Technical Feasibility of Teleassessments for Rehabilitation

William K. Durfee, Member, IEEE, Lynda Savard, and Samantha Weinstein

Abstract—Background: technical feasibility was evaluated for conducting standard motor assessment instruments in a remote setting. Remote assessment was compared to co-located assessment for five clinical evaluation instruments: joint range-of-motion (ROM), manual muscle test (MMT), Berg sit-to-stand, Berg forward reach, and timed up and go (TUG). Methods: co-located and remote rooms were in the same building connected by broadband video and audio. Ten subjects without impairments participated, but were given simulated impairments to mimic the patient population commonly seen in rehabilitation clinics. One therapist performed all co-located testing while another performed all remote assessments. Measurements followed standard clinical methods. Data were analyzed using repeated measures ANOVA and paired t-tests. Results: no differences were found between co-located and remote assessments except for some cases using screen-based ROM measures. Remote ROM tests using snapshots and a virtual goniometer were preferred. A digital dynamometer added no additional information to a visually-based remote MMT assessment.

Index Terms—Assessment, telehealth, telerehabilitation.

I. INTRODUCTION

TELEHEALTH is experiencing rapid growth with new clinical applications and new technology products appearing daily. Telerehabilitation, a subset of telehealth, is the provision of rehabilitation services delivered at a distance using videoconferencing and other telehealth technologies. These services include evaluation and treatment, as well as education, consultation, and coordination of care. Telerehabilitation remains in the formative stage with a growing body of research but with little impact to date on clinical practice. The state of telerehabilitation has been reviewed several times [1]–[4] and has been the subject of two special issues [5], [6]. The reviews describe the promise of telerehabilitation and list a growing number of research projects, but do not cite examples where telerehabilitation is used on a regular clinical basis. This mirrors the broader picture of telehealth where a recent metareview found only a few controlled trials comparing telemedicine with face-to-face patient care, and while those studies demonstrated technical feasibility, they did not show an obvious clinical benefit for telehealth [7]. However, because there is great potential for telerehabilitation to make a significant impact on healthcare, and potential for telerehabilitation to save money [8], there is a continuing need for studies that explore new ways of delivering mainstream rehabilitation assessment and therapies using telerehabilitation technologies.

Measuring the effectiveness of rehabilitative intervention depends on outcome assessment tests that are reliable and valid. Assessment tools are readily available to rehabilitation clinicians in the traditional on-site setting. However, one may not assume that standard assessment measures can be applied effectively using telehealth methods. To date, only a few studies have investigated standard assessment tests applied to telerehabilitation. In one pilot study, the Kohlman Evaluation in Living Skills and the Canadian Occupational Performance Measure were reported to have agreement when completed in-person and remotely [9]. The National Institutes of Health stroke scale was shown to have good reliability when administered remotely [10]. Internet-based assessment of motor speech disorders, including the Frechay dysarthria assessment and the assessment of intelligibility of dysarthric speech instruments were shown to generally agree with face-to-face assessments [11]. Measuring knee angle remotely from captured still pictures using a computer program that calculated joint angles showed good agreement with measuring knee angle in person with a universal goniometer [12].

Teleassessment could be an effective method for conducting functional assessments on remote populations and on patients for whom transportation is a significant barrier. A valid, reliable teleassessment service would be particularly valuable for patients who need periodic assessments to determine treatment progress or to determine if they qualify to be admitted into a specialized rehabilitation program. The long range goal of this research is to determine if standard assessment instruments used by physical therapists can be conducted with the patient located far from the therapist. The specific objective of this study was to investigate the technical feasibility of teleassessment by determining whether the results of standard assessments of motor function conducted remotely are the same as those conducted in-person.

