Remote Outcome Tracking after Facial Reanimation Surgery

Dr Jordan Fuzi (BMed)

A thesis submitted to fulfil requirements for the degree of Master of Philosophy

Faculty of Medicine and Health

The University of Sydney

Awarded 2023

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. The work presented here was produced by myself with assistance from my supervisors.I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Furthermore, this thesis contains material that was created during the thesis production and has either been published or is under review (Listed on page 8). Chapter 1 of this thesis is currently under review with the journal Facial Plastic Surgery and Aesthetic Medicine. As the corresponding author I designed this review, undertook the literature search, collected and synthesised the data and wrote the manuscript. Chapter 2 of this thesis has been published in the journal Facial Plastic Surgery and Aesthetic Medicine. I designed this study, obtained ethics, collected and analysed the data and wrote the manuscript. I am the corresponding author of this work.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Dr Jordan Fuzi 29/12/22

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Assoc Prof. Sydney Ch'ng 29/12/22

Table of Contents

Acknowledgements

I wish to offer my appreciation and gratitude to my supervisors. I am incredibly thankful for the guidance and support they have given me. Their teaching and supervision has been invaluable and allowed me to be able to produce this thesis.

I would also like to acknowledge my family and loved ones who have provided endless encouragement and support throughout this journey. Without them, I would never have been able to complete such an undertaking.

Thesis Abstract

Introduction

Facial reanimation surgery encompasses several different surgical approaches to facial palsy (FP). Goals of surgery are dependent on patient priorities and disease related loss of function. A major frustration in this field is the lack of a universal outcome measure that can aid clinicians in gaining a clear understanding of the true benefits of surgery. Advancements in computer vision technology may provide clinicians with an accurate and automated outcome assessment tool.

Aims

This thesis has two main goals. Firstly, to describe current and proposed future methods of facial palsy assessment, focusing on newly implemented automated methods. The second goal is to detail and validate the development of a novel outcome measurement tool designed to provide an automated and objective assessment of smile in FP patients.

Methods

The first chapter of this thesis will involve a narrative review detailing the current literature available for outcome measurement in facial palsy. The second chapter validates a newly developed mobile phone app 'SmileCheck' by comparing app-based smile analysis with clinician observation in FP patients. A pilot trial was undertaken prior to the second study to

validate quality assurance within the application by determining optimal usage environment.

Results

Multiple new facial palsy assessment systems utilising automated computer-based approaches have been developed. These tools either provide entirely new outcome measurement approaches or automate previously established subjective tools. Some of these have been proven effective in the clinical setting however have yet to be widely implemented. In the main study, the 'SmileCheck' application demonstrated a high degree of agreement with clinician based observation. Repeatability of measurements was confirmed through a test-retest approach.

Conclusions

Recent advances in computer vision technology provide the possibility of the development of objective smile assessment tools in facial palsy. This will increase clinician understanding into postoperative outcomes and provide a common outcome language amongst institutions. The new 'SmileCheck' application has been demonstrated to be a quick and accurate tool to quantify the quality of spontaneous smile in the facial palsy population.

List of Publications

Fuzi J, Meller C, Ch'ng S, Hadlock TM, Dusseldorp J. Voluntary and Spontaneous Smile Quantification in Facial Palsy Patients: Validation of a Novel Mobile Application. Facial Plast Surg Aesthet Med. 2022 Oct 21. doi: 10.1089/fpsam.2022.0104. Epub ahead of print. PMID: 36269609

Fuzi J, Meller C, Ch'ng S, TM, Dusseldorp J. Current and novel approaches to outcome measurement in facial reanimation surgery (Under Review)

List of Presentations

Remote Outcome Tracking in Facial Reanimation Surgery – Validation of the Novel "SmileCheck" Mobile Application. ASPN 2022 Scientific Meeting, ePoster

Chapter 1 - *Current and novel approaches to outcome measurement in facial reanimation surgery*

1.0 Background

There are significant difficulties associated with outcome measurement in facial reanimation surgery. Not only do facial palsy patients require long term follow up due to the protracted nature of postoperative recovery. Patients are also often located far from large facial nerve centres and frequent follow up can be both difficult and impractical. Whilst many forms of facial palsy assessment are available, what is clear is that a perfect outcome measurement solution has yet to be established worldwide.

This thesis was designed to achieve two main goals. The first aim is to describe the current methods of facial palsy assessment utilised in the field of facial reanimation internationally. The second aim is to describe and validate a newly developed application, 'SmileCheck', that has been designed to harness machine learning technology to provide an automated and objective smile assessment in facial palsy patients. This tool may fill some of the gaps in the literature highlighted in the first chapter.

This first chapter reviews the current literature on established and newly developed methods of outcome measurement in facial reanimation surgery. Historically, clinicians have relied upon subjective methods of quantifying disease severity and recovery in facial palsy. With the advancement of facial reanimation surgery, clinicians are in need of more objective means of assessing change in facial function overtime. This chapter forms a literature review which is currently under review in the peer-reviewed journal Facial Plastic Surgery and Aesthetic Medicine (Mary-ann Liebert inc. Publishers) (submission can be found in the appendix). The author aims for this to be an easy to reference guide for clinicians to aid in their choice of outcome measurement regime.

1.1 Abstract

Importance: Accurate quantification of postoperative outcomes after facial reanimation surgery is essential for the individualisation of treatment and ensuring that each patient's unique needs are met. The objective of this review is to detail established and emerging approaches to outcome measurement.

Observations: Currently established patient reported outcome measures provide key insights into disease related patient experience. Clinician graded scales have historically been used to quantify disease severity however are limited by their subjective nature and inter-observer variability. More recently, computer-based approaches leveraging machine learning(ML) technology have been developed to improve the objectivity of disease severity assessment. This has shown promise both in improving the reliability of well-established clinical scales and in the production of newer automated disease severity scales. Whilst restoration of spontaneous emotional activity is a key goal for patient and surgeon, a practical method of accurately quantifying smile spontaneity has yet to be established.

Conclusions: Whilst a multi-modal approach has been proposed to capture the wide array of pertinent outcomes, the exact tools have yet to be agreed upon. ML technology has shown increasing promise in providing automated solutions that can increase both the reliability of measurements and practicality in implementation.

1.2 Introduction

Facial reanimation surgery compromises various surgical procedures aimed at restoring facial symmetry and dynamic facial function in patients with facial palsy (FP). Whilst improving both static and dynamic facial symmetry is of great benefit, a key goal for both patient and surgeon is the restoration of spontaneous smile and for the patient to appear normal and joyous during organic social interaction(1). This was demonstrated through a discrete choice experiment where healthy volunteers were shown to be accepting of a significantly higher chance of treatment failure, necessitating additional surgery, to achieve spontaneous smile over voluntary smiling (2). In the acute and subacute setting, approaches to facial nerve reconstruction include direct nerve repair, utilisation of interposition or nerve transfer grafts or via direct muscle neurotisation (3–5). For long standing paralysis, free muscle transfer is required and reinnervated either with the contralateral facial nerve, nearby non-facial cranial nerves, or both(6). With regards to spontaneity, utilisation of a cross facial nerve graft (CFNG) has historically been considered the optimal neural source for restoration of emotional smile. However, it has long been recognised that cross facial nerve grafting has a higher risk of treatment failure particularly in older patients. As a result, numerous other nerve sources have been utilised including the nerve to masseter (NTM), hypoglossal, spinal accessory or combinations of multiple nerve sources. Various small volume case series of these techniques have reported promising spontaneity results however their methods of outcome measurement varied greatly. A recent systematic review of all papers found that 63% of studies did not report their methodology of assessment and of those that did the majority used clinical observation or patient reporting (7). Herein lies a major obstacle in the field of facial reanimation surgery, a lack of a universal spontaneity outcome measurement tool. A universal and objective measurement tool would allow for accurate characterisation of disease severity and treatment response both within and amongst institutions. Current methods of assessment of facial palsy include patient reported outcome measures (PROMs) and clinician-based grading systems (8). Whilst these are easy to implement and provide important information on the psychosocial impact and patient perceived disease burden, these assessments are subjective in nature and susceptible to intra- and interobserver bias. Several methods for objective spontaneity assessment have been suggested. In recent years, computer based approaches are becoming more widely employed to study facial palsy(9,10). Initially, these approaches relied upon manual identification of facial landmarks with subsequent software analysis of relevant distances and angles of movement. Such techniques provide accurate quantification of landmark movement and symmetry, however do not provide spontaneity or layperson measurements. Other techniques relied upon more complex software algorithms and 3-dimensional technologies which limited their widespread use (11–16). More recently, machine learning (ML) based computer algorithms have been demonstrated to accurately predict the position of facial landmarks without the need for manual human input (17,18). These algorithms are now being utilised within the facial palsy domain and have been demonstrated to accurately measure the emotionality and quality of spontaneous smiles (19). Herein, this review aims to outline current methods of subjective facial palsy assessment (*Table 1*) and discuss emerging computer based tools designed to provide an objective outcome measure for facial reanimation surgery (*Table 2*).

1.3 Subjective Assessment

1.3.1 Patient Reported Outcome Measures

Whilst the physical sequalae of facial palsy (FP) are easily identifiable by the layperson and clinician, the negative psychosocial effects of this disease must not be overlooked. It has been well established that facial palsy has a significantly negative impact on a patient’s emotional and psychosocial wellbeing (20). This impact is often complex and individualised to the patient with recent studies recognising that the degree of functional impairment does not always correlate with patient distress levels (21,22). As such, patient reported outcome measures (PROMs) play an integral role in evaluating the individualised psychosocial effects of facial palsy and their response to treatment (8). The first PROM designed to evaluate quality of life in facial palsy specifically was the Facial Disability Index (FDI) (23). First published in 1996, this questionnaire assesses psychosocial wellbeing and physical function through 5 questions each. During validation, FDI subscales were associated with each other and produced reliable measurements for patient focused disability. The Facial Clinimetric Evaluation (FACE) scale was subsequently developed in 2001 with an aim to measure both quality of life and facial impairment (24). This survey utilises 15 questions on a 5 point likert scale to evaluate facial movement, facial comfort, oral function, eye comfort, lacrimal control and social function. A recent review suggests that the FACE scale may better evaluate for psychological outcomes than the FDI (25). Neither scale however, evaluates for the impact of synkinesis prompting the development of the Synkinesis Assessment Questionnaire (SAQ) (26). Whilst not directly measuring quality of life, the SAQ demonstrates patient perception of the presence and impact of synkinesis producing a total score from 20 (worst) to 100 (best). A newer module of the FACE-Q questionnaire has since been developed to directly assess the outcomes of functional facial differences on children and young adults (27). This outcme measure has shown promising results in the evaluation of FP in children and young adults however requires ongoing clinical evaluation (28). Most recently, a Canadian group developed the 25 item Alberta Facial clinical evaluation scale aimed at assessing key domains of concern from the patients perspective (29). Whilst this was developed through interviewing FP patients on their key functional and psychological concerns, this questionnaire has yet to undergo rigorous testing.

1.3.2 Clinician Graded Scoring Systems

For many years clinician based grading tools have been utilised as the most objective means of facial nerve functional assessment. The most widely used of these tools is the House-Brackmann Grading Scale (HBGS) (30). This six point scale is useful in measuring large changes in function over time and has high inter- and intra-observer reliability (30–33). However, the lack of regional subdomain scores does limit this scales ability to act as an outcome measurement tool in facial reanimation where surgeons are interested in subtle changes in function of specific facial regions(8). Terzis and Noah proposed a 5 tier classification system to describe post facial reanimation outcomes based on facial symmetry and contraction (34). It is able to quantify changes over time especially after surgery however has yet to undergo rigorous testing. The Sunnybrook Facial Grading Scale (SFGS) was created as a regional-based scoring system that directly assesses facial symmetry at rest, symmetry with voluntary movements and synkinesis (35). Through assessing specific facial regions, the SFGS is more sensitive to changes in facial function than the HBGS with similarly high reliability (31,32,36). Whilst comprehensive, this tool is cumbersome to employ and requires subjective interpretation of facial function. The Sydney Facial Grading System similarly employs a regional based scoring approach by directly scoring the five main branches of the facial nerve (32). This scale is practical and intuitive however has yet to undergo rigorous reliability testing or be validated for repeated measures over time (8). The Facial Nerve Grading System 2.0 was designed to assess four distinct facial regions for degree of movement and synkinesis and can be converted to an equivalent HBGS score (37). This scale has been shown to be reliable between observers and has a moderate agreement with the HBGS in evaluating change in function over time (37,38). Most recently, Banks and colleagues developed the Electronic Facial Paralysis Assessment tool (eFACE) aimed at providing clinicians with a digital solution to facial nerve assessment (39). The clinician scores 15 facial domains with regards to static appearance, dynamic movement and synkinesis and an easy to interpret total sum score between 0-100 is produced, with 100 representing normal facial function. The eFACE has been demonstrated to be useful in assessing synkinesis and change in function overtime with high interobserver, intraobserver and test-retest reliability(39,40). Whilst each of these tools carry their own merit, it is important to recognise that they are by nature subjective tools and are prone to bias and human error.

1.4 Computer Based Assessment

Computer based approaches to smile assessment have long been studied as potential objective outcome measures for facial reanimation. Most early techniques were designed to quantify smile based on degree and synchronicity of oral commissure excursion. These approaches demonstrated a high degree of accuracy in measurements however their widespread use across centres was limited by their reliance upon specialty equipment such as 3D cameras, facial markers or handheld scanners (11,13,41–44). To overcome this, in 2012 Hadlock and colleagues developed the Facial Assessment by Computer Evaluation (FACEgram) software (45). This freely available Java based software program was demonstrated to accurately quantify facial movement and did not require complex equipment. However, this tool has not been widely adopted as it requires manual user identification of facial landmarks, which is time-consuming and may vary greatly between users. Facial analysis algorithms based off machine learning technology may overcome both the need for specialty equipment or human input by performing automated facial landmark localisation, quantification of facial movements and prediction of emotional expressions(46–48). These algorithms have been implemented within facial palsy research to produce automated scoring of existing grading scales, production of new automated scales and emotional expressivity analysis.

1.4.1 Automated Facial Landmarking and Symmetry Analysis

Multiple centres have harnessed ML based automatic facial landmarking to quantify facial symmetry both in repose and during standardised expressions. The 'Emotrics' software was developed to employ ML technology to automatically identify facial landmarks from a set of standard frontal still photographs (17). Multiple photographs can be analysed and position, distance and symmetry of landmarks is automatically calculated and used to characterise severity of palsy and reflect quality of smile. The algorithms used in Emotrics were initially trained on a large database of healthy facial photographs raising concerns around the accuracy of landmark localisation in facial palsy faces. Emotrics was subsequently retrained with a facial palsy dataset, greatly increasing accuracy of landmark localisation(49). Lee and colleagues developed the PC based Facial Asymmetry Assessment Program (PC-FAAP). This smartphone based program automatically calculates a novel FP grading system based off ML analysis of still photographs of three key facial expressions (50). Mouth, eyebrow and eye closure asymmetry ratios are determined and summated to develop a composite 'FNP grading scale (FGS)'. Through validation, the FGS was shown to be sensitive to change in severity of FP and was more consistent than subjective assessment however it has yet to be widely adopted across centres. Hidaka and colleagues describe a novel grading system that provides real-time facial symmetry analysis (51). This approach automatically identifies 68 facial landmarks on video recordings and calculates eyebrow and oral symmetry through representation of ‘displacement ratios’. This tool may aid clinicians both in tailoring and quantifying outcomes of reanimation however has yet to undergo rigorous testing. Lastly, Monini and colleagues have developed a markerless videosystem for staging facial palsy by comparing total movement of forehead frowning and smiling between healthy and palsy hemi-face (44,52). This smartphone based system utilises the ‘Emotrics’ algorithm(17) to identify facial landmarks in videos and calculates degree of movement of the oral commissure and eyebrow. Within this study, automatic derived facial grading was consistent with previously used subjective methods.

1.4.2 Automated Clinician-Graded Scales

Whilst limited by the subjectivity and variability of observer interpretation, clinician graded scales do provide a common language assessment of facial palsy severity. With the advent of ML technology, multiple centres have leveraged these algorithms to eliminate the need for clinician input to improve both the objectivity and consistency of these scales. O’reilly first developed the ‘Facogram’ in 2010 which uses trained artificial neural networks (ANN) to produce both objective HBGS scores and regional facial grading in facial palsy patients (53). This java-based system requires a video camera for data capture and standard laptop for data processing. ‘Facogram’ based assessment had good agreement with subjective assessment and demonstrated consistency on repeated testing. Expanding on previous work, Mothes and colleagues developed a convolutional neural network (CNN) based ML program to automatically grade facial palsy based on HBGS, SFGS and Stennert index (54,55). Fair agreement between subjective and automated grading of SFGS was found however no agreement was demonstrated for HBGS or Stennert index. In 2021, the Auto-eFACE tool was created by combining the 'Emotrics' algorithm with the commonly used eFACE clinician graded scoring system (56). This easy-to-use software can provide rapid and automated eFACE assessments and only requires readily available commercial computers. Most recently, Jirawatnotai and colleagues developed the 'SBface' mobile phone application to provide computerised FP assessments based of the SFGS (57). Facial landmarks are identified using the 'VNFaceObservation' (Apple Inc, USA) image analysis technology and distances between these landmarks are calculated during the 6 standardised facial expressions listed in the SFGS. Distances between landmarks are correlated to the SFGS subdomains and a composite total score is derived. Whilst repeatability of app-derived scores was demonstrated, inconsistent correlation between app and human observation was found.

