Original Article

A Cleft-Customized Occlusal Rating System to Assess Orthodontic Occlusal Improvement in Patients With Unilateral Cleft Lip and Palate

Fabio Henrique de Sa Leitao Pinheiro, DDS, MSc, PhD[1], Carolina Martins Frota, DDS, MSc[2], Daniela Gamba Garib, DDS, MSc, PhD[3], Renata Sathler, DDS, MSc, PhD[2], Terumi Okada Ozawa, DDS, MSc, PhD[2], Rita de Cassia Moura Carvalho Lauris, DDS, MSc, PhD[2], Renata Mayumi Kato, DDS, MSc[2], and Érika Tiemi Kurimori, DDS, MSc[2]

Abstract

Objective: This study aimed to develop a new method to quantify occlusal improvement in patients with unilateral cleft lip and palate (UCLP) who had undergone orthodontic treatment and to evaluate its reproducibility.

Design: A panel of orthodontists decided on the relevance of different occlusal features to score initial and final 3-dimensional study models and panoramic radiographs. A subsequent subjective analysis was later performed by a local orthodontic panel.

Setting: The sample was obtained from the orthodontic clinical archives of a hospital known for the treatment of patients with craniofacial differences.

Patients: Thirty-one nonsyndromic patients, 17 males and 14 females, were randomly selected according to preestablished inclusion/exclusion criteria.

Interventions: The records corresponded to the period during which the patients were treated with conventional multibracket mechanics and adjunctive restorative procedures.

Main Outcome/Measures: The intraclass correlation coefficient measured intraexaminer and interexaminer agreements. The Spearman correlation test assessed the relationship between the local orthodontic panel perception and the improvement scores.

Results: Inter- and intra-rater ICCs varied between fair/good to excellent. There was a strong correlation between the Cleft-Customized Occlusal Rating system classification of occlusal improvement and the local orthodontic panel's perception, thereby enabling the utilization of the interpretation scale by the panel.

Conclusions: The method showed to be a useful tool in quantifying and classifying occlusal improvement in this specific population. As any other method, some limitations apply and need to be accounted for.

Keywords

dental occlusion, orthodontics, nonsyndromic cleft, dental arch

[1] Division of Orthodontics, Department of Preventive Dental Science, University of Manitoba, Winnipeg, Manitoba, Canada
[2] Department of Orthodontics, Hospital for Rehabilitation of Craniofacial Anomalies, University of Sao Paulo, Bauru, São Paulo, Brazil
[3] Department of Orthodontics, Hospital for Rehabilitation of Craniofacial Anomalies and Bauru Dental School, University of Sao Paulo, Bauru, São Paulo, Brazil

Corresponding Author:
Fabio Henrique de Sa Leitao Pinheiro, Dr. Gerald Niznick College of Dentistry, Rady Faculty of Health Sciences, Division of Orthodontics, Department of Preventive Dental Science, University of Manitoba, D344G – 790 Bannatyne Avenue, Winnipeg, Manitoba R3E 0W2, Canada.
Email: fabio.pinheiro@umanitoba.ca

Introduction

Intercenter studies and randomized controlled clinical trials are invaluable methods to assess the impact of cleft team protocols. The most relevant studies (Molsted et al., 2005; Hathaway et al., 2011; Heliovaara et al., 2017) were conducted on patients with unilateral cleft lip and palate (UCLP) whose prevalence can be 2 (Genisca et al., 2009) to 3 (Shapira et al., 1999) times higher than in bilateral cleft lip and palate (BCLP).