There are hundreds of methods and instruments used to assess motor function [13]. A small number of standard assessment instruments were selected to investigate. Criteria for selection required that the instrument be: 1) a published measurement tool, 2) reliable and valid, 3) used widely by physical therapists, and 4) supported by standard instructions for administration and scoring.
II. METHODS

A. Subjects

Ten healthy subjects (two male, eight female, ages 18–35) participated as simulated patients. An additional ten healthy subjects (four male, six female, ages 18–35) participated as simulated caregivers and were present in the examination room during remote assessment sessions. Subjects were assigned randomly to either the simulated patient or simulated impairment group. Subjects were given the simulated impairment described in the following section with the level of simulated impairment different for each subject. Two licensed physical therapists each with over 10 years of clinical experience conducted the assessments. One completed all co-located measurements and the other conducted all remote measurements. The project was approved by the appropriate Institutional Review Boards, and informed consent was obtained from subjects.

B. Remote Assessment Technology

The study was conducted in one facility with the remote location being a different room in the same building (Fig. 1). For the remote assessments, the evaluating therapist was in one room (remote site), connected to the other room (co-located site) by a videoconferencing system (Polycomm, ViewStation) that provided two-way, high-quality voice and video over a broadband, T1 network. The therapist could remotely pan, tilt, and zoom the camera on the patient’s end. A video capture unit (Hauptauge, USB-Live) captured the video image on the therapist’s computer for taking and storing snapshots. A custom digital dynamometer was connected to the serial port of the local PC through a sensor interface unit. As the caregiver pressed on the patient’s limb with the dynamometer, a bar graph on the remote computer indicated the force level in real-time.

C. Assessments

Five assessments were studied, range-of-motion (ROM) of the shoulder and knee, manual muscle test (MMT) of the biceps and quadriceps, item 1 (sit-to-stand) and item 8 (forward reach) components from the Berg balance test, and the timed up and go (TUG) (Fig. 2). The selection included hands-on and hands-off assessments and required a range of instruction complexity. Each assessment was conducted using a traditional co-located method, and again using one or more teleassessment methods. For the remote assessments, the therapist gave scripted, verbal instructions to the patient and caregiver on how the test should be performed.

ROM assessment followed the methods in [14]. Joint motions assessed were shoulder abduction, shoulder external rotation and knee flexion. The joint angle was set to a randomly assigned position with a measurement jig anchored to sites proximal and distal to the joint being measured Fig. 2(a). The jig ensured consistency of joint position. Angles were randomly selected as a percentage of the subject’s full available, pain-free range in increments of 10%. Three shoulder abduction, two shoulder external rotation, and two knee flexion angles were selected. The joint positions were different for each subject but the same for co-located and remote measuring.

Co-located joint position was measured by the therapist using a universal goniometer (Jamar EZ Read). Four methods of remote ROM assessment were implemented. For methods 1 and 2, the therapist instructed the caregiver on where to place the goniometer while observing via videoconferencing. For method 1, the physical therapist (PT) zoomed the camera and read the goniometer directly. For method 2, the caregiver read the number off the goniometer and reported to the therapist. For methods 3 and 4, the PT captured the image from the camera for offline analysis. Capturing the image meant the subject did not have to hold still during goniometer manipulation. For method 3, the joint angle was determined using a virtual goniometer image processing program developed for the study (Fig. 3). Video snapshots were taken by the therapist of the subject and the snapshots were later imported into the angle analysis program where the therapist could click on the subject’s bony landmarks to position a virtual goniometer. When the therapist was satisfied with goniometer placement, the program calculated the joint angle. For method 4, the PT held a universal goniometer up to the image on the computer screen. Both physical therapists independently measured the angles in methods 3 and 4. Joint angles were measured in degrees rounded to integers.

In-person biceps and quadriceps MMT assessment followed the methods in [14]. Two methods of remote MMT assessment were implemented. For both, the therapist observed the testing through the videoconferencing system and instructed the subject and caregiver on how to conduct the test. For method 1, the
therapist scored the MMT based solely on observing the subject push against the caregiver. Method 2 added a digital dynamometer which the caregiver placed between his hand and the tested limb so that the force exerted by the patient appeared as a moving bar on the therapist’s computer. MMT was scored from 0 to 5 according to standard procedure [14], including the assignment of plus and minus. In the subsequent analysis, a plus added 0.33 to the integer score and a minus subtracted 0.33. Subjects were provided with a simulated impairment for co-located and remote MMT assessment. Weight was added distally to mimic muscle weakness. For biceps testing zero to 60 lb were added to the distal forearm and for quadriceps testing zero to 70 lb were added just proximal to the ankle. The amount of attached weight was randomized across subjects, but for each subject the same weight was used for co-located and remote assessments.

