Title: FACE-TO-FACE VERSUS VIDEO ASSESSMENT OF FACIAL PARALYSIS: IMPLICATIONS FOR TELEMEDICINE

Authors:

1. Jian Rong Tan  
   BAppSc (Physiotherapy) (Honours)  
   Faculty of Health Sciences  
   University of Sydney

2. Dr Susan Coulson  
   PhD (Physiotherapy)  
   MAppSc (Exercise and Sport Science)  
   BAppSc (Physiotherapy)  
   University of Sydney

3. Dr Melanie Keep  
   PhD (Psychology)  
   MEd (Higher Education)  
   BPsysch (Honours)  
   University of Sydney

Correspondence: Name: Dr Susan Coulson  
Department: Physiotherapy  
Institution: University of Sydney  
Address: The University of Sydney  
Cumberland Campus  
75 East Street, Building O  
Lidcombe, NSW 2141  
Country: Australia  
Phone: +61 2 9351 9416  
Fax: +61 2 9351 9278  
Email: susan.coulson@sydney.edu.au

Abbreviated title: Implications for telemedicine in FNP

Word Count: 294 words (Abstract)
Disclosure statement: None of the authors have a financial interest or received financial support in relation to the generation of this submission.
ABSTRACT

Background: Patients with facial nerve paralysis (FNP) experience challenges to accessing healthcare that could potentially be overcome by telemedicine. The reliability of telemedicine, however, has yet to be established in this field.

Objective: This study aimed to investigate the consistency between face-to-face and video assessment of patients with FNP by experienced clinicians.

Methods: A repeated-measures design was used. Seven clinicians assessed the FNP of 28 patients in a face-to-face clinic using standardised grading systems (the House-Brackmann, Sydney, and Sunnybrook Facial Grading Systems). Three months later, the same grading systems were used to assess facial palsy in video recordings of the same patients.

Results: The House-Brackmann system in video assessment had excellent reliability and agreement (ICC=0.780; PCA=87.5%), similar to face-to-face assessment (ICC=0.686; PCA=79.2%). Reliability of the Sydney system was good to excellent, with excellent agreement face-to-face (ICC=0.633 to 0.834; PCA=81.0% to 95.2%). However, video assessment of the cervical branch and synkinesis had fair reliability and good agreement (ICC=0.437 to 0.597; PCA=71.4%), while that of other branches had good to excellent reliability, and excellent agreement (ICC=0.625 to 0.862; PCA=85.7% to 100.0%). Reliability of the Sunnybrook system was poor to fair for resting symmetry (ICC=0.195 to 0.498; PCA=91.3% to 100.0%), and synkinesis (ICC=-0.037 to 0.637; PCA=69.6% to 87.0%), but was good to excellent for voluntary movement (ICC=0.601 to 0.906; PCA=56.5% to 91.3%) in face-to-face and video assessments. Bland-Altman plots indicated normal limits of agreement within +/-1 between face-to-face and video-assessed scores only for the temporal and buccal branches of the Sydney system and for resting symmetry in the Sunnybrook system.

Conclusions: Video assessment of FNP with the House-Brackmann and Sunnybrook systems was as reliable as face-to-face but with insufficient agreement. However, video assessment does not account for the impact of real-time interactions which occur during tele-assessment sessions.

KEYWORDS
Telemedicine, e-Health, Facial Paralysis, Facial Nerve, Outcome Assessment, Video Recording, Facial Expression, Smiling, Quality of Life, Bell Palsy
INTRODUCTION

Facial nerve paralysis (FNP) occurs when there is damage to the seventh cranial nerve (CN VII), the facial nerve. This affects an individual’s ability to express emotions via facial movements (e.g. frowning, smiling, or pouting) thereby compromising interpersonal communication [1-4]. The effects of FNP are not only limited to physical and aesthetic deficits caused by weakness, synkinesis (involuntary muscle movements accompanying voluntary movements) and asymmetry of the resting and moving face [1-3, 5-10]. FNP is also associated with functional effects such as oral incontinence, speech deficits, dry eye and subsequent corneal damage, and psychosocial effects such as decreased self-esteem, psychological distress, depression, and reduced quality of life [1-3, 5-10]. It is therefore desirable that individuals with FNP receive appropriate help for their condition to reduce deficits and improve functional outcomes.

