Internet-based physical assessment of people with Parkinson disease is accurate and reliable: A pilot study

Trevor G. Russell, PhD; Tammy C. Hoffmann, PhD; Mark Nelson, BPHTY(Hons); Leah Thompson, BOccThy (Hons); Amy Vincent, BOccThy (Hons)

Divisions of Physiotherapy and Occupational Therapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia

Abstract—Telerehabilitation may be an alternative service delivery model for people with Parkinson disease (PD) who live in areas where traditional rehabilitation services are not readily accessible. The aim of this study was to determine the accuracy and reliability of performing remote physical assessments of people with PD via telerehabilitation when compared with traditional face-to-face assessments. Twelve subjects were simultaneously examined by a face-to-face investigator and a remote investigator via the eHAB tele rehabilitation system. The outcome measures evaluated included the timed stance test, Timed “Up and Go” test, step test, steps in 360 degree turn, Berg Balance Scale, and lateral and functional reach tests. Limits of agreements intervals and weighted kappa statistics demonstrated the telerehabilitation assessments to be accurate within clinically acceptable limits. A high level of inter- and intrarater reliability was demonstrated across all telerehabilitation assessments. This study indicates that it is possible to assess the physical ability of people with PD via telerehabilitation systems and provides support for the further development of telerehabilitation applications for patients with neurological disorders.

Key words: accuracy, assessment, functional, Internet, Parkinson disease, physiotherapy, reliability, teledermatology, telerehabilitation, videoconference.

INTRODUCTION

Parkinson disease (PD) is a chronic progressive neurological condition that manifests in a disturbance of voluntary motor control resulting in resting tremor, rigidity, bradykinesia, and postural instability [1]. Physiotherapy is an integral component in the multidisciplinary management of people with PD and assists people to function to their maximum potential. Evidence exists for the positive effects of physiotherapy intervention for people with PD. A systematic review of 25 physiotherapy intervention studies by Gage and Storey found that physiotherapy treatment resulted in improvements across a host of outcome measures, including mobility, functional status, motor performance, gait, and activities of daily living [2]. This was further demonstrated in a review by Allen et al. [3]. The successful management of PD depends not only on the quality of the intervention provided, but also on the timing and frequency of the intervention. For many people with PD, regular attendance at health centers for appointments is difficult because of the debilitating effects that the disease has on their mobility, the dependence on carers for transportation, and a lack of appropriate services in close proximity. This is especially

Abbreviations: ICC = intraclass correlation coefficient, MAD = mean absolute difference, PD = Parkinson disease, SD = standard deviation, TUG = Timed “Up and Go” test.

*Address all correspondence to Trevor G. Russell, PhD; Division of Physiotherapy, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane, Queensland, Australia 4072; +61-7-3346-9633; fax: +61-7-3365-1622. Email: t.russell@shrs.uq.edu.au
http://dx.doi.org/10.1682/JRRD.2012.08.0148
rural and remote areas where the tyranny of distance and the reduced number and distribution of health services is such that regular appointments are not feasible [4]. One possible solution to this service delivery issue is to use telerehabilitation technologies to enable the delivery of allied health services at a distance, directly into the homes of patients.

Evidence exists supporting the use of telerehabilitation for people with PD in regards to the treatment of speech disorders [5] and assessing activities of daily living and hand function [6]. However, there is no evidence supporting the use of telerehabilitation for physical assessments of people with PD. Telerehabilitation studies in physiotherapy have included the successful development of various applications to enable physical rehabilitation consultations [7], seating assessment and wheelchair evaluations [8], ankle-foot orthotic assessments [9], and gait analysis [7]. A research team at the University of Queensland, Australia, has developed a computer-based videoconferencing system call eHAB that combines high-quality videoconferencing with a battery of measurement tools that can objectively measure various aspects of functional performance across an Internet link. Studies demonstrating the validity and reliability of this system have been performed to examine physical outcome measurements [6,10], observational kinematic gait analysis [11], upper-limb range of motion in stroke patients [12], lower-limb range of motion in an orthopedic population [13], the effectiveness of rehabilitation following total knee replacement surgery [14–16], and diagnostic accuracy of ankle injuries [17]. Other studies using similar videoconferencing equipment have demonstrated that telerehabilitation assessments of Functional Reach, the European Stroke Scale, and the Functioning Everyday with a Wheelchair-Capacity can be performed accurately compared with face-to-face encounters [18–19].