The Berg balance scale is a comprehensive, objective measure of balance abilities [15]–[17]. Only two of the 14 Berg items were measured to save time, but were representative of the assessment techniques needed for the full Berg test. The methods described by Berg [15] were followed. For remote assessment, the therapist provided instruction to the subject and observed the results. Item 1, sit-to-stand, was scored on a 0–4 scale following Berg instructions. Subjects were given a simulated balance impairment by standing on one or two 14 in air-filled rubber disks (Exertools Dynadisc, Novato CA). The inflation pressure was adjusted and the disks were stacked in one or two to adjust the difficulty of maintaining balance. For Item 8, forward reach, subjects reached parallel to a board with vertical rule lines that could be read by the remote therapist using the camera pan and zoom. Reach was recorded in inches and was not converted into a 0–4 Berg score. For forward reach, the first trial was with no disk and the second with one disk.

The timed up and go test measures the time to rise from a chair, walk 10 ft, return and sit, and is a quick, practical method for testing basic mobility [18], [19]. The TUG methods in [18] were followed. Subjects were given a simulated impairment by walking on top of a 12-ft-long, 4-in-wide balance beam whose top surface was 5 in from the floor. Timing was accomplished with a stopwatch. Remote assessment was performed by therapist observation.

D. Protocol

Testing for a subject was done in a single session that lasted about 90 min. Prior to data collection, the subject practiced the tests with the imposed impairments to minimize learning effects. For the first five subjects, the co-located tests were conducted before the remote tests while for the remaining five the order was reversed to avoid order effects. A consistent test order was maintained so that learning or fatigue effects would impact co-located and remote assessments equally. The test order was
as follows: 1) MMT R quads with first weight; 2) MMT L quads with second weight and digital dynamometer; 3) MMT R biceps with first weight; 4) MMT L biceps with second weight and digital dynamometer; 5) Berg #1 on one disk; 6) Berg #1 on two disks; 7) Berg #8 unimpaired; 8) Berg #8 on one disk; 9) TUG on balance beam; 10) repeat TUG on balance beam; 11) L shoulder abduction in three positions while sitting facing the camera and the arm moving in the coronal plane with the elbow straight; 12) L shoulder external rotation in two positions with the subject supine and the elbow at 90°; 13) L knee flexion in two positions Fig. 2(a).

To control repeatability of the angle setting for the ROM measurements, co-located and remote assessments were done without the subject changing position between trials. The joint angle was first set by the experimenter using the jig. The local therapist measured the angle with the goniometer, then left the room. The remote therapist enabled the videoconferencing system, adjusted the camera and took a photo of the joint. Next, the remote therapist instructed the caregiver on how to place the goniometer and then zoomed in to read and record the joint angle. The remote therapist asked the caregiver to verbally report the goniometer reading which was recorded. The process was repeated for the next angle. The joint position photographs were measured off line with both therapists separately conducting the two on-screen measurement methods. Data were collected from subjects 5–10 for the ROM protocol. The first four subjects followed a different protocol where the joint angle was not guaranteed to be the same for all seven measurement methods resulting in unreliable data that could not be analyzed. The data from subjects 1–4 were included in the analysis of the other four assessments.