Management of facial nerve disorders is a highly specialized field with a limited number and availability of specialists. For example, in Australia, there is only one major facial nerve centre – the Sydney Facial Nerve Service. In addition to limited access to relevant health professionals, individuals with FNP face common challenges to healthcare access, which include the direct and indirect cost of care, distance, transportation, mobility, and time constraints [11-15].

One potential way of improving access to care for patients who experience FNP is telemedicine. Telemedicine is healthcare delivered via telecommunications technology over a distance [11, 15-18], allowing improved access to care in various physical and mental health conditions by overcoming geographical and temporal barriers [11, 13-21]. Telemedicine has also been associated with increased access and compliance to pulmonary rehabilitation [22], and reduced hospitalization rates, emergency department visits, and length of hospital stay for patients with chronic obstructive pulmonary disease (COPD) [23]. Patients have reported similar or better quality of life compared to usual care [23]. In addition, Internet-based interventions have been found to significantly decrease pain intensity, depression, anxiety and stress, and increase quality of life compared with in-person psychotherapy in patients with medically unexplained pain [20], in youth with depression and anxiety [21], in military personnel with PTSD and depression [17, 24], and in patients with neurofibromatosis [15].
Limitations of telemedicine include technical difficulties, such as poor video resolution, connectivity and sound issues [17, 18, 20, 21, 25]. These could potentially compromise the quality of the service provided, and are commonly cited reasons for not implementing tele-health [17, 18, 20, 21, 25].

Although telemedicine has not yet been formally studied in FNP, it has, anecdotally, been implemented by some clinicians to provide services for patients experiencing geographical barriers to treatment. The above evidence suggests that telemedicine may play a role in increasing access to medical care for patients with FNP. As such, studies are required to establish the reliability of telemedicine as compared to face-to-face healthcare of FNP.

Tele-assessment of facial function using the appropriate facial grading systems is an integral part of management of FNP since FNP management requires initial and ongoing assessment of patients’ facial function. The clinician’s overall impression of the patient is taken as standard and is assumed to represent the degree of abnormality of facial function [1, 2, 4-6, 8, 26-36]. This enables clinicians to keep track of patients’ recovery and evaluate the efficacy of treatment [4, 28, 31, 37]. FNP is assessed using grading scales that typically measure patients’ facial function, grade symmetry of the face at rest, and measure displacement of facial features during voluntary movement, as well as the presence of synkinesis. Previous research has shown that greatest displacement during facial expressions occurs in the vertical axis of the frontal plane, followed by the anterior-posterior axis of the sagittal plane and the horizontal axis of the frontal plane [27]. This raises the question of whether tele-assessment of the patient, which would present the patient in a two-dimensional (2D) view, would be as reliable and valid as a face-to-face assessment, which presents the patient in a three-dimensional (3D) view.

Despite the significant body of literature on the applications of telemedicine, there has been a lack of research conducted on its use for the management of FNP. This exploratory study aims to compare the reliability of three commonly used FNP grading systems when administered via video and face-to-face. Face-to-face assessment presents the patient in a 3D view, while the frontal video shows the patient in a 2D view that would not take into account the anterior-posterior axis of the sagittal plane. Therefore, it was hypothesised that reliability across the three grading systems would differ between the face-to-face and video assessments.
in movements where there was significant anterior-posterior axis displacement, for example, the pout, smile, and snarl.

METHODS
Study design
A repeated measures design was undertaken, in which the face-to-face and video assessment were conducted on the same patients and assessed by the same assessors. There was a three-month delay between the face-to-face and video assessment to minimize retention of gradings used by the assessors. The three-month period was based on a similar study by Banks et al. [31] which compared in-person and video assessment of facial mimetic function using the newly developed eFACE facial grading system.

Participants
Participants comprised clients at the Sydney Facial Nerve Clinic (SFNC) who presented with a diagnosed FNP between August 2016 and March 2017. Assessment was conducted in a standard, routine manner and the video assessment segment of the study did not require additional involvement by the participants, nor did it preclude them from receiving treatment or other appropriate interventions. Ethical approval was granted by the Ethics Review Committee (RPAH Zone) of the Sydney Local Health District (Protocol No X17-0013 ERC).

Procedure
Participants were assessed face-to-face at the SFNC between August 2016 and March 2017. Videos in the frontal view of the patients performing a protocol of facial movements which included eyebrow raise, eye closure, smile, snarl, and pucker were recorded on the same day. Three months later, each assessor rated the video footage taken at the initial face to face session. The order in which the videos were shown was randomized to control for rater fatigue. Each assessor graded each patient face-to-face with the House-Brackmann system, the Sunnybrook system, and the Sydney system, and then again in the video assessment.