The aim of this study was to determine if remote physiotherapy assessment of people with PD could be conducted via the Internet. Specifically, the study aimed to (1) evaluate the level of agreement between assessments performed via the telerehabilitation system and assessments performed in the traditional face-to-face manner and (2) establish the intra- and interrater reliability of the telerehabilitation assessments. We hypothesized that remote assessment of people with PD via telerehabilitation would be both valid and reliable.

**METHODS**

**Subjects**

Following approval from the appropriate institutional review board, clients of a neurological disorders, aging, and balance clinic who had a current diagnosis of PD were invited to participate in the study. Thirteen participants expressed interest and were screened for inclusion criteria eligibility. Participants were included if they had a PD severity rating between stages I and IV on the Hoehn and Yahr Score [20], were independently mobile, and had a communication and cognitive status adequate to complete assessment tasks. Participants were excluded if they had any comorbid conditions, such as a chronic orthopedic complaint that would prevent them from safely completing assessment tasks. Following screening, 12 participants (6 male and 6 female) were included in the study after giving written informed consent.

**Telerehabilitation Assessments**

Telerehabilitation assessments were performed using the eHAB telerehabilitation system. The eHAB telerehabilitation system enables videoconferencing at 320 × 240 pixel resolution and real-scale measurement of the patient’s performance on assessment items (such as balance, joint range of motion, muscle strength, and gait) through a suite of optical calibrated assessment tools. This system has been described elsewhere in the literature [11,13–17,21]. The systems were connected via a 3G network connection (Telstra Next G; Melbourne, Australia) with theoretical downlink speeds of 7.2 Mbps and uplink speeds of 5.7 Mbps. During this project, the eHAB system consumed approximately 80 Kbits symmetrically for the videoconference image. Telerehabilitation assessments were performed in a room isolated from the testing room where the participant was located. The system was located at the end of the room and was switched on prior to the participants entering. Once the participant was in the room, the videoconference was initiated by the remote investigator dialing the system. Face-to-face investigators were present at all times to ensure the safety of the participants while performing balance assessment items and transfers.

**Procedure**

Investigators consisted of one final-year physiotherapy and two occupational therapy students who simultaneously assessed each study participant. All investigators
received extensive training from experienced clinicians in both the application of the outcome measures used in the study and the use of the telerehabilitation system. Investigator 1 performed the assessments via the telerehabilitation system while investigator 2 was in the room with the participant scoring all test items as an observer. The third investigator was also present in the testing room and scored all test items as an observer, enabling the collection of data for interrater reliability testing of the face-to-face assessments. There was no interaction between the investigators during the assessment, and each investigator was blinded from the other investigators’ scoring to prevent bias. We considered it necessary that all assessments be performed simultaneously to control for any participant variance that may have been introduced if the assessments were performed in a sequential manner (i.e., participants performing the battery of assessment tasks three times). To enable the reliability testing of the telerehabilitation assessments, video recordings of each assessment were made using a function of the telerehabilitation system. These clips were reviewed on the system by the third investigator 2 mo later to provide data for intrarater reliability testing of the telerehabilitation system. Intrarater reliability data for the telerehabilitation assessments were produced by investigator 1 reviewing and rerating the video clips. This was conducted 2 mo after the initial testing period to limit test retest bias.

**Outcome Measures**

A variety of assessments were performed simultaneously in both the telerehabilitation and face-to-face environments for each participant, including the timed stance test [22–23], Timed “Up and Go” test (TUG) [24], step test [25–26], steps in 360° turn [27], Berg Balance Scale [28], and lateral [29] and functional reach [30]. These assessments are typical and appropriate for determining the physical functioning of people with PD.

Telerehabilitation of the timed stance test and TUG were assessed in real time by observing the participant via the videoconference and timing with a conventional stopwatch. The step test and steps in 360° turn were performed in the same manner, with the number of steps in each assessment counted. For the Berg Balance Scale, participants performed and were scored on each item while being observed through the telerehabilitation system. Both lateral and functional reach required the participant to perform the assessment while a video recording was made using the telerehabilitation system. The investigator reviewed the video recording and used a calibrated assessment tool to determine the functional reach distances. All face-to-face measures were performed according to conventional clinical practice. Lateral and functional reach measures were determined by the face-to-face investigator using pins on a cork board to mark the start and end position of the reach and a tape measure to determine the reach distance. The telerehabilitation investigator was blinded to the face-to-face measures by obscuring the view of the tape measure while performing the measure.

