion in movement</td>
</tr>
<tr>
<td>Dysrhythmia</td>
<td>Disturbance of rhythm</td>
</tr>
<tr>
<td>Idiopathic</td>
<td>Disease of unknown origin</td>
</tr>
<tr>
<td>Gait</td>
<td>Manner of walking</td>
</tr>
<tr>
<td>Levodopa</td>
<td>Amino acid, which is metabolized to dopamine</td>
</tr>
<tr>
<td>Off</td>
<td>PD symptoms present</td>
</tr>
<tr>
<td>On</td>
<td>PD symptoms relieved</td>
</tr>
<tr>
<td>Mentation</td>
<td>Mental activity, process of thinking</td>
</tr>
<tr>
<td>Nasoduodenal</td>
<td>From the nose to the intestine (duodenum)</td>
</tr>
<tr>
<td>Tremor</td>
<td>Involuntary shaking movement</td>
</tr>
</tbody>
</table>
Clinical studies

Study 1
12 patients were included in a randomized crossover trial. Baseline characteristics are shown in Table 1 (no 101-112). Levodopa/carbidopa was administered as oral sustained-release tablets and by nasoduodenal continuous intestinal infusion for periods of 3 weeks of each treatment. Plasma levodopa concentrations and motor performance were evaluated every 30 minutes during 3 test days of each treatment period. (Nyholm et al. 2003). Data from the study were used in paper I.

Study 2
18 patients with motor fluctuations and dyskinesia fulfilled a randomized crossover trial to compare individualized conventional treatment and intestinal levodopa infusion for 3+3 weeks. Baseline characteristics are shown in Table 1 (no 201-218). Video scoring of motor function was assessed by blinded assessors on a global Treatment Response Scale from -3 to 0 to +3 (from severe off to on to on with severe dyskinesia). Patient self-assessment of motor performance and quality of life was done using an e-diary. (Nyholm et al.2005). Data from the study were used in paper I.

DAPHNE
The test battery is currently being used by 65 patients with advanced PD at nine clinics around Sweden. DAPHNE (Duodopa in Advanced Parkinson’s: Health Outcomes & Net Economic Impact, EudraCT No. 2005-002654-21) is an open and observational, prospective 48-month study (12 months recruitment, 36 months treatment) on health outcomes and cost impact of intestinal levodopa infusion treatment in PD, approved by the Swedish Medical Products Agency. The original items in the test battery are secondary outcome measures in the study. On inclusion, the patients were either treated with infusion, or they were candidates for receiving this treatment. The subjects were assessed at baseline (infusion-naïve group only) and then every three months during the first year and then every six
months for the remaining two years. Baseline characteristics are shown in Table 3. The study is planned to continue until 2011. Interim results have been published as poster abstracts and study data were used in papers II, III and IV.

**TEVAL**

A validation study was performed in order to assess compliance with and usability of the test battery; the internal consistency of the test battery’s test items; the test-retest reliability of the test battery’s combined output (overall test score); convergent validity to UPDRS and PDQ-39; and known-groups validity. This study (35 patients) was completed in collaboration with Prof. Angelo Antonini and Dr. Mauro Schiavella at Parkinson Institute, Istituti Clinici di Perfezionamento, Milan, Italy and results are presented in paper V. All subjects were unpaid volunteers and informed consent was obtained after full explanation of the study purpose and design and approval from the Institutional Ethics Committee at ICP Milan. Principles of the Declaration of Helsinki were followed. Baseline characteristics of the patients are shown in Table 4.

**PEDAL**

A small pilot study (three patients) aimed at identification of a model relating dose to body-distribution to effect of levodopa/carbidopa intestinal gel, was performed in collaboration with Dr. Sven Påhlhagen at Karolinska Institute, Huddinge. The study protocol was accepted by the ethics committee of the Karolinska Institute, Sweden and the patients gave informed consent in accordance with the Helsinki declaration. Baseline characteristics are presented in Table 5. The data was combined with data from two other studies (Nyholm et al. 2003, Nyholm et al. 2010, unpublished data) and modeling was performed in collaboration with Prof. Mats Karlsson, Pharmaceutical Biosciences, Pharmacokinetics and Drug therapy, Uppsala University. Results from PEDAL are used in the pharmacokinetic-pharmacodynamic model for the levodopa/carbidopa intestinal gel in paper VI.
Acknowledgements

This work was performed in the framework of the IDOL and MOVISTAR and E-MOTIONS research projects in collaboration between industry and academy. Half the resources were contributed by the Swedish Knowledge Foundation (KK-stiftelsen) and half by the industry partners Solvay Pharmaceuticals (now Abbott) and Nordforce Technology. As an outcome of the work, Jerker Westin, Mark Dougherty, Dag Nyholm and Torgny Groth have filed a patent application (MOVEMENT DISORDER MONITORING, international publication number WO/2006/088415) and founded a company (Jemardator AB).

I wish to express my gratitude to the following people for their contributions to this work:

Torgny Groth, Mark Dougherty, Sven Pålhagen and Marja-Liisa Dahl – My supervisors, for your guidance

Stefan Åsberg and Ulf Bergqvist and the other staff at Solvay/Abbott and Nordforce – For assisting and facilitating in so many ways

Dag Nyholm – For your ideas, feedback and support

Thomas Ericson – For initiating it

Angelo Antonini and Mauro Schiavella – For your work with the validation study

Peter Larsson – For your work with the PDA software

Mats Karlsson – For your work and guidance with modeling

Mevludin Memedi – For your work with the web application

Anders Johansson – For rating the spiral drawings

Samira Ghiamati – For your work with the discrete wavelet transforms
Jan Kowalski, Marcus Thuresson, Moudud Alam, Mengjie Han and Kenneth Carling – For your help with statistics

The pilot patients – For helping in developing the test battery

The study nurses and investigators in the DAPHNE study – For your efforts and feedback

The students doing master thesis work with me – For useful contributions to much of the work. I am sorry for not always giving you the time and attention that you deserve

My parents in law Roland and Eva – For your help with practical issues

My parents Birger and Gunilla – For your love and support throughout my life

My wife Kristina and my sons Amandus and Rasmus – For your love, support and never-ending patience

Thank you for making this possible!
References


Goetz CG, Stebbins GT, Shale HM, Lang AE, Chernik DA, Chmura TA, Ahlskog JE, Dorflinger EE. Utility of an objective dyskinesia


Likert R A technique for the measurement of attitudes. Arch Psychol. 1932; 140: 1-55.


Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. British J Psychiatry 1979;134:382-89.


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 621

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-132635