*British Journal of Plastic Surgery* (1993), **46**, 7–12
© 1993 The British Association of Plastic Surgeons

BRITISH JOURNAL OF PLASTIC SURGERY

# A panel based assessment of early versus no nasal correction of the cleft lip nose

P. D. Cussons, M. S. C. Murison, A. E. L. Fernandez and R. W. Pigott

*Department of Plastic and Reconstructive Surgery, Frenchay Hospital, Frenchay, Bristol*

*SUMMARY.* **There is a need to be able to assess the overall result in a significant series of cases of a method of management of the cleft lip and nose deformity in order to avoid "best case" reporting often used to introduce new techniques. The present study was performed by a panel placing standardised base view photographs in rank order. The photographs were of 10-year-old subjects of whom 15 were normal controls, 22 were from the Rikshospitalet, Oslo, all of whom had no primary nasal correction and 25 from Frenchay Hospital, Bristol, who all had radical primary nasal correction.**

**Ranking was performed for upper nasal perimeter symmetry, nostril outline symmetry and for overall aesthetic appearance. Analysis of the results showed a significant difference between the three groups, with the corrected noses showing better symmetry. Inter and intraobserver correlations were very close.**

**The limitations of a ranking and marking method of panel assessment are discussed, and a computerised method is presented in subsequent papers.**

Comparison of two surgical methods or the results of two surgical centres presents numerous problems. The most common method defines an endpoint event such as death or 5-year survival and compares percentages for the two groups. Such a method is not suitable to determine which method gives the better appearance, which is one of the things that concerns the cleft surgeon.

The earliest demonstrations of a technique relied upon publishing selected representative results, or even selected best results to commend an innovation. However, the unpredictable nature of the results of cleft lip repair ensures that all techniques produce some very good and some very bad results.

In order to perform a valid comparison between the results of two techniques or of two centres, it is therefore necessary to compare large numbers of cases. The cases should ideally be sequential, with no omissions. Such a number of cases is only amassed over several years and results at a particular age can therefore only be compared by means of standardised photographs. Comparison requires a panel of "blind" observers who do not know to which group the cases belong and it requires normal controls.

The panel then awards marks for the quality of chosen features for each photograph and comparison is made from the marks. Alternatively, the panel ranks the photographs in order from best to worst for the chosen features and comparison is made from the rankings.

Marking is convenient for a panel since photographs are viewed sequentially. Maximal resolution is small if marks are awarded on a small scale (typically 1–5 or 1–10 for each feature) resulting in many ties unless the panel is large. To control for drift in marking standards during one marking session, reference standard photographs should be interspersed repeatedly in the sequence of photographs to be marked. Alternatively, marking should be repeated with the photographs randomised into a different order. Analysis of results requires parametric statistics.

In contrast, ranking is cumbersome. Each individual member of the panel needs to view all the photographs simultaneously, which limits the number which can be handled. If no ties are allowed, the resolution is equal to the number of photographs, even for individual panel members ranking alone and no standardisation is required within each ranking session. Panel ranking can easily be compared with computer ranking. Non-parametric statistics are required for analysis of panel rankings.

Marking methods have been used most frequently despite the lack of reference standards in the majority of series to date. Panels have comprised experts (Ward, 1979; Larson and Nilsson, 1983), lay observers (Saxby and Palmer, 1986; James *et al.*, 1991) and mixed expert and lay observers (Williams, 1968). Opinion has favoured lay panels for being less biased than panels of experts, but Poole *et al.* (1991) show a good correlation between lay and expert panels so long as the operating surgeon, who will recognise the cases, is excluded.

Roberts-Harry and Stephens (1992) have compared the ranking results of lay and expert observers and found good correlation, suggesting that the composition of a panel is not critical.

Correction of the cleft lip nasal deformity continues to pose problems in the achievement of symmetry and in the timing and extent of repair. For this study, the effect of surgery to alleviate this deformity was selected for analysis.

The base view of the nose was used since it allows an observation to be made on the degree to which a


Fig. 1

**Figure 1**—(A, B) A pair of oblique views of the nose. This pair of photographs includes similar information to that required to construct a stereoscopic image of the nose and thence a comparison of the prominence of the alar domes. However, standardisation, measurement and comparison present such problems that the inferior view was adopted.