1.4.3 Emotional Expressivity Analysis

The majority of previously described grading systems utilise facial symmetry both as a measure of disease severity and post-operative outcome. Recognising a strong desire of patients to appear 'normal' to the layperson, Dusseldorp and colleagues describe a new approach to reanimation outcome measurement by quantifying the emotional expressivity of smiles in FP patients (1). This group used the commercially available 'Affdex' (Affectiva, Boston) emotional analysis software to measure changes in the emotional expressivity of spontaneous smile in FP patients after reanimation. This study demonstrated a dramatic improvement in expressivity in postoperative flaccid FP patients and a reduction in negative emotion expression in the synkinetic cohort. Extending on this, the same group has since developed an automated solution to spontaneous smile analysis by combining this approach with the 'Emotrics' technology. Time points of healthy oral commissure movements as detected by 'Emotrics' are subsequently analysed by the 'Affdex' algorithm to automatically provide an emotional expressivity value for smiles (19). This approach was utilised to measure spontaneous smile emotional expression after different reanimation approaches however has yet to gain widespread use due to the need for powerful computer processing power.

1.5 Discussion

It is well known that the sequalae of facial palsy include both physical and psychosocial effects that vary greatly in severity between patients (20). As such, disease severity and surgical outcome measurement requires a multi-modal approach that can evaluate the many facets of this disease (58). Patient reported outcome measurements are limited by their subjective nature but provide key insights into the psychosocial detriment of the paralysis and of the patient's disease-related experience. The validity and reliability of these scales have been optimised over the years and a clear consensus on which scale to use between centres would greatly improve the relatability of outcome data worldwide. Clinician graded scales vary significantly with regards to complexity but do provide easy to understand measurements of disease severity. Similar to PROMs, these scales still require subjective observer-based grading, limiting the relatability between observers and institutions. Over the past decade, machine learning algorithms have shown promise in overcoming these limitations by providing objective measurement solutions. The creation of automated methods of calculating clinician graded scales may be able to overcome current concerns regarding subjectivity and provide a common language tool amongst centres. This technology has also resulted in the advent of a new method of outcome measurement by measuring emotional expression to quantify quality of smile. For these approaches to gain widespread use they need to be able to be performed on readily available hardware in any location. This would prevent the limitation of accessibility between institutions and demographics and provide remote outcome measurement solutions.

1.6 Conclusion

Outcome measurement in facial reanimation requires a multimodal approach that encompasses evaluation of all facets of FP. Computer vision based artificial intelligence has shown promising results in both improving current outcome measurement methods and providing new solutions. This will hopefully contribute to a greater understanding of the benefits of surgery and guide clinicians in optimising patient care.

1.7 Chapter Summary

There are many different domains of facial palsy assessment and a multi-modal approach is required to assess the many sequelae of facial palsy. For dynamic reanimation procedures where the goal is to restore movement to the affected hemi-face, there is yet to be a widely accepted tool to assess postoperative outcomes. Current forms of clinician or patient reported measures lack the objectivity required to compare results between surgeons and institutions. In recent years computer-based approaches have shown promise in potentially providing automated and objective systems.

Outcome measurement in facial reanimation requires a multi-modal approach that captures the many sequelae of facial paralysis. Dusseldorp and colleagues suggest the 'P.A.L.Sy' approach including **P**atient-reported outcome measures, **A**utomated and clinician-graded scoring systems, **L**ayperson assessment equivalent and **S**pontaneous smile anal**y**sis(58). Both patient reported and clinician graded assessments have historically been the cornerstone of facial palsy assessment. These tools have been widely implemented over the years across institutions however their inter-observer reliability is somewhat limited by their inherent subjective nature. Automated approaches to facial palsy assessment may improve consistency in outcomes assessment both within and across institutions. As previously discussed, a key priority for patients is to appear normal to the casual observer and for return of organic emotional conveyance during social interaction. This is achieved through improving resting symmetry and restoring spontaneous emotional smile. The ability to smile spontaneously is a domain of facial reanimation outcomes assessment that has yet to be well quantified. Some studies have shown promising results at leveraging artificial intelligence systems to provide automated methods of evaluating this. Initial approaches have used quantification of oral commissure synchronicity as a marker for smile success. As yet, none of these solutions have gained widespread implementation. Two separate groups have proposed calculating emotional expressivity as a measure of smile quality, with successful smiles able to convey higher probabilities of perceived joy (1,59). The premise of this approach is that during interaction, humans are focused on emotional expression rather than symmetry of smile.

Outcome measurement has advanced significantly since the advent of facial reanimation. However, the final prong of the proposed ‘P.A.L.Sy’ method still lacks an effective and universally agreed upon solution. A welcome tool would be able to provide automated and objective spontaneous smile quantification that can be easily implemented across institutions.

1.8 References

1. Dusseldorp JR, Guarin DL, van Veen MM, Jowett N, Hadlock TA. In the Eye of the Beholder: Changes in Perceived Emotion Expression after Smile Reanimation. Plast Reconstr Surg. 2019;144(2):457–71.

2. Dusseldorp JR, Naunheim MR, Quatela O, Fortier E, Hadlock TA, Jowett N. Neurotization Preferences in Smile Reanimation: A Discrete Choice Experiment. Plast Reconstr Surg. 2021 Sep;148(3):407e.

3. Sun DQ, Andresen NS, Gantz BJ. Surgical Management of Acute Facial Palsy. Otolaryngol Clin North Am. 2018;51(6):1077–92.

4. Klebuc M, Shenaq SM. Donor nerve selection in facial reanimation surgery. Semin Plast Surg. 2004 Feb;18(1):53–60.

5. Garcia RM, Hadlock TA, Klebuc MJ, Simpson RL, Zenn MR, Marcus JR. Contemporary solutions for the treatment of facial nerve paralysis. Plast Reconstr Surg. 2015;135(6):1025e–46e.

6. Natghian H, Fransén J, Rozen SM, Rodriguez-Lorenzo A. Qualitative and Quantitative Analysis of Smile Excursion in Facial Reanimation: A Systematic Review and Meta-analysis of 1- versus 2-stage Procedures. Plast Reconstr Surg – Glob Open. 2017 Dec;5(12):e1621.

7. van Veen T. Facial palsy: treatment, quality of life, and assessment [Thesis fully internal (DIV)]. University of Groningen; 2020.

8. Dusseldorp JR, Veen MM van, Mohan S, Hadlock TA. Outcome Tracking in Facial Palsy. Otolaryngol Clin North Am. 2018 Dec 1;51(6):1033–50.

9. Gerós A, Horta R, Aguiar P. Facegram – Objective quantitative analysis in facial reconstructive surgery. J Biomed Inform. 2016 Jun 1;61:1–9.

10. Bray D, Henstrom DK, Cheney ML, Hadlock TA. Assessing Outcomes in Facial Reanimation: Evaluation and Validation of the SMILE System for Measuring Lip Excursion During Smiling. Arch Facial Plast Surg. 2010 Sep;12(5):352–4.

11. Frey M, Giovanoli P, Gerber H, Slameczka M, Stussi E. Three-dimensional video analysis of facial movements: a new method to assess the quantity and quality of the smile. Plast Reconstr Surg. 1999;104(7):2032–9.

12. Coulson SE, Croxson GR, Gilleard WL. Three-dimensional quantification of “still” points during normal facial movement. Ann Otol Rhinol Laryngol. 1999 Mar;108(3):265–8.

13. Frey M, Jenny A, Giovanoli P, Stüssi E. Development of a new documentation system for facial movements as a basis for the international registry for neuromuscular reconstruction in the face. Plast Reconstr Surg. 1994 Jun;93(7):1334–49.

14. Meier-Gallati V, Scriba H, Fisch U. Objective scaling of facial nerve function based on area analysis (OSCAR). Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 1998;118(4):545–50.

15. Neely JG, Cheung JY, Wood M, Byers J, Rogerson A. Computerized quantitative dynamic analysis of facial motion in the paralyzed and synkinetic face. Am J Otol. 1992 Mar;13(2):97–107.

16. Yuen K, Inokuchi I, Maeta M, Kawakami SI, Masuda Y. Evaluation of facial palsy by moiré topography index. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 1997 Nov;117(5):567–72.

17. Guarin DL, Dusseldorp J, Hadlock TA, Jowett N. A Machine Learning Approach for Automated Facial Measurements in Facial Palsy. JAMA Facial Plast Surg. 2018 Jul;20(4):335–7.

18. Fan H, Zhou E. Approaching human level facial landmark localization by deep learning. Image Vis Comput. 2016 Mar 1;47:27–35.

19. Dusseldorp JR, Guarin DL, van Veen MM, Miller M, Jowett N, Hadlock TA. Automated Spontaneity Assessment after Smile Reanimation: A Machine Learning Approach. Plast Reconstr Surg. 2022 Jun 1;149(6):1393–402.

20. Fuzi J, Taylor A, Sideris A, Meller C. Does Botulinum Toxin Therapy Improve Quality of Life in Patients with Facial Palsy?. Aesthetic Plast Surg. 2020;44(5):1811–9.

21. Gyori E, Przestrzelski C, Pona I, Hagmann M, Rath T, Radtke C, et al. Quality of life and functional assessment of facial palsy patients: A questionnaire study. Int J Surg Lond Engl. 2018;55(101228232):92–7.

22. Bruins TE, van Veen MM, Werker PMN, Dijkstra PU, Broekstra DC. Associations Between Clinician-Graded Facial Function and Patient-Reported Quality of Life in Adults With Peripheral Facial Palsy: A Systematic Review and Meta-analysis. JAMA Otolaryngol--Head Neck Surg. 2021;147(8):717–28.

23. VanSwearingen JM, Brach JS. The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. Phys Ther. 1996 Dec;76(12):1288–98; discussion 1298-1300.

24. Kahn JB, Gliklich RE, Boyev KP, Stewart MG, Metson RB, McKenna MJ. Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. The Laryngoscope. 2001;111(3):387–98.

25. Ho AL, Scott AM, Klassen AF, Cano SJ, Pusic AL, Van Laeken N. Measuring quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient-reported outcome measures. Plast Reconstr Surg. 2012;130(1):91–9.

26. Mehta RP, WernickRobinson M, Hadlock TA. Validation of the Synkinesis Assessment Questionnaire. The Laryngoscope. 2007;117(5):923–6.

27. Longmire NM, Wong Riff KWY, O'Hara JL, Aggarwala S, Allen GC, Bulstrode NW, et al. Development of a New Module of the FACE-Q for Children and Young Adults with Diverse Conditions Associated with Visible and/or Functional Facial Differences. Facial Plast Surg FPS. 2017;33(5):499–508.

28. Klassen AF, Rae C, Gallo L, Norris JH, Bogart K, Johnson D, et al. Psychometric Validation of the FACE-Q Craniofacial Module for Facial Nerve Paralysis. Facial Plast Surg Aesthetic Med. 2022 Feb;24(1):1–7.

29. Hamour AF, Mendez AI, Biron VL, Seemann R, Ansari K, Liu R, et al. Development of the Alberta facial clinical evaluation scale: A patient-centred outcomes instrument for facial nerve paralysis. Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg. 2019;44(6):919–26.

30. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 1985 Apr;93(2):146–7.

31. Berg T, Jonsson L, Engström M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol. 2004 Nov;25(6):1020–6.

32. Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 2005 Apr;132(4):543–9.

33. Reitzen S.D., Babb J.S., Lalwani A.K. Significance and reliability of the House-Brackmann grading system for regional facial nerve function. Otolaryngol Head Neck Surg. 2009;140(2):154–8.

34. Terzis JK, Noah ME. Analysis of 100 cases of free-muscle transplantation for facial paralysis. Plast Reconstr Surg. 1997;99(7):1905–21.

35. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 1996 Mar;114(3):380–6.

36. Kanerva M, Jonsson L, Berg T, Axelsson S, Stjernquist-Desatnik A, Engström M, et al. Sunnybrook and House-Brackmann Systems in 5397 Facial Gradings. Otolaryngol Neck Surg. 2011 Apr 1;144(4):570–4.

37. Vrabec JT, Backous DD, Djalilian HR, Gidley PW, Leonetti JP, Marzo SJ, et al. Facial Nerve Grading System 2.0. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg. 2009 Apr;140(4):445–50.

38. Lee HY, Park MS, Byun JY, Chung JH, Na SY, Yeo SG. Agreement between the Facial Nerve Grading System 2.0 and the House-Brackmann Grading System in Patients with Bell Palsy. Clin Exp Otorhinolaryngol. 2013 Sep;6(3):135–9.

39. Banks CA, Bhama PK, Park J, Hadlock CR, Hadlock TA. Clinician-Graded Electronic Facial Paralysis Assessment: The eFACE. Plast Reconstr Surg. 2015 Aug;136(2):223e–30e.

40. Malka R, Miller M, Guarin D, Fullerton Z, Hadlock T, Banks C. Reliability Between In-Person and Still Photograph Assessment of Facial Function in Facial Paralysis Using the eFACE Facial Grading System. Facial Plast Surg Aesthetic Med. 2021 Sep;23(5):344–9.

41. Hontanilla B, Auba C. Automatic three-dimensional quantitative analysis for evaluation of facial movement. J Plast Reconstr Aesthetic Surg JPRAS. 2008;61(1):18–30.

42. Sforza C, Galante D, Shirai YF, Ferrario VF. A three-dimensional study of facial mimicry in healthy young adults. J Cranio-Maxillofac Surg. 2010 Sep 1;38(6):409–15.

43. Ozsoy U, Uysal H, Hizay A, Sekerci R, Yildirim Y. Three-dimensional objective evaluation of facial palsy and follow-up of recovery with a handheld scanner. J Plast Reconstr Aesthetic Surg JPRAS. 2021;74(12):3404–14.

44. Monini S, Marinozzi F, Atturo F, Bini F, Marchelletta S, Barbara M. Proposal of a Video-recording System for the Assessment of Bell's Palsy: Methodology and Preliminary Results. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol. 2017;38(8):1178–85.

45. Hadlock TA, Urban LS. Toward a universal, automated facial measurement tool in facial reanimation. Arch Facial Plast Surg. 2012;14(4):277–82.

46. Bartlett MS, Littlewort G, Lainscsek C, Fasel I, Movellan J. Machine learning methods for fully automatic recognition of facial expressions and facial actions. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat No04CH37583). 2004. p. 592–7 vol.1.

47. Stöckli S, Schulte-Mecklenbeck M, Borer S, Samson AC. Facial expression analysis with AFFDEX and FACET: A validation study. Behav Res Methods. 2018 Aug 1;50(4):1446–60.

48. Baskar A, Gireesh Kumar T. Facial Expression Classification Using Machine Learning Approach: A Review. In: Satapathy SC, Bhateja V, Raju KS, Janakiramaiah B, editors. Data Engineering and Intelligent Computing. Singapore: Springer; 2018. p. 337–45. (Advances in Intelligent Systems and Computing).

49. Guarin DL, Yunusova Y, Taati B, Dusseldorp JR, Mohan S, Tavares J, et al. Toward an Automatic System for Computer-Aided Assessment in Facial Palsy. Facial Plast Surg Aesthetic Med. 2020 Feb;22(1):42–9.

50. Lee DY, Kim HS, Kim SY, Park KS, Kim YH. Comparison between Subjective Scoring and Computer-Based Asymmetry Assessment in Facial Nerve Palsy. J Audiol Otol. 2019 Jan;23(1):53–8.

51. Hidaka T, Kurita M, Ogawa K, Tomioka Y, Okazaki M. Application of Artificial Intelligence for Real-Time Facial Asymmetry Analysis. Plast Reconstr Surg. 2020;146(2):243e–5e.

52. Monini S, Ripoli S, Filippi C, Fatuzzo I, Salerno G, Covelli E, et al. An objective, markerless videosystem for staging facial palsy. Eur Arch Otorhinolaryngol. 2021;278(9):3541–50.

53. O'Reilly BF, Soraghan JJ, McGrenary S, He S. Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol. 2010 Apr;31(3):486–91.

54. Haase D, Minnigerode L, Volk GF, Denzler J, Guntinas-Lichius O. Automated and objective action coding of facial expressions in patients with acute facial palsy. Eur Arch Otorhinolaryngol. 2015 May 1;272(5):1259–67.

55. Mothes O, Modersohn L, Volk GF, Klingner C, Witte OW, Schlattmann P, et al. Automated objective and marker-free facial grading using photographs of patients with facial palsy. Eur Arch Otorhinolaryngol. 2019 Dec;276(12):3335–43.

56. Miller MQ, Hadlock TA, Fortier E, Guarin DL. The Auto-eFACE: Machine Learning-Enhanced Program Yields Automated Facial Palsy Assessment Tool. Plast Reconstr Surg. 2021;147(2):467–74.

57. Jirawatnotai S, Jomkoh P, Voravitvet TY, Tirakotai W, Somboonsap N. Computerized Sunnybrook facial grading scale (SBface) application for facial paralysis evaluation. Arch Plast Surg. 2021 May;48(3):269–77.

58. Dusseldorp J, Meares C, Hadlock T. A Multi-modal Approach to Outcome Tracking in Facial Palsy. In 2021. p. 23–30.

59. Boonipat T, Asaad M, Lin J, Glass GE, Mardini S, Stotland M. Using Artificial Intelligence to Measure Facial Expression following Facial Reanimation Surgery. Plast Reconstr Surg [Internet]. 2020;146(5). Available from:

https://journals.lww.com/plasreconsurg/Fulltext/2020/11000/Using_Artificial_Intelligence_to_Measure_Facial.43.aspx

1.9 Tables and Figures

Subjective Assessment	
Patient Reported Outcome Measures	**Clinican Graded Scales**
	House Brackmann Grading Scale (HBGS)
Facial Disability Index (FDI)	Terzis and Noah
Facial Clinimetric Evaluation (FACE)	Sunnybrook Facial Grading Scale (SFGS)
Synkinesis Assessment Questionnaire (SAQ)	Sydney Facial Grading System
FACE-Q	Facial Nerve Grading System 2.0 (FNGS 2.0)
Alberta Facial Clinical Evaluation Scale	Electronic Facial Paralysis Assessment Tool (eFACE)

Table 1: Subjective assessment methods

Computer Based Approaches		
Facial Landmarking	**Automated Clinician-graded Scales**	**Emotional Expressivity**

Facial Assessment by Computer Evaluation (FACEgram)	Facogram	Dusseldorp Affdex Tool
Emotrics	Mothes Automated Marker-Free Grading Tool	
PC based Facial Asymmetry Assessment Program (PC-FAAP)	Auto-eFACE	
Monini Markerless Videosystem	Sbface	

Table 2: Computer based assessment approaches

Chapter 2 - Voluntary and spontaneous smile quantification in facial palsy patients– validation of a novel mobile application

2.0 Background

The impact of facial palsy is both complex and multifaceted and it has been well established that outcome measurement will require a multi-modal approach. With regards to restoration of spontaneous smile and for the patient to once again appear 'normal' to the layperson, an agreed upon measurement tool has yet to be developed. For such a tool to be implemented worldwide across institutions it needs to be highly accessible to all demographics. To increase its use, this tool should be quick and cheap to implement and only require readily available commercial hardware.