How the assessors graded the participants in the face-to-face assessment was the standard against which the video assessment was compared. The reliability of video assessment was examined against face-to-face assessment to enable inferences to be made about tele-assessment potentially substituting for face-to-face assessment.
**House-Brackmann Facial Grading System**
The House-Brackmann Facial Grading System is the most commonly used system to quantify FNP [2-5, 8, 26-28, 30, 31, 33, 34, 37]. It is a gross scale ranked from 1 to 6, with each grade giving an overall impression of facial function, resting symmetry, and synkinesis [4, 5, 28, 34]. However, the House-Brackmann system is limited in that it does not allow for regional assessment, the range of scores does not reflect clinically important change and lacks sufficient classification of synkinesis [8, 26, 30, 32, 34-36]. This has led to the administration of a range of alternative systems being used by practitioners around the world.

**Sunnybrook Facial Grading System**
One such system is the Sunnybrook Facial Grading System which is a regional weighted system that assesses resting symmetry, voluntary movements, and synkinesis of the face, after which a composite score on a 100-point scale is computed [28, 34]. It grades patients in a more objective and continuous manner than the House-Brackmann system, and each component of the system is sensitive to change and contributes to a change in the composite score [1, 3, 8, 26, 28, 30, 32, 34, 35].

**Sydney Facial Grading System**
The Sydney Facial Grading System, which assesses voluntary movement of the five branches of the facial nerve and overall synkinesis, has also been used regionally in Australia [1, 6, 26] and reported in International Facial Nerve Symposium Conference Proceedings over a 20-year period.

**Analyses**
Participants’ data was de-identified before analyses. Quantitative analyses were conducted using the Statistical Package for the Social Sciences (SPSS). Descriptive statistics were used to summarise demographic information. A significance level of \( p \leq 0.5 \) and corresponding 95% confidence intervals (CI) were used for all inferential statistics.

Agreement between face-to-face and video-assessed House-Brackmann grades, Sydney scores, and Sunnybrook scores was assessed using the one-way random, single measure intraclass correlation coefficient (ICC \([1,1]\)) and two-tailed repeat measures t-tests. ICC \((1,1)\) was selected because each participant was assessed by a different set of assessors. For each
participant, scores given by three randomly selected assessors were used to compute the ICC, hence the rater was a random effect [38, 39]. Portney and Watkins [38] noted that “ICC ranges between 0.00 and 1.00, with values closer to 1.00 representing stronger reliability”, but there were “no standard values for acceptable reliability using the ICC”, and that the researcher should determine the level of reliability needed to justify the use of the tool being assessed. Based on the study by Banks et al. [31], we defined reliability as poor for ICC values less than 0.40, fair for values between 0.40 and 0.59, good for values between 0.60 and 0.74, and excellent for values 0.75 and above. The 95% CI indicates that there is 95% chance that any score would lie within the range [39].

As the ICC could be affected by outlying data, the percentage exact agreement (PEA) and percentage close agreement (PCA) were calculated to make up for the potential deficiency [40]. PEA is the proportion of cases where all three assessors gave the same grade for each participant. PCA is similar to PEA but includes cases where there was a maximum difference of one grade between assessors [40]. Agreement of PEA with ICC suggests the data is sound [40]. However, as a one-grade and 17-point difference between assessors had been determined to be reasonable for the grading systems and the Sunnybrook composite score respectively [4, 32], we will be comparing the ICC with the PCA instead. Similar to the ICC, agreement was defined as poor for PCA values less than 40%, fair for values between 40% and 59%, good for values between 60% and 74%, and excellent for PCA values 75% and above.

Repeat measures t-tests were administered on a per-protocol analysis. Missing data (i.e. no matched pairs of face-to-face and video assessment grades) was discarded.

Bland-Altman plots were constructed where differences between face-to-face and video assessment were plotted against the means of the two assessment modes to determine agreement between the two and whether video assessment can substitute face-to-face assessment with an acceptable degree of error [38, 41, 42]. Horizontal lines were drawn at the mean difference, and at the upper and lower limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation (SD) of the differences respectively [38, 41, 42]. Assuming normal distribution, between the limits of agreement is an agreement interval within which 95% of the differences of video assessment, compared to face-to-face assessment, fall [38, 43]. The Bland-Altman method only defines the intervals of
agreement but not whether those limits are acceptable [43]. We have therefore defined a
difference of 1 grade to be acceptable between House-Brackmann grades and components of
the Sydney and Sunnybrook systems, and a difference of 17 points in the Sunnybrook
composite score.