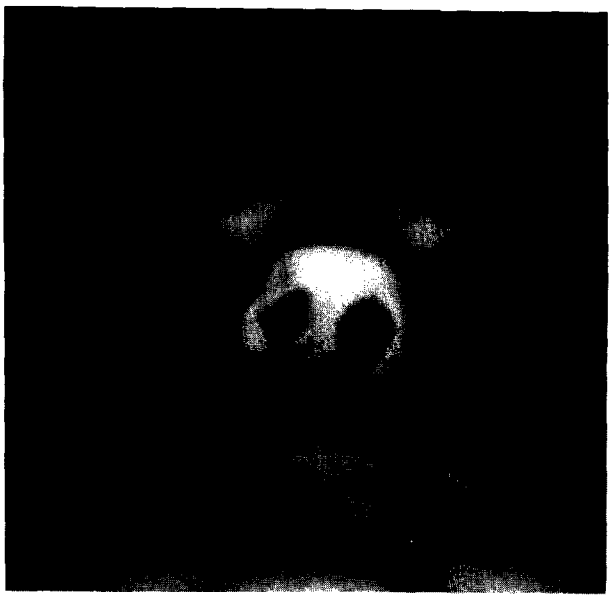

Fig. 2

**Figure 2**—(A) Standardised photograph in the inferior view to show the upper nasal perimeter below the eyebrows and above the canthi. The inner canthi, upper eyelids and pupils are included. (B) Tracing from (A), showing nostril outline and upper nasal perimeter but ignoring interalar base.

radical nasal correction has achieved the objective of advancing the alar dome anteriorly, to match the opposite side. While this view is rarely seen in everyday contact, it provides the same information that is available to the observer with stereoscopic vision during natural movements of the head of the subject. The two "three quarter profiles" or oblique views provide similar information but are almost impossible to standardise, measure or compare (Fig. 1).

In order to find out whether panel assessment can reliably separate groups of patients treated by different methods, the extreme cases of subjects who had undergone no nasal correction (NNC) at the time of assessment were compared with another group who had undergone radical nasal correction (RNC) and both groups were compared with normal subjects.

## Materials and methods

### *Photographic records*

A standard inferior photographic view was required for valid comparisons to be made. This view is shown in Figure 2 and was obtained by tilting the subject's head back to bring the upper nasal perimeter to a level


Fig. 3

**Figure 3** – (A, B) The difficulty in defining the interalar base in the inferior view is shown. Both the lines drawn on the inferior view in (A) might represent the interalar base but when the same lines are seen in the anterior view (B) the margin of error is seen to be great.

below the eyebrows and above the canthi in order to obtain a true skyline image (Pigott, 1985). This view was required in order to demonstrate the shape of the upper nasal perimeter, since this was the feature corrected by one method but left uncorrected by the other. Either both inner canthi, both upper eyelids or both pupils were included in this view for the purpose of orientation for future computerised analysis.

*Subjects*

All subjects had been born with unilateral complete bony cleft lip and palate. They were photographed at age 10 years plus or minus 1 year. For the purpose of comparison, standardised clinical photographs were obtained from the following three sources.

*Radical Nasal Correction Group (RNC).* Photographs were obtained retrospectively from the personal records of the senior author (RWP). These patients had undergone Millard lip repair and Alar Leapfrog radical nose correction (Pigott, 1985) at the age of 3 months. Of 35 sequential cases photographed within 1 year of their tenth birthdays, 25 had photographs which fulfilled the criteria which would allow future computerised analysis as defined above, and these were selected for panel assessment.

Three cases had undergone secondary surgery further to correct alar slump, and seven cases had undergone alveolar bone grafting before their 10-year photograph. Two cases had Simonart bands.

*No Nasal Correction Group (NNC).* A sequential series of photographs was supplied from the records of the Cleft Palate Team, Rikshospitalet, Oslo, by kind permission of Dr Gunvor Semb. These patients had undergone lip correction at the age of 3 months, with no nasal correction. 22 cases fulfilled the above criteria and were selected. Thirteen cases had already undergone alveolar bone grafting before the age of 10 years and six cases originally had Simonart bands. Nasal correction was planned for these patients during their teenage years.