In recent years, computer vision based artificial intelligence algorithms have been utilised to develop facial analysis tools. These algorithms have been used extensively in market research and have been shown to be effective in automatically evaluating emotional expressivity of respondents while watching humorous content. Additionally, these algorithms require simple and readily available hardware to run. Recognising this, we developed a novel mobile phone application 'SmileCheck' designed to utilise computer vision technology to automatically evaluate the emotional expressivity of a facial palsy patient's smile. The application was designed to be used on a patient's smartphone and to overcome two main limitations of other tools. Firstly, using the patient's phone no complex hardware is required and patients can be followed up long term remotely. This negates the need for in person evaluation to increase the possibility of repeated measurements over time. Secondly, an entirely automated and objective solution can overcome the subjective bias and variability of observer-based approaches and increase the inter-observer reliability of data across institutions.

The following chapter includes a study undertaken to validate the newly developed mobile phone application: SmileCheck. The validation process involved two steps. Firstly, we compared the accuracy of application measurements to clinician-based observations. After this, the repeatability of application measurements was determined using a test-retest approach. Prior to undergoing the validation process, a pilot study was performed to ensure quality assurance of application measurements. As part of the development process the application was tested to determine limits of movement of the user relative to the mobile phone. This was done to ensure high quality of measurements and is detailed in the appendix.

2.1 Abstract

Introduction

Restoration of spontaneous smiling is a key goal in facial reanimation. A major obstacle to quantifying recovery of spontaneous smiling is the current lack of a uniform and objective means of smile quantification.

Objective

This study aims to measure the accuracy of a novel mobile phone application designed to provide automated smile quantification within the facial palsy population.

Methods

Video recordings of 25 patients with unilateral facial palsy watching humorous videos were utilised. Application-derived smile timestamping was compared to manual observer interpretation. Internal reliability of measurements was evaluated through a test-retest approach.

Results

Application-derived smile identification demonstrated almost perfect agreement with manual interpretation (Kappa 0.861, $p<0.001$). There was no statistically significant difference in mean number of smiles between detection method ($p=0.354$). Automated smile identification demonstrated a high degree of specificity (95.4%), accuracy (93.1%), positive-predictive value (94.7%) and negative-predictive value (91.8%). This method demonstrated a high degree of reliability (Kappa 0.864, $p<0.01$).

Conclusion

The novel 'SmileCheck' mobile phone application performed accurate and reliable smile quantification in facial palsy patients in comparison to manual observation.

2.2 Introduction

Facial palsy (FP) can result in severe oro-facial muscle dysfunction and loss of common emotional facial expressions, especially joy. This dysfunction can dramatically impact quality of life and psychosocial well-being [1,2]. Facial reanimation surgeries aim to enable normal appearance during commonplace facial movements such as smiling, laughing, talking, or eating. Recent studies have suggested that a primary goal to most people is the restoration of spontaneous or emotional smiling [3]. However, a major barrier to the identification of treatments that reliably restore emotional smiling is the lack of an objective method to quantify spontaneity. Currently, outcome assessment following smile reanimation surgery is reliant on the perception of experts, patients and society[45–8]. However, due to the complexity and ensuing prolonged rehabilitation of smile reanimation surgeries, both clinician and patient are prone to be affected by confirmation bias. Although rates of smile spontaneity are extremely high with one study demonstrating rates of 70% [9] other studies have shown no correlation between objectively measured synchronous smiles and clinician-determined smile spontaneity [10]. Over the past 50 years, significant effort has been placed on the development of clinician-rated facial grading scales[11–13], however these subjective assessments can be cumbersome and inherently susceptible to inter-/intra-observer variability [14]. Objective clinical tools have also been utilised including sensor-based techniques, electromyography or functional magnetic resonance imaging, yet these remain limited by availability and cost [15,16]. A welcome tool in facial reanimation surgery would permit rapid, inexpensive, and objective voluntary and spontaneous smile assessment. This would aid clinicians in quantifying post-operative outcomes, thus yielding greater insight into the success rates of different smile reanimation techniques. Recent advancements in artificial intelligence have prompted research into the utility of novel techniques in machine learning and computer vision for automatic oro-facial assessment and emotion detection [17181920212223]. Such tools provide automatic localisation of facial landmarks and automatic generation of emotion expression probabilities from both still frames and videos [3]. This study aims to compare the accuracy of automated application derived smile quantification with manual observer interpretation of facial palsy smiles. A second objective is to quantify the reliability of application derived measurements.

2.3 Methods and Materials

Ethics approval to undertake this study was obtained through the University of Sydney, Australia.

2.3.1 'SmileCheck' app (Mass Eye and Ear Infirmary, USA)

This study utilised the newly developed mobile App 'SmileCheck' designed to provide automatic assessment of voluntary and spontaneous smile ability in FP patients. Assessment of smiling in these patients requires 1. Elicitation of an emotional smile in the subject, 2. Identification ("timestamping") of moments when smiles occur and 3. Analysis of the ability of the patient to look happy during these smile events. To achieve this the user first nominates which side of their face is affected by FP and instructions are provided on optimal environment for use and correct head position. The user is then prompted to perform three voluntary smiles before being shown customisable humorous video content such as the previously validated spontaneous smile assay (SSA) [24]. Whilst the videos are being displayed the front-facing camera on the device records the user's facial expressions (Full-Face Recording (FFR)). To timestamp the smile events the App uses a hemi-facial mirroring technique to create a second video of a composite whole face that only includes the movements of the healthy side of the face (Mirrored-Face recording (MFR)). The mirroring technique creates a perfectly symmetrical composite smile which can be detected by an emotional analysis algorithm Figure 1. The application then rapidly processes both FFR and MFR through the 'Affdex' (Affectiva - Boston, MA) emotional analysis algorithm to automatically timestamp smile events from the MFR and correlate this to the degree of joy emotion perceived from FFR (Figure 2).

As per prior methodology, a smile event was defined as occurring when probability of smile in the MFR was perceived as >=90% for at least 1 second [10]. This minimises the potential for analysis of non-meaningful facial movements such as facial flickers or synkinetic movements. Outputs produced by the App are demonstrated in Appendix 1. The amount of time taken for the App to produce the resultant data files is 10 seconds (Figure 3) and requires 2 minutes of the user's time.

Two steps of Validation were undertaken to ensure the App achieved the above functionality consistently.

2.3.2 Step 1: Comparison of Automatic to Manual Timestamping Smile Events

App derived automated timestamping of spontaneous smile events (SSEs) were compared to manual classification of SSEs using our previously described methods [10]. Accuracy metrics were then determined by comparing number of SSEs detected and timepoints of smile timestamps between the manual and automatic approaches. Agreement was determined using Cohen's kappa.

2.3.3 Step 2: Determination of Internal Consistency

The reliability of application derived smile timestamping was evaluated through a test-retest approach. Each recording was analysed twice by the App. Time-stamping of smile events were compared between each evaluation. Cohen's Kappa was utilised to determine agreement between the two rounds of analysis.

2.3.4 Statistical Analysis

Post-processing and statistical analysis within this study was undertaken using 'R Project' [25].

2.4 Results

2.4.1 Step 1: Automatic vs Manual Time stamping Approaches

25 Recordings of 25 patients with a unilateral FP were analysed. Demographic data and aetiology of FP is detailed in Table 1. Application derived smile identification demonstrated an almost perfect agreement with manual interpretation (Kappa 0.861, $p<0.001$). There was

no statistically significant difference in mean number of smiles between automated or manual timestamping approaches (5.24(4.37,6.11,95%CI) vs 5.48(4.63,6.32,95% CI), p=0.354) (Table 2). In 5/25patients, an additional 7 smiles were detected by the App than by manual observation. Each of these smile events were subsequently reanalysed manually and it was determined that these small or short smiles were missed by the manual timestamping in 6/7 cases.

2.4.2 Step 2: Test-Retest Reliability

A total of 50 recordings of 25 patients were evaluated in a test-retest fashion. App derived smile identification demonstrated a high degree of consistency between recordings (Kappa 0.864, p<0.01). There was no significant difference in mean number of smiles detected between each round of recordings (p>0.05) (Table 3)

2.5 Discussion

Herein, we validated an entirely automated, objective and uniform approach to the assessment of smiling against human observations. By combining an emotional analysis algorithm with a novel hemifacial mirroring technique this app prevents the need for multiple artificial intelligence (AI) tools and complex hardware. Within this FP population, App derived automatic timestamping was highly sensitive and specific when compared to gold standard manual observer timestamping. In each recording, at least one smile was accurately detected by the App, highlighting its success in each use. Additionally, throughout the analysis not a single manually observed smile was missed by App interpretation and at least one smile was simultaneously detected by both manual and automated methods in 100% of cases. This automated smile identification demonstrated a high internal reliability during a test-retest approach.

This study is an extension of work performed by a group at MEEI (Boston, USA) aimed at overcoming the current limitations of clinician or patient reported smile assessment methods[9,10,26–29]. First, a 1.5 minute film clip with superimposed audible laughter (SSA) was developed to elicit at least one spontaneous smile in more than 95% of both normal and FP subjects[24]. Dusseldorp et al then undertook a manual approach to the analysis of patients watching such videos to determine the synchronicity of oral commissure excursion after facial reanimation surgery[10]. Whilst effective, this approach was time consuming and impractical for routine clinical use. This prompted further research into the use of AI to provide an automated solution to spontaneous smile quantification. A concern with traditional algorithms is based on their training on large datasets of 'healthy' faces and the potential for inaccurate analysis of the asymmetrical facial palsy smile[30]. Considering this, the same group then developed a two-stage automated approach to spontaneous smile analysis utilising two AI algorithms[31]. The 'Emotrics' app (MEEI, USA) was utilised to timestamp SSEs based on healthy side oral commissure movement and the same 'Affdex' algorithm employed in this study then determined the probabilities of various emotion categories perceived during each smile. Whilst the feasibility of this approach was demonstrated it required complex software and powerful computing and would be impractical for everyday clinical use. The 'SmileCheck' app is an extension of this prior work aiming to provide a single step automatic solution to smile assessment. Employing MFRs negated the need for two separate algorithms whilst utilising the same premise of smile assessment as above.

The ability to smile spontaneously is a key component of human social interaction [32], and a loss of smile can impart a significant detriment to quality of life and psychosocial well-being in FP patients[1]. Accordingly, restoration of spontaneous smile and for patients to appear 'normal' to the layperson during organic social interaction is a key goal for facial reanimation surgeons and patients alike[3,33].

Within the expanding field of facial reanimation surgery, there is ongoing debate as to the optimal approach for the restoration of smile spontaneity. Whilst utilisation of the ipsilateral facial nerve stump or cross facial nerve graft is considered the optimal approaches to achieve spontaneous smiling[26,27,33–37] , many small volume case series of various techniques report extremely high spontaneity rates. However, in a systematic review of all such papers it was found that 63% did not report the methodology of assessment or who was responsible for determining the presence of spontaneity [4]. Of those studies that did report, the majority used clinical observation and the remainder used patient reporting. No study, other than our prior work[10] has sought to quantify spontaneity objectively. This lack of a universal spontaneity outcome measure hinders accurate quantification of both degree and quality of spontaneous smile, especially over protracted postoperative intervals[10,28].

The ‘SmileCheck’ app was designed specifically to overcome this limitation in the field by providing a tool that allows for rapid, objective and remote smile assessment that can be performed by any patient in any setting. It requires minimal user input and takes little time to perform a smile assessment. Through presentation of the emotionality of each smile, we believe this application offers clinicians an easy to interpret understanding of the functional ability of a voluntary and spontaneous smile to organically express joy. We believe this application can be utilised as a means of quantifying the overall success of a facial reanimation surgery and may provide a universal language for future research projects.

The main limitation within this study is the potential influence of user environment on

the App’s accuracy. In particular, the degree of light in the user environment and the movement of the device or user during app use. To overcome this the user is first prompted with a set of instructions on ways to optimise their environment. Secondly, the application continually measures relative movement between the device and user and will stop recording if movement breaches permissible limits. These limits were developed in a pilot study detailed within the supplementary content.

Overall, this study demonstrated the potential utility of ‘SmileCheck’ app as a smile assessment tool in facial palsy patients. Future research into the use of this app as a means of measuring response to treatment in these patients would be of particular value.

2.6 Conclusion

Accurate and uniform evaluation of spontaneous smile is critical for developing a true understanding of the success of various forms of smile reanimation. This will aid clinicians in their selection of approach and will improve patient understanding of the true likelihood of return of spontaneous smile after various treatment pathways. Within this study, the novel 'SmileCheck' App has been demonstrated to accurately and reliably identify smile in facial palsy patients. This tool may allow for quick, accurate and repeatable evaluation of smile and can be utilised in further studies to provide a common success outcome measure amongst various facial reanimation institutions.

2.7 Chapter Summary

This chapter demonstrates the potential utility of the 'SmileCheck' mobile phone application in providing facial reanimation surgeons with a universal spontaneous smile outcome measure.

During application development it was noticed that user head position relative to the phone intimately affected app recognition of facial midline and subsequent facial mirroring. Large degrees of head movement off centre often resulted in skewed and unrecognisable mirrored images. To ensure app accuracy and prevent interpretation of misrepresented images a pilot study was undertaken to determine the permissible amount of movement between the user and the mobile phone.

This testing demonstrated that facial mirroring was largely unaffected by movement until an absolute amount of 15 degrees of movement. However, upon reviewing the produced mirrored images of different degrees of movement a maximum permissible amount of 10 degrees was selected by the developing authors. This was to ensure that each mirrored image being analysed was high quality and a true reflection of the activity of the user's healthy hemi-face. This limit of 10 degrees was coded into the application so that whenever this movement limit was breeched, data recording was stopped, and the user was prompted to return to an acceptable head position.

The validation process of the application was subsequently undertaken after the above quality assurance measures were implemented. Identification of spontaneous smiling based on hemi-facial mirroring had a high degree of concordance with clinician-based observation. With every patient at least one smile was captured by both measurement methods. To the authors, this indicates app success as only one smile is needed to quantify smile quality during clinical follow up. Furthermore no manually identified smile was missed by the app, with the app identifying 7 smiles in 5 patients that were missed by the clinician observer. Reassuringly, the repeatability of app measurements was confirmed through the retest approach.

Overall, within this facial palsy cohort the ‘SmileCheck’ app provided accurate and easy to interpret data on the quality of these patients’ spontaneous smiles.

2.8 References

1. Fuzi J, Taylor A, Sideris A, et al. Does Botulinum Toxin Therapy Improve Quality of Life in Patients with Facial Palsy? Aesthetic Plast Surg 2020;44(5):1811–1819; doi: 10.1007/s00266-020-01870-4.

2. Faris C, Tessler O, Heiser A, et al. Evaluation of Societal Health Utility of Facial Palsy and Facial Reanimation. JAMA Facial Plast Surg 2018;20(6):480–487; doi: 10.1001/jamafacial.2018.0866.

3. Dusseldorp JR, Guarin DL, van Veen MM, et al. In the Eye of the Beholder: Changes in Perceived Emotion Expression after Smile Reanimation. Plast Reconstr Surg 2019;144(2):457–471; doi: 10.1097/PRS.0000000000005865.

4. van Veen T. Facial Palsy: Treatment, Quality of Life, and Assessment. Thesis fully internal (DIV). University of Groningen; 2020; doi: 10.33612/diss.131756313.

5. Dey JK, Ishii LE, Nellis JC, et al. Comparing Patient, Casual Observer, and Expert Perception of Permanent Unilateral Facial Paralysis. JAMA Facial Plast Surg 2017;19(6):476–483; doi: 10.1001/jamafacial.2016.1630.

6. Dey JK, Ishii LE, Byrne PJ, et al. Seeing Is Believing: Objectively Evaluating the Impact of Facial Reanimation Surgery on Social Perception. The Laryngoscope 2014;124(11):2489–97; doi: 10.1002/lary.24801.

7. Dey JK, Ishii M, Boahene KDO, et al. Changing Perception: Facial Reanimation Surgery Improves Attractiveness and Decreases Negative Facial Perception. The Laryngoscope 2014;124(1):84–90; doi: 10.1002/lary.24262.

8. Ishii L, Dey J, Boahene KDO, et al. The Social Distraction of Facial Paralysis: Objective Measurement of Social Attention Using Eye-Tracking. The Laryngoscope 2016;126(2):334–339; doi: 10.1002/lary.25324.

9. Sforza C, Frigerio A, Mapelli A, et al. Double-Powered Free Gracilis Muscle Transfer for Smile Reanimation: A Longitudinal Optoelectronic Study. J Plast Reconstr Aesthet Surg 2015;68(7):930–939; doi: 10.1016/j.bjps.2015.03.029.

