Quantifying short term dynamics of Parkinson's disease using self-reported symptom data from an Internet social network

Max A. Little[1,2], Paul Wicks[3], Timothy E. Vaughan[3], Alex 'Sandy' Pentland[1]

[1]MIT Media Lab, Cambridge, MA, USA; [2]Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK; [3]Research and Development, PatientsLikeMe Inc., Cambridge, MA , USA

Abstract

Background: Parkinson's disease (PD) is an incurable neurological disease with around 0.3% prevalence. The hallmark symptom is gradual movement deterioration. Current scientific consensus about disease progression holds that symptoms will worsen smoothly over time unless treated. Accurate information about symptom dynamics is of critical importance to patients, caregivers, and the scientific community, for the design of new treatments, clinical decision-making, and individual disease management. Long-term studies characterize the typical time course of the disease as an early, linear progression, gradually reaching a plateau in later stages. However, symptom dynamics over durations of days to weeks remain unquantified. There is currently a scarcity of objective clinical information about symptom dynamics at intervals shorter than 3 months stretching over several years, but Internet-based patient self-report platforms may change this.

Objective: To assess the clinical value of online, self-reported PD symptom data recorded by users of the health-focused Internet social research platform PatientsLikeMe (PLM), in which patients quantify their symptoms on a regular basis on a subset of the Unified Parkinson's Disease Ratings Scale (UPDRS). Analyzing this data we aim for a scientific window on the nature of symptom dynamics for assessment intervals shorter than 3 months, over durations of several years.

Methods: Online self-reported data was validated against the "gold standard" PD-DOC database, containing clinical symptom data at intervals greater than 3 months. The data were compared visually using quantile-quantle plots, and numerically using the Kolmogorov-Smirnov test. Using a simple piecewise linear trend estimation algorithm, the PLM data was smoothed to separate random fluctuations from continuous symptom dynamics. Subtracting the trends from the original data revealed random fluctuations in symptom severity. The average magnitude of fluctuations versus time since diagnosis was modeled using a gamma generalized linear model.

Results: Distributions of ages at diagnosis and UPDRS in the PLM and PD-DOC databases were broadly consistent. PLM patients were systematically younger than PD-DOC patients, and showed increased symptom severity in the PD "off" state. The average fluctuation in symptoms was 2.6 (Part I and II) UPDRS points at the time of diagnosis, rising to 5.9 points 16 years after diagnosis. This fluctuation exceeds the estimated minimal and moderate clinically important differences, respectively. Not all patients conformed to the current clinical picture of gradual, smooth changes: many patients had 'regimes' where symptom severity varied in an unpredictable manner, or underwent large, rapid changes in otherwise more stable progression.

Conclusions: This information about short term PD symptom dynamics contributes new scientific understanding about disease progression, currently very costly to obtain without self-administered, internet-based reporting. This understanding should have implications for the optimization of clinical trials into new treatments, and for the choice of treatment decision timescales.

Keywords: Parkinson's disease, social networks, medical informatics, symptoms, pharmacodynamics

Introduction

Parkinson's disease (PD) is a relatively common, progressive neurological disorder affecting around 0.3% of the general population in industrialized countries [1]. It generally affects people over 60 years, but rarely, younger subjects under the age of 40 also develop the disease. PD is generally considered a *movement disorder*, that is, it affects the ability to perform normal voluntary motion, but subjects also experience cognitive impairment and emotional/mood disturbances. The classic movement symptoms of PD include exaggerated tremor, rigidity and slow or hesitant motion. These movement problems often have a substantial negative impact on the ability of the

patient to perform essential everyday activities such as bathing, dressing, turning in bed, walking unaided, and getting up from a sitting position. The cause of PD is currently thought to be the loss of dopaminergic neurons in an area of the brain known as the *substantia nigra*. PD is incurable, and there are no absolutely conclusive diagnostic tests. The most accurate diagnosis based on behavioural symptoms achieves at best 90% accuracy when compared to post-mortem pathological examination [2].

The mortality rate of subjects with the disease is significantly increased relative to healthy people [1]. There are a few approaches to treating the symptoms of PD: the first line of defense is the drug levodopa which replenishes dopamine in the substantia nigra, which in turn reduces the severity of movement symptoms. However, this drug tends to become less effective over time, and can also lead to severe side effects such as involuntary movements (*dyskinesias*). Surgical treatments such as deep brain stimulation have been shown to be effective for many subjects who do not, or have ceased to respond to, drug treatments. Current scientific understanding holds that the severity of PD symptoms will smoothly increase over time, faster at first and often leveling out in the later stages [3].

Trials for new treatments and assessing the effectiveness of treatments require objective data about symptom severity. A coarse quantitative measure of symptom severity is the *Hoehn and Yahr* (H&Y) ordinal scale [4] which assigns a number from 0 to 5, with 0 being healthy and 5 denoting severe disability. This has been largely supplanted by the ordinal *Unified Parkinson's Disease Rating Scale* (UPDRS) (here we refer to version 3.0) [5] and associated tests [6], which is more time consuming and expensive to administer, but is much more precise – the most commonly used parts of UPDRS (Parts I, II and III) range on the scale 0 (healthy) to 176 (severe disability) [5], although a simple and accurate formula exists to predict H&Y from UPDRS [7].

UPDRS values have been collected for patients at all stages of the disease, and there is substantial research data available on PD symptom progression quantified on this scale. This kind of data has been used to calibrate models of PD symptom progression over the course of years to decades [3]. However, the full UPDRS is a complex test that requires expertise to administer (even if that expertise can be taught to general medical personnel [5]), attendance of the patient in the clinic, and the average time for administration of the full test is around 17 minutes [8]. Unfortunately, these difficulties mean that it is, usually, prohibitive to objectively score PD symptom severity on time scales shorter than 3 months (*low-frequency*). Since most longitudinal UPDRS data is low-frequency, objective information about symptom dynamics occurring on a shorter time scale than 3 months (*high-frequency* data) is lacking.

There are many clinical situations in which high-frequency symptom dynamics would be useful. For example, in testing new drug treatments, there is a trade-off between minimizing exposure to the novel drug to reduce the risk of unknown side effects, and maximizing the opportunity to detect significant changes in symptoms. This temporal trade-off cannot be optimized on a quantitative basis without high-frequency data upon which to base the statistical analysis. Similar issues arise in diagnosis where PD is suspected. If, in conjunction with movement symptoms on one side of the body only, taking levodopa leads to a reduction in symptom severity, the patient is highly likely to have PD [2]. However, there is still a non-negligible chance that the patient has some other neurological disorder with PD-like symptoms, such as progressive supranuclear palsy. This disease can progress very rapidly, so it is important to diagnose this quickly. Thus the window of this "exploratory" prescription of levodopa for differential diagnosis must be made as short as possible. However, it should not be so short that rapid, natural fluctuations in symptoms confound proper diagnosis.