*Control Group.* A sample of 15 normal 10 year old school children with no facial deformity was photographed as specified above and used as controls for comparison.

*Standardisation of photographs*

All three sets of photographs (62 in all) were rephotographed onto 35 mm colour slide film with the magnification individually adjusted such that on every slide the alar base width measured 12 mm plus or minus 1 mm. Each slide was mounted identically and then numbered randomly and uniquely in the series 1–62. The code to their group of origin was kept secret from the observers.

*The panel*

The panel comprised 3 junior members of the Plastic Surgery Unit at Frenchay Hospital, Bristol. At the time the rankings were performed, none of the members of the panel had experience of cleft lip nose surgery in Bristol as assessed by this trial.

*Ranking*

All 62 slides were placed in rank order from best (Rank no. 1) to worst (Rank no. 62) by arranging the slides on

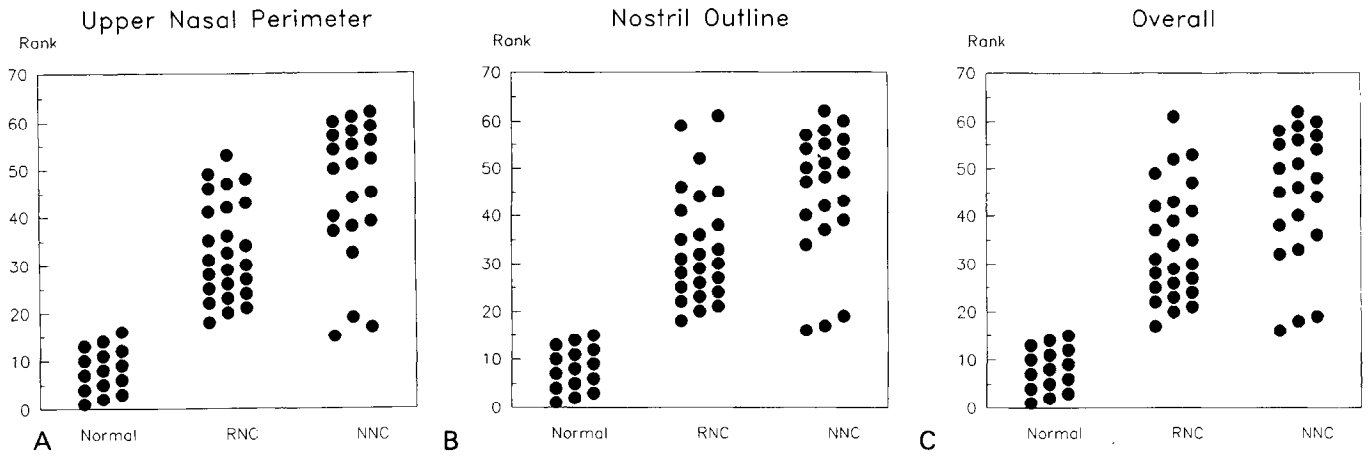


Fig. 4

**Figure 4**—Final rankings for the 15 Normal, 25 Radical Nasal Correction (RNC) and 22 No Nasal Correction (NNC) subjects. (A) Upper nasal perimeter symmetry. (B) Nostril outline symmetry. (C) Overall aesthetic appearance.

a light box. Comparisons were made by rank for upper nasal perimeter symmetry, nostril outline symmetry and for overall aesthetic appearance. The panel were instructed to rank solely on these features. They were instructed specifically to ignore the lip and eyes as irrelevant to the study, and interalar base of the nose since this is ill defined in the standardised inferior view (Pigott, 1988). This is illustrated in Figure 3.

Separate rankings for upper nasal perimeter symmetry, nostril outline symmetry and overall aesthetic appearance were repeated three times by each observer individually and three times by all observers ranking by consensus, giving 12 sets of rankings in all. The order of the slides was randomised between each ranking, and the entire series was spread over 2 months to minimise learning of the slides by the observers.

