Parkinsonism and Related Disorders 19 (2013) 232–237

Contents lists available at SciVerse ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Validating an objective video-based dyskinesia severity score in Parkinson's disease patients

Anusha Sathyanarayanan Rao [a], Benoit M. Dawant [a], Robert E. Bodenheimer [a], Rui Li [a], John Fang [b], Fenna Phibbs [b], Peter Hedera [b], Thomas Davis [b,*]

[a] Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37232, USA
[b] Department of Neurology, School of Medicine, Vanderbilt University, Nashville, TN 37232, USA

ARTICLE INFO

Article history:
Received 20 January 2012
Received in revised form
16 September 2012
Accepted 26 October 2012

Keywords:
Parkinson's disease
Rating scales
Quantify dyskinesia
Rater variability
Severity ranking

ABSTRACT

Dyskinesia is a common side effect of prolonged dopaminergic therapy in Parkinson's disease patients. Assessing the severity of dyskinesia could help develop better pharmacological and surgical interventions. We have developed a semi-automatic video-based objective dyskinesia quantifying measure called the severity score (SVS) that was evaluated on 35 patient videos. We present a study to evaluate the utility of our severity score and compare its performance to clinical ratings of neurologists. In addition to the Unified Dyskinesia Rating Scale (UDysRS) score for each video, four neurologists provided three sets of time lapsed ratings and rankings of the 35 videos using a specifically developed protocol. The statistical analysis of our data using Kendall's tau-b and intra-class correlations shows that (a) ranking patient videos based on severity is suitable for studying the utility of the SVS, and (b) SVS exhibits moderate utility to quantify dyskinesia severity when compared to manual assessment of dyskinesia by neurologists using the UDysRS. These results support the effective use of SVS as an objective measure to quantify dyskinesia and the rationale for a ranking system that complements traditional rating scales.

Published by Elsevier Ltd.

1. Introduction

Levodopa therapy in Parkinson's disease (PD) patients results in drug-induced dyskinesia characterized by hyperkinetic involuntary movements that may often interfere with activities of daily living [1]. Despite current treatment measures, the disabling symptoms of dyskinesia continue to challenge the development of better pharmacological and surgical interventions. In this context, rating scales have been the most established and widely used means of assessment of the severity of dyskinesia. The key attributes of dyskinesia evaluated include anatomical distribution, phenomenology, duration, intensity, disability, and patient perception [2]. Different scales base their severity ratings on different sets of attributes of dyskinesia. Some of the widely used ratings scales are the Abnormal Involuntary Movement Scale (AIMS) [3], the Lang Fahn activities of daily living scale [4], the Rush Dyskinesia Rating Scale [5], the Parkinson's disease Dyskinesia Scale [6], and the Clinical Dyskinesia Rating Scale (CDRS) [7]. The most recently developed scale is the Unified Dyskinesia Rating Scale (UDysRS), which may become the standardized dyskinesia rating scale equivalent to the UPDRS scale for PD symptoms [2]. The UDysRS is a combination of several rating scales in such a way that all attributes of dyskinesia are assessed using a single rating scale. The results of the clinimetric testing of this scale over a range of 70 patients indicated an inter-rater and intra-rater reliability with correlation coefficients ranging from 0.37 to 0.87 for various tasks. Further validation and responsiveness testing is underway [2].

Though rating scales are the conventional assessment tool, there are several disadvantages to their use [8]. First, they are subjective and often require intensive training to obtain acceptable intra- and inter-rater reliability. Second, scales such as the CDRS and UDysRS include patient questionnaires, which may not represent the severity of dyskinesia accurately. The rating scales often rely on a discrete five point scale. This lack of resolution leads to the possibility of misclassifying patients with symptoms that fall in between two rating intervals. These factors encourage the development of quantitative assessment techniques, which has been our primary research interest. Accelerometers and gyroscopes have achieved moderate success in quantifying the severity of dyskinesia in the recent past [9–12]. The disadvantages of these techniques are the use of expensive and dedicated devices that require complex software, and the inconvenience to the patients wearing the devices.