Recently, health-focused Internet websites have become established that allow users to track their disease progression using, for example, surveys and other remote monitoring devices. We obtained UPDRS data from PD users of the PatientsLikeMe website [9], which has, to-date, recruited over 6000 PD patients worldwide since 2007. Some of these patients are particularly dedicated diarists who have, over a number of years, documented their symptoms on an extremely regular basis. The result is an unprecedented, high-frequency symptom data set that has the potential to be used to address some of the shortcomings of existing, low-frequency clinical data. For example, if the data is sufficiently accurate, it could be used to supplement in-clinic checkups between visits. In another neurological disease, amyotrophic lateral sclerosis, similar data was used to refute the idea that lithium carbonate slowed the progression of that disease [10]. The purpose of this study is an exploratory investigation into the high frequency dynamics and other properties of this novel data set, to assess the clinical value of these data.

Methods

Patient recruitment and data collection

The main outcomes of this study were quantified using UPDRS. This scale consists of five parts: Part I covering cognitive, behavioral, and mood symptoms; Part II evaluating activities of daily living; Part III measuring motor symptom severity; Parts IV and V contain H&Y stage and an evaluation of daily living activities on the Schwab and England scale [6]. Parts I – III contain separate sections each with a score ranging from 0 (no symptoms) to 4 (severe symptoms). Part I has four sections, and Part II, 13 sections.

Two data sources were used: the PatientsLikeMe (PLM), and the Parkinson's Disease Data and Organizing Center (PD-DOC) data sets [11]. The PLM data were used to provide long-term quantification of individual symptoms occurring on a time scale shorter than 3 months. The data are entirely self-reported. Users connect to a website, into which they can enter demographic details, information about their disease course and symptoms, and their treatment history. In particular, we collected age, gender, treatment status, H&Y staging, and UPDRS, Parts I (mentation, behavior and mood) and II (activities of daily living). UPDRS Part III (motor symptoms) was excluded, as the collection of this data was deemed not suitable for self-report. Not all symptom self-reports were accompanied by treatment status indications.

At the time of preparation of this manuscript, the PLM data set contained 6074 PD patients, of which 2931 completed at least one UPDRS survey, and entered their birth date and date of diagnosis. Patients were included in this study if they reported at least 15 UPDRS scores with maximum average UPDRS reporting interval of at most 65 days between reports. This led to 100 patients being included in this study, and 29±14 (mean ± standard deviation) symptom self-reports per patient (total of 2896 reports), with reporting intervals of 45±12 days. The age of the selected patients was 54±9 years at diagnosis, of which 52 were female, 48 male. Patients began self-reporting symptoms approximately one year, on average, after diagnosis. The total time interval covered by self-reporting per patient, from the first report to the last, was 3.1±0.8 years, and all reports were prospective (after date of joining the website). Patients contributing to the PLM data agreed to the terms and conditions of the website when they enrolled, which included granting permission to PatientsLikeMe to use their medical data for research purposes [12]. Qualitatively, the PLM data set represents a large number of frequent Part I and II UPDRS reports and treatment status across a medium size cohort of young to mid-aged patients.

The PD-DOC data set contains data on PD patients from multiple clinical centers in the US across several trials, data collected by clinicians over the period 2006-2011, to aid the process of statistical analysis of PD, and for the design and planning of clinical trials into treatments. In this study, it was used as a reference data set to verify the PLM data and to provide "background" data on PD. Data collection was co-ordinated by the University of Rochester, NY. The set represents UPDRS symptom reports from 564 individuals with PD, of which 200 were female and 364 male, aged 59±10 years at diagnosis. In the "on" state, 1612 UPDRS scores were recorded, 354 recorded in the "off" state. There were 2.9±0.9 symptom reports, covering an average of 1.9±0.9 years, per patient. Ethical approval was obtained from the IRBs of each US medical center contributing patient details to the data set. By contrast to the PLM data set then, PD-DOC can be described as data from a large number of mid-age to older patients with clinical UPDRS reports collected on an infrequent basis.

Validating the PatientsLikeMe data set

At the outset, the concept of symptom self-reporting may raise data reliability questions, primarily because it could be suspected that untrained, non-clinical raters may be more prone to certain systematic errors or biases than trained, clinical raters. For example, they may tend to be biased towards repeating previous measurements, or may have more inconsistent interpretations of specific questions across tests than trained clinical staff. Previous research has shown that when PD patients without dementia self-report Part I and II scores, the scores are consistent with those assessed by the neurologist assessing them [13]. To our knowledge, there have been no similar assessments into the reliability of self-reported Part I and II scoring conducted online under non-clinical circumstances.

To address this issue, we compared the PLM data set against the PD-DOC data which we considered to be a "gold-standard" clinical reference set. The distributions of UPDRS Part I and II values and ages at diagnosis, were compared visually on quantile-quantile (*q-q*) plots: if the distributions are of the same form (that is, the same up to a transformation of location and scale – typically the mean and standard deviation), then on the q-q plot, the data will lie, approximately, on a straight line [14]. If, in addition, the location and scale parameters are the same, the data will lie on a line with slope 1. Numerical comparisons were made using the two-sample Kolmogorov-Smirnov test applied to the z-scored data (that is, data where the mean had been removed and then divided

through by the standard deviation). This high-precision test was applied to quantify the results of the visual q-q plot analysis.

Trend estimation

In order to analyze the dynamics of PD over short time periods, it is necessary to remove the effect of trends that occurred due to the natural progression of the disease over that timescale. One widespread approach to modeling disease progression is the use of *hierarchical mixed-effects models* [15]. These are very commonly applied in pharmacodynamics studies [16]. Considerable effort over the preceding decades has increased the sophistication of these models from their origins in simple linear mixed-effects models by incorporating additional features such as smooth [15], or abrupt, nonlinearities in progression [3], nonparametric progression curves [17], and more recently, clustering of individuals into arbitrary groupings using nonparametric Bayes techniques [18]. In PD, pharmacodynamic studies have fitted the smooth, *Gompertz sigmoidal curve* as a model for progression over the lifetime of the patient with parameters estimated on low-frequency data [3].

A predominant feature of these models is *pooling* data between subjects. Drawing on specific knowledge about underlying physiological processes – for example, in virology, the mechanism of viral infection of cell populations – can give a biologically plausible functional form for the curve. Then the problem becomes one of estimating the parameters for the curve, also known as a *regression problem*. When there is insufficient progression data about each individual to get reliable (low variance) individual parameter estimates, a global model that fits the data pooled over all individuals can be more reliable, but biased with respect to each individual. By assuming that the individual regression parameters are random variables, it is possible to form compromise parameter estimates using an appropriate mix of the global and individual models: this is the main premise of (two-level) hierarchical modeling.