### *Statistical analysis*

The code was broken after all the ranks had been obtained. A final ranking was calculated by summing the ranks and then ranking the sums (Siegel, 1956) for upper nasal perimeter symmetry, and then for nostril outline symmetry and next for overall aesthetic appearance. The significance of any observed difference between the Normal, RNC and NNC subjects was assessed by the Mann–Whitney U test. Sizes of subgroups (such as the presence of a Simonart band) were insufficient for statistical analysis. Inter and intra observer ranking correlation was assessed by the Hotelling-Pabst test (Hotelling and Pabst, 1936). Every ranking was compared with every other ranking for each feature (upper nasal perimeter symmetry, nostril outline symmetry and overall aesthetic appearance).

## Results

The normal controls were consistently ranked best ($\alpha < 0.001$). As shown in Figure 4, the RNC group results were, as a population, better than those of the NNC group, although there were some excellent results in this group without nasal correction. In all three rankings for upper nasal perimeter symmetry, nostril outline symmetry and overall aesthetic appearance, the difference between the means of the RNC and NNC populations was significant ($\alpha < 0.001$). The distributions of the RNC and NNC populations were also significantly different (Upper nasal perimeter symmetry $\alpha = 0.002$; Nostril outline symmetry $\alpha = 0.006$; Overall aesthetic appearance $\alpha = 0.015$).

### *Observer variability*

Interobserver correlation was very close ($p < 0.001$) for all possible pairs of ranks within each group for upper nasal perimeter symmetry, nostril outline symmetry and overall aesthetic appearance. A representative example of interobserver correlation is shown in Figure 5 ($p < 0.001$, $R = 0.96$). Intraobserver correlation was equally close and any learning element in the comparison procedure could not be detected.

## Discussion

The method described permits valid comparison of two techniques of repair of the unilateral cleft lip. It shows a significant difference between the results of the RNC population with a radical nasal correction, and the NNC population with no nasal correction at the time of lip repair. The method also further emphasises the point that despite the general trend, the NNC population had some very good results and the RNC population had some very bad results.

The sample sizes in this study were not large enough to show any significant influence of alveolar bone grafting, Simonart bands, or further surgery for nasal appearance in either the RNC or the NNC populations.

In order to yield the least equivocal comparisons, this method requires large samples of consecutive standardised photographs at standard age taken over

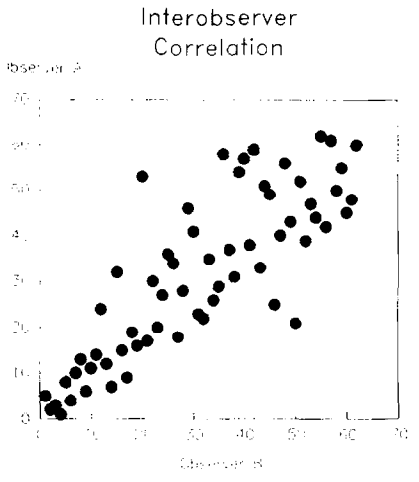


Fig. 5

**Figure 5** -Comparison of the ranks assigned by two members of the panel.

several years. For the purpose of future audit, it is suggested that photography should be undertaken before and after lip repair, and at 5, 10, 15 and 20 years. These ages are, of course, arbitrary and do not relate to specific episodes of patient management. Such secondary surgery must be recorded and taken into account to make any comparison valid. The presence of such subgroups emphasises the difficulty of collecting adequate sized groups for study.

*Problems with panel assessment*

Comparison of large numbers of photographs by ranking demands the investment of a lot of time, but in this study the number of photographs was small enough for them to be manageable. Ranking has the advantage that it is capable of greater resolution with a smaller panel than marking methods (Saxby and Palmer, 1986; Poole *et al.*, 1991) and the results of ranking are easily analysed by non-parametric statistical methods. Comparison of panel rankings and of subsequent computer rankings was also easy to perform. It has frequently been assumed that observers need to be as independent as possible in order to be unbiased by preconceptions of the merits of different techniques of cleft lip nose repair (James *et al.*, 1991). However, this has not proved to be the case, and professional and lay observers have been shown to correlate well in their ranking and marking of photographs (Poole *et al.*, 1991; Roberts-Harry and Stephens, 1992).

The need to repeat rankings several times is not proved by this assessment since rank correlations were so close in all cases.