10. Dusseldorp JR, van Veen MM, Guarin DL, et al. Spontaneity Assessment in Dually Innervated Gracilis Smile Reanimation Surgery. JAMA Facial Plast Surg 2019;21(6):551–557; doi: 10.1001/jamafacial.2019.1090.

11. House JW and Brackmann DE. Facial Nerve Grading System. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1985;93(2):146–147; doi: 10.1177/019459988509300202.

12. Ross BG, Fradet G and Nedzelski JM. Development of a Sensitive Clinical Facial Grading System. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1996;114(3):380–386; doi: 10.1016/s0194-5998(96)70206-1.

13. Banks CA, Bhama PK, Park J, et al. Clinician-Graded Electronic Facial Paralysis Assessment: The EFACE. Plast Reconstr Surg 2015;136(2):223e–230e; doi: 10.1097/PRS.0000000000001447.

14. Niziol R, Henry FP, Leckenby JI, et al. Is There an Ideal Outcome Scoring System for Facial Reanimation Surgery? A Review of Current Methods and Suggestions for Future Publications. J Plast Reconstr Aesthetic Surg JPRAS 2015;68(4):447–456; doi: 10.1016/j.bjps.2014.12.015.

15. Bandini A, Green JR, Wang J, et al. Kinematic Features of Jaw and Lips Distinguish Symptomatic From Presymptomatic Stages of Bulbar Decline in Amyotrophic Lateral Sclerosis. J Speech Lang Hear Res JSLHR 2018;61(5):1118–1129; doi: 10.1044/2018_JSLHR-S-17-0262.

16. Bandini A, Rezaei S, Guarin D, et al. A New Dataset for Facial Motion Analysis in Individuals with Neurological Disorders. IEEE J Biomed Health Inform 2020;PP; doi: 10.1109/JBHI.2020.3019242.

17. Thevenot J, Lopez MB and Hadid A. A Survey on Computer Vision for Assistive Medical Diagnosis From Faces. IEEE J Biomed Health Inform 2018;22(5):1497–1511; doi: 10.1109/JBHI.2017.2754861.

18. Guarin DL, Dusseldorp J, Hadlock TA, et al. A Machine Learning Approach for Automated Facial Measurements in Facial Palsy. JAMA Facial Plast Surg 2018;20(4):335–337; doi: 10.1001/jamafacial.2018.0030.

19. Guarin DL, Yunusova Y, Taati B, et al. Toward an Automatic System for Computer-Aided Assessment in Facial Palsy. Facial Plast Surg Aesthetic Med 2020;22(1):42–49; doi: 10.1089/fpsam.2019.29000.gua.

20. Bandini A, Orlandi S, Escalante HJ, et al. Analysis of Facial Expressions in Parkinson's Disease through Video-Based Automatic Methods. J Neurosci Methods 2017;281:7–20; doi: 10.1016/j.jneumeth.2017.02.006.

21. O'Reilly BF, Soraghan JJ, McGrenary S, et al. Objective Method of Assessing and Presenting the House-Brackmann and Regional Grades of Facial Palsy by Production of a

Facogram. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol 2010;31(3):486–491; doi: 10.1097/MAO.0b013e3181c993dc.

22. Wang T, Zhang S, Dong J, et al. Automatic Evaluation of the Degree of Facial Nerve Paralysis. Multimed Tools Appl 2016;75(19):11893–11908; doi: 10.1007/s11042-015-2696-0.

23. Guo Z, Dan G, Xiang J, et al. An Unobtrusive Computerized Assessment Framework for Unilateral Peripheral Facial Paralysis. IEEE J Biomed Health Inform 2018;22(3):835–841; doi: 10.1109/JBHI.2017.2707588.

24. Iacolucci CM, Banks C, Jowett N, et al. Development and Validation of a Spontaneous Smile Assay. JAMA Facial Plast Surg 2015;17(3):191–196; doi: 10.1001/jamafacial.2015.0083.

25. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria; 2021.

26. Biglioli F, Bayoudh W, Colombo V, et al. [Double innervation (facial/masseter) on the gracilis flap, in the middle face reanimation in the management of facial paralysis: a new concept]. Ann Chir Plast Esthet 2013;58(2):89–95; doi: 10.1016/j.anplas.2012.12.001.

27. Cardenas-Mejia A, Covarrubias-Ramirez JV, Bello-Margolis A, et al. Double Innervated Free Functional Muscle Transfer for Facial Reanimation. J Plast Surg Hand Surg 2015;49(3):183–188; doi: 10.3109/2000656X.2014.988218.

28. Dusseldorp JR, Veen MM van, Mohan S, et al. Outcome Tracking in Facial Palsy. Otolaryngol Clin North Am 2018;51(6):1033–1050; doi: 10.1016/j.otc.2018.07.003.

29. Chuang DC-C, Lu JC-Y and Anesti K. One-Stage Procedure Using Spinal Accessory Nerve (XI)–Innervated Free Muscle for Facial Paralysis Reconstruction. Plast Reconstr Surg 2013;132(1):117e; doi: 10.1097/PRS.0b013e318290f8cd.

30. Guarin D, Taati B, Hadlock T, et al. Automatic Facial Landmark Localization in Clinical Populations - Improving Model Performance with a Small Dataset. 2020; doi: 10.21203/rs.3.rs-19405/v1.

31. Dusseldorp JR. Automated Spontaneity Assessment after Smile Reanimation: A Machine Learning Approach. 2022.

32. Kraut R and Johnston R. Social and Emotional Messages of Smiling. J Pers Soc Psychol 1979;37:1539–1553; doi: 10.1037/0022-3514.37.9.1539.

33. Dusseldorp JR, Naunheim MR, Quatela O, et al. Neurotization Preferences in Smile Reanimation: A Discrete Choice Experiment. Plast Reconstr Surg 2021;148(3):407e; doi: 10.1097/PRS.0000000000008302.

34. Bhama PK, Weinberg JS, Lindsay RW, et al. Objective Outcomes Analysis Following Microvascular Gracilis Transfer for Facial Reanimation: A Review of 10 Years' Experience. JAMA Facial Plast Surg 2014;16(2):85–92; doi: 10.1001/jamafacial.2013.2463.

35. Hontanilla B, Marre D and Cabello Á. Facial Reanimation with Gracilis Muscle Transfer Neurotized to Cross-Facial Nerve Graft versus Masseteric Nerve: A Comparative Study Using the FACIAL CLIMA Evaluating System. Plast Reconstr Surg 2013;131(6):1241–1252; doi: 10.1097/PRS.0b013e31828bd4da.

36. Bae Y-C, Zuker RM, Manktelow RT, et al. A Comparison of Commissure Excursion Following Gracilis Muscle Transplantation for Facial Paralysis Using a Cross-Face Nerve Graft

versus the Motor Nerve to the Masseter Nerve. Plast Reconstr Surg 2006;117(7):2407–2413; doi: 10.1097/01.prs.0000218798.95027.21.

37. Watanabe Y, Akizuki T, Ozawa T, et al. Dual Innervation Method Using One-Stage Reconstruction with Free Latissimus Dorsi Muscle Transfer for Re-Animation of Established Facial Paralysis: Simultaneous Reinnervation of the Ipsilateral Masseter Motor Nerve and the Contralateral Facial Nerve to Improve the Quality of Smile and Emotional Facial Expressions. J Plast Reconstr Aesthetic Surg JPRAS 2009;62(12):1589–1597; doi: 10.1016/j.bjps.2008.07.025.

2.9 Tables and Figures

Figure 1. 'SmileCheck' screenshot demonstrating hemifacial mirroring before (A) and after (B) facial reanimation.

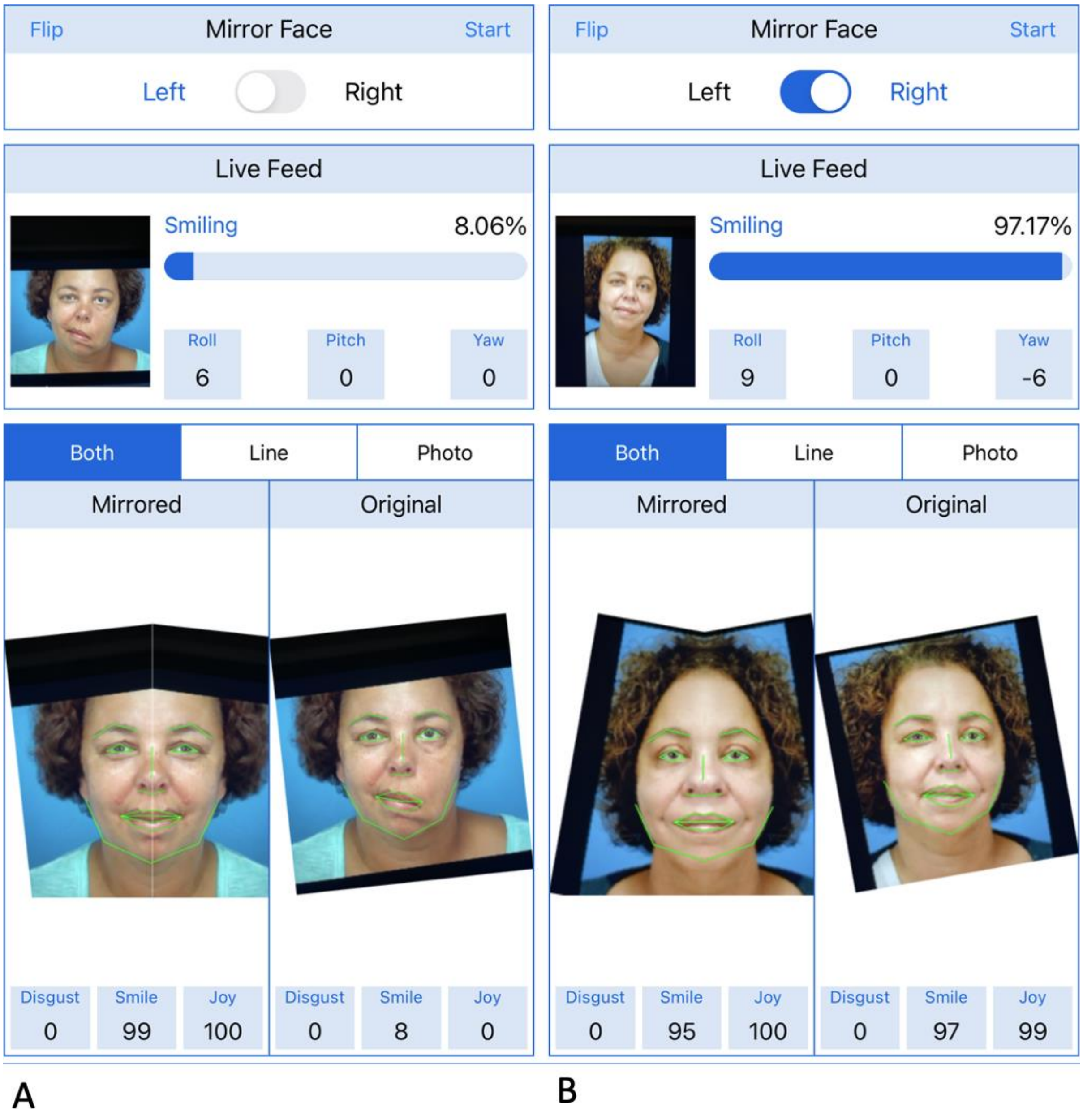

Figure 2: App user interface

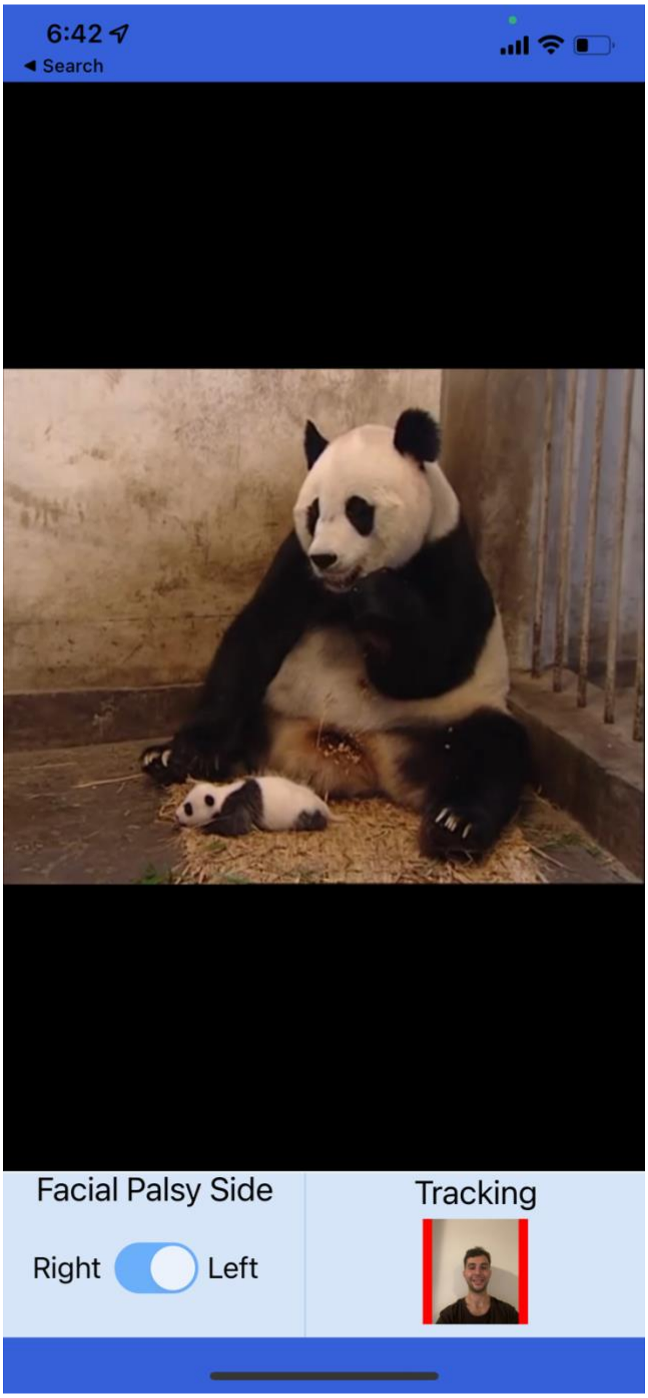
6:42
Search
Facial Palsy Side
Right
Left
Tracking

Figure 3: App derived scorecard detailing voluntary and spontaneous smile events in a healthy test user

Smile_event	Start_time	End_time	Duration	Avg_mirror_smile	Avg_mirror_joy	Avg_mirror_disgust	Mirror_emq	Avg_orig_smile	Avg_orig_joy	Avg_orig_dis	Orig_emq
					Voluntary						
1	9.007389	10.814827	1.8074379	99.44	99.85	0	99.85 /- 0.00	70.14	68.4	0.05	68.40 /- 0.05
2	13.625403	16.034628	2.4092245	99.66	99.82	0	99.82 /- 0.00	74.26	71.61	0.03	71.61 /- 0.03
3	18.645018	20.653511	2.0084934	99.94	99.88	0	99.88 /- 0.00	76.93	72.56	0.07	72.56 /- 0.07
					Spontaneous						
1	3.780606	9.600983	5.8203764	99.89	99.93	0	99.93 /- 0.00	95.17	94.51	0.02	94.51 /- 0.02
2	29.479548	34.496593	5.017046	99.8	99.92	0	99.92 /- 0.00	92.2	91.88	0.1	91.88 /- 0.10
3	37.70924	39.314228	1.6049881	98.97	99.9	0	99.90 /- 0.00	82.94	79.44	0.06	79.44 /- 0.06
4	39.721745	42.126263	2.4045181	99.06	99.91	0	99.91 /- 0.00	99.99	99.92	0	99.92 /- 0.00
5	42.536243	45.739388	3.203144	99.7	99.92	0	99.92 /- 0.00	99.96	99.75	0.02	99.75 /- 0.02
6	50.364613	52.767883	2.4032707	99.86	99.93	0	99.93 /- 0.00	99.95	99.93	0.01	99.93 /- 0.01
7	73.64619	78.263466	4.617279	99.63	99.88	0	99.88 /- 0.00	99.97	99.88	0	99.88 /- 0.00

Table 1: Demographics of facial palsy population

Number of Participants	25
Mean Age (SD)	47.2 (19.6)
Gender, M:F (n (%))	9 (36) : 16 (64)
Flaccid : Non-Flaccid (n (%))	17 (68) : 8 (32)
Aetiology of Palsy (n (%))	
Parotid Tumour	4 (16)
Brainstem Tumour	4 (16)
CNS tumour	2 (8)
Iatrogenic	4 (16)

Acoustic Neuroma	7 (28)
Congenital	1 (4)
Other	3 (12)

Table 2: Accuracy metrics for smile interpretation in the test population

	Manual Smiles	Automatic Smiles	
Patients (n)	25	25	
Mean	5.48	5.25	p=0.354*
95% CI	(4.64 – 6.32)	(4.37 – 6.11)	
Sensitivity			90.5 %
Specificity			95.4 %
Accuracy			93.1 %
Prevalence			47.4 %
Positive Predictive Value			94.7 %
Negative Predictive Value			91.8%
Agreement		0.861	p<0.001
*Wilcoxon signed rank test			

Table 3: Reliability testing results

	Test	Retest

Patients (n)	25	25	
Mean	5.12	5.04	$p>0.05$ *
95% CI	(4.28 – 5.96)	(4.12 – 5.96)	

*Wilcoxon signed rank test

Chapter 3 – Conclusions

3.1 Summary of Findings

An optimal approach to outcome measurement in facial reanimation surgery has yet to be established within the facial palsy domain. This is of critical importance within the field as multiple surgical techniques exist and each aims to address the many different sequalae associated with facial palsy. A robust understanding of the success and limitations of each surgical approach will allow facial reanimation surgeons to more accurately tailor their approach to meet the individual needs of the patient. It will also allow for improved preoperative counselling and patient education, ensuring patients are well informed prior to consenting to treatment. The aim of this thesis was to investigate the currently available and emerging outcome measurement approaches, as well as evaluate the potential of a newly developed tool.