**Data Analysis**

To determine the accuracy of the telerehabilitation assessments, different methods were used for continuous and ordinal data. For continuous data, differences obtained between face-to-face and telerehabilitation methods of assessment were examined using the limits of agreement statistic described by Bland and Altman [31] along with the mean absolute difference (MAD) of the assessments. The limits of agreement statistic provides an estimate of the agreement between the face-to-face and telerehabilitation measurements by expressing a range within which 95 percent of all possible differences in the assessments will lie. Provided this agreement range is within a clinically acceptable limit, the new measurement tool is deemed to be valid and can be used in the clinical situation [31]. The clinically acceptable limit was determined for each assessment item at the commencement of the study. For functional and lateral reach measures, a clinically acceptable limit of 4.74 cm was set according to the mean variance of reach scores found within each Fugal-Meyer stroke severity level as reported by Smith et al. [32]. It is reasonable to assume that if the limits of agreement for the reach measures is within this clinical limit, that the telerehabilitation assessment can be said to have sufficient accuracy to determine a patient’s functional status. Clinical criteria for the TUG outcome measure was set at 5 s according to inter- and intrarater reliability reported by Podsiadlo and Richardson [24]. For the timed stance test, a level of accuracy of 1 s per item was considered adequate. This is reasonable because the majority of studies reporting timed stance measures report the values in whole seconds only. The test clinical criteria was set at 8 s because the test contains eight items.

For ordinal data, weighted kappa scores along with percent exact agreement and percent exact agreement within one point on the ordinal scale were used. Kappa
values of 0.81 to 1.00 were interpreted to represent almost perfect agreement, 0.61 to 0.80 to indicate substantial agreement, 0.41 to 0.60 to indicate moderate agreement, 0.20 to 0.40 to indicate fair agreement, 0.00 to 0.20 to indicate slight agreement, and <0.001 to indicate poor agreement as defined by Landis and Koch [33].

The inter- and intrarater reliability of telerehabilitation assessments and the interrater reliability of the face-to-face assessments was evaluated using intraclass correlation coefficients (ICCs) (2,1). Interpretation of ICC values were based on the categories suggested by Fleiss [34] (>0.75 = excellent reliability, 0.40–0.75 = fair to good reliability, and <0.40 = poor reliability).

RESULTS

The mean age of participants included in the study was 66.1 yr (standard deviation [SD] = 8.5; range 45–76 yr) and the mean age at time of diagnosis was 53.5 yr (SD = 9.0; range 38–69 yr). The average number of years since participants had been diagnosed with PD was 6.8 yr (SD = 4.4; range 2–15 yr), and each sex was equally represented (6 male, 6 female).

The limits of agreement, mean difference, SD of the difference, and MAD for the difference between face-to-face and telerehabilitation assessments for all continuous data items are presented in Table 1. The limits of agreement for all items fell within the clinically acceptable criteria for adequate agreement.

Weighted kappa scores, percent exact agreement, and percentage agreement within one point on the ordinal scale are presented for all ordinal data items in Table 2. Weighted Kappa scores were all above 0.90; however, percent exact agreement scores were lower, with three of the five items being below 75 percent. When using percent agreement within one point on the ordinal scale, four of the five items had agreement levels of above 80 percent, with only the Berg Balance Scale falling below this level with an agreement of 75 percent. When individual items from the Berg Balance Scale were analyzed (Table 3), 13 of the 14 items had an exact agreement of 75 percent or above, with standing on one leg (50%) scoring lower. All 14 items had an agreement of above 80 percent when considering percent agreement within one point on the ordinal scale. Of the 14 individual items, 12 had weighted kappa scores of above 0.80, with tandem standing (0.77) and standing on one leg (0.70) scoring lower.

Intrarater ICCs (2,1) were ≥0.96 for all assessment items in both face-to-face and telerehabilitation environments. Intrarater reliability for the telerehabilitation assessments was also very high with all ICCs (2,1) ≥ 0.98.

DISCUSSION

The results from this study indicate that it is possible to perform online physical assessments of people with PD using the eHAB telerehabilitation system with high levels of accuracy and reliability. The accuracy of the telerehabilitation assessments was established by comparing these assessments with assessments performed in the traditional face-to-face manner. The limits of agreement between these assessments for both the functional reach (−2.71 to 0.69 cm) and the lateral reach (−2.09 to 0.51 cm) were found to lie within the clinically acceptable level set for this study of 4.74 cm. Similarly, the limits of agreement values for the TUG (−1.25 to 1.24 s) and the Timed Stance test (−4.17 to 5.06 s) were within the clinically acceptable level of 5 and 8 s, respectively. This suggests that, for these test items, the assessments performed via the telerehabilitation system were of sufficient accuracy be used in the clinical setting. Despite this, the limits of agreement for the

Table 1.
Limits of agreement, mean difference, standard deviation (SD) of difference, and mean absolute difference (MAD) for difference between face-to-face and telerehabilitation assessments for all continuous data variables.