Our work is an example of a video-based, marker-less, model-free human motion tracking using the standardized clinical

* Corresponding author. Tel.: +1 615 936 0060; fax: +1 615 936 1229.
E-mail address: thomas.l.davis@Vanderbilt.Edu (T. Davis).

1353-8020/$ – see front matter Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.parkreldis.2012.10.015

videos of PD patients that are often a part of the patient's clinical records. Using this technique, we have developed a severity score (SVS) to quantify the severity of dyskinesia exhibited by the patient [13]. As a continuous variable, SVS cannot be directly compared to the discrete rating scales. We therefore developed a rating based ranking protocol to validate the utility of SVS to quantify dyskinesia. Each patient in the study had the following parameters: neurologist ratings and rankings, UDySRS rankings and SVS ranking. Three studies were performed using these parameters to establish the utility of the SVS and the ranking protocol – (a) Assessment of intra- and inter-neurologist consistency using the ranking protocol; (b) Comparison of SVS with the UDysRS and the neurologists; and (c) Effect of ratings vs. rankings. Our results indicate that SVS correlates well with neurologists' and the UDysRS rankings. The validity of the SVS was studied by evaluating it using different video segments of some of the patient videos. This longitudinal study on the videos was performed to observe if 10 s of video data was sufficient to quantify dyskinesia severity in patients.

2. Methods

Our analysis used samples of 35 patient videos with varying dyskinesia severity obtained as part of the extensive clinimetric testing of the UdysRS [2]. The videos were captured in a controlled environment with plain backgrounds in a well-illuminated room with no occluding furniture. Details regarding the video protocol and the informed consent obtained have been previously published [2]. The patients were rated based on four tasks that are activities of daily living (ADL). Part IV UDysRS scores for each task was also available. Our task of interest was the communication task, where the patients were asked to read while seated on a chair. The communication task was the simplest task to track using our semi-automatic technique. Though speech disorders were primarily rated using this task, the patients also exhibited movement of the face, head, neck, hands and legs and it was observed that the patients with severely impaired speech also showed dyskinetic movements of these body parts. The average length of the communication task was one minute. A 10 s excerpt from the middle of each 60 s sequence was analyzed using our semi-automatic technique. Our previous work had determined that 10 s of video provided the optimal registration results without significant tracking delay. We avoided any starting and stopping movement effects by not analyzing the beginning or end of the video. The movements of patient's head, shoulders, chest, forearms, knees, feet, and the reading material were semi-automatically tracked using the Adaptive Bases Algorithm (ABA) that is an intensity based non-rigid image registration algorithm [14]. We analyzed the tracked anatomical points of interests by applying principal component analysis (PCA) on the cluster of points from every frame of the video sequence as described in our prior work [15]. A severity score was computed for each video sequence using the parameters obtained from the PCA analysis

$$SVS = TV \times NSM/STDEV$$

TV: total variance of all eigen modes, where the total variance is the sum of the magnitude of all the eigenvalues.

NSM: Number of significant modes of variation, which defines the number of modes of variations that capture 90% of the variations in the patient movements.

STDEV: standard deviation of the percentage contribution of the eigenvalues to the total variance, which represents the rate of fall of eigenvalues. A gradual fall of eigenvalues indicates complex patient movements in various directions and a steep fall indicates simple movements in fewer directions.

The 10 s sequences were ranked in the increasing order of SVS. These videos were also ranked based on the increasing order of the UDysRS part IV communication task scores obtained from the UDysRS study.

Four movement disorder neurologists, N1, N2, N3, and N4, ranked the 35 10 s video sequences using a ranking protocol that included amplitude and speed of dyskinetic movement, anatomical distribution of dyskinesia and the extent of disability seen in the patient. Intelligibility of speech was not considered in the ranking. Each neurologist independently rated and ranked the 35 video segments using this protocol. Three sets of ratings and rankings on the same dataset were obtained at monthly intervals to ensure they were not voluntarily repeated. Thus, each neurologist had three sets of ratings and rankings – ratings: Set 1R, Set 2R, and Set 3R and rankings: Set 1r, Set 2r, and Set 3r.