RESULTS
Demographics
Twenty-eight participants and seven assessors were invited to take part in the study by
provision of participant information statements. The mean (SD) age of the 28 patient
participants (12 males and 16 females) was 41.7 (15.4) years, with a mean (SD) of 6.5 (6.9)
years since FNP onset. There has been no prior research comparing face-to-face and video
assessment of FNP so the sample size (n=28) was based on a related study with 21
participants where there was sufficient power to detect differences between the House-
Brackmann, Sydney and Sunnybrook systems [26]. The etiologies of the FNP are presented
in Table 1.

Table 1: Etiology of FNP in study participants

<table>
<thead>
<tr>
<th>Diagnoses</th>
<th>n (%)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trauma/ Iatrogenic</td>
<td>13 (46%)</td>
<td>Removal of acoustic neuromas, parotid gland tumours, hemangioma, basal cell carcinoma</td>
</tr>
<tr>
<td>Bell’s Palsy</td>
<td>5 (18%)</td>
<td>Idiopathic</td>
</tr>
<tr>
<td>Tumour</td>
<td>4 (14%)</td>
<td>Facial nerve neuroma</td>
</tr>
<tr>
<td>Herpes Zoster Oticus</td>
<td>3 (11%)</td>
<td>Viral (Ramsay Hunt Syndrome)</td>
</tr>
<tr>
<td>Congenital</td>
<td>2 (7%)</td>
<td>Developmental</td>
</tr>
<tr>
<td>Trauma/ Accidental</td>
<td>1 (4%)</td>
<td>Gunshot wound</td>
</tr>
</tbody>
</table>

The assessors comprised ear, nose and throat (ENT) specialists, head and neck surgeons, and
plastic surgeons who regularly attended the SFNC and had many years of experience in
managing facial nerve disorders.

Face-to-Face vs Video-Assessed House-Brackmann Grades
The House-Brackmann system administered face-to-face had good reliability and excellent
agreement, while that administered over a video had excellent reliability and agreement
(Table 2). Significant between-subject effects denoted by significant p-value in an F-test with
true value 0 are indicated with an asterisk (*) in the tables. When this effect is significant ($p<.05$), participants are then different from each other, which is a necessary condition for reliability testing [38].

Table 2: Reliability and agreement of House-Brackmann Scale across assessment modes

<table>
<thead>
<tr>
<th>Assessment Mode</th>
<th>ICC (95% CI)</th>
<th>PEA (%)</th>
<th>PCA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face-to-Face</td>
<td>0.686 (0.488 to 0.835)*</td>
<td>41.7</td>
<td>79.2</td>
</tr>
<tr>
<td>Video</td>
<td>0.780 (0.623 to 0.889)*</td>
<td>25.0</td>
<td>87.5</td>
</tr>
</tbody>
</table>

(*F-test for between-subjects effect significant at $p<.05$)

The Bland-Altman plot indicated that the mean difference between House-Brackmann grades given in the face-to-face and video assessment was 0.05, but limits of agreement were between -1.79 to +1.89.

**Face-to-Face vs Video-Assessed Sydney System Scores**

Reliability in assessing the five branches of the facial nerve face-to-face using the Sydney system was good to excellent, with excellent agreement (Table 3). However, video assessment of the cervical branch and synkinesis had only fair reliability and good agreement while video assessment of the other branches had similarly good to excellent reliability and excellent agreement.