<table>
<thead>
<tr>
<th>Test Item</th>
<th>Limits of Agreement</th>
<th>Clinically Acceptable Limit</th>
<th>Mean Difference</th>
<th>SD</th>
<th>MAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional Reach (cm)</td>
<td>−2.71 to 0.69</td>
<td>4.74</td>
<td>−1.01</td>
<td>0.87</td>
<td>1.01</td>
</tr>
<tr>
<td>Lateral Reach (cm)</td>
<td>−2.09 to 0.51</td>
<td>4.74</td>
<td>−0.79</td>
<td>0.66</td>
<td>0.82</td>
</tr>
<tr>
<td>TUG (s)</td>
<td>−1.25 to 1.24</td>
<td>5.00</td>
<td>−0.01</td>
<td>0.63</td>
<td>0.47</td>
</tr>
<tr>
<td>Timed Stance Test (s)</td>
<td>−4.17 to 5.06</td>
<td>8.00</td>
<td>0.44</td>
<td>2.35</td>
<td>1.58</td>
</tr>
</tbody>
</table>

TUG = Timed “Up and Go” test.
Table 2.
Weighted kappa, percent exact agreement (%EA), and percent agreement within one point on ordinal scale (%A ± 1) for difference between face-to-face and telerehabilitation assessments for all ordinal data variables.

<table>
<thead>
<tr>
<th>Test Item</th>
<th>Kappa</th>
<th>%EA</th>
<th>%A ± 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step Test Right Foot</td>
<td>0.97</td>
<td>75.0</td>
<td>83.3</td>
</tr>
<tr>
<td>Step Test Left Foot</td>
<td>0.95</td>
<td>66.7</td>
<td>83.3</td>
</tr>
<tr>
<td>Steps in 360° Turn Right</td>
<td>0.98</td>
<td>75.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Steps in 360° Turn Left</td>
<td>0.97</td>
<td>66.7</td>
<td>91.7</td>
</tr>
<tr>
<td>Total Berg Balance Scale Score</td>
<td>0.94</td>
<td>16.7</td>
<td>75.0</td>
</tr>
</tbody>
</table>

Timed Stance test were relatively large. On closer inspection of the data related to this variable, two subjects demonstrated a considerable difference between telerehabilitation scores and face-to-face scores (difference > 4 s). In each of these cases, a large disagreement occurred on only one or two of the eight timed stance items, with the telerehabilitation scores being higher in both cases. A potential explanation for this observation may relate to the quality of real-time videoconference image, which made it difficult for the online investigator to detect subtle signs such as participants opening their eyes during an eyes closed test or slightly touching the ground during a single leg stance test, which would normally terminate the timing of these tests. A way of avoiding this in the future may be to utilize a carer or family member at the patient’s end of the consultation to monitor for these subtle signs and inform the treating clinician.

The weighted kappa scores for all ordinal scale items were above 0.90, indicating a high level of agreement between telerehabilitation and face-to-face assessments. This is consistent with the percent agreement within one point on the ordinal scale data, for which all assessments demonstrated higher than 75 percent agreement. Percent exact agreement scores were much lower, ranging from 16.7 to 75 percent exact agreement. These observations suggest that although assessment items carried out via the telerehabilitation system were not always scored identically to face-to-face assessments, they were quite consistently scored to within one point on an ordinal scale. We contend that this level of agreement is reasonable given that there is some investigator subjectivity associated with the clinical use of these assessment tools.

Table 3.
Weighted kappa, percent exact agreement (%EA), and percent agreement within one point on ordinal scale (%A ± 1) for difference between face-to-face and telerehabilitation assessments for individual Berg Balance Scale items.

<table>
<thead>
<tr>
<th>Berg Balance Scale Item</th>
<th>Kappa</th>
<th>%EA</th>
<th>%A ± 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Transfers</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2. Standing to sitting</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>3. Sitting unsupported</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>4. Sitting to standing</td>
<td>0.81</td>
<td>75.0</td>
<td>91.6</td>
</tr>
<tr>
<td>5. Standing unsupported</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>6. Standing unsupported eyes closed</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>7. Standing unsupported feet together</td>
<td>&gt;0.99</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>8. Tandem standing</td>
<td>0.77</td>
<td>83.3</td>
<td>100.0</td>
</tr>
<tr>
<td>9. Standing on one leg</td>
<td>0.70</td>
<td>50.0</td>
<td>83.3</td>
</tr>
<tr>
<td>10. Turns 360° degrees</td>
<td>0.90</td>
<td>75.0</td>
<td>100.0</td>
</tr>
<tr>
<td>11. Turning trunk</td>
<td>0.85</td>
<td>75.0</td>
<td>91.7</td>
</tr>
<tr>
<td>12. Pick up object from floor</td>
<td>0.92</td>
<td>91.7</td>
<td>100.0</td>
</tr>
<tr>
<td>13. Stool stepping</td>
<td>0.96</td>
<td>91.7</td>
<td>100.0</td>
</tr>
<tr>
<td>14. Reaching forward while standing</td>
<td>0.81</td>
<td>75.0</td>
<td>91.7</td>
</tr>
</tbody>
</table>