In our case, we wish to perform an exploratory smoothing of the PLM data that makes as few assumptions as possible, and has easily traceable logic from underlying assumptions to the results, obtained by a simple statistical inference procedure. Also, because we have adequate data at the individual level, we do not actually need a pooled model. These considerations mean that existing mixed-effects models are not suited to our application: they require complex inference schemes that involve approximations (because nonlinear models are generally analytically intractable) that obscure the interpretation of the results, and would be biased from the perspective of the individual [15-17].

We use a *piecewise linear convex regression smoothing* approach (see Appendix), which can approximate smooth, nonlinear progression as a series of lines, and can also naturally model abrupt changes in progression. The only assumption about the resulting curve is that it has minimal total absolute curvature (second derivative against time) given a fixed total mean squared error with respect to the individual's PLM data. Note that this model is related to, but much simpler than, the nonparametric spline mixed-effects model of Rice and Wu [17]. By contrast to the Rice and Wu model however, the inference problem is *convex* (it has a verifiable optimum solution) which is solved by stable computations whose convergence properties are guaranteed [19].

Residual modeling

The trend identified above is subtracted from the UPDRS data to obtain the *residuals*. Modeling these fluctuations allows us to quantify the high-frequency dynamics of PD symptoms. Trends in the size of these fluctuating residuals can be detected using a variety of methods, but due to the specifics of the trend estimation algorithm described above, we modeled the size of the residuals against time since diagnosis using a *gamma generalized linear model* (see Appendix).

All analysis was carried out using specialized software written for the MATLAB platform version R2007a (MathWorks Inc., 3 Apple Hill Drive, Apple Hill Drive, Natick, MA, USA). Creative Commons-licensed trend estimation software is distributed with this publication.

Results

PatientsLikeMe data set validation

Our findings in this regard are encouraging: the PLM and PD-DOC data sets agree in terms of the broad shape of the distribution of UPDRS values and ages at diagnosis (Figure 1). There are some systematic differences (see discussion section), but the fact that the PLM and PD-DOC distributions are broadly similar in distribution is

good evidence that the online PLM data set is as reliable as objective, clinical data about patient's symptom severity.

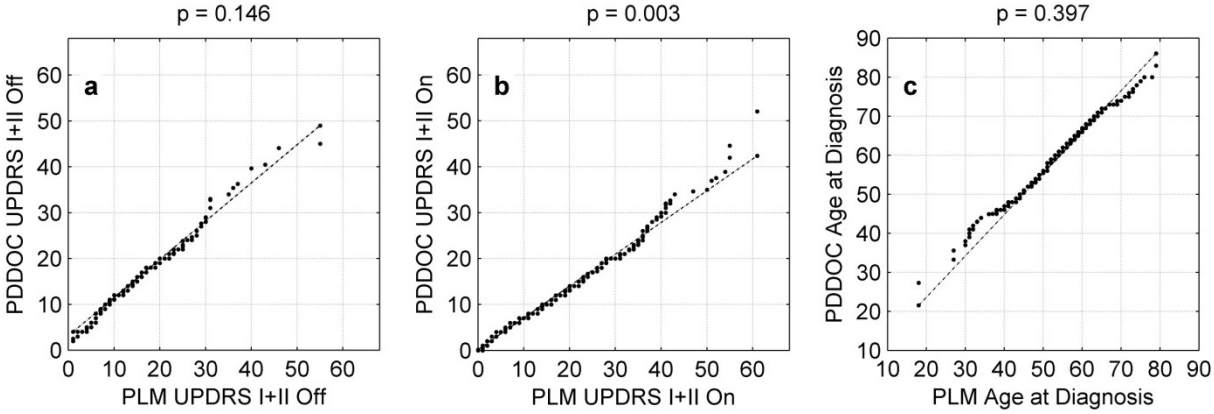

Figure 1: Validating the online, self-reported PatientsLikeMe (PLM) data set against the clinically-scored, PD-DOC reference data set. Visual comparisons using quantile-quantile plots, statistical comparisons using the two-sample Kolmogorov-Smirnov (K-S) test applied to z-scored data (when $p<0.05$, the null hypothesis that the z-scored data come from the same distribution can be rejected at the 95% level). (a) Unified Parkinson's Disease Rating Scale (UPDRS) values (sum of Parts I and II) for values labeled as "off" treatment in the PLM data set against values labeled as "off state" in the PD-DOC data. (b) As with (a), except for the "on" treatment/state labels. (c) Ages at diagnosis. K-S test results (p-values above graphs) indicate that, up to a change in standard deviation and mean, "off" UPDRS values and ages appear come from the same distribution, whereas "on" UPDRS values do not.

Trend estimation

After performing trend estimation (Figure 2 illustrates the selection of the regularization constant and the resulting trend), our next finding is that, whilst a majority of patients do have smooth progression in symptom severity over time with small to moderate short-term variability (Figure 3), there are an interesting and important minority who do not (Figure 4). In the former group, we find patients with very predictable increases in symptom severity, increases that slow over time (Figure 3a,c,d). We also see patients responding well to treatment with gradually decreasing symptom severity, which eventually reaches a plateau (Figure 3b). These patients all conform to the current consensus picture of smooth, long-term symptom changes (for example, following the smooth Gompertz curve [3]). However, in the "non-conforming" group, we find evidence for unpredictable medium-term changes (Figure 4a,b), and occasional, rapid increases (*outliers*) in otherwise smooth progression (Figure 4c,d).