Table 3: Reliability and agreement of Sydney System facial nerve branches across assessment modes

<table>
<thead>
<tr>
<th>Sydney System Facial Nerve Branches</th>
<th>ICC (95% CI)</th>
<th>PEA (%)</th>
<th>PCA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal</td>
<td>Face-to-Face</td>
<td>0.789 (0.624 to 0.900)*</td>
<td>052.4</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>0.857 (0.734 to 0.934)*</td>
<td>66.7</td>
</tr>
<tr>
<td>Zygomatic</td>
<td>Face-to-Face</td>
<td>0.633 (0.402 to 0.813)*</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>0.625 (0.392 to 0.809)*</td>
<td>42.9</td>
</tr>
<tr>
<td>Buccal</td>
<td>Face-to-Face</td>
<td>0.834 (0.695 to 0.922)*</td>
<td>61.9</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>0.862 (0.742 to 0.974)*</td>
<td>61.9</td>
</tr>
</tbody>
</table>
Inspection of the Bland-Altman plots showed that the mean differences between face-to-face and video assessment were close to zero when assessing all facial nerve branches using the Sydney system. However, limits of agreement between the two modes of assessment were only close to +/-1 in assessment of the temporal and buccal branches, and a patient could potentially be graded more than 1 grade apart between the face-to-face and video assessment of the zygomatic, marginal mandibular, and cervical branches and synkinesis.

**Face-to-Face vs Video-Assessed Sunnybrook Scores**

The reliability of face-to-face assessment using the Sunnybrook system was similar to video assessment across most parameters. The reliability was generally poor to fair when assessing resting symmetry and synkinesis, but was good to excellent when assessing voluntary movement (Table 4). Agreement was generally good to excellent across parameters for both face-to-face and video assessments. The Sunnybrook composite score had excellent reliability in both face-to-face and video assessments.

The F-test for between-subjects effect was not significant in the face-to-face assessment of resting symmetry of the eye, and in both modes of assessment of synkinesis in forehead wrinkle, indicating no significant difference amongst patients’ resting symmetry of the eye and presence of synkinesis in the forehead wrinkle.
No PCAs were calculated for the weighted total resting symmetry and weighted total voluntary movement scores. As the raw scores were multiplied by 5 and 4 respectively to determine the weighted total scores, there would not be cases where scores differed by only one point. No PEAs and PCAs were calculated as well for the composite score, which was derived from a formula in which resting symmetry, voluntary movement and synkinesis were weighted differently. Two patients with similar Sunnybrook composite scores might therefore present differently, with one patient scoring better in resting symmetry and the other scoring better in voluntary movement or synkinesis, so agreement in scores might not necessarily reflect similar facial function anyway.

Table 4: Reliability and agreement of Sunnybrook System parameters across assessment modes

<table>
<thead>
<tr>
<th>Sunnybrook System Components</th>
<th>ICC (95% CI)</th>
<th>PEA (%)</th>
<th>PCA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resting Symmetry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye</td>
<td>Face-to-Face</td>
<td>0.195 (-0.045 to 0.480)</td>
<td>043.5</td>
</tr>
<tr>
<td>Video</td>
<td>0.367 (0.113 to 0.625)*</td>
<td>073.9</td>
<td>100.0</td>
</tr>
<tr>
<td>Cheek</td>
<td>Face-to-Face</td>
<td>0.332 (0.078 to 0.597)*</td>
<td>043.5</td>
</tr>
<tr>
<td>Video</td>
<td>0.498 (0.250 to 0.719)*</td>
<td>52.2</td>
<td>95.7</td>
</tr>
<tr>
<td>Mouth</td>
<td>Face-to-Face</td>
<td>0.235 (-0.011 to 0.516)*</td>
<td>43.5</td>
</tr>
<tr>
<td>Video</td>
<td>0.314 (0.061 to 0.582)*</td>
<td>47.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Totalx5</td>
<td>Face-to-Face</td>
<td>0.548 (0.308 to 0.753)*</td>
<td>34.8</td>
</tr>
<tr>
<td>Video</td>
<td>0.544 (0.303 to 0.750)*</td>
<td>26.1</td>
<td></td>
</tr>
<tr>
<td>Voluntary Movement</td>
<td>Forehead Wrinkle</td>
<td>Face-to-Face</td>
<td>0.859 (0.745 to 0.932)*</td>
</tr>
<tr>
<td>Video</td>
<td>0.906 (0.824 to 0.955)*</td>
<td>60.9</td>
<td>87.0</td>
</tr>
<tr>
<td>Eye Closure</td>
<td>Face-to-Face</td>
<td>0.601 (0.367 to 0.790)*</td>
<td>22.7</td>
</tr>
<tr>
<td>Video</td>
<td>0.688 (0.481 to 0.842)*</td>
<td>18.2</td>
<td>77.3</td>
</tr>
<tr>
<td></td>
<td>Face-to-Face</td>
<td>Video</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>-------------</td>
<td>-------</td>
<td>----</td>
</tr>
<tr>
<td>Smile</td>
<td>0.741 (0.561 to 0.869)*</td>
<td>0.818 (0.677 to 0.911)*</td>
<td>26.1</td>
</tr>
<tr>
<td>Snarl</td>
<td>0.707 (0.512 to 0.850)*</td>
<td>0.806 (0.659 to 0.905)*</td>
<td>21.7</td>
</tr>
<tr>
<td>Lip Pucker</td>
<td>0.690 (0.489 to 0.840)*</td>
<td>0.826 (0.691 to 0.915)*</td>
<td>34.8</td>
</tr>
<tr>
<td>Totalx4</td>
<td>0.812 (0.668 to 0.908)*</td>
<td>0.912 (0.834 to 0.958)*</td>
<td>8.7</td>
</tr>
<tr>
<td>Synkinesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forehead Wrinkle</td>
<td>-0.006 (-0.203 to 0.276)</td>
<td>-0.037 (-0.225 to 0.241)</td>
<td>52.2</td>
</tr>
<tr>
<td>Eye Closure</td>
<td>0.248 (0.000 to 0.527)*</td>
<td>0.299 (0.047 to 0.570)*</td>
<td>56.5</td>
</tr>
<tr>
<td>Smile</td>
<td>0.562 (0.325 to 0.762)*</td>
<td>0.527 (0.284 to 0.739)*</td>
<td>52.2</td>
</tr>
<tr>
<td>Snarl</td>
<td>0.475 (0.225 to 0.703)*</td>
<td>0.564 (0.327 to 0.763)*</td>
<td>52.2</td>
</tr>
<tr>
<td>Lip Pucker</td>
<td>0.459 (0.208 to 0.692)*</td>
<td>0.637 (0.418 to 0.809)*</td>
<td>43.5</td>
</tr>
<tr>
<td>Total</td>
<td>0.534 (0.292 to 0.744)*</td>
<td>0.475 (0.225 to 0.809)*</td>
<td>34.8</td>
</tr>
</tbody>
</table>
### Table 1: ICC Values for Face-to-Face vs Video-Assessed Symmetry Assessment