For the same reason, it is not possible to show that repeated ranking led to the observers learning the photographs. Objectivity of assessment cannot be quantified in this study, but it is the opinion of the authors that a poor nose on a pretty face will score better than an equally poor nose on an ugly face, despite instructions to ignore all but specified features. Roberts-Harry and Stephens (1992) report that professional and lay observers are equally influenced in this respect. For this reason, the study has been continued in an attempt to perform less subjective assessment by computerised methods. Reports of these results follow (Coghlan *et al.*, 1992; Laitung *et al.*, 1992).

It is emphasised that this paper is intended to demonstrate the potential of the method to separate differently managed populations of patients. Valid comparison of the outcome of management protocols must await assessment in the late teens, by which time secondary correction will have been performed as appropriate. By that stage, gross differences in overall facial growth could outweigh any differences in management of an individual feature.

It remains to be seen whether the method is capable of separating groups of patients who have undergone less extreme differences in management.

## Acknowledgements

The authors wish to express their grateful thanks to the members of the Oslo cleft palate team, who have most generously allowed their cases to be included in this analysis.

## References

**Coghlan, B. A., Laitung, J. K. G. and Pigott, R. W.** (1993). A computer aided method of measuring nasal symmetry in the cleft lip nose. *British Journal of Plastic Surgery*, **46**, 13–17.

**Hotelling, H. and Pabst, M. R.** (1936). Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, **7**, 29.

**James, N. K., Mercer, N. S. G., Peat, B., Pigott, R. W. and McComb, H.** (1991). Nasal symmetry: a 10 year comparison between the Pigott and McComb nasal correction. *British Journal of Plastic Surgery*, **44**, 562.

**Laitung, J. K. G., Coghlan, B. A. and Pigott, R. W.** (1993). A comparison of computer versus panel assessment of two groups of patients with cleft lip and palate. *British Journal of Plastic Surgery*, **46**, 18–21.

**Larson, L. and Nilsson, B.** (1983). Early bone grafting in complete cleft lip and palate cases following maxillofacial orthopaedics. VI. Assessments from photographs and anthropometric measurements. *Scandinavian Journal of Plastic and Reconstructive Surgery*, **17**, 209.

**Pigott, R. W.** (1985). Alar Leapfrog. A technique for repositioning the total alar cartilage at primary cleft lip repair. *Clinics in Plastic Surgery*, **12**, 643.

**Pigott, R. W.** (1988). Aesthetic considerations related to repair of the bilateral cleft lip nasal deformity. *British Journal of Plastic Surgery*, **41**, 593.

**Poole, M. D., Walters, E., Beardsworth, E. and Griffiths, P.** (1991). Orbital dystopia: attempts to evaluate the results of surgery. *British Journal of Plastic Surgery*, **44**, 81.

**Roberts-Harry, D. P. and Stephens, C. D.** (1992). The intra and inter examiner agreement in assessing facial attractiveness from photographs. *European Journal of Orthodontics*, in press.

**Saxby, P. J. and Palmer, J. H.** (1986). The use of an independent panel to assess the long term results of cleft lip repair. *British Journal of Plastic Surgery*, **39**, 373.

**Siegel, S.** (1956). *Non-parametric statistics for the behavioural sciences*. Tokyo, McGraw Hill Kogakusha Ltd, p. 23.

**Ward, C. M.** (1979). An analysis, from photographs, of the results of four approaches to elongating the columella after repair of bilateral cleft lip. *Plastic and Reconstructive Surgery*, **64**, 68.

**Williams, H. B.** (1968). A method of assessing cleft lip repairs: comparison of Le Mesurier and Millard techniques. *Plastic and Reconstructive Surgery*, **41**, 103.

## The Authors

**P. D. Cussons, DPhil, FRCS,** Registrar
**M. S. C. Murison, FRCS,** Senior House Officer
**A. E. L. Fernandez, FRACS,** Registrar
**R. W. Pigott, FRCS, FRCSI,** Consultant Plastic Surgeon

Department of Plastic and Reconstructive Surgery, Frenchay Hospital, Frenchay, Bristol BS16 1LE.

Requests for reprints to Mr R. W. Pigott.

Paper received 15 October 1991.
Accepted 23 March 1992, after revision.