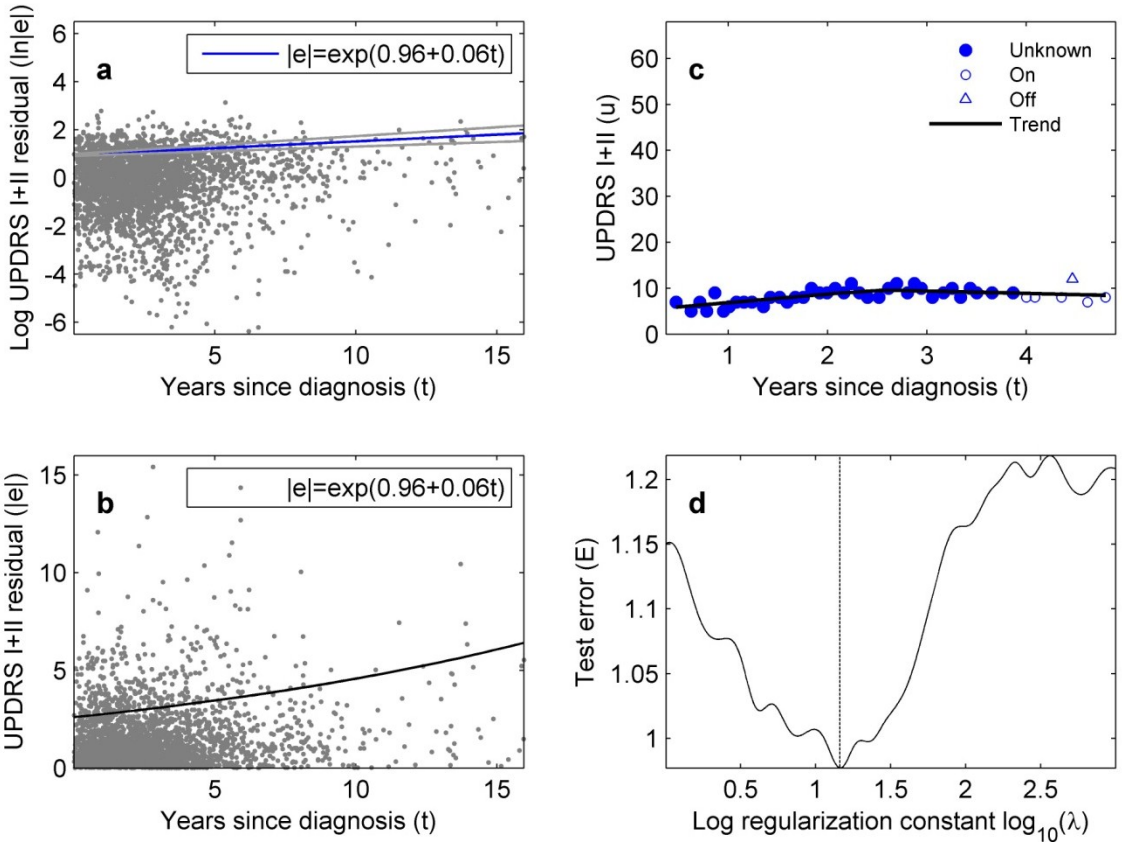

Figure 2: Trend fitting and residual modeling of the self-reported Unified Parkinson's Disease Rating Scale (UPDRS) values (Parts I and II) in the PatientsLikeMe (PLM) data set. (a) Absolute values of residuals obtained by subtracting the long- to medium-term trend from the raw values (natural logarithmic vertical scale), plotted against time since diagnosis in years. The blue line (formula inset) shows the estimated most likely relationship between time since diagnosis and average absolute residual value. The gray lines are the 95% confidence interval for the relationship. (b) The relationship between average absolute residual and time since diagnosis in (a) shown on a linear vertical scale. (c) UPDRS trend, used to calculate residuals, estimated from an example patient. (d) Choice of trend regularization constant for (c), occurring at the smallest value of the cross-validated trend test error.

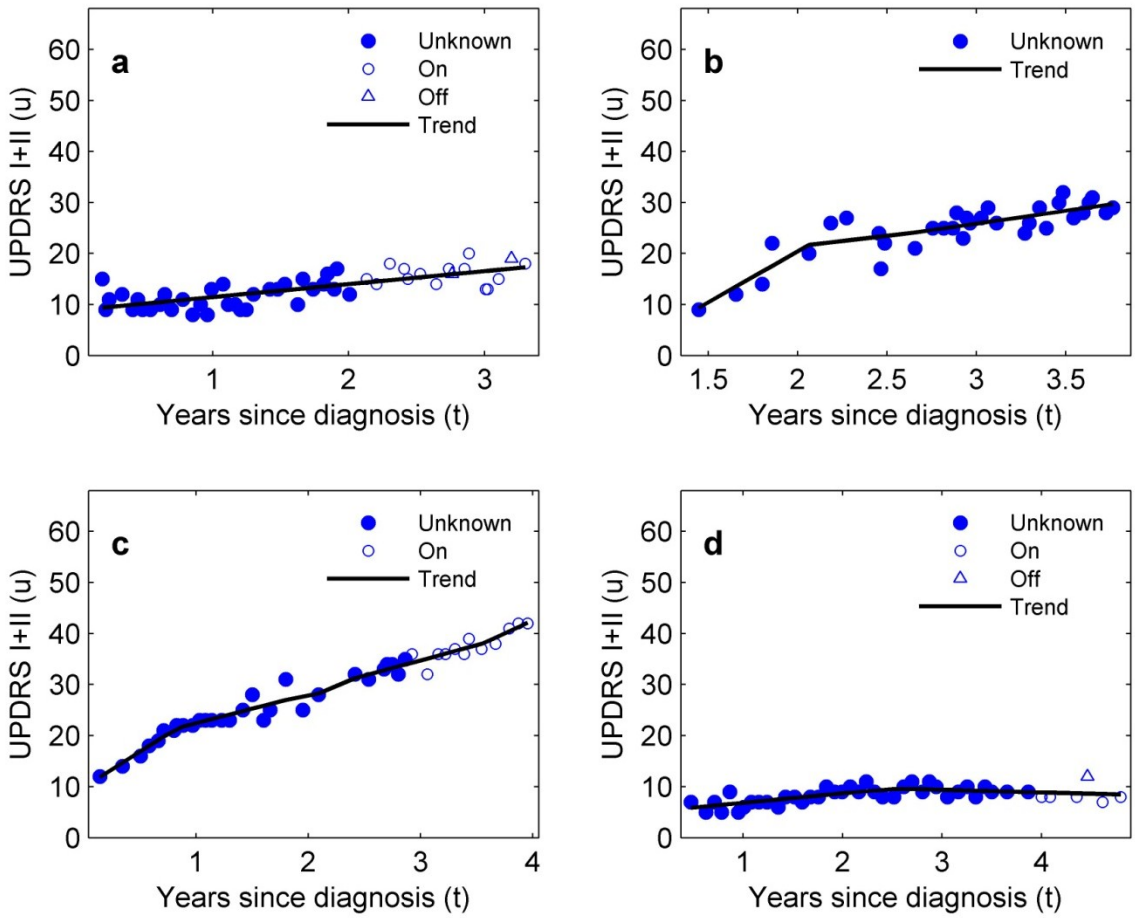

Figure 3: Four examples of patients in the PatientsLikeMe data set whose self-reported symptom dynamics conform to the current consensus picture of slow, predictable Parkinson's disease symptom progression. Increase is generally smooth, variation around the trend (residuals) are generally small. "Unknown" refers to data where the patient did not state whether they were on treatment ("On") or off treatment ("Off") at the time of the symptom report.