Firstly, a literature search was undertaken to evaluate current methods of postoperative outcome measurement in facial palsy. As highlighted within Chapter 1, multiple different solutions have been implemented however none of these have yet to gain widespread implementation or confirmed to be gold-standard. Most established approaches rely on either patient reporting or clinician grading, both of which are limited by variability and subjectivity. This hampers the inter-observer reliability of measurements both amongst clinicians and between institutions. To overcome this, many centres began to develop computer based tools in the hopes of creating a more objective and common language outcome measurement approach. Initially most solutions were limited by the need for complex equipment that was not readily available. This required institutions to invest in such equipment and for patients to undergo follow up in-person. Newer solutions were designed to utilise everyday equipment and demonstrate promising feasibility however have yet to undergo rigorous testing. What is clear from the literature is that no one approach will sufficiently capture all the data necessary for clinicians to truly gauge the success of their surgery. A multimodal approach using an array of outcome measurement tools has recently been suggested to ensure that all physical and psychological domains of treatment are evaluated. Patient reported outcome measurements are subjective in nature however do provide key insights into the psychosocial aspects of facial palsy and the patients perception of disease severity. The introduction of automated computer-based approaches to clinician grading scales can provide meaningful data on initial disease severity and overall change in function after treatment. What remains elusive is a robust means of quantifying postoperative smile spontaneity. For dynamic facial reanimation surgery, a key goal for surgeons is to restore spontaneous smile to allow for patients to appear 'normal' during organic social reanimation. Methods of spontaneous smile quantification have only recently been developed and most rely on either scarcely available equipment or time-consuming manual input.

The lack of a tool that allows clinicians to quickly and reliably quantify quality of spontaneous smile is a major obstacle within the facial reanimation field. This prompted the development of a novel mobile phone app – 'SmileCheck'. This app was designed to automatically evaluate spontaneous smiling in postoperative patients by recording patients whilst displaying humorous content. The app then harnesses an artificial intelligence algorithm to evaluate the recordings and quantify the quality of smiles. During development, ease of use was a key focus in order to promote usage and repeated measurements amongst patients. It is a python (Python Software Foundation, USA) based mobile application that can be used on every-day mobile phones and tablets. This increases the accessibility of the application for patients and allows for outcome measurement to be performed remotely and wherever easiest for the patient. The app provides easy to understand prompts that guide the user through the process. Each testing takes a total of 2 minutes, and all content can be updated to remain fresh and humorous for repeat users. The aim is for patients to have this app downloaded on their mobile phones and to get interval reminders to undergo testing. Results of each measurement is then sent to clinicians and provides vital long term outcome data which is pivotal in facial reanimation where protracted recovery is expected.

Within Chapter 2, the ability for the 'SmileCheck' application to automatically quantify spontaneous smile within a facial palsy cohort was demonstrated. App based spontaneous smile recognition compared well with manual observation and provided rapid and easy to understand results. During the study, the app was able to successfully detect a smile in every testing episode and demonstrated a high degree of accuracy when compared to manual observation. The results of this study suggest that this app may provide a feasible method of spontaneity analysis that could be utilised across multiple institutions.

3.2 Future Avenues

In this thesis, the ‘SmileCheck’ app was only tested within one moment in time. Further studies evaluating the app’s ability to measure changes over time and particularly after reanimation would be of great value. This would test the app’s utility as a serial outcome measurement tool. If subsequent testing confirmed this, implementing this tool across multiple institutions would create a large dataset of postoperative outcomes and provide clinicians with accurate understandings of expected postoperative outcomes from various facial reanimation techniques.

Appendix

Pilot Study: Determining Movement Limit Thresholds

Introduction

The 'SmileCheck' app is a novel mobile phone application designed to perform rapid, objective and uniform smile assessment in facial palsy patients. It employs a novel hemi-facial mirroring technique to provide a composite symmetrical smile that can be detected by emotional analysis algorithms. The relative position of the user's face to the device can greatly impact on the app's identification of midline and thus produced a skewed and unrecognisable mirrored image (Figure 1). This pilot study aimed to determine permissible limits of movement of the device relative to the user by analysing accuracy metrics in healthy volunteers testing the app functionality.

Methods

Hemi-facial mirroring relies on recordings being taken from the frontal aspect to produce a symmetrical and interpretable composite face. Relative movement of the user or device can result in an off-centre mirror and produce a skewed and unrecognisable mirrored face (Figure 1). Permissible limits of device movement in the pitch and yaw axes were determined in this preliminary validation study. [Data generated to determine the limits of movement are available in Figure 2]. Frame by frame data from 60 recordings of 30 healthy volunteers were analysed through the mobile application. Accuracy within each degree of movement was calculated by comparing the application determined presence of smile between each corresponding frame of the simultaneous frontal facial and mirrored facial recording videos. Movement thresholds were then determined based on clinical utility and ease of App use.

The App compensates for roll (third axis) and as such roll was fixed at a maximum absolute value of 15 degrees.

Results

60 spontaneous smile recordings of 30 healthy volunteers were utilised for analysis. Accuracy data for each degree of movement in both pitch and yaw planes is demonstrated in supplemental digital content. Overall, balanced accuracy, specificity and negative predictive value varied little between 0 and 15 degrees. An absolute limit of 10 degrees of movement in these planes was selected based on empiric usage demonstrating increased mirror image skewedness beyond this.

Conclusion

The 'SmileCheck' app's hemifacial mirroring function demonstrated consistent accuracy when the position of the device relative to user remained within 15 degrees. After empiric usage, a maximum limit of 10 degrees of movement has been assigned to the app's quality control measures.

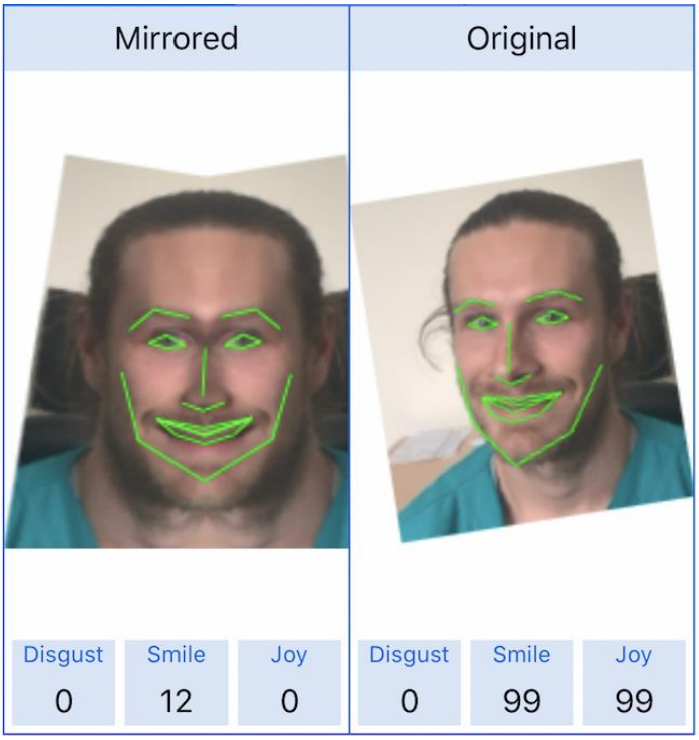

Figure 1: Skewed and unrecognisable mirrored image due to excessive relative movement of user and device

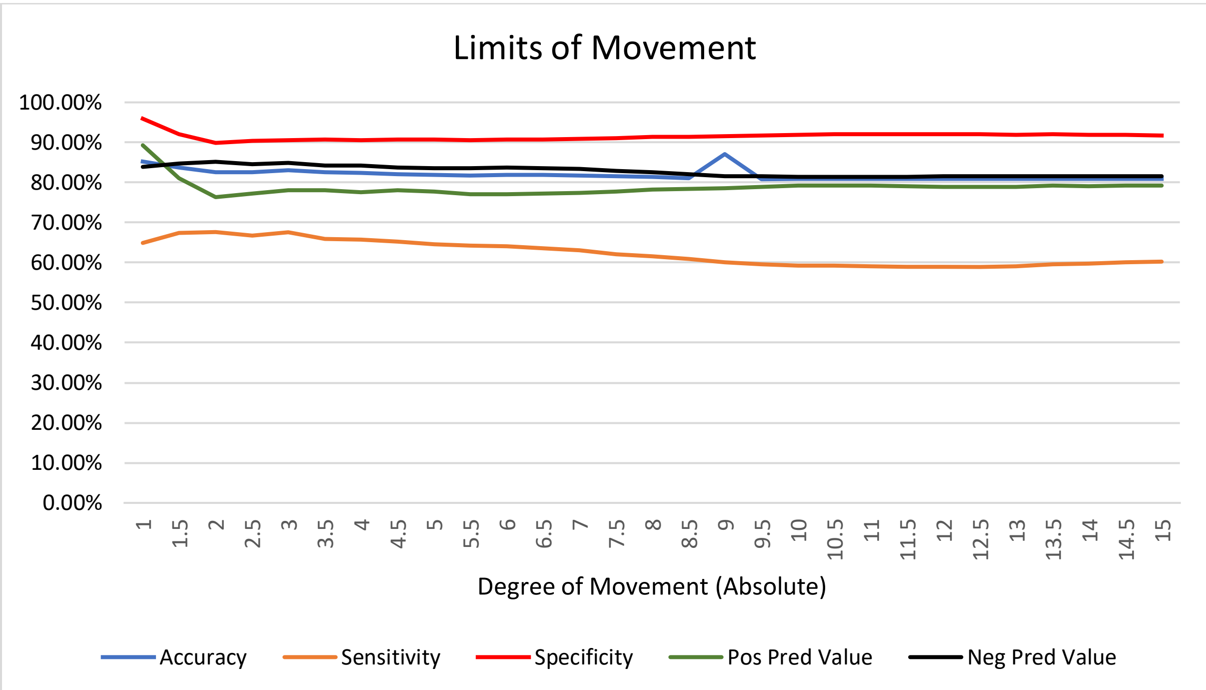

FIgure 2: Accuracy metrics at differing degrees of movement

Ethics Submission

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Monday, 14 February 2022

Assoc Prof Sydney Ch'ng
Central Clinical School: Surgery; Faculty of Medicine and Health
Email: Sydney.chng@sydney.edu.au

Dear Sydney,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.:	**2022/011**
Project Title:	**Remote outcome tracking in facial reanimation surgery validation of a novel mobile application**
Authorised Personnel:	**Ch'ng Sydney; Dusseldorp Joseph; Fuzi Jordan; Meller Catherine; Hadlock Tessa;**
Approval Period:	**14/02/2022** to **14/02/2026**
First Annual Report Due:	**14/02/2023**

Documents Approved:

Date Uploaded	Version Number	Document Name
28/01/2022	Version 2	participation email V2
28/01/2022	Version 3	PCF/PIS updated V3

Special Condition/s of Approval

- PIS

Please check the PIS for typographical errors and correct these (eg. devlopment and conact)

- Consent

We don't require a witness signature on the consent forms, we would recommend you remove the requirement for the consent to be witnessed.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.

- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely,

Associate Professor **Haryana Dillon**
Chair
Human Research Ethics Committee (HREC 3)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2018) and the NHMRC's Australian Code for the Responsible Conduct of Research (2018)

Journal Submission 1 – Under Review

Mary Ann Liebert, Inc.

Journal Submissions

Facial Plastic Surgery & Aesthetic Medicine

Current and novel approaches to outcome measurement in facial reanimation surgery - A review

Journal:	*Facial Plastic Surgery & Aesthetic Medicine*
Manuscript ID	Draft
Manuscript Type:	Narrative Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Fuzi, Jordan; Prince of Wales Hospital and Community Health Services, Otolaryngology Head & Neck Surgery; The University of Sydney Faculty of Medicine and Health Meller, Catherine; Prince of Wales Hospital and Community Health Services Ch'ng, Sydney; The University of Sydney Faculty of Medicine and Health; Chris O'Brien Lifehouse Dusseldorp, Joseph; The University of Sydney Faculty of Medicine and Health; Chris O'Brien Lifehouse; Concord Repatriation General Hospital
Primary Subject:	Facial Nerve
Manuscript Keywords (Search Terms):	Facial reanimation, Facial Palsy, Outcome Measurement, Artificial intelligence, Machine Learning
Secondary Subject:	Facial Symmetry, Outcomes, Surgical Management, Reconstruction
Abstract:	Importance: Accurate quantification of postoperative outcomes after facial reanimation surgery is essential for the individualisation of treatment and ensuring that each patient's unique needs are met. The objective of this review is to detail established and emerging approaches to outcome measurement. Observations: Currently established patient reported outcome measures provide key insights into disease related patient experience. Clinician graded scales have historically been used to quantify disease severity however are limited by their subjective nature and inter-observer variability. More recently, computer-based approaches leveraging machine learning(ML) technology have been developed to improve the objectivity of disease severity assessment. This has shown promise both in improving the reliability of well-established clinical scales and in the production of newer automated disease severity scales. Whilst restoration of spontaneous emotional activity is a key goal for patient and surgeon, a practical method of accurately quantifying smile spontaneity has yet to be established. Conclusions: Whilst a multi-modal approach has been proposed to capture the wide array of pertinent outcomes, the exact tools have yet to be agreed upon. ML technology has shown increasing promise in providing automated solutions that can increase both the reliability of measurements and practicality in implementation.

Current and novel approaches to outcome measurement in facial reanimation surgery

- ***A review***

Dr Jordan Fuzi[1,2]

Dr Catherine Meller[1]

Prof Sydney Ch'ng[2,3]

Dr Joseph Dusseldorp[2,3,4]

1. Prince of Wales Hospital, Randwick, NSW, Australia
2. University of Sydney, Camperdown, NSW, Australia
3. Chris O'Brien Lifehouse, Camperdown, NSW, Australia
4. Concord Repatriation General Hospital, Concord, NSW, Australia

Corresponding Author

Dr Jordan Fuzi

Email: fuzijordan@gmail.com

Running Title

Outcome measurement in facial reanimation surgery

Word count – Abstract: 193 words, Main body: 2882

Abstract

Importance: Accurate quantification of postoperative outcomes after facial reanimation surgery is essential for the individualisation of treatment and ensuring that each patient's unique needs are met. The objective of this review is to detail established and emerging approaches to outcome measurement.

Observations: Currently established patient reported outcome measures provide key insights into disease related patient experience. Clinician graded scales have historically been used to quantify disease severity however are limited by their subjective nature and inter-observer variability. More recently, computer-based approaches leveraging machine learning(ML) technology have been developed to improve the objectivity of disease severity assessment. This has shown promise both in improving the reliability of well-established clinical scales and in the production of newer automated disease severity scales. Whilst restoration of spontaneous emotional activity is a key goal for patient and surgeon, a practical method of accurately quantifying smile spontaneity has yet to be established.

Conclusions: Whilst a multi-modal approach has been proposed to capture the wide array of pertinent outcomes, the exact tools have yet to be agreed upon. ML technology has shown increasing promise in providing automated solutions that can increase both the reliability of measurements and practicality in implementation.

Introduction

Facial reanimation surgery compromises various surgical procedures aimed at restoring facial symmetry and dynamic facial function in patients with facial palsy (FP). Whilst improving both static and dynamic facial symmetry is of great benefit, a key goal for both patient and surgeon is the restoration of spontaneous smile and for the patient to appear normal and joyous during organic social interaction[1]. This was demonstrated through a discrete choice experiment where healthy volunteers were shown to be accepting of a significantly higher chance of treatment failure, necessitating additional surgery, to achieve spontaneous smile over voluntary smiling [2]. In the acute and subacute setting, approaches to facial nerve reconstruction include direct nerve repair, utilisation of interposition or nerve transfer grafts or via direct muscle neurotisation [3–5]. For long standing paralysis, free muscle transfer is required and reinnervated either with the contralateral facial nerve, nearby non-facial cranial nerves, or both[6]. With regards to spontaneity, utilisation of a cross facial nerve graft (CFNG) has historically been considered the optimal neural source for restoration of emotional smile. However, it has long been recognised that cross facial nerve grafting has a higher risk of treatment failure particularly in older patients. As a result, numerous other nerve sources have been utilised including the nerve to masseter (NTM), hypoglossal, spinal accessory or combinations of multiple nerve sources. Various small volume case series of these techniques have reported promising spontaneity results however their methods of outcome measurement varied greatly. A recent systematic review of all papers found that 63% of studies did not report their methodology of assessment and of those that did the majority used clinical observation or patient reporting [7]. Herein lies a major obstacle in the field of facial reanimation surgery, a lack of a universal spontaneity outcome measurement tool. A universal and objective measurement tool would allow for accurate characterisation of disease severity and treatment response both within and amongst institutions. Current methods of assessment of facial palsy include patient reported outcome measures (PROMs) and clinician-based grading systems [8]. Whilst these are easy to implement and provide important information on the psychosocial impact and patient perceived disease burden, these assessments are subjective in nature and susceptible to intra- and interobserver bias. Several methods for objective spontaneity assessment have been suggested. In recent years, computer based approaches are becoming more widely employed to study facial palsy[9,10]. Initially, these approaches relied upon manual identification of facial landmarks with subsequent software analysis of relevant distances and angles of movement. Such techniques provide accurate quantification of landmark movement and symmetry, however do not provide spontaneity or layperson measurements. Other techniques relied upon more complex software algorithms and 3-dimensional technologies which limited their widespread use [11–16]. More recently, machine learning (ML) based computer algorithms have been demonstrated to accurately predict the position of facial landmarks without the need for manual human input [17,18]. These algorithms are now being utilised within the facial palsy domain and have been demonstrated to accurately measure the emotionality and quality of spontaneous smiles [19]. Herein, this review aims to outline current methods of subjective facial palsy assessment (*Table 1*) and discuss emerging computer based tools designed to provide an objective outcome measure for facial reanimation surgery (*Table 2*).

Subjective Assessment

Patient Reported Outcome Measures

Whilst the physical sequalae of facial palsy (FP) are easily identifiable by the layperson and clinician, the negative psychosocial effects of this disease must not be overlooked. It has been well established that facial palsy has a significantly negative impact on a patient's emotional and psychosocial wellbeing [20]. This impact is often complex and individualised to the patient with recent studies recognising that the degree of functional impairment does not always correlate with patient distress levels [21,22]. As such, patient reported outcome measures (PROMs) play an integral role in evaluating the individualised psychosocial effects of facial palsy and their response to treatment [8]. The first PROM designed to evaluate quality of life in facial palsy specifically was the Facial Disability Index (FDI) [23]. First published in 1996, this questionnaire assesses psychosocial wellbeing and physical function through 5 questions each. During validation, FDI subscales were associated with each other and produced reliable measurements for patient focused disability. The Facial Clinimetric Evaluation (FACE) scale was subsequently developed in 2001 with an aim to measure both quality of life and facial impairment [24]. This survey utilises 15 questions on a 5 point likert scale to evaluate facial movement, facial comfort, oral function, eye comfort, lacrimal control and social function. A recent review suggests that the FACE scale may better evaluate for psychological outcomes than the FDI [25]. Neither scale however, evaluates for the impact of synkinesis prompting the development of the Synkinesis Assessment Questionnaire (SAQ) [26]. Whilst not directly measuring quality of life, the SAQ demonstrates patient perception of the presence and impact of synkinesis producing a total score from 20 (worst) to 100 (best). A newer module of the FACE-Q questionnaire has since been developed to directly assess the outcomes of functional facial differences on children and young adults [27]. This outcme measure has shown promising results in the evaluation of FP in children and young adults however requires ongoing clinical evaluation [28]. Most recently, a

Canadian group developed the 25 item Alberta Facial clinical evaluation scale aimed at assessing key domains of concern from the patients perspective [29]. Whilst this was developed through interviewing FP patients on their key functional and psychological concerns, this questionnaire has yet to undergo rigorous testing.

<u>Clinician Graded Scoring Systems</u>

For many years clinician based grading tools have been utilised as the most objective means of facial nerve functional assessment. The most widely used of these tools is the House-Brackmann Grading Scale (HBGS) [30]. This six point scale is useful in measuring large changes in function over time and has high inter- and intra-observer reliability [30–33]. However, the lack of regional subdomain scores does limit this scales ability to act as an outcome measurement tool in facial reanimation where surgeons are interested in subtle changes in function of specific facial regions[8]. Terzis and Noah proposed a 5 tier classification system to describe post facial reanimation outcomes based on facial symmetry and contraction [34]. It is able to quantify changes over time especially after surgery however has yet to undergo rigorous testing. The Sunnybrook Facial Grading Scale (SFGS) was created as a regional-based scoring system that directly assesses facial symmetry at rest, symmetry with voluntary movements and synkinesis [35]. Through assessing specific facial regions, the SFGS is more sensitive to changes in facial function than the HBGS with similarly high reliability [31,32,36]. Whilst comprehensive, this tool is cumbersome to employ and requires subjective interpretation of facial function. The Sydney Facial Grading System similarly employs a regional based scoring approach by directly scoring the five main branches of the facial nerve [32]. This scale is practical and intuitive however has yet to undergo rigorous reliability testing or be validated for repeated measures over time [8]. The Facial Nerve Grading System 2.0 was designed to assess four distinct facial regions for degree of movement and synkinesis and can be converted to an equivalent HBGS score [37]. This scale has been shown to be reliable between observers and has a moderate agreement with the HBGS in evaluating change in function over time [37,38]. Most recently, Banks and colleagues developed the Electronic Facial Paralysis Assessment tool (eFACE) aimed at providing clinicians with a digital solution to facial nerve assessment [39]. The clinician scores 15 facial domains with regards to static appearance, dynamic movement and synkinesis and an easy to interpret total sum score between 0-100 is produced, with 100 representing normal facial function. The eFACE has been demonstrated to be useful in assessing synkinesis and change in function overtime with high interobserver, intraobserver and test-retest reliability[39,40]. Whilst each of these tools carry their own merit, it is important to recognise that they are by nature subjective tools and are prone to bias and human error.

Computer Based Assessment

Computer based approaches to smile assessment have long been studied as potential objective outcome measures for facial reanimation. Most early techniques were designed to quantify smile based on degree and synchronicity of oral commissure excursion. These approaches demonstrated a high degree of accuracy in measurements however their widespread use across centres was limited by their reliance upon specialty equipment such as 3D cameras, facial markers or handheld scanners [11,13,41–44]. To overcome this, in 2012 Hadlock and colleagues developed the Facial Assessment by Computer Evaluation (FACEgram) software [45]. This freely available Java based software program was demonstrated to accurately quantify facial movement and did not require complex equipment. However, this tool has not been widely adopted as it requires manual user identification of facial landmarks, which is time-consuming and may vary greatly between users. Facial analysis algorithms based off machine learning technology may overcome both the need for specialty equipment or human input by performing automated facial landmark localisation, quantification of facial movements and prediction of emotional expressions[46–48]. These algorithms have been implemented within facial palsy research to produce automated scoring of existing grading scales, production of new automated scales and emotional expressivity analysis.

Automated Facial Landmarking and Symmetry Analysis

Multiple centres have harnessed ML based automatic facial landmarking to quantify facial symmetry both in repose and during standardised expressions. The 'Emotrics' software was developed to employ ML technology to automatically identify facial landmarks from a set of standard frontal still photographs [17]. Multiple photographs can be analysed and position, distance and symmetry of landmarks is automatically calculated and used to characterise severity of palsy and reflect quality of smile. The algorithms used in Emotrics were initially trained on a large database of healthy facial photographs raising concerns around the accuracy of landmark localisation in facial palsy faces. Emotrics was subsequently retrained with a facial palsy dataset, greatly increasing accuracy of landmark localisation[49]. Lee and colleagues developed the PC based Facial Asymmetry Assessment Program (PC-FAAP). This smartphone based program automatically calculates a novel FP grading system based off ML analysis of still photographs of three key facial expressions [50]. Mouth, eyebrow and eye closure asymmetry ratios are determined and summated to develop a composite 'FNP grading scale (FGS)'. Through validation, the FGS was shown to be sensitive to change in severity of FP and was more consistent than subjective assessment however it has yet to be widely adopted across centres. Hidaka and colleagues describe a novel grading system that provides real-time facial symmetry analysis [51]. This approach automatically identifies 68 facial landmarks on video recordings and calculates eyebrow and oral symmetry through representation of 'displacement ratios'. This tool may aid clinicians both in tailoring and quantifying outcomes of reanimation however has yet to undergo rigorous testing. Lastly, Monini and colleagues have developed a markerless videosystem for staging facial palsy by comparing total movement of forehead frowning and smiling between healthy and palsy hemi-face [44,52]. This smartphone based system utilises the 'Emotrics' algorithm[17] to identify facial landmarks in videos and calculates degree of movement of the oral commissure and eyebrow. Within this study, automatic derived facial grading was consistent with previously used subjective methods.

<u>Automated Clinician-Graded Scales</u>

Whilst limited by the subjectivity and variability of observer interpretation, clinician graded scales do provide a common language assessment of facial palsy severity. With the advent of ML technology, multiple centres have leveraged these algorithms to eliminate the need for clinician input to improve both the objectivity and consistency of these scales. O'reilly first developed the 'Facogram' in 2010 which uses trained artificial neural networks (ANN) to produce both objective HBGS scores and regional facial grading in facial palsy patients [53].

This java-based system requires a video camera for data capture and standard laptop for data processing. 'Facogram' based assessment had good agreement with subjective assessment and demonstrated consistency on repeated testing. Expanding on previous work, Mothes and colleagues developed a convolutional neural network (CNN) based ML program to automatically grade facial palsy based on HBGS, SFGS and Stennert index [54,55]. Fair agreement between subjective and automated grading of SFGS was found however no agreement was demonstrated for HBGS or Stennert index. In 2021, the Auto-eFACE tool was created by combining the 'Emotrics' algorithm with the commonly used eFACE clinician graded scoring system [56]. This easy-to-use software can provide rapid and automated eFACE assessments and only requires readily available commercial computers. Most recently, Jirawatnotai and colleagues developed the 'SBface' mobile phone application to provide computerised FP assessments based of the SFGS [57]. Facial landmarks are identified using the 'VNFaceObservation' (Apple Inc, USA) image analysis technology and distances between these landmarks are calculated during the 6 standardised facial expressions listed in the SFGS. Distances between landmarks are correlated to the SFGS subdomains and a composite total score is derived. Whilst repeatability of app-derived scores was demonstrated, inconsistent correlation between app and human observation was found.

Emotional Expressivity Analysis

The majority of previously described grading systems utilise facial symmetry both as a measure of disease severity and post-operative outcome. Recognising a strong desire of patients to appear 'normal' to the layperson, Dusseldorp and colleagues describe a new approach to reanimation outcome measurement by quantifying the emotional expressivity of smiles in FP patients [1]. This group used the commercially available 'Affdex' (Affectiva, Boston) emotional analysis software to measure changes in the emotional expressivity of spontaneous smile in FP patients after reanimation. This study demonstrated a dramatic improvement in expressivity in postoperative flaccid FP patients and a reduction in negative emotion expression in the synkinetic cohort. Extending on this, the same group has since developed an automated solution to spontaneous smile analysis by combining this approach with the 'Emotrics' technology. Time points of healthy oral commissure movements as detected by 'Emotrics' are subsequently analysed by the 'Affdex' algorithm to automatically provide an emotional expressivity value for smiles [19]. This approach was utilised to measure spontaneous smile emotional expression after different reanimation approaches however has yet to gain widespread use due to the need for powerful computer processing power.

Discussion

It is well known that the sequalae of facial palsy include both physical and psychosocial effects that vary greatly in severity between patients [20]. As such, disease severity and surgical outcome measurement requires a multi-pronged approach that can evaluate the many facets of this disease [58]. Patient reported outcome measurements are limited by their subjective nature but provide key insights into the psychosocial detriment of the paralysis and of the patient's disease-related experience. The validity and reliability of these scales have been optimised over the years and a clear consensus on which scale to use between centres would greatly improve the relatability of outcome data worldwide. Clinician graded scales vary significantly with regards to complexity but do provide easy to understand measurements of disease severity. Similar to PROMs, these scales still require subjective observer-based grading, limiting the relatability between observers and institutions. Over the past decade, machine learning algorithms have shown promise in overcoming these limitations by providing objective measurement solutions. The creation of automated methods of calculating clinician graded scales may be able to overcome current concerns regarding subjectivity and provide a common language tool amongst centres. This technology has also resulted in the advent of a new method of outcome measurement by measuring emotional expression to quantify quality of smile. For these approaches to gain widespread use they need to be able to be performed on readily available hardware in any location. This would prevent the limitation of accessibility between institutions and demographics and provide remote outcome measurement solutions.

Conclusion

Outcome measurement in facial reanimation requires a multimodal approach that encompasses evaluation of all facets of FP. Computer vision based artificial intelligence has shown promising results in both improving current outcome measurement methods and providing new solutions. This will hopefully contribute to a greater understanding of the benefits of surgery and guide clinicians in optimising patient care.

Acknowledgements

Nil

Conflict of Interest and Funding Statement

No funding was obtained to undertake this study. These Authors declare no conflicts of interest.

Authorship Confirmation/Contribution/Statement

JF: Conceptualisation(supporting) , data curation (lead) , formal analysis (lead), writing - 0riginal draft (lead), writing (equal), methodology (lead), **CM:** writing – review & editing (equal), supervision (supporting). **SC**: writing – review & editing (equal), supervision (supporting). **JD:** conceptulatisation (lead), Supervision (lead),, writing – review & editing (lead).

References

1. Dusseldorp JR, Guarin DL, van Veen MM, et al. In the Eye of the Beholder: Changes in Perceived Emotion Expression after Smile Reanimation. Plast Reconstr Surg 2019;144(2):457–471; doi: 10.1097/PRS.0000000000005865.

2. Dusseldorp JR, Naunheim MR, Quatela O, et al. Neurotization Preferences in Smile Reanimation: A Discrete Choice Experiment. Plast Reconstr Surg 2021;148(3):407e; doi: 10.1097/PRS.0000000000008302.

3. Sun DQ, Andresen NS, Gantz BJ. Surgical Management of Acute Facial Palsy. Otolaryngol Clin North Am 2018;51(6):1077–1092; doi: 10.1016/j.otc.2018.07.005.

4. Klebuc M, Shenaq SM. Donor nerve selection in facial reanimation surgery. Semin Plast Surg 2004;18(1):53–60; doi: 10.1055/s-2004-823124.

5. Garcia RM, Hadlock TA, Klebuc MJ, et al. Contemporary solutions for the treatment of facial nerve paralysis. Plast Reconstr Surg 2015;135(6):1025e–1046e; doi: 10.1097/PRS.0000000000001273.

6. Natghian H, Fransén J, Rozen SM, et al. Qualitative and Quantitative Analysis of Smile Excursion in Facial Reanimation: A Systematic Review and Meta-analysis of 1- versus 2-stage Procedures. Plast Reconstr Surg – Glob Open 2017;5(12):e1621; doi: 10.1097/GOX.0000000000001621.

7. van Veen T. Facial Palsy: Treatment, Quality of Life, and Assessment. Thesis fully internal (DIV). University of Groningen; 2020; doi: 10.33612/diss.131756313.

8. Dusseldorp JR, Veen MM van, Mohan S, et al. Outcome Tracking in Facial Palsy. Otolaryngol Clin North Am 2018;51(6):1033–1050; doi: 10.1016/j.otc.2018.07.003.

9. Gerós A, Horta R, Aguiar P. Facegram – Objective quantitative analysis in facial reconstructive surgery. J Biomed Inform 2016;61:1–9; doi: 10.1016/j.jbi.2016.03.011.

10. Bray D, Henstrom DK, Cheney ML, et al. Assessing Outcomes in Facial Reanimation: Evaluation and Validation of the SMILE System for Measuring Lip Excursion During Smiling. Arch Facial Plast Surg 2010;12(5):352–354; doi: 10.1001/archfaci.2010.69.

11. Frey M, Giovanoli P, Gerber H, et al. Three-dimensional video analysis of facial movements: a new method to assess the quantity and quality of the smile. Plast Reconstr Surg 1999;104(7):2032–9.

12. Coulson SE, Croxson GR, Gilleard WL. Three-dimensional quantification of "still" points during normal facial movement. Ann Otol Rhinol Laryngol 1999;108(3):265–268; doi: 10.1177/000348949910800308.

13. Frey M, Jenny A, Giovanoli P, et al. Development of a new documentation system for facial movements as a basis for the international registry for neuromuscular reconstruction in the face. Plast Reconstr Surg 1994;93(7):1334–1349; doi: 10.1097/00006534-199406000-00002.

14. Meier-Gallati V, Scriba H, Fisch U. Objective scaling of facial nerve function based on area analysis (OSCAR). Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1998;118(4):545–50.

15. Neely JG, Cheung JY, Wood M, et al. Computerized quantitative dynamic analysis of facial motion in the paralyzed and synkinetic face. Am J Otol 1992;13(2):97–107.

16. Yuen K, Inokuchi I, Maeta M, et al. Evaluation of facial palsy by moiré topography index. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1997;117(5):567–572; doi: 10.1016/s0194-5998(97)70032-9.

17. Guarin DL, Dusseldorp J, Hadlock TA, et al. A Machine Learning Approach for Automated Facial Measurements in Facial Palsy. JAMA Facial Plast Surg 2018;20(4):335–337; doi: 10.1001/jamafacial.2018.0030.

18. Fan H, Zhou E. Approaching human level facial landmark localization by deep learning. Image Vis Comput 2016;47:27–35; doi: 10.1016/j.imavis.2015.11.004.

19. Dusseldorp JR, Guarin DL, van Veen MM, et al. Automated Spontaneity Assessment after Smile Reanimation: A Machine Learning Approach. Plast Reconstr Surg 2022;149(6):1393–1402; doi: 10.1097/prs.0000000000009167.

20. Fuzi J, Taylor A, Sideris A, et al. Does Botulinum Toxin Therapy Improve Quality of Life in Patients with Facial Palsy?. Aesthetic Plast Surg 2020;44(5):1811–1819; doi: 10.1007/s00266-020-01870-4.

21. Gyori E, Przestrzelski C, Pona I, et al. Quality of life and functional assessment of facial palsy patients: A questionnaire study. Int J Surg Lond Engl 2018;55(101228232):92–97; doi: 10.1016/j.ijsu.2018.04.061.

22. Bruins TE, van Veen MM, Werker PMN, et al. Associations Between Clinician-Graded Facial Function and Patient-Reported Quality of Life in Adults With Peripheral Facial Palsy: A Systematic Review and Meta-analysis. JAMA Otolaryngol-- Head Neck Surg 2021;147(8):717–728; doi: 10.1001/jamaoto.2021.1290.

23. VanSwearingen JM, Brach JS. The Facial Disability Index: reliability and validity of a disability assessment instrument for disorders of the facial neuromuscular system. Phys Ther 1996;76(12):1288–1298; discussion 1298-1300; doi: 10.1093/ptj/76.12.1288.

24. Kahn JB, Gliklich RE, Boyev KP, et al. Validation of a patient-graded instrument for facial nerve paralysis: the FaCE scale. The Laryngoscope 2001;111(3):387–98.

25. Ho AL, Scott AM, Klassen AF, et al. Measuring quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient-reported outcome measures. Plast Reconstr Surg 2012;130(1):91–99; doi: 10.1097/PRS.0b013e318254b08d.

26. Mehta RP, WernickRobinson M, Hadlock TA. Validation of the Synkinesis Assessment Questionnaire. The Laryngoscope 2007;117(5):923–6.

27. Longmire NM, Wong Riff KWY, O'Hara JL, et al. Development of a New Module of the FACE-Q for Children and Young Adults with Diverse Conditions Associated with Visible and/or Functional Facial Differences. Facial Plast Surg FPS 2017;33(5):499–508; doi: 10.1055/s-0037-1606361.

28. Klassen AF, Rae C, Gallo L, et al. Psychometric Validation of the FACE-Q Craniofacial Module for Facial Nerve Paralysis. Facial Plast Surg Aesthetic Med 2022;24(1):1–7; doi: 10.1089/fpsam.2020.0575.

29. Hamour AF, Mendez AI, Biron VL, et al. Development of the Alberta facial clinical evaluation scale: A patient-centred outcomes instrument for facial nerve paralysis. Clin Otolaryngol Off J ENT-UK Off J Neth Soc Oto-Rhino-Laryngol Cervico-Facial Surg 2019;44(6):919–926; doi: 10.1111/coa.13400.

30. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1985;93(2):146–147; doi: 10.1177/019459988509300202.

31. Berg T, Jonsson L, Engström M. Agreement between the Sunnybrook, House-Brackmann, and Yanagihara facial nerve grading systems in Bell's palsy. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol 2004;25(6):1020–1026; doi: 10.1097/00129492-200411000-00027.

32. Coulson SE, Croxson GR, Adams RD, et al. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and

synkinesis after facial nerve paralysis. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 2005;132(4):543–549; doi: 10.1016/j.otohns.2005.01.027.

33. Reitzen S.D., Babb J.S., Lalwani A.K. Significance and reliability of the House-Brackmann grading system for regional facial nerve function. Otolaryngol Head Neck Surg 2009;140(2):154–158; doi: 10.1016/j.otohns.2008.11.021.

34. Terzis JK, Noah ME. Analysis of 100 cases of free-muscle transplantation for facial paralysis. Plast Reconstr Surg 1997;99(7):1905–21.

35. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 1996;114(3):380–386; doi: 10.1016/s0194-5998(96)70206-1.

36. Kanerva M, Jonsson L, Berg T, et al. Sunnybrook and House-Brackmann Systems in 5397 Facial Gradings. Otolaryngol Neck Surg 2011;144(4):570–574; doi: 10.1177/0194599810397497.

37. Vrabec JT, Backous DD, Djalilian HR, et al. Facial Nerve Grading System 2.0. Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg 2009;140(4):445–450; doi: 10.1016/j.otohns.2008.12.031.

38. Lee HY, Park MS, Byun JY, et al. Agreement between the Facial Nerve Grading System 2.0 and the House-Brackmann Grading System in Patients with Bell Palsy. Clin Exp Otorhinolaryngol 2013;6(3):135–139; doi: 10.3342/ceo.2013.6.3.135.

39. Banks CA, Bhama PK, Park J, et al. Clinician-Graded Electronic Facial Paralysis Assessment: The eFACE. Plast Reconstr Surg 2015;136(2):223e–230e; doi: 10.1097/PRS.0000000000001447.

40. Malka R, Miller M, Guarin D, et al. Reliability Between In-Person and Still Photograph Assessment of Facial Function in Facial Paralysis Using the eFACE Facial Grading System. Facial Plast Surg Aesthetic Med 2021;23(5):344–349; doi: 10.1089/fpsam.2020.0300.

41. Hontanilla B, Auba C. Automatic three-dimensional quantitative analysis for evaluation of facial movement. J Plast Reconstr Aesthetic Surg JPRAS 2008;61(1):18–30.

42. Sforza C, Galante D, Shirai YF, et al. A three-dimensional study of facial mimicry in healthy young adults. J Cranio-Maxillofac Surg 2010;38(6):409–415; doi: 10.1016/j.jcms.2009.10.018.

43. Ozsoy U, Uysal H, Hizay A, et al. Three-dimensional objective evaluation of facial palsy and follow-up of recovery with a handheld scanner. J Plast Reconstr Aesthetic Surg JPRAS 2021;74(12):3404–3414; doi: 10.1016/j.bjps.2021.05.003.

44. Monini S, Marinozzi F, Atturo F, et al. Proposal of a Video-recording System for the Assessment of Bell's Palsy: Methodology and Preliminary Results. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol 2017;38(8):1178–1185; doi: 10.1097/MAO.0000000000001498.

45. Hadlock TA, Urban LS. Toward a universal, automated facial measurement tool in facial reanimation. Arch Facial Plast Surg 2012;14(4):277–82; doi: 10.1001/archfacial.2012.111.

46. Bartlett MS, Littlewort G, Lainscsek C, et al. Machine Learning Methods for Fully Automatic Recognition of Facial Expressions and Facial Actions. In: 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583) 2004; pp. 592–597 vol.1; doi: 10.1109/ICSMC.2004.1398364.

47. Stöckli S, Schulte-Mecklenbeck M, Borer S, et al. Facial expression analysis with AFFDEX and FACET: A validation study. Behav Res Methods 2018;50(4):1446–1460; doi: 10.3758/s13428-017-0996-1.

48. Baskar A, Gireesh Kumar T. Facial Expression Classification Using Machine Learning Approach: A Review. In: Data Engineering and Intelligent Computing. (Satapathy SC, Bhateja V, Raju KS, et al. eds). Advances in Intelligent Systems and Computing Springer: Singapore; 2018; pp. 337–345; doi: 10.1007/978-981-10-3223-3_32.

49. Guarin DL, Yunusova Y, Taati B, et al. Toward an Automatic System for Computer-Aided Assessment in Facial Palsy. Facial Plast Surg Aesthetic Med 2020;22(1):42–49; doi: 10.1089/fpsam.2019.29000.gua.

50. Lee DY, Kim HS, Kim SY, et al. Comparison between Subjective Scoring and Computer-Based Asymmetry Assessment in Facial Nerve Palsy. J Audiol Otol 2019;23(1):53–58; doi: 10.7874/jao.2018.00318.

51. Hidaka T, Kurita M, Ogawa K, et al. Application of Artificial Intelligence for Real-Time Facial Asymmetry Analysis. Plast Reconstr Surg 2020;146(2):243e–245e; doi: 10.1097/PRS.0000000000007035.

52. Monini S, Ripoli S, Filippi C, et al. An objective, markerless videosystem for staging facial palsy. Eur Arch Otorhinolaryngol 2021;278(9):3541–3550; doi: 10.1007/s00405-021-06682-z.

53. O'Reilly BF, Soraghan JJ, McGrenary S, et al. Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol 2010;31(3):486–491; doi: 10.1097/MAO.0b013e3181c993dc.

54. Haase D, Minnigerode L, Volk GF, et al. Automated and objective action coding of facial expressions in patients with acute facial palsy. Eur Arch Otorhinolaryngol 2015;272(5):1259–1267; doi: 10.1007/s00405-014-3385-8.

55. Mothes O, Modersohn L, Volk GF, et al. Automated objective and marker-free facial grading using photographs of patients with facial palsy. Eur Arch Otorhinolaryngol 2019;276(12):3335–3343; doi: 10.1007/s00405-019-05647-7.

56. Miller MQ, Hadlock TA, Fortier E, et al. The Auto-eFACE: Machine Learning-Enhanced Program Yields Automated Facial Palsy Assessment Tool. Plast Reconstr Surg 2021;147(2):467–474; doi: 10.1097/PRS.0000000000007572.

57. Jirawatnotai S, Jomkoh P, Voravitvet TY, et al. Computerized Sunnybrook facial grading scale (SBface) application for facial paralysis evaluation. Arch Plast Surg 2021;48(3):269–277; doi: 10.5999/aps.2020.01844.

58. Dusseldorp J, Meares C, Hadlock T. A Multi-Modal Approach to Outcome Tracking in Facial Palsy. 2021; pp. 23–30; doi: 10.1007/978-3-030-50784-8_2.

Tables and Figures

Subjective Assessment	
Patient Reported Outcome Measures	**Clinican Graded Scales**
	House Brackmann Grading Scale (HBGS)
Facial Disability Index (FDI)	Terzis and Noah
Facial Clinimetric Evaluation (FACE)	Sunnybrook Facial Grading Scale (SFGS)
Synkinesis Assessment Questionnaire (SAQ)	Sydney Facial Grading System
FACE-Q	Facial Nerve Grading System 2.0 (FNGS 2.0)
Alberta Facial Clinical Evaluation Scale	Electronic Facial Paralysis Assessment Tool (eFACE)

Table 1: Subjective assessment methods

Computer Based Approaches		
Facial Landmarking	**Automated Clinician-graded Scales**	**Emotional Expressivity**
Facial Assessment by Computer Evaluation (FACEgram)	Facogram	Dusseldorp Affdex Tool
Emotrics	Mothes Automated Marker-Free Grading Tool	

PC based Facial Asymmetry Assessment Program (PC-FAAP)	Auto-eFACE	
Monini Markerless Videosystem	Sbface	

Table 2: Computer based assessment approaches

Journal Submission 2 – Published Paper

Facial Plastic Surgery & Aesthetic Medicine
Volume 00, Number 0, 2022

DOI: 10.1089/fpsam.2022.0104

Facial Plastic Surgery & Aesthetic Medicine

Voluntary and Spontaneous Smile Quantification in Facial Palsy Patients: Validation of a Novel Mobile Application

Jordan Fuzi, BMed,[1,2,*] Catherine Meller, MBBS(Hons), FRACS(OHNS), BAppSc(NMT), MCL, MBA,[1]
Sydney Ch'ng, MBBS, PhD, FRACS,[2,3] Tessa M. Hadlock, MD,[4] and Joseph Dusseldorp, MBBS, MS, FRACS[2,3,5]

Abstract

Introduction: Restoration of spontaneous smiling is a key goal in facial reanimation. A major obstacle to quantifying recovery of spontaneous smiling is the current lack of a uniform and objective means of smile quantification.

Objective: To compare the facial movements during smiling in patients with facial paralysis as measured by an automated application and human observers.

Methods: Video recordings of 25 patients with unilateral facial palsy (FP) watching humorous videos were utilized. Application-derived smile timestamping was compared with manual observer interpretation. Internal reliability of measurements was evaluated through a test–retest approach.

Results: Application-derived smile identification demonstrated almost perfect agreement with manual interpretation (kappa 0.861, $p < 0.001$). There was no statistically significant difference in mean number of smiles between detection method ($p = 0.354$). Automated smile identification demonstrated a high degree of specificity (95.4%), accuracy (93.1%), positive-predictive value (94.7%), and negative-predictive value (91.8%). This method demonstrated a high degree of reliability (kappa 0.864, $p < 0.01$).

Conclusion: The novel "SmileCheck" mobile phone application performed accurate and reliable smile quantification in FP patients in comparison with manual observation.

Introduction

Facial palsy (FP) can result in severe orofacial muscle dysfunction and loss of common emotional facial expressions, especially joy. This dysfunction can dramatically impact quality of life and psychosocial well-being.[1,2] Facial reanimation surgeries aim to enable normal appearance during commonplace facial movements such as smiling, laughing, talking, or eating. Recent studies have suggested that a primary goal to most people is the restoration of spontaneous or emotional smiling.[3] However, a major barrier to the identification of treatments that reliably restore emotional smiling is the lack of an objective method to quantify spontaneity. Currently, outcome assessment after smile reanimation surgery is reliant on the perception of experts, patients, and society.[4,5–8]

[1]*Prince of Wales Hospital, Randwick, Australia.*
[2]*University of Sydney, Camperdown, Australia.*
[3]*Chris O'Brien Lifehouse, Camperdown, Australia.*
[4]*Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA.*
[5]*Concord Repatriation General Hospital, Concord, Australia.*

*Address correspondence to: *Jordan Fuzi, BMed, Prince of Wales Hospital, Randwick 2031, Australia*, Email: fuzijordan@gmail.com

KEY POINTS

Question: Can the novel "SmileCheck" application provide an accurate and reliable tool to automatically evaluate smile in facial palsy patients?

Findings: The application's ability to detect smile in facial palsy patients was similar to a manual observer's interpretation.

Meaning: This study has shown that the new "SmileCheck" application can be an accurate tool in automatically assess the quality of a facial palsy patient's smile.

However, due to the complexity and ensuing prolonged rehabilitation of smile reanimation surgeries, both clinician and patient are prone to be affected by confirmation bias. Although rates of smile spontaneity are extremely high with one study demonstrating rates of 70%,[9] other studies have shown no correlation between objectively measured synchronous smiles and clinician-determined smile spontaneity.[10] For the past 50 years, significant effort has been placed on the development of clinician-rated facial grading scales[11–13]; however, these subjective assessments can be cumbersome and inherently susceptible to inter-/intraobserver variability.[14] Objective clinical tools have also been utilized, including sensor-based techniques, electromyography, or functional magnetic resonance imaging, yet these remain limited by availability and cost.[15,16] A welcome tool in facial reanimation surgery would permit rapid, inexpensive, and objective voluntary and spontaneous smile assessment.

This would aid clinicians in quantifying postoperative outcomes, thus yielding greater insight into the success rates of different smile reanimation techniques. Recent advancements in artificial intelligence (AI) have prompted research into the utility of novel techniques in machine learning and computer vision for automatic orofacial assessment and emotion detection.[17–23] Such tools provide automatic localization of facial landmarks and automatic generation of emotion expression probabilities from both still frames and videos.[3] Among patients with unilateral facial paralysis when smiling, this study aims to demonstrate if application-based measurement of facial movement is accurate when compared with observer-based analysis. A second objective is to quantify the reliability of application-derived measurements.

Materials and Methods

Ethics approval to undertake this study was obtained through the University of Sydney, Australia.

"SmileCheck" app (Mass Eye and Ear Infirmary, USA)

This study utilized the newly developed mobile app ''SmileCheck'' designed to provide automatic assessment of voluntary and spontaneous smile ability in FP patients. Assessment of smiling in these patients requires (1) elicitation of an emotional smile in the subject, (2) identification (''timestamping'') of moments when smiles occur, and (3) analysis of the ability of the patient to look

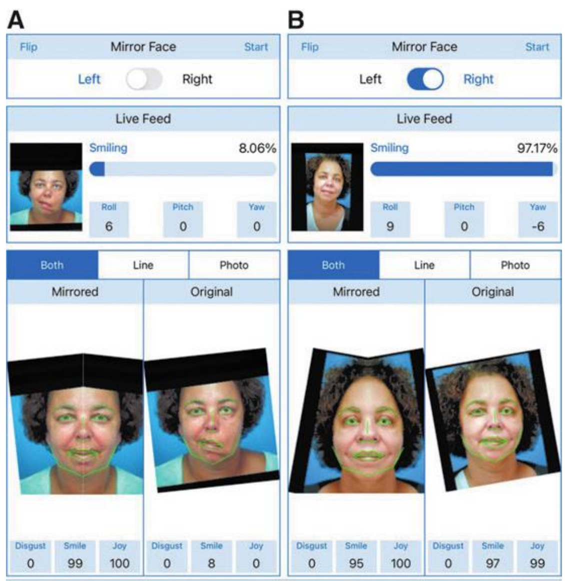

Fig. 1. "SmileCheck" screenshot demonstrating hemifacial mirroring before **(A)** and after **(B)** facial reanimation.

happy during these smile events. To achieve this the user first nominates which side of their face is affected by FP and instructions are provided on optimal environment for use and correct head position. The user is then prompted to perform three voluntary smiles before being shown customizable humorous video content such as the previously validated spontaneous smile assay (SSA).[24]

While the videos are being displayed, the front-facing camera on the device records the user's facial expressions (full-face recording [FFR]). To timestamp the smile events the app uses a hemifacial mirroring technique to create a second video of a composite whole face that only includes the movements of the healthy side of the face (mirrored-face recording [MFR]). The mirroring technique creates a perfectly symmetrical composite smile that can be detected by an emotional analysis algorithm (Fig. 1). The application then rapidly processes both FFR and MFR through the ''Affdex'' (Affectiva, Boston, MA) emotional analysis algorithm to automatically timestamp smile events from the MFR and correlate this to the degree of joy emotion perceived from FFR (Fig. 2).

As per prior methodology, a smile event was defined as occurring when probability of smile in the MFR was perceived as ≥90% for at least 1 s.[10] This minimizes the potential for analysis of nonmeaningful facial movements such as facial flickers or synkinetic movements. The amount of time taken for the app to produce the resultant data files is 10 s (Table. 1) and requires 2 min of the user's time.

Two steps of validation were undertaken to ensure the app achieved the aforementioned functionality consistently.

Step 1: comparison of automatic with manual timestamping smile events

App-derived automated timestamping of spontaneous smile events (SSEs) were compared with manual classification of SSEs using our previously described methods.[10] Accuracy metrics were then determined by comparing number of SSEs detected and time points of smile timestamps between the manual and automatic approaches. Agreement was determined using Cohen's kappa.

Step 2: determination of internal consistency

The reliability of application-derived smile timestamping was evaluated through a test–retest approach. Each recording was analyzed twice by the app. Timestamping of smile events were compared between each evaluation. Cohen's kappa was utilized to determine agreement between the two rounds of analysis.

Statistical analysis

Postprocessing and statistical analysis within this study was undertaken using ''R Project.''[25]

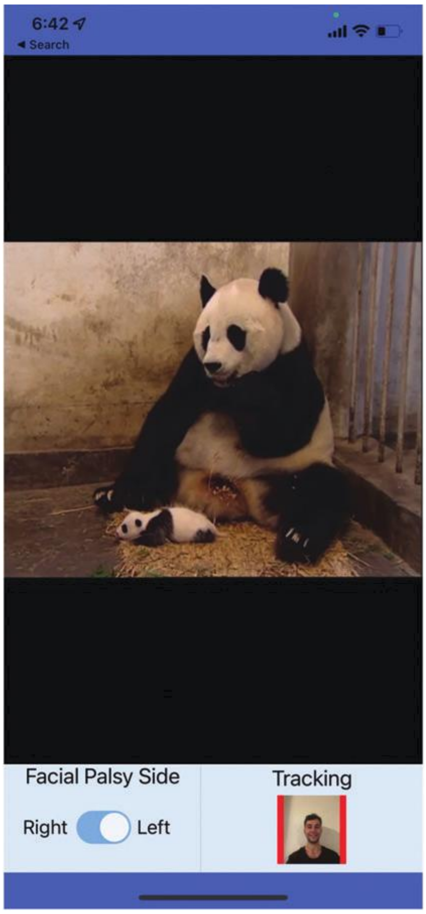

Fig. 2. App user interface.

Results

Step 1: automatic versus manual timestamping approaches

Twenty-five recordings of 25 patients with a unilateral FP were analyzed. Demographic data and etiology of FP are detailed in Table 2. Application-derived smile identification demonstrated an almost perfect agreement with manual interpretation (kappa 0.861, $p < 0.001$). There was no statistically significant difference in mean number of smiles between automated or manual timestamping approaches (5.24 confidence interval [95% CI 4.37–6.11] vs. 5.48 [95% CI 4.63–6.32], $p = 0.354$) (Table 3). In 5/25 patients, an additional seven smiles were detected by the app than by manual observation. Each of these

Table 1. App-derived scorecard detailing voluntary and spontaneous smile events in a healthy test user

Smile-event	*Start_time*	*End_time*	*Duration*	*Aug_ mirror_ smile*	*Aug_ mirror_ joy*	*Aug_ mirror_ disgust*	*Mirror_emq*	*Aug_ orig_ smile*	*Aug_ orig_ joy*	*Aug_ orig_ dis*	*Orig_emq*
Voluntary											
1	9.007389	10.814827	1.8074379	99.44	99.85	0	99.85/−0.00	70.14	68.4	0.05	68.40/−0.05
2	13.625403	16.034628	2.4092245	99.66	99.82	0	99.82/−0.00	74.26	71.61	0.03	71.61/−0.03
3	18.645018	20.653511	2.0084934	99.94	99.88	0	99.88/−0.00	76.93	72.56	0.07	72.56/−0.07
Spontaneous											
1	3.780606	9.600983	5.8203764	99.89	99.93	0	99.93/−0.00	95.17	94.51	0.02	94.51/−0.02
2	29.479548	34.496593	5.017046	99.8	99.92	0	99.92/−0.00	92.2	91.88	0.1	91.88/−0.10
3	37.70924	39.314228	1.6049881	98.97	99.9	0	99.90/−0.00	82.94	79.44	0.06	79.44/−0.06
4	39.721745	42.126263	2.4045181	99.06	99.91	0	99.91/−0.00	99.99	99.92	0	99.92/−0.00
5	42.536243	45.739388	3.203144	99.7	99.92	0	99.92/−0.00	99.96	99.75	0.02	99.75/−0.02
6	50.364613	52.767883	2.4032707	99.86	99.93	0	99.93/−0.00	99.95	99.93	0.01	99.93/−0.01
7	73.64619	78.263466	4.617279	99.63	99.88	0	99.88/−0.00	99.97	99.88	0	99.88/−0.00

smile events were subsequently reanalyzed manually, and it was determined that these small or short smiles were missed by the manual timestamping in six out of seven cases.

Step 2: test–retest reliability

A total of 50 recordings of 25 patients were evaluated in a test–retest manner. App-derived smile identification demonstrated a high degree of consistency between recordings (kappa 0.864, $p < 0.01$). There was no significant difference in mean number of smiles detected between each round of recordings ($p > 0.05$) (Table 4).

Discussion

Herein, we validated an entirely automated objective and uniform approach to the assessment of smiling against human observations. By combining an emotional analysis algorithm with a novel hemifacial mirroring technique, this app prevents the need for multiple AI tools and complex hardware. Within this FP population, app-derived automatic timestamping was highly sensitive and specific when compared with gold standard manual observer timestamping. In each recording, at least one smile was accurately detected by the app, highlighting its success in each use. In addition, throughout the analysis not a single manually observed smile was missed by app interpretation and at least one smile was simultaneously detected by both manual and automated methods in 100% of cases. This automated smile identification demonstrated a high internal reliability during a test–retest approach.

This study is an extension of work performed by a group at MEEI (Boston) aimed at overcoming the current limitations of clinician- or patient-reported smile assessment methods.[9,10,26–29] First, a 1.5 min film clip with superimposed audible laughter (SSA) was developed to elicit at least one spontaneous smile in >95% of both normal and FP subjects.[24] Dusseldorp et al. then undertook a manual approach to the analysis of patients watching such videos to determine the synchronicity of oral commissure excursion after facial reanimation surgery.[10] Although effective, this approach was time consuming and impractical for routine clinical use. This prompted further research into the use of AI to provide an automated solution to spontaneous smile quantification. A concern with traditional algorithms is based on their training on large data sets of ''healthy'' faces and the potential for inaccurate analysis of the asymmetrical FP smile.[30]

Considering this, the same group then developed a two-stage automated approach to spontaneous smile analysis utilizing two AI algorithms.[31] The ''Emotrics'' app (MEEI) was utilized to timestamp SSEs based on healthy

Table 2. Demographics of facial palsy population

No. of participants	25
Mean age (SD)	47.2 (19.6)
Gender, M:F, *n* (%)	9 (36):16 (64)
Flaccid:nonflaccid, *n* (%)	17 (68):8 (32)
Etiology of palsy, *n* (%)	
Parotid tumor	4 (16)
Brainstem tumor	4 (16)
CNS tumor	2 (8)
Iatrogenic	4 (16)
Acoustic neuroma	7 (28)
Congenital	1 (4)
Other	3 (12)

Table 3. Accuracy metrics for smile interpretation in the test population

	Manual smiles	*Automatic smiles*	
Patients (*n*)	25	25	
Mean	5.48	5.25	$p = 0.354$[a]
95% CI	(4.64–6.32)	(4.37–6.11)	
Sensitivity			90.5%
Specificity			95.4%
Accuracy			93.1%
Prevalence			47.4%
Positive predictive value			94.7%
Negative predictive value			91.8%
Agreement		0.861	$p < 0.001$

[a]Wilcoxon signed rank test.
CI, confidence interval.

Table 4. Reliability testing results

	Test	*Retest*	
Patients (*n*)	25	25	
Mean	5.12	5.04	$p > 0.05$[a]
95% CI	(4.28–5.96)	(4.12–5.96)	

[a]Wilcoxon signed rank test.

side oral commissure movement and the same ''Affdex'' algorithm employed in this study then determined the probabilities of various emotion categories perceived during each smile. While the feasibility of this approach was demonstrated it required complex software and powerful computing and would be impractical for everyday clinical use. The ''SmileCheck'' app is an extension of this prior work aiming to provide a single-step automatic solution to smile assessment. Employing MFRs negated the need for two separate algorithms while utilizing the same premise of smile assessment as earlier.

The ability to smile spontaneously is a key component of human social interaction,[32] and a loss of smile can impart a significant detriment to quality of life and psychosocial well-being in FP patients.[1] Accordingly, restoration of spontaneous smile and for patients to appear ''normal'' to the layperson during organic social interaction is a key goal for facial reanimation surgeons and patients alike.[3,33]

Within the expanding field of facial reanimation surgery, there is ongoing debate as to the optimal approach for the restoration of smile spontaneity. Although utilization of the ipsilateral facial nerve stump or cross facial nerve graft is considered the optimal approaches to achieve spontaneous smiling,[26,27,33–37] many small volume case series of various techniques report extremely high spontaneity rates. However, in a systematic review of all such articles it was found that 63% did not report the methodology of assessment or who was responsible for determining the presence of spontaneity.[4] Of those studies that did report, the majority used clinical observation and the remainder used patient reporting. No study, other than our prior work,[10] has sought to quantify spontaneity objectively. This lack of a universal spontaneity outcome measure hinders accurate quantification of both degree and quality of spontaneous smile, especially over protracted postoperative intervals.[10,28]

The ''SmileCheck'' app was designed specifically to overcome this limitation in the field by providing a tool that allows for rapid, objective, and remote smile assessment that can be performed by any patient in any setting. It requires minimal user input and takes little time to perform a smile assessment. Through presentation of the emotionality of each smile, we believe this application offers clinicians an easy-to-interpret understanding of the functional ability of a voluntary and spontaneous smile to organically express joy. We believe this application can be utilized as a means of quantifying the overall success of a facial reanimation surgery and may provide a universal language for future research projects.

The main limitation within this study is the potential influence of user environment on the app's accuracy. In particular, the degree of light in the user environment and the movement of the device or user during app use. To overcome this the user is first prompted with a set of instructions on ways to optimize their environment. Second, the application continually measures relative movement between the device and user and will stop recording if movement breaches permissible limits. These limits were developed in a pilot study detailed within the Supplementary Data.

Overall, this study demonstrated the potential utility of ''SmileCheck'' app as a smile assessment tool in FP patients. Future research into the use of this app as a means of measuring response to treatment in these patients would be of particular value.

Conclusion

Accurate and uniform evaluation of spontaneous smile is critical for developing a true understanding of the success of various forms of smile reanimation. This will aid clinicians in their selection of approach and will improve patient understanding of the true likelihood of return of spontaneous smile after various treatment pathways. Within this study, the novel ''SmileCheck'' app has been demonstrated to accurately and reliably identify smile in FP patients. This tool may allow for quick, accurate, and repeatable evaluation of smile and can be utilized in further studies to provide a common success outcome measure among various facial reanimation institutions.

Authors' Contributions

Conceptualization (supporting), data curation (lead), formal analysis (lead), writing—original draft (lead), writing (equal), methodology (lead), and investigation (lead) by J.F. Writing—review and editing (equal) and supervision (supporting) by C.M. and S.C. Conceptualization (supporting), writing—review and editing (equal), and validation (equal) by T.M.H. Conceptualization (lead), methodology (supporting), supervision (lead), validation (equal), and writing—review and editing (lead) by J.D. All coauthors have reviewed and approved this article before submission.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this study.

Supplementary Material

Supplementary Data

References

1. Fuzi J, Taylor A, Sideris A, et al. Does botulinum toxin therapy improve quality of life in patients with facial palsy? *Aesthetic Plast Surg* 2020;44(5):1811–1819.
2. Faris C, Tessler O, Heiser A, et al. Evaluation of societal health utility of facial palsy and facial reanimation. *JAMA Facial Plast Surg* 2018;20(6):480–487.
3. Dusseldorp JR, Guarin DL, van Veen MM, et al. In the eye of the beholder: changes in perceived emotion expression after smile reanimation. *Plast Reconstr Surg* 2019;144(2):457–471.
4. van Veen T. *Facial Palsy: Treatment, Quality of Life, and Assessment*. Thesis fully internal (DIV). University of Groningen; 2020; doi: 10.33612/diss.131756313.
5. Dey JK, Ishii LE, Nellis JC, et al. Comparing patient, casual observer, and expert perception of permanent unilateral facial paralysis. *JAMA Facial Plast Surg* 2017;19(6):476–483.
6. Dey JK, Ishii LE, Byrne PJ, et al. Seeing is believing: objectively evaluating the impact of facial reanimation surgery on social perception. *Laryngoscope* 2014;124(11):2489–97.
7. Dey JK, Ishii M, Boahene KDO, et al. Changing perception: facial reanimation surgery improves attractiveness and decreases negative facial perception. *Laryngoscope* 2014;124(1):84–90.
8. Ishii L, Dey J, Boahene KDO, et al. The social distraction of facial paralysis: objective measurement of social attention using eye-tracking. *Laryngoscope* 2016;126(2):334–339.
9. Sforza C, Frigerio A, Mapelli A, et al. Double-powered free gracilis muscle transfer for smile reanimation: a longitudinal optoelectronic study. *J Plast Reconstr Aesthet Surg* 2015;68(7):930–939.
10. Dusseldorp JR, van Veen MM, Guarin DL, et al. Spontaneity assessment in dually innervated gracilis smile reanimation surgery. *JAMA Facial Plast Surg* 2019;21(6):551–557.
11. House JW, Brackmann DE. Facial nerve grading system. *Otolaryngol Head Neck Surg* 1985;93(2):146–147.
12. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. *Otolaryngol Head Neck Surg* 1996;114(3):380–386.
13. Banks CA, Bhama PK, Park J, et al. Clinician-graded electronic facial paralysis assessment: the EFACE. *Plast Reconstr Surg* 2015;136(2):223e–230e.
14. Niziol R, Henry FP, Leckenby JI, et al. Is there an ideal outcome scoring system for facial reanimation surgery? A review of current methods and suggestions for future publications. *J Plast Reconstr Aesthetic Surg* 2015;68(4):447–456.
15. Bandini A, Green JR, Wang J, et al. Kinematic features of jaw and lips distinguish symptomatic from presymptomatic stages of bulbar decline in amyotrophic lateral sclerosis. *J Speech Lang Hear Res* 2018;61(5):1118–1129; doi: 10.1044/2018_JSLHR-S-17-0262.
16. Bandini A, Rezaei S, Guarin D, et al. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE J Biomed Health Inform* 2020. [Epub ahead of print]; doi: 10.1109/JBHI.2020.3019242.
17. Thevenot J, Lopez MB, Hadid A. A survey on computer vision for assistive medical diagnosis from faces. *IEEE J Biomed Health Inform* 2018;22(5):1497–1511.
18. Guarin DL, Dusseldorp J, Hadlock TA, et al. A machine learning approach for automated facial measurements in facial palsy. *JAMA Facial Plast Surg* 2018;20(4):335–337.
19. Guarin DL, Yunusova Y, Taati B, et al. Toward an automatic system for computer-aided assessment in facial palsy. *Facial Plast Surg Aesthetic Med* 2020;22(1):42–49.
20. Bandini A, Orlandi S, Escalante HJ, et al. Analysis of facial expressions in Parkinson's disease through video-based automatic methods. *J Neurosci Methods* 2017;281:7–20.
21. O'Reilly BF, Soraghan JJ, McGrenary S, et al. Objective method of assessing and presenting the House-Brackmann and regional grades of facial palsy by production of a facogram. *Otol Neurotol* 2010;31(3):486–491.
22. Wang T, Zhang S, Dong J, et al. Automatic evaluation of the degree of facial nerve paralysis. *Multimed Tools Appl* 2016;75(19):11893–11908.
23. Guo Z, Dan G, Xiang J, et al. An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis. *IEEE J Biomed Health Inform* 2018;22(3):835–841.
24. Iacolucci CM, Banks C, Jowett N, et al. Development and validation of a spontaneous smile assay. *JAMA Facial Plast Surg* 2015;17(3):191–196.
25. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2021.
26. Biglioli F, Bayoudh W, Colombo V, et al. [Double innervation (facial/masseter) on the gracilis flap, in the middle face reanimation in the management of facial paralysis: a new concept]. *Ann Chir Plast Esthet* 2013;58(2):89–95.
27. Cardenas-Mejia A, Covarrubias-Ramirez JV, Bello-Margolis A, et al. Double innervated free functional muscle transfer for facial reanimation. *J Plast Surg Hand Surg* 2015;49(3):183–188.
28. Dusseldorp JR, Veen MM van, Mohan S, et al. Outcome tracking in facial palsy. *Otolaryngol Clin North Am* 2018;51(6):1033–1050; doi: 10.1016/j.otc.2018.07.003.
29. Chuang DC-C, Lu JC-Y, Anesti K. One-stage procedure using spinal accessory nerve (XI)–innervated free muscle for facial paralysis reconstruction. *Plast Reconstr Surg* 2013;132(1):117e.
30. Guarin D, Taati B, Hadlock T, et al. Automatic facial landmark localization in clinical populations—improving model performance with a small dataset. 2020. Research square. [Epub ahead of print]; doi: 10.21203/rs.3.rs-19405/v1.
31. Dusseldorp JR, Guarin DL, van Veen MM, et al. Automated spontaneity assessment after smile reanimation: A machine learning approach. *Plast Reconstr Surg* 2022;149(6):1393–1402.
32. Kraut R, Johnston R. Social and emotional messages of smiling. *J Pers Soc Psychol* 1979;37:1539–1553.
33. Dusseldorp JR, Naunheim MR, Quatela O, et al. Neurotization preferences in smile reanimation: a discrete choice experiment. Plast Reconstr Surg 2021;148(3):407e.
34. Bhama PK, Weinberg JS, Lindsay RW, et al. Objective outcomes analysis following microvascular gracilis transfer for facial reanimation: a review of 10 years' experience. *JAMA Facial Plast Surg* 2014;16(2):85–92.
35. Hontanilla B, Marre D, Cabello Á. Facial reanimation with gracilis muscle transfer neurotized to cross-facial nerve graft versus masseteric nerve: a comparative study using the FACIAL CLIMA evaluating system. *Plast Reconstr Surg* 2013;131(6):1241–1252.
36. Bae Y-C, Zuker RM, Manktelow RT, et al. A comparison of commissure excursion following gracilis muscle transplantation for facial paralysis using a cross-face nerve graft versus the motor nerve to the masseter nerve. *Plast Reconstr Surg* 2006;117(7):2407–2413.
37. Watanabe Y, Akizuki T, Ozawa T, et al. Dual innervation method using one-stage reconstruction with free latissimus dorsi muscle transfer for re-animation of established facial paralysis: simultaneous reinnervation of the ipsilateral masseter motor nerve and the contralateral facial nerve to improve the quality of smile and emotional facial expressions. *J Plast Reconstr Aesthetic Surg JPRAS* 2009;62(12):1589–1597.