<table>
<thead>
<tr>
<th></th>
<th>Face-to-Face</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>0.703*</td>
<td>0.811 (0.666 to 0.907)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.845 (0.720 to 0.925)*</td>
</tr>
</tbody>
</table>

(*F-test for between-subjects effect significant at p < .05)

The Bland-Altman plots demonstrated mean differences of close to zero between face-to-face and video assessment of resting symmetry when using the Sunnybrook system. Limits of agreement between the two modes of assessment of resting symmetry were overall close to +/-1. Assessment of voluntary movement with the Sunnybrook system saw mean differences of close to zero between face-to-face and video assessment. Limits of agreement between the two modes of assessment of voluntary movement were largely more than +/-1 and closer to +/-2. Assessment of synkinesis with the Sunnybrook system saw mean differences of close to zero between face-to-face and video assessment. Limits of agreement between the two modes of assessment of synkinesis were overall more than +/-1 and closer to +/-2. The Bland-Altman plots also indicated a mean difference of 1.52 points in the Sunnybrook composite score between face-to-face and video assessment, which was within the previously defined reasonable limits of 17 points, but with limits of agreement greater than +/-17 points.

Missing data was removed and not included in the results.

### DISCUSSION

Reliability of the House-Brackmann, Sydney, and Sunnybrook systems was largely similar between face-to-face assessment and video assessment of FNP, despite the 3D nature of face-to-face assessment versus the 2D nature of video assessment. There was, however, poor reliability of assessment of synkinesis.

#### Face-to-Face vs Video-Assessed House-Brackmann Grades

Reliability of the House-Brackmann scale was good in the face-to-face assessment and excellent in the video assessment, with excellent agreement within both modes of assessment. The high level of reliability supports the findings of Evans et al. [4], House and Brackmann [33], Coulson et al. [26] and Kanerva et al. [32]. The lack of an anterior-posterior axis in the video assessment did not appear to affect reliability of the House-Brackmann Scale negatively, as implied by the higher ICC value than in face-to-face assessment.
However, the Bland-Altman plot indicated that patients could potentially be given more than 1 grade difference between the face-to-face than video assessment. One grade of difference reflects significant variation in facial function, hence using the House-Brackmann scale in tele-assessment may give the clinician an erroneous impression of the patient’s condition. Therefore, although the reliability of the House-Brackmann scale indicated that its implementation in a video assessment may yield comparable grades to face-to-face assessment, potential large differences between the two assessment modes meant that video assessment using the House-Brackmann scale cannot necessarily substitute face-to-face assessment.

**Face-to-Face vs Video-Assessed Sydney System Scores**

Reliability in assessing the five branches of the facial nerve face-to-face using the Sydney system was good to excellent, with excellent agreement. Reliability and agreement were largely similar between face-to-face and video assessments, except for video assessment of the cervical branch and synkinesis, which were less reliable and had lower agreement than the face-to-face assessment. This supports Coulson et al.’s [26] findings of good reliability in assessing function of the facial nerve branches although the ICC values reported were slightly lower, and poor reliability in assessment of synkinesis. The higher ICC values in this study could be attributed to the slightly larger number of participants (n=28, vs n=21) and possibly fewer outliers, which is supported by the excellent agreement. A further explanation for this finding may be in the high familiarity and clinical usage of the Sydney System amongst assessors.

There are a few factors that could have affected reliability of video assessment of the cervical branch. First, the cervical branch of the facial nerve innervates the platysma muscle, which tightens the anterior neck region and contributes to downward movement of the lower lip [44]. Downward movement of the lips has been found to have significantly more anterior-posterior axis than horizontal axis displacement, so the 2D nature of video assessment of the cervical branch could have affected reliability [27]. Furthermore, two assessors had noted poor visibility of the neck region during video assessment of patients who were wearing a hijab or jewellery around the neck. As the neck region is where the platysma muscle is situated, poor visibility of the area would have affected video assessment of the cervical
branch. In a face-to-face assessment or tele-assessment over real-time videoconference, the assessor would be able to request the patient to remove any obstructions.

Consistently poor reliability for synkinesis assessment was observed between face-to-face and video assessments. As synkinesis is a multidimensional involuntary movement that occurs simultaneously in an area of the face different to the area being examined for voluntary movement (for example, during smiling, there could be synkinesis in the eye squinting) [26], it could be missed by the clinician who is directing attention to the area being examined for voluntary movement, and that could happen in both face-to-face and tele-assessments. In video-based tele-assessment, though, there is potential to pause and rewind footage to check for synkinesis.

The Bland-Altman plot indicated that patients could potentially be graded within 1 grade apart between face-to-face and video assessment of the temporal and buccal branches but more than 1 grade apart for that of the zygomatic, marginal mandibular, and cervical branches and synkinesis. This was not surprising as these movements (e.g. forehead raise, nose wrinkle, smile, eye closure, and movement of the lower lips) produce significant displacement in the anterior-posterior axis [27], which is difficult to adequately assessed during a video assessment in the frontal plane.

During a face-to-face assessment or tele-assessment over a videoconference media such as Skype, the clinician would have the opportunity to request a profile view of the patient, which would take into account movement occurring in the anterior-posterior axis. The clinician may also be able to request removal of any obstruction for better visibility of the area being examined in real-time tele-assessment.

**Face-to-Face vs Video-Assessed Sunnybrook System Scores**

When implemented in face-to-face settings, reliability and agreement of the Sunnybrook system was similar to when used over a video. Reliability was poor to fair when assessing resting symmetry and synkinesis, but good to excellent when assessing voluntary movement. Overall, agreement was generally good to excellent across all parameters. Similar to this study, previous research [26] has recorded good ICC reliability scores for the assessment of
voluntary movement when using the Sunnybrook system, however, poor ICC reliability scores have been found for the assessment of synkinesis.

There are a few reasons that could account for this. First, as discussed earlier, synkinesis is difficult to measure, and clinicians may have (naturally) asked patients to repeat movements in the face-to-face setting to clarify their assessment, but were less likely to rewind video assessments. This empirical question about practice norms for face-to-face and mediated assessments requires further investigation. Second, studies on FNP have found patients’ quality of life and emotional health to be affected to a larger extent by functional deficits caused by asymmetry in voluntary movement, such as the inability to smile symmetrically and oral incontinence, but to a smaller extent by synkinesis and resting symmetry [1, 2, 9, 37]. Hence, assessors might have prioritized the rating of voluntary movement over resting symmetry and synkinesis as a product of their own personal experience in dealing with what patients have perceived to be their primary impairments. Third, the low ICC and fairly high PEA and PCA were indicators of outliers skewing the reliability measure, or a restriction of range in the grades [40].

Again, reliability was similar, and poor, between face-to-face and video assessments. In addition, Kayhan et al. [45] had found “substantial” reliability while using the Sunnybrook system to assess videos of patients, while Banks et al. [31] had found excellent reliability when using the eFACE system to rate videos of participants. The poor reliability for synkinesis assessment may therefore be attributed to subjectivity in determining the degree of synkinesis amongst assessors instead of the 2D nature of video assessment.

In this study, the Sunnybrook composite score had excellent reliability in both face-to-face and video assessments. Hu et al. [46], Ross and Nedzelski [35], and Kanerva et al. [32] reported similar ICCs for the composite score indicating excellent reliability. However, the components of the Sunnybrook system comprising resting symmetry, voluntary movement, and synkinesis would give more information about a patient’s facial function during management of FNP than the composite score alone.

Bland-Altman plots indicated that video-assessed mean Sunnybrook scores for resting symmetry, voluntary movement, and synkinesis were similar to the face-to-face assessment scores. However, limits of agreement were too large when comparing video assessment of
voluntary movement, synkinesis, and composite scores to face-to-face assessment, although they were reasonable for assessment of resting symmetry. This was expected as there is no displacement in the face at rest when no facial expression is performed, so most of what a clinician would be assessing would be in the vertical and horizontal axes of the frontal plane, and the lack of an anterior-posterior axis is not expected to make a difference in video assessment.

Overall, video-assessed Sunnybrook scores were generally as reliable as face-to-face assessed scores but limits of agreement were too large between video and face-to-face assessments. This suggests that the Sunnybrook system could reasonably be implemented over a video but video assessment would not necessarily substitute face-to-face assessment until there is further evidence proving reasonable limits of agreement between face-to-face and video assessment.

Limitations and Future Research
Although this exploratory study presents interesting findings about the potential for computer mediated assessments of FNP, there are some limitations. Firstly, the use of recorded videos, rather than real-time videoconferencing, may have limited ecological validity. In tele-health settings, assessment is interactive; patients can be asked to repeat movements and move the camera for closer or different views of particular movements. This is an opportunity for future research, to compare teleassessment in real time, and to understand the impact of clinician-patient interactions on teleassessment accuracy and reliability.

Secondly, synkynesia was consistently poorly assessed. To improve reliability of assessing synkynesia overall regardless of grading system used, protocols on assessing synkynesia could be improved and standardized, with additional notes about the need to review for synkynesia in the same area potentially multiple times. There could also be more specific and relevant training provided towards the protocols [26].

Implications for tele-assessment of FNP
The current study suggests that further work is required for the 2D flat screen interface used in tele-health to be comparable to face-to-face assessment of facial nerve disorders. Whilst
video assessment of FNP using the House-Brackmann Facial Grading Scale and Sunnybrook Facial Grading System was generally as reliable as face-to-face assessment, though there was insufficient agreement between video and face-to-face assessment. This study also showed that the differences in reliability between face-to-face and video assessment using the Sydney system could likely be attributed to the lack of an anterior-posterior axis in the video assessment and generally poor reliability in assessment of synkinesis. Video assessment, however, does not take into account the opportunity for real-time clinician-patient interaction that would present in tele-assessment, which would allow clinicians to request patients to turn to their sides to show a profile view, to repeat certain movements, and to remove viewing obstructions such as face and neck coverings and jewellery, thereby potentially improving agreement between face-to-face and tele-assessment in all three facial grading systems. As cost of care is an additional challenge for patients with FNP, future research could also compare the direct and indirect costs of tele-practice with face-to-face management of FNP.

CONCLUSION
Our findings highlight the need for further research into the use of the House-Brackmann, Sydney and Sunnybrook systems in tele-assessment of patients’ facial function. Although reliability of face-to-face scores is similar to video-assessed scores, there is insufficient agreement between assessments in both modalities. This suggests a potential for tele-assessment, but the impact of synchronicity and clinician-patient interactions on the reliability of assessment needs to be empirically investigated. If the results are positive, the use of tele-health could increase access to specialized services for individuals with FNP thereby improving quality of care and rehabilitative outcomes [12, 18, 19, 23, 25].

DISCLOSURE STATEMENT
There is no financial interest nor financial support received in relation to the generation of this submission.
REFERENCES