E. Data Analysis

The ROM experiment was a single factor design with seven levels: (M1) co-located, (M2) remote with caregiver reporting goniometer setting, (M3) remote with therapist zooming in and reading the goniometer directly, (M4) remote photo with angle measured by one therapist with a virtual goniometer, (M5) remote photo with angle measured by the second therapist with a virtual goniometer, (M6) remote photo with angle measured by one therapist using a universal goniometer held up to the snapshot, (M7) remote photo with angle measured by the second therapist using a universal goniometer. The data were analyzed separately for each joint as a repeated measures, within-subjects design [20], [21] with one block being one joint setting in one subject resulting in 18 blocks for shoulder abduction, 11 blocks for shoulder rotation (one block was discarded because one data point was erroneous) and 12 blocks for knee flexion. The null hypothesis was that measurements of a single joint position with the seven different measurement methods did not differ significantly. An alpha of 0.05 was used for the critical F value in the ANOVA analysis. If the analysis showed a significant difference, a post-hoc Tukey’s multiple comparison [21] was used to determine which pairs of methods differed.

Three MMT methods were compared: co-located, remote with dynamometer assist, and remote by observation only. Because only two methods were used for any one muscle and subject, the data were analyzed by paired t-tests comparing co-located with remote-dynamometer and co-located with remote-observation. The data were also analyzed by a Wilcoxon signed-ranks test. Berg sit-to-stand, Berg reach, and TUG were analyzed with paired t-tests comparing co-located and remote scores. For TUG, trials 1 and 2 were averaged before performing the t-test.

III. RESULTS

The pooled ROM data for shoulder abduction showed a significant difference between the measurement methods (F = 5.49, Fcrit(0.05) = 2.19, p = .00), as did the data for knee flexion (F = 4.37, Fcrit(0.05) = 2.24, p = .00), while the data for shoulder rotation showed no significant difference between measurement methods (F = 0.43, Fcrit(0.05) = 2.25, p = .86). Tukey’s multiple comparison procedure showed that for the shoulder abduction tests, M6 differed from M1, M3, and M5, and M7 differed from M1, while for the knee flexion test M5 differed from M2, M3, and M6. A power analysis [21] showed the power of the experiment to detect a clinically significant angle measurement method difference of 5° was 100% for all three joints and the power to detect a difference of 2° was 90%, 95%, and 70% for shoulder abduction, shoulder rotation and knee flexion.

To determine if one method had a consistently high or low bias, the difference of the seven methods about the mean for each angle setting on each subject was computed (Fig. 4). The plot shows no consistent trends or patterns for any particular method. The seven methods had an average deviation from the mean between −3.3 and 4.5° and approximately equal scatter with a standard deviation about the mean between 2.2° and 5.32°.

One issue was whether there was a difference in goniometer angle readings based on whether the caregiver or the remote PT performed the reading. A paired t-test for all 10 subjects showed...
there was no significant difference between the two methods, 
\[ t = 1.15, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 1.99, p = .25. \]
For 68 of the 69 observations, the two methods differed by 0 or 1°. For one observation, the methods differed by 2°. Another issue was whether using a virtual goniometer was different than holding a universal goniometer to the screen for remote assessment of digital snapshots taken during the test session. A paired \( t \)-test showed no significant difference between the methods when the pairing was blocked for the therapist, \( n = 83, t = 1.01, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 1.98, p = .27. \) and over all observations, there was a mean absolute difference between the two methods of 4.4 degrees (\( SD = 3.6 \)).

MMT data were pooled across the 10 subjects and all applied weights. Paired \( t \)-tests showed no significant difference between co-located assessment and remote assessment by visual observation, \( t = .21, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.09, p = .83, \) and no significant difference between co-located assessment and remote assessment by visual observation augmented with the digital dynamometer, \( t = .39, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.09, p = .09. \) The power of the test to detect differences in MMT scores of .33 (a plus/minus increment) and 1 was 60% and 100% for co-located versus remote observation, and 51% and 100% for co-located and remote dynamometer. The Wilcoxon Signed-Ranks Test showed no difference between co-located and remote assessments.

Berg sit-to-stand data were pooled across all subjects. Data from trial 1 of each subject were not analyzed as all Berg scale scores were 4 which indicates the activity was performed safely and independently by the subject. For trial 2, 13 of 20 observations scored a 4 with some differences between remote and co-located for the rest. A paired \( t \)-test of the trial 2 data showed no significant difference between co-located and remote measures, \( t = 2.12, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.26, p = .06. \) Four subjects received scores in trial 2 that differed between co-located and remote. Of these, three had a remote score of zero which meant the subject stepped off the disc during the remote observation.