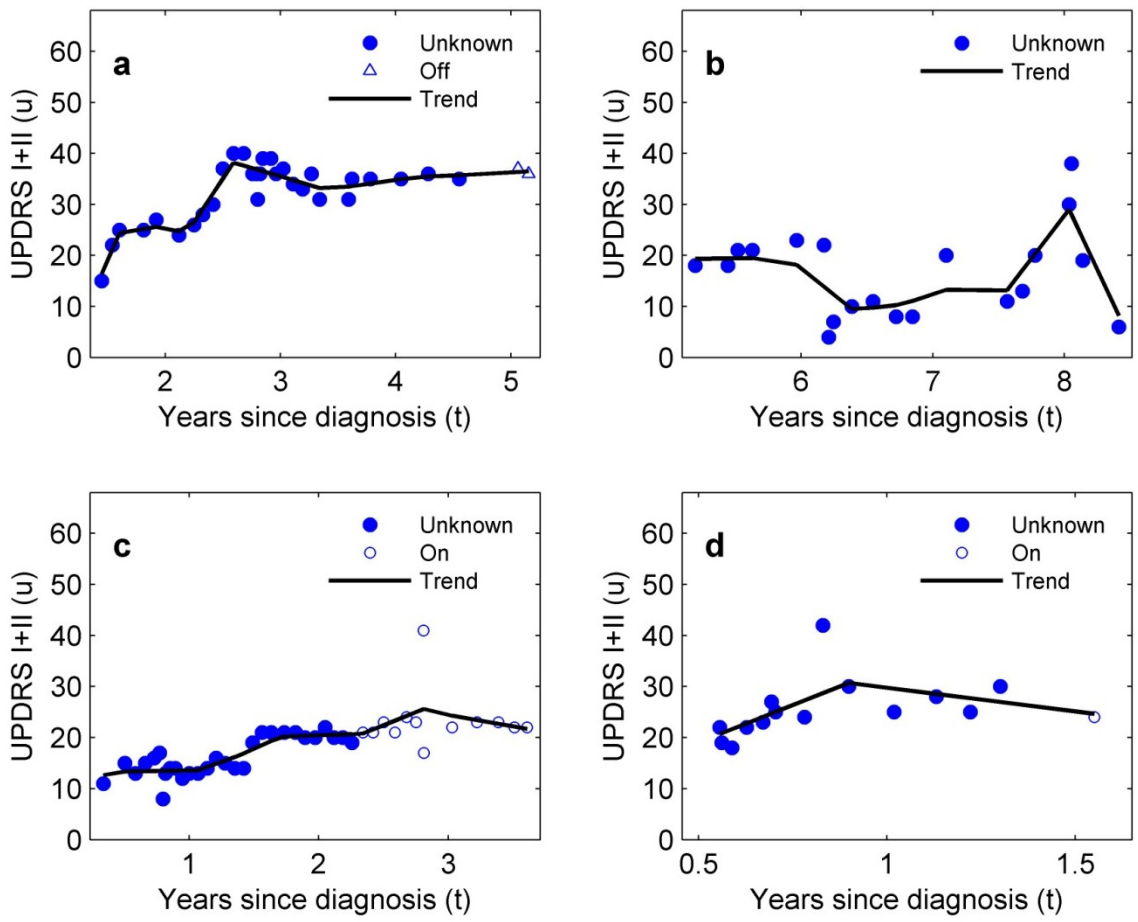

Figure 4: Four examples of patients in the PatientsLikeMe data set whose self-reported symptom dynamics diverge from the current consensus picture of smooth, gradual Parkinson's disease progression (in contrast to Figure 3). (a,b) Large, abrupt changes in the trend of symptom severity occurring over periods of a few weeks or months. (c,d) Examples of single, large deviations from the trend in otherwise smooth progression. "Unknown" refers to data where the patient did not state whether they were on treatment ("On") or off treatment ("Off") at the time of the symptom report.

Residual modeling

Residuals quantifying short-term fluctuations in symptom severity on the scale of days to weeks (which affect all patients to a lessor or greater extent), increase steadily in amplitude with time since diagnosis (Figure 2a,b). At diagnosis, average symptom severity variation is 2.6 points, rising to 5.9 points 16 years later. This finding is not simply a systematic consequence of the non-negative UPDRS scale: if the short-term variation is to be symmetric about the long-term trend, then as the score becomes small, the residuals must get smaller to avoid negative UPDRS values. However, we find that the residuals are significantly positively skewed (t-test against zero skewness rejected with $p < 0.001$, based on 1000-replicate bootstrapped skewness values). A similar argument would hold for very large UPDRS values, since the scale has a maximum value of 68. The PLM data does not contain sufficient information about severely disabled patients with very high Part I and II UPDRS values (less than 10% of the symptom reports in the database have stage 4 or 5 H&Y), and so this argument cannot be tested with the data available to this study.

Discussion

Summary of results

This study addressed the topic of quantifying trends and variability in PD symptoms that occurs on a time scale shorter than 3 months. A data set of 100 PD patients with symptoms, self-reported on a standard clinical scale, was analyzed. Although we have found examples of specialized studies collecting weekly UPDRS values in the literature (see e.g. Goetz et al. [20] recording motor UPDRS weekly for 8 weeks from 16 patients to assess the effect of switching dopamine agonists), to our knowledge, this rapid self-reporting of PD symptoms stretching over many years in the PLM data is unprecedented amongst existing reference clinical data sets. With appropriate feedback and social network community engagement, this data set has the potential to grow quickly at little marginal cost per patient, because the usefulness of the network grows with the square of the network size (an observation known as Metcalfe's law).

Validation demonstrated the high-frequency self-reported data is consistent with a low-frequency clinical data set in common use in clinical PD studies. The distributions of PLM to PD-DOC "off" scores are essentially the same (Figure 1a). One systematic difference is that the mean age of PLM patients is approximately six years younger than the mean age of PD-DOC patients (Figure 1c) which is most likely a sociological effect: younger patients are generally more technologically aware, able, or willing to share their personal data. Similar patterns have been identified in other conditions such as multiple sclerosis [21], where the average online patient was four years younger than patients in a comparable clinical reference data set.

Another systematic difference is that the symptom scores for PLM patients in the "on" state are biased upwards by comparison to the PD-DOC data (Figure 1b). Furthermore, the largest symptom scores (greater than 40 UPDRS points) for PLM patients are much more common than in the PD-DOC data set (this is the reason why, even after z-scoring, the K-S test fails). The most plausible explanation for this difference in "on" scores is due to differences in interpretation of the meaning of "on", as discussed next.

In PD, during the day there will be "on" periods where the symptoms of PD are suppressed, to a greater or lesser degree, by the treatment, and "off" periods where the full symptoms reoccur, even whilst taking treatment. The "on/off" terminology therefore has this somewhat specialized clinical meaning. In the PLM data set, when completing UPDRS self-reports, patients are presented with the following question: "When you answer these questions, are you thinking about how you are on-treatment or off-treatment?", and they can respond by selecting either "On-treatment" or "Off-treatment". In the PD-DOC data set which is collected by trained clinical staff, it can be assumed that the "on/off" terminology is used according to the clinically-accepted definition above. By contrast, with self-reporting in the PLM data set, it is more likely that the "on/off" labels refer to *taking ("on") treatment versus not taking ("off") treatment*, and it is unclear whether patients are generally aware of the accepted clinical meaning of the "on/off" terms. PLM self-reporters can indicate that this is their UPDRS value while "on treatment", and this would partly concur with the clinical "on" state. Similarly, PLM self-reports indicating "off treatment" might, only partly, overlap with the clinical "off" state in the PD-DOC data set.

It is likely that the PLM "on" label includes many scores that would be considered, clinically as "off" instead, because they refer to unsuppressed symptoms occurring while the patient is actively taking treatments (the clinical "off" condition). This would lead to the increased scores we observe here.

To identify trends in symptom progression, a cross-validated, convex piecewise linear smoothing technique was applied to the self-reported data. After subtracting the trend from the self-reported scores, the remaining residual variations appeared to increase with time since diagnosis. Furthermore, a minority of patients were shown to deviate quite considerably from the existing consensus understanding which proposes smooth, gradual change in symptoms over time. Our conclusion is that these residuals are naturally heteroscedastic.

The variations in symptom severity we detect are unlikely to simply be clinically irrelevant fluctuation: previous studies have estimated the minimal *clinically important difference* (CID) in *total* (Parts I – III) UPDRS values is around 4.1 to 4.5 points [22]. The maximum value measuring only Parts I and II is 68, whereas the total UPDRS value is 176 points. From this, we can get a rough estimate of $68/176 \times 4.3 = 1.7$ as the minimal CID for the data in this study, which implies that at the time of diagnosis, the average residual variation of 2.6 around the trend that we find is larger than the minimal variation in symptoms needed to trigger clinical decisions. Later, at 16 years after diagnosis, the same calculation shows a "moderate" CID of 3.3 points, so the average variation we find here (5.9 points) could be very misleading if taken out of context in for example, a clinical trial for a new drug treatment.

Limitations

This study collected self-reported data about cognitive, behavioral, mood symptoms, and impairment in activities of daily living. PD is primarily a movement disorder so that it is important to be able to quantify movement symptoms in addition. Nonetheless, activities of daily living are significantly impaired by motor deterioration, so this section of the UPDRS measures motor symptoms indirectly. Since the UPDRS is additive, even if motor symptoms increase smoothly in severity according to a long-term trend, the *total* UPDRS score (integrating Parts I – III) would still show the effect of variability in Parts I and II that we observe here. Previous, low-frequency studies show evidence of the kind of variability that we find here in motor symptoms such as bradykinesia, rigidity and tremor [3]. Thus we have some confidence that the explicit inclusion of a direct quantification of motor symptoms, while an important addition that would alter our assessment of the specific numerical results presented here, will not fundamentally alter our conclusions.

Since the PLM website has no mechanism to require patients to return to the site and enter new symptom reports, it is possible that many patients only return to the site to enter their symptoms when they have experienced a symptom fluctuation. However, patients are unlikely to agree on what level of change in UPDRS constitutes a reportable fluctuation, and so, we would expect to see fluctuations of all sizes, and differing reporting intervals, in the data set. We would also not expect to find regular time intervals between reports (if symptom fluctuations are indeed random). Therefore, there is no reason to believe that such fluctuation-triggered reporting is a significant source of bias in our results.

The standard CID calculations in UPDRS are performed on cross-sectional data, and refer to the symptom variation around the average across all individuals in the PD population [22]. Therefore, in order to draw meaningful comparisons against this literature we have performed the equivalent pooling across all individuals. These CID calculations therefore make the (implicit) statistical assumption that the patients all come from a homogenous group sharing the same UPDRS distribution. Our findings here probably indicate that this assumption may not be statistically accurate, because we have found quite significant differences in symptom progression and magnitude of variation. Further statistical analysis may be needed to identify the nature of any systematic differences or sub-groups in residual distribution.

Implications for Parkinson's disease research and clinical practice

What we detect here is fundamental variability in symptoms on time scales less than 3 months, which all patients at all stages of the disease seem to show. We note that the variability captured by the residuals we see here is not the same as the variability usually associated with 'fluctuators': the clinical term used to refer to patients with severe symptoms, usually in the later stages of the disease, who experience intermittent responsiveness to drug treatments [23].

Typical of many eHealth studies [24], we find a large attrition rate: of the more than 6000 PD patients registered to the site, the fraction of sufficiently committed users is very small (less than 2%). It can be estimated that entering 30 symptom reports would require on average around 7.5 hours of patients' time in total, using timing information derived from self-administered paper data entry [5]. This is a lot of time to dedicate to entering data into a computer if there is no obvious reward (such as financial compensation, as is frequently the case with clinical trials), even if spread over nearly 3 years, and is one plausible explanation for this severe attrition rate. It is possible that patients who are this dedicated are a select group which may introduce some, as yet, unknown bias into the results. Nonetheless, aside from this group being younger than typical clinical populations, we are not aware of any particular reason why the results we present here would be biased by focusing on a core of more dedicated symptom diarists.

We found that altering the inclusion/exclusion criteria from the PLM data set did not lead to significant changes in the residual model.

The fundamental variability we detect here represents a critical factor in clinical decision-making: knowing what sort of variability to expect is important because it determines how long to wait to detect a significant improvement in symptoms following a change in treatment regime, for example. The explicit information provided here could also be used to build improved progression models, for example, knowledge of the distribution of the residuals can be used to derive more accurate, statistical model-fitting algorithms.

It is difficult to speculate on the origins of such heterogeneity in progression, but other studies have identified different clinical 'subtypes' of PD [25]. It is possible that this might also be reflected in different progression profiles. Future research using this kind of high-frequency data might be able identify different 'progression subtypes'.

The main issue we identify with existing clinical PD symptom data is that it is an *undersampling* of the high-frequency data we present here [26]: that is, because the sampling frequency is so low, it does not adequately represent the kind of symptom fluctuations that most patients experience on time scales shorter than 3 months.

The existence of such non-conforming patients is of critical importance to trial design, where it is typically assumed, based on current understanding of PD symptom progression, that symptoms will change slowly over the duration of the trial. However, this is not always true (e.g. Figure 4a,b). Recruiting patients into trials with the expectation that symptoms will necessarily change slowly over that period, may lead to questionable results, including the failure of trial statistics to show clinically significant outcomes, not as a consequence of the failure of the treatment under test, but because of a failure to incorporate such non-conforming progression into the statistical procedures used to analyze the data.

We see this study as a prelude to the next logical step of increasing the frequency of objective symptom measurement even further. For example, we envisage these results being of utility of in the design of novel non-invasive, objective symptom severity quantification algorithms. Methods based on voice [27] or accelerometry [28], particularly using smartphones, seem promising, because they offer the potential to track the effectiveness, in real-time, of choices in drug dosage and timing. These new methods will require high-frequency reference symptom data for verification, and current clinical reference data such as the PD-DOC database are insufficiently detailed for this purpose.

The ability to self-administer tests for PD symptom severity data remotely offers considerable cost reductions for most clinical applications, for example, reducing the cost of clinical staff time and transport for patients during routine checkups, and lowering the costs of recruitment and tracking of patients in clinical trials. Finally, there is the potential to use this kind of high-frequency data to fit models that can be used for prognostics: that is to predict each patient's future symptom severity.

Acknowledgements

Max Little is funded by a Wellcome Trust-MIT postdoctoral fellowship, grant number WT090651MF.

Conflicts of interest

PW & TV are employees of PatientsLikeMe and own stock options in the company. The PatientsLikeMe R&D team has received research funding from Abbott, Accorda, Avanir, Biogen, Genzyme, Merck, Novartis, Sanofi and UCB.

Appendix

Trend estimation by piecewise linear regression smoothing

In this section we describe the regression smoothing method used to detect the medium- and long-term trends in the self-reported UPDRS data. The method used is an adaptation of L_1 *trend filtering* [29], a technique that has shown to be useful in a range of smoothing problems. The following notation is used: the time since diagnosis for each patient is t_n, $n = 1,2 \ldots N$, where N is the number of observations for each patient, and the combined total Part I and Part II UPDRS values is $u(t_n)$. The regression smoothing is achieved by minimizing the following functional with respect to the estimated trend $v(t_n)$:

$$E = \frac{1}{2}\|u - v\|_2^2 + \lambda\|\mathrm{D}^2 v\|_1 \qquad (1)$$

Here, the notation $\|.\|_q$ is the L_q-norm. The parameter λ is the *regularization constant*. For each value of λ, the output can be shown to consist of a series of straight lines joined together at their ends, i.e. it is a *piecewise linear spline* [29]. When $\lambda = 0$, the first term in the Eq. (1) dominates, so the output $v(t_n)$ is the same as the input $u(t_n)$. As λ increases, the output $v(t_n)$ becomes progressively smoother. It can be shown that there is a maximum useful value of the regularization constant $\lambda_{\max}$: if the regularization constant is equal to or larger than this, the output consists of a single, least squares straight line fit going through the data $u(t_n)$.

The matrix D^2 is a second derivative matrix that takes into account the non-uniform time spacing of the UPDRS data points. It is a tridiagonal matrix encoding a second-order accurate finite difference approximation [30]:

$$\frac{\mathrm{d}^2 v}{\mathrm{d}t^2}(t_n) \approx 2\left[\left(h_n(h_n + h_{n+1})\right)^{-1} v(t_{n-1}) - (h_n h_{n+1})^{-1} v(t_n) + \left(h_{n+1}(h_n + h_{n+1})\right)^{-1} v(t_{n+1})\right] \qquad (2)$$

where $h_n = t_n - t_{n-1}$ is the local temporal difference. After minimization of Eq. (1) to obtain $v(t_n)$, the error *residual* $e(t_n) = u(t_n) - v(t_n)$ is further analyzed. Because Eq. (1) is in the form of a *quadratic program*, it is a *convex optimization problem* for which a unique, globally optimal solution is guaranteed to exist. Special optimization algorithms have been developed for such functionals, here, we use an efficient version of the *primal-dual interior-point* algorithm [29].

The regularization constant λ determines the smoothness of the output $v(t_n)$, and so must be chosen appropriately. In this study we use *cross-validation* to choose this parameter [31]. This involves a uniformly random partition of the data for each patient into a training set (80% of the data) and a testing set (the remaining 20%). Eq. (1) is optimized on the training set, then the mean absolute *test error* is calculated:

$$E(\lambda) = \left(\frac{1}{|Q|}\right) \sum_{m \in Q} |u(t_m) - \hat{v}(t_m)| \quad (3)$$

where Q is the set of indices, and $|Q|$ is the number of data points, in the test partition. The test set values $\hat{v}(t_m)$ are obtained by linear interpolation/extrapolation from the smooth output points $v(t_n)$ closest in time to the test time t_m; this interpolation is justified by the piecewise linear nature of the smoothing operation. The optimal λ is the value that minimizes the test error $E(\lambda)$ (note that this is generally unique to each patient). In order to find this optimal value, we sweep across a wide range of values of λ and calculate the curve $E(\lambda)$. In order to reduce the effects of random partition sampling variation in this curve, we smooth the curve using kernel regression with Gaussian kernel of bandwidth set to 100. This makes it straightforward to find the optimal degree of smoothing for each patient. In this study, we sample 2500 values of the regularization constant over the range [0, 1000].

Gamma generalized linear modeling of residuals

In modeling the residuals of the piecewise linear regression smoothing described in section A.1 above, it is important to take into consideration the distribution of these residuals. It can be shown that minimizing Eq. (1) leads to residuals that are increasingly *Laplacian*, that is, they become Laplace distributed as λ increases [19,32]. From this, it follows that the absolute residuals $|e(t_n)| = |u(t_n) - v(t_n)|$ are approximately exponentially distributed (because the Laplace distribution is a symmetric, two-sided exponential). We are interested in modeling systematic variations of the absolute residuals with respect to the time since diagnosis. The sufficient statistic for the exponential distribution is the mean. Therefore, regressing the mean of the absolute residuals on the time since diagnosis allows us to make predictions about the change in distribution of the residuals over the lifetime of the patient's illness.

Least-squares linear regression is the simplest approach, but this method assumes that the residuals are Gaussian distributed, which contradicts what we know about the residuals. However, we can perform linear regression using *generalized linear modeling* (GLM), which allows the residuals to come from the more general class of *exponential family* distributions [33]. This class includes the exponential and *gamma distributions* as special cases. It also allows a monotonic nonlinearity (known as the *link function*) as part of the regression. In this study, we use the natural logarithm link function because this is the *canonical* choice for the gamma distribution which includes the exponential distribution as a special case:

$$\ln \mu = b_0 + b_1 t \quad (4)$$

where μ refers to the mean of the absolute residuals $|e(t_n)|$. Finding the values for the regression coefficients b_0, b_1 that maximize the likelihood of the data given the coefficients, is a convex optimization problem solvable by *iteratively reweighted least squares* [33].

References

1. de Lau, LM, Breteler, MM. Epidemiology of Parkinson's disease. Lancet Neurol 2006;5(6):525-35. DOI:10.1016/S1474-4422(06)70471-9
2. Wirdefeldt, K, Adami, HO, Cole, P, Trichopoulos, D, Mandel, J. Epidemiology and etiology of Parkinson's disease: a review of the evidence. Eur J Epidemiol 2011;26 Suppl 1(S1-58. PMID: 21626386
3. Vu, TC, Nutt, JG, Holford, NH. Progression of Motor and Non-Motor Features of Parkinson's Disease and Their Response to Treatment. Br J Clin Pharmacol 2012. PMID: 22283961 (online ahead of print)
4. Hoehn, MM, Yahr, MD. Parkinsonism: onset, progression and mortality. Neurology 1967;17(5):427-42. PMID:6067254
5. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. Mov Disord 2003;18(7):738-50. PMID: 12815652

6. Ramaker, C, Marinus, J, Stiggelbout, AM, Van Hilten, BJ. Systematic evaluation of rating scales for impairment and disability in Parkinson's disease. Mov Disord 2002;17(5):867-76. PMID:12360535
7. Tsanas, A, Little, MA, McSharry, PE, Scanlon, BK, Papapetropoulos, S. Statistical analysis and mapping of the unified Parkinson's Disease rating scale to Hoehn and Yahr staging. Parkinsonism Relat Disord 2012. PMID: 22321863 (online ahead of print)
8. Martinez-Martin, P, Gil-Nagel, A, Gracia, LM, Gomez, JB, Martinez-Sarries, J, Bermejo, F. Unified Parkinson's Disease Rating Scale characteristics and structure. The Cooperative Multicentric Group. Mov Disord 1994;9(1):76-83. PMID: 8139608
9. http://www.patientslikeme.com. Accessed:2012-03-04. (Archived by WebCite® at http://www.webcitation.org/65vSjbR4n)
10. Wicks, P, Vaughan, TE, Massagli, MP, Heywood, J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotechnol 2011;29(5):411-4. PMID: 21516084
11. Kurlan, R, Murphy, D. Parkinson's disease data and organizing center. Mov Disord 2007;22(6):904-5. PMCID:PMC2584802
12. http://www.patientslikeme.com/about/privacy. Accessed:2012-03-07. (Archived by WebCite® at http://www.webcitation.org/65zQdX8tJ)
13. Louis, ED, Lynch, T, Marder, K, Fahn, S. Reliability of patient completion of the historical section of the Unified Parkinson's Disease Rating Scale. Mov Disord 1996;11(2):185-92. PMID: 8684390
14. Wasserman, L. All of statistics : a concise course in statistical inference. Springer texts in statistics. New York: Springer; 2004. Isbn:0387402721
15. Davidian, M, Giltinan, DM. Nonlinear Models for Repeated Measurement Data: An Overview and Update. J Agri Bio Env Stat 2003;8(4):387-419. DOI: 10.1198/1085711032697
16. Bonate, PL. Recommended reading in population pharmacokinetic pharmacodynamics. AAPS J 2005;7(2):E363-73. DOI: 10.1208/aapsj070237
17. Rice, JA, Wu, CO. Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 2001;57(1):253-9. DOI: 10.1111/j.0006-341X.2001.00253.x
18. Muller, P, Rosner, GL, De Iorio, M, MacEachern, S. A nonparametric Bayesian model for inference in related longitudinal studies. J Roy Stat Soc C: Appl Stat 2005;54(3):611-626. DOI: 10.1111/j.1467-9876.2005.05475.x
19. Boyd, SP, Vandenberghe, L. Convex optimization. Cambridge, UK: Cambridge University Press; 2004. Isbn:0521833787
20. Goetz, CG, Blasucci, L, Stebbins, GT. Switching dopamine agonists in advanced Parkinson's disease: is rapid titration preferable to slow? Neurology 1999;52(6):1227-9. PMID:10214748
21. Wicks, P, Massagli, M, Kulkarni, A, Dastani, H. Use of an online community to develop patient-reported outcome instruments: the Multiple Sclerosis Treatment Adherence Questionnaire (MS-TAQ). J Med Internet Res 2011;13(1):e12. PMID: 21266318
22. Shulman, LM, Gruber-Baldini, AL, Anderson, KE, Fishman, PS, Reich, SG, Weiner, WJ. The clinically important difference on the unified Parkinson's disease rating scale. Arch Neurol 2010;67(1):64-70. PMID:20065131
23. Weiner, WJ. Motor fluctuations in Parkinson's disease. Rev Neurol Dis 2006;3(3):101-8. DOI:10.1016/S0022-510X(99)00052-0
24. Eysenbach, G. The law of attrition. J Med Internet Res 2005;7(1):e11. PMID: 15829473
25. van Rooden, SM, Colas, F, Martinez-Martin, P, Visser, M, Verbaan, D, Marinus, J, Chaudhuri, RK, Kok, JN, van Hilten, JJ. Clinical subtypes of Parkinson's disease. Mov Disord 2011;26(1):51-8. DOI: 10.1002/mds.23346
26. Clifford, GD, Clifton, D. Wireless technology in disease management and medicine. Annu Rev Med 2011;63(479-92. DOI: 10.1146/annurev-med-051210-114650
27. Tsanas, A, Little, MA, McSharry, PE, Ramig, LO. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. J R Soc Interface 2010;8(59):842-55. PMID:21084338
28. Kostikis, N, Hristu-Varsakelis, D, Arnaoutoglou, M, Kotsavasiloglou, C, Baloyiannis, S. Towards remote evaluation of movement disorders via smartphones. Conf Proc IEEE Eng Med Biol Soc 2011:5240-3; 2011. PMID:22255519
29. Kim, SJ, Koh, K, Boyd, S, Gorinevsky, D. L1 trend filtering. SIAM Review 2009;51(2):339-360. DOI:10.1137/070690274

30. Sundqvist, H, Veronis, G. A simple finite-difference grid with non-constant intervals. Tellus 1970;22(1):26-31. DOI:10.1111/j.2153-3490.1970.tb01933.x
31. Hastie, T, Tibshirani, R, Friedman, JH. The elements of statistical learning: data mining, inference, and prediction. Springer Series in Statistics. New York: Springer; 2001. Isbn:0387952845
32. Kubin, G. On the nonlinearity of linear prediction. IXth European Signal Processing Conference EUSIPCO'98. Rhodes, Greece; 1998.
33. McCullagh, P, Nelder, JA. Generalized linear models. Monographs on statistics and applied probability ; 37. London ; New York: Chapman and Hall; 1989. Isbn:0412317605