Previous studies investigating the face-to-face reliability of the assessments used in this study. For instance, previous studies reports ICC values of 0.89 for functional reach [30], 0.99 for lateral reach [29], 0.83 for components of the timed stance test, 0.92 for the TUG [24], 0.88 for step test [25,26], and 0.99 for the Berg Balance Scale [28]. There may be a number of explanations for the high levels of reliability found in this study. These include the fact that the telerehabilitation system enabled the video recording of the patient performing the various tests. Therefore, the performance could be viewed multiple times, leading to a more accurate assessment. The telerehabilitation tools that were used to perform the assessments (such as the functional reach measure) have been designed to limit the measurement error that is inherent with face-to-face measurements. Because the scoring of some of the measures was performed after the consultation, when the participant was not present, more of the investigators’ attention could be directed toward the assessment. High levels of interrater reliability were also seen for face-to-face assessments for all tests (all ICC (2,1) > 0.99) in this study and may relate to the detailed training that the investigators received in administering the tests prior to the commencement of the study. This high level of training may predispose the results in this study to be slightly inflated when compared with general clinical use where specific training may not be available.
While a high level of accuracy and reliability were observed for the telerehabilitation assessments performed in the study, a number of system factors could be improved. The most significant factor effecting the quality of the telerehabilitation assessment was the low audio level experienced with softly spoken participants. Hypophonia is a symptom of PD, and at times it became difficult to hear participants and therefore direct the assessments across the Internet link. While lapel microphones were used to try to limit this problem, occasionally the participants could not be heard across the link and had to be asked to repeat themselves. This increased the time required to perform the assessment; however, it did not appear to interfere with the accuracy of results. Future studies should evaluate preamplifying the microphone at the patient end of the consultation or using higher quality microphones.

A number of limitations are present in this study that must be acknowledged. First, the inclusion of only 12 subjects restricts the generalizability of the results to the wider population of people with PD. Second, the extensive training provided in the use of the assessment tools may have inflated the reliability results reported. Future studies should also include patient populations located in rural and remote areas since these areas are the most likely to benefit from telerehabilitation interventions. This study provides evidence to support the use of telerehabilitation technologies to evaluate the physical functioning of people with PD and is a positive step in developing a means of managing these people in their own communities.

CONCLUSIONS

It is possible to perform accurate and reliable online physical assessments of people with PD using the eHAB telerehabilitation system. Improvements in audio quality could further enhance these results. Despite the small number of participants, this study provides evidence to support the use of telerehabilitation technologies to evaluate the physical functioning of remote PD clients.

ACKNOWLEDGMENTS

Author Contributions:
Study concept and design: T. G. Russell, T. C. Hoffmann.

Statistical analysis design: T. G. Russell.
Statistical analysis: M. Nelson, L. Thompson, A. Vincent.
Statistical analysis review and critique: T. G. Russell, T. C. Hoffmann.
Drafting of manuscript: T. G. Russell, M. Nelson.
Critical revision of manuscript for important intellectual content: T. G. Russell, T. C. Hoffmann, L. Thompson, A. Vincent.
Financial Disclosures: The primary author has a material role in the commercialization of the eHAB system. To maintain independence, this author was not directly involved in the collection or analysis of the data in this study. No competing financial interests exist for the remaining authors.
Funding/Support: This material was unfunded at the time of manuscript preparation.
Institutional Review: This study was approved by the University of Queensland Human Research Ethics Committee. All participants gave written informed consent.
Participant Follow-Up: The authors have no plans to notify the study participants of the publication of this article because of a lack of contact information.

REFERENCES


ResearcherID/ORCID: Trevor Russell, PhD: F-6888-2010; Tammy C. Hoffman, PhD: E-8029-2010

Submitted for publication August 13, 2012. Accepted in revised form November 28, 2012.

This article and any supplementary material should be cited as follows: