JPRAS
An International Journal of
Surgical Reconstruction
www.JPRASurg.com

Facial aesthetic outcome analysis in unilateral cleft lip and palate surgery using web-based extended panel assessment ☆

CrossMark

H. Bella [a,1], N.S.S. Kornmann [a,1], J.T. Hardwicke [a,b],
K.L. Wallis [a], C. Wearn [a,b], T.-L. Su [c], B.M. Richard [a,*]

[a] *Birmingham Institute for Paediatric Plastic Surgery (BIPPS), Birmingham Children's Hospital NHS Foundation Trust, Steelhouse Lane, Birmingham, UK*
[b] *School of Clinical and Experimental Medicine, University of Birmingham, Edgbaston, Birmingham, UK*
[c] *School of Dentistry, University of Manchester, Manchester, UK*

Summary *Background and aim:* The reproducible measurement of aesthetic outcomes after cleft lip and palate (CLP) surgery remains elusive, and there is no internationally recognised system. The aim of this pilot study was to better understand how humans rate postoperative aesthetic outcome after unilateral cleft lip and palate (UCLP) repair using a novel web-based rating platform with an extended panel of surgeon raters.
Methods: Cropped images of 5-year-old UCLP patients were arranged in a randomly generated sequence within a web-based aesthetic scoring tool as part of an agreement/reliability study. Assessors rated the appearances of patients using a five-point Likert-type scale on two occasions. A mixed-effect statistical model was adopted to analyse the effects of rater, image and timing.
Results: Images of 76 patients were scored by 29 UK-based cleft surgeons. Intra-rater variability was found, and the linear weighted kappa was 0.56. This allowed identification of the most and least consistent raters. The random image effect ($p < 0.001$) suggested that a broad range of aesthetic outcomes were included in the current study. Surgeon raters in this study were likely to score the images more preferably at the second assessment.
Conclusions: A web-based scoring system provides extended data capture, and mixed-effect statistical modelling reveals the effect that time, image and rater have on the scorings. The selection and training of raters, in combination with an exemplary yardstick, might improve inter- and intra-rater agreement. The development of objective measures based upon digital

facial recognition can replace the highly variable subjective human influence on rating the aesthetic outcome.

Introduction

The measurement of aesthetic outcomes after cleft lip and palate (CLP) surgery remains elusive; despite several attempts at devising a scoring method, there is still no internationally recognised system.[1] There is a need for a simple and reliable method of rating photographs that manages the intrinsic subjective nature of human assessment and produces a valid and reproducible result. In order to establish a valid measure, understanding what fluctuates the system is essential: what is it that the individual rater sees and how is this interpreted? These variables may include *rater-related factors* including the number of raters, timing of rating and profession; *image-related factors* including whole or cropped aesthetic units, types of views and two- or three-dimensional (2D or 3D) formats; *subject-related factors* including number, ethnicity and laterality of the cleft; or *scoring-related factors* including five- or seven-point Likert scales, visual analogue scales or the use of exemplary comparators.[2–6]

The validated measurement of outcomes has become an important factor in the evolution of current clinical practice: In 1998, the UK Clinical Standards Advisory Group on Cleft Lip and Palate (CSAG) recommended a centralisation of service provision for CLP patients to allow protocol-driven management strategies.[7] These recommendations were based on the findings of long-term studies based on outcome scoring systems for facial growth (using the Goslon Yardstick)[8] and speech (using the Cleft Audit Protocol for Speech — Augmented).[9] With the addition of a scoring system for facial aesthetic outcome, optimised cleft management protocols could be further developed to allow the standardisation of best practice. Several large studies including the CSAG study, and the Eurocleft and Americleft studies, have used Asher-McDade's system to assess facial aesthetics.[10–13] Whilst many studies using this system state that it is a reliable assessment of the aesthetic outcome, they quote relatively low agreement between raters and use small numbers of mixed-professional raters, usually between four and six.[10–14]

Whilst computerised 3D imaging modalities are expected to produce a valid outcome measure for cleft aesthetics,[15,16] no such mechanism exists to date for either 2D or 3D images. In 2010, Pigott and Pigott introduced Sym-Nose, a computer program designed to analyse clinical photographs by measuring the symmetry of the lip and nose, as a surrogate for aesthetic outcome in unilateral cleft lip and palate (UCLP) patients.[17] Although this computer program enables rapid semi-objective comparison of these features, it remains unclear to what extent the symmetry corresponds with a subjective aesthetic result.

For the past 8 years, the Tri-Centre Group in the UK (West Midlands, South West and Wales Regional Cleft Centres) have used the Asher-McDade-style system to evaluate cropped photographs for internal audit of practice. The aim of this pilot study was to better understand how humans rate postoperative aesthetic outcome after UCLP repair: Specifically, we aim to study the inter- and intra-rater variability for an extended group of professional human raters; to characterise the images in terms of their relationship to the five-point Likert scale; and to study the side-cleft effect on an image being rated. All information was stored and carried out on a novel custom web-based rating portal.

Materials and methods

A retrospective analysis of 2D clinical photographs was conducted and presented according to the Guidelines for Reporting Reliability and Agreement Studies (GRRAS).[18] Standardised anteroposterior (AP) images taken at 5 years of age were obtained from the Tri-Centre Cleft database of patients with UCLP born between 1 January 2000 and 31 December 2005. Exclusion criteria were patients with any type of incomplete CLP, bilateral CLP or a visible Simonart's band on their preoperative photographs. All images were screened and poor-quality images, which could confound aesthetic scoring, were rejected: The quality was considered poor when the image resolution was <100 dots per inch; when saliva or mucous was obstructing view of the scar, nose or lip; when the patient was smiling; or if there was no true AP view photographed.

Image processing

All AP images were cropped with a polygonal lasso to a trapezoid shape using Photoshop Elements software (Adobe Systems Incorporated, San Jose, CA, USA). In summary, the images were initially rotated and levelled to the pupils. Horizontals were approximated to both the superior corneal limbi and the mental crease, with verticals set at both pupils. The trapezoidal crop was completed from the inferior transection of the horizontal and vertical lines, to the superior horizontal at a point corresponding to the medial canthus (Figure 1a and b). This technique was expanded from previously published data.[3] Hair, ears and irises were excluded from the assessment of photographs as they may influence the rating.[3,19,20]

Web-based aesthetic scoring

The cropped images were arranged in a randomly generated sequence within a web-based aesthetic scoring portal on the Birmingham Institute of Paediatric Plastic Surgery secure website (Figure 1c). The invited assessors were given a personalised secure logon to access the scoring exercise, and they proceeded to rate the aesthetic appearances of