(a) Ratings: The videos were first rated on a scale of one to four with one – no dyskinesia, two – mild dyskinesia, three – moderate dyskinesia and four – severe dyskinesia.
(b) Rankings: The videos in each rating category, except the no dyskinesia category, were viewed simultaneously on a single screen and ranked according to increasing order of severity within that category.
(c) The first two and the final two videos in each rating category were then compared with the correspondingly ranked videos of the immediately next category to confirm if these rankings were still valid. Thus, the neurologist could view cross category videos to finalize their ranks.
(d) In case of rank changes, steps (b) and (c) were repeated until ranks were finalized and the corresponding rating categories in Step (a) were also modified to ensure coherence between ratings and ranking.

Each patient in the study had the following parameters: three sets of neurologist ratings and rankings, UDysRS ranking and SVS ranking.

The longitudinal study on the videos was performed by computing the SVS using the above technique on two more video segments on patients who had good longitudinal video data available. Five patients had good video data for 30–60 s of the communication task and two 10 s segments apart from the original segment were used for this study.

2.1. Data analyses

Three statistical studies were conducted from the data obtained using the above methods.

(a) Validation of ranking protocol: Evaluation of intra- and inter-neurologist ranking consistency.
(b) SVS Utility: Comparison of SVS rankings with UDysRS rankings and the neurologists rankings
(c) Effect of ratings vs. rankings: Comparison of SVS with neurologists' ratings and rankings.

The original rankings obtained from neurologists were modified as follows to permit statistical analysis because the number of non-dyskinetic patients (neurologist rating of 1) was different in both the inter- and intra-neurologist ratings. Hence, the total number of patients ranked by each neurologist was not necessarily equal. To ensure statistical consistency in the analyses, for each set of rankings, two types of rank data sets were developed.

(1) Type I: All 35 video sequences were part of this dataset. A tied rank was assigned to non-dyskinetic patients such that its value is the average of the ranks the patients would have received if there were given distinct ranks [16]. This process ensured the maximum ranking in each ranking was 35, but the minimum rank would depend on the number of non-dyskinetic patients.
(2) Type II: Seven patients were consistently labeled non-dyskinetic by the senior neurologist in all the three rank sets. These patients were uniformly eliminated from the original rank sets of all the neurologists and the remaining 28 patients were re-ranked keeping the order unchanged. UDysRS and SVS rankings were also modified accordingly.

2.1.1. Study I: validation of ranking protocol – intra- and inter-neurologist agreement

Step (a) of the ranking protocol was based on the clinical definition of dyskinesia and not a specific rating scale. The ranking protocol was developed to compare the discrete neurologists' ratings in Step (a) to the continuous SVS score. By evaluating the intra- and inter-neurologist consistency in using the ranking protocol, the validity of the protocol can be determined. A high intra- and inter-neurologists consistency indicates that the protocol, based on clinical definition of dyskinesia, can be used to rank severity of dyskinesia by neurologists and in turn can be used to evaluate the utility of SVS. Independent analyses were performed to observe intra- and inter-neurologist agreement on the Type I and Type II ranking datasets. Kendall's tau-b correlation coefficient was computed pairwise between the four neurologists in each type to evaluate the inter-neurologist agreement. Intra-class coefficient for each neurologist across Set 1r, Set 2r, and Set 3r was computed [17].

2.1.2. Study II: SVS utility

The goal of this study was to assess if the SVS was a suitable score to quantify dyskinesia in clinical settings. Hence, we compared it to the UDysRS which was used as a gold standard by computing the Kendall's Tau-b correlation coefficient between SVS rankings and the UDysRS rankings. The SVS was also compared to neurologist's performance by computing the Kendall's Tau-b correlation coefficient between the SVS rankings and the neurologists' rankings. The statistical analysis was performed on Type I and Type II rankings of Set 1r, Set 2r and Set 3r. A good correlation would indicate that the SVS can quantify dyskinesia as well as neurologists and can complement UDysRS scores by providing an objective dimension to it.

2.1.3. Study III: effect of ratings vs. rankings

We propose that ranking the severity of dyskinesia within each rating category of mild, moderate or severe dyskinesia assists in quantifying the differences between patients at a finer level. It is easier for neurologists to use a discrete five point rating scale which allows them to assign more than one patient in a single