Berg forward reach data were pooled across all subjects and both trials. A paired \( t \)-test showed no significant difference between co-located and remote methods, \( t = 1.18, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.09, p = .85. \) The power of the test to detect a difference of one inch and two inches was 48% and 95%. Over all the data, the average absolute value of the difference between co-located and remote measures was 1.8 inches (\( SD = 1.5 \)). The absolute difference for trial 2 data, for which subjects stood on an air disk for a balance impairment, (\( M = 2.2, SD = 1.8 \)) was almost double that of trial 1 (\( M = 1.4, SD = 1.2 \)). In trial 1 subjects reached an average of 15.5 in but only 11.3 for trial 2, a significant difference, \( t = 14, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.26, p = .00. \)

For TUG, a paired \( t \)-test of the averaged trials showed no significant difference between co-located and remote methods of timing, \( t = 1.37, \text{trid}(0.05, 2 \rightarrow \text{tail}) = 2.26, p = .20. \) The power of the test to detect differences of 1 s and 2 s was 32% and 85%. When the data were resorted by observation order, they showed that most subjects improved their TUG score on the second trial, demonstrating the existence of a learning effect.

### IV. Discussion

#### A. ROM

Remote measurement of joint angles is feasible as the results showed that with some exceptions, remote measures did not differ from co-located measures. For shoulder abduction and knee flexion, M5 differed from M6 which means there was a difference between one therapist holding a goniometer up to a screen shot and the second therapist using the virtual goniometer on the screen shot. The difference is likely interrater because the secondary analysis showed that virtual and real goniometer yielded no difference when compared within the rater. For shoulder abduction, M6 also differed from M3, M5, and M7 and for knee flexion, M5 also differed from M2 and M3 which implies that additional training may be required for the screen shot methods.

Caregivers can be instructed to place and read a universal goniometer. The method of reading the goniometer by zooming was less useful as it required a pan/tilt/zoom camera with good optics, and the process of zooming is time consuming. The method of taking a snapshot for later measurement is technically feasible, assuming that the camera can be positioned so that its axis aligned with that of the patient’s joint to avoid parallax errors. Subjects sat or lay on a motorized plinth that could be raised and lowered for optimum positioning for viewing by the camera. Motorized hi-lo mats are not available in the home which will make camera positioning more challenging.

The therapists reported no difficulty in identifying anatomic landmarks for angle measurements from digital snapshots. Although using real and virtual goniometers to measure the photos were equivalent within therapists, the virtual was easier to use and required fewer instructions to the patient. Identifying anatomic landmarks in obese patients will be more challenging.

#### B. MMT

The remote therapist was able to score manual muscle tests reliably based on observation alone. MMT scores are differentiated by the ability to maintain position against “maximal,” “moderate,” and “minimal” resistance by the therapist [14] which could be detected by the therapist by observing the patient and the caregiver to estimate exertion. While the digital dynamometer did not appear to impact the ability to measure, it may give the remote therapist a sense of being more in control and eliminates the need for the therapist to estimate the resistance force being applied by the caregiver.

#### C. Balance

Scoring of the Berg Item #1 (sit-to-stand) is based on visual observation, and the resulting close match between co-located and remote scores was expected. All of the subjects were able to stand quietly on a single disk which explains why all trial 1 scores were the same for remote and local assessment. The difficulty with making inferences from the second set of trials is that there was no guarantee that the remote and local therapists were observing the same performance as standing on two balance disks is extremely difficult. Thus, variations in the sit-to-stand scores most likely resulted from trial-to-trial variation.
Video observation was effective in assessing Berg Item #8 (Functional Reach). While a 1.8 in difference in reach can be clinically significant, the therapists reported that the remote measuring scheme was accurate so that the differences seen was likely caused by subject trial to trial variation. The larger variation of trial 2 data resulted from subjects being given a balance impairment which likely resulted in a larger variation in their performance between trials. Additional technology could eliminate the ruled board and camera zooming.

D. TUG

It was relatively easy for the remote therapist to observe and time the TUG after the camera was oriented for a good view of the chair and walkway. The power to detect differences of one second was low but this was because the subjects walked faster on each successive trial as they learned how to balance on the beam.

E. Communication Bandwidth

A high quality, real-time audio link was essential for communicating instructions between the therapist and the caregiver and patient. A one-way video link was needed so that the therapist could visually confirm that instructions were being carried out and to record observations based on assessment measures. Video in the other direction was not required but may help in patient interactions.

The quality of the teleconferencing video and audio impacts the accuracy of remote assessment [22]. This study used ideal television quality audio and video and a pan/tilt/zoom camera. For remote assessments done in the home, the quality of the video will degrade based on what the phone lines can transmit unless the home has wired or wireless broadband access. The effect of video quality on remote measurement methods is unknown. For example, the administration of the gait assessment rating scale has been reported to be as accurate using store-and-forward video clips transmitted at 128 kbps and 18 kbps compared to the full-resolution video camera clips [23].

F. Technology

Other than video conferencing, little technology was used. More technology could easily be added to support more accurate measurement while keeping the essentials of the measurement instrument the same. The digital dynamometer is one example. Recording the data in pounds from the digital dynamometer might help the therapist to assign an MMT score, but the actual resistance in pounds is only a part of the MMT score. It might help the therapist to assign an MMT score, but the actual resistance in pounds is only a part of the MMT score. The digital dynamometer is one example. Recording the data in pounds from the digital dynamometer might help the therapist to assign an MMT score, but the actual resistance in pounds is only a part of the MMT score. If this information can be gathered through observation, adding a digital force reading in pounds could assist the therapist in assigning a 0–5 MMT score.

A pressure sensor on the chair could detect the start and end of a TUG test, and a range sensor could record forward reach. While these sensors could be small, light, wireless and easy to use, they would require additional training for the therapist and would increase system cost. Whether this would lead to faster, simpler, and more reliable remote assessments remains to be seen.

G. Limitations

Subjects were nonimpaired and their simulated impairments were clearly different from the impairments seen in actual patients. For example, subjects with simulated impairments could fall off the Dynadiscs with no consequences while elderly patients with balance impairments must be caught by the caregiver. Subjects acting as patients and caregivers were highly educated and had no trouble following instructions which may not be the case for real patients and their caregivers, particularly for patients with cognitive impairments. For the purpose of the study, the simulated impairments were successful in eliciting the range of assessment scores that would be expected in a clinical population. Subjects were not in their own homes which may have changed their behavior and performance. Trial-to-trial variation in task performance may have contaminated the results. It was not possible to conduct the co-located and remote assessments simultaneously as therapist instructions are part of the comparison and each therapist must be blinded to what the other is doing. Inter-rater reliability for teleassessment must be determined. In this study, one physical therapist conducted all of the co-located assessments, and a second conducted all of the remote assessments. A future study should include several therapists to measure reliability. The small number of subjects limited the power to detect Type II errors, however, it was adequate to investigate technical feasibility of the teleassessment methods. Cost was not addressed although cost has a dominating influence on reimbursement and adoption. A financial feasibility analysis would be complicated by the rapidly changing costs of technology and broadband communications, and changing attitudes towards reimbursing teleconsults.

V. Conclusion

Remote rehabilitation assessment using traditional assessment instruments is technically feasible. Assessments administered remotely were conducted without difficulty and yielded approximately the same results as when administered locally. A high quality, two-way audio link, as least as good as a standard telephone connection, is required for effective verbal communication. A clinical study with rehabilitation patients is needed to determine the efficacy and reliability of the test instruments with a patient population and additional work is needed to determine the inter-rater reliability of the screen shot angle measuring methods. No conclusions can be drawn from this study about whether teleassessment can be done safely at home with only the patient and caregiver present.

REFERENCES