Since the 1970s, different methods have been developed to score occlusal disorders (Summers, 1971; Eismann, 1974; Gottlieb, 1975; Berg, 1979; Eismann, 1980). The Goslon yardstick (Mars et al., 1987) and the modified Bauru yardstick (Ozawa et al., 2011) are considered gold standards to assess the impact of surgical interventions on dental arch relationships of patients with UCLP and BCLP, respectively. The Goslon yardstick was developed for late mixed and early permanent dentitions, whereas the modified Bauru yardstick for deciduous, mixed, and early permanent dentitions. Despite their ability to grade dental and skeletal features, occlusal improvement resulting from orthodontic finishing and detailing was not their priority. In addition, they both use ordinal scores such as in the modified Huddart and Bodenham scoring system (Dobbyn et al., 2015), thus making finer discrimination less feasible than with continuous numeric scores. The modified Huddart and Bodenham scoring system was developed for any type of cleft and all dentitions. It only evaluates the incisal anteroposterior relationship and the transverse relationship of canines and molars, not including a vertical assessment and the anteroposterior appraisal of the buccal segments. Among the ones developed for patients without cleft, the Peer Assessment Rating (PAR) index (Richmond et al., 1992) and the American Board of Orthodontics (ABO) scoring index (Casko et al., 1998) are the most popular systems to assess the permanent dentition. Although the PAR index (Richmond et al., 1992) can quantify the amount of postorthodontic occlusal improvement, this measuring system was said to lack enough precision to discriminate between the minor inadequacies of tooth position that are found in ABO case reports (Casko et al., 1998). It has also been criticized for its leniency with residual extraction spacing, unfavorable incisor inclinations, and rotations (Hinman, 1995). On the other hand, although the ABO scoring index can be very detailed in evaluating orthodontic finishing, it was not designed to calculate the amount of occlusal improvement such as with the PAR index. Another drawback of both the PAR index and the ABO scoring index is the fact that they both rely on measurement tools. In addition, their high standard requirements also make them unsuitable for patients with severe cleft of the lip and palate who usually present with hypodontia, crown anatomical variations, and the presence of supernumeraries (Tan et al., 2018). Alveolar bone defects (Enemark et al., 1985), considerable scar tissue formation (Ayoub et al., 2011), and severe skeletal discrepancies (Semb, 1991) also render treatment more challenging and longer.

In 2018, the Commission on Dental Accreditation (American Dental Association, 2018) outlined the standards for the clinical fellowship training programs in craniofacial and special care orthodontics. As their graduates usually work full time in specialized centers, hospitals, or universities, their patient population differs from the one in regular orthodontic practices. Therefore, it seemed important to develop a method capable of distinguishing fine differences in finishing, yet simple and taking into account the limitations inherent to treating patients with cleft. This study aimed to develop and evaluate the reproducibility of a method devoid of measurement tools capable of quantifying occlusal improvement due to orthodontic treatment in patients with UCLP.

Methods

General Approach

Six cleft orthodontists (F.P., T.O., D.G., R.L., R.S., C.F.), 4 of them with over 15 years of clinical experience, met via videoconference to discuss and refine a rating system draft, designed to assess orthodontic finishing outcomes in patients with UCLP. For descriptive purposes, they will be referred here as the "local orthodontic panel." The meeting agenda consisted in, but was not limited to, making decisions on scores and weights, contemplating ways to increase simplicity, and identifying flaws or limitations.

Four raters (F.P., C.F., E.K., R.K.) tested the final version of the Cleft-Customized Occlusal Rating (COR) system by rating 31 pretreatment and post-treatment 3-dimensional (3-D) models and panoramic radiographs of patients with UCLP. A comparison of the scores enabled to calculate intra- and inter-rater levels of agreement.

The local orthodontic panel was requested to assess the perceived occlusal improvement. After confirming a correlation between the COR system mean improvement scores and the medians generated by the panel's subjective evaluation, an interpretation scale was created based on the total mean improvement score $\pm$ SD.

Records

Following approval by the local ethics committee (file number: 87080518.9.0000.5441), dental models and panoramic radiographs were randomly selected from the archives of the Hospital for Rehabilitation of Craniofacial Anomalies, University of Sao Paulo, Brazil. The inclusion criteria were complete unilateral cleft, late mixed or permanent dentitions with or without erupted second molars, and good-quality records obtained before and after completion of comprehensive orthodontic treatment. The exclusion criteria were syndromes, cognitive impairment, and lack of proper records. There was no exclusion based on ethnicity since there would be no intergroup comparisons to be made. The dental models were digitalized using a D700 3Shape scanner and stored for later visualization in 3-D format using 3-D Viewer – Orthoanalyzer 1.4 software for Windows version 7.0.

Thirty-one individuals with UCLP, 14 (45.16%) females and 17 (54.83%) males, fulfilled the inclusion criteria. All individuals were multiracial such as most people in their country of origin.

Development Process of the Rating System

The primary focus of the initial draft was on the occlusion, rather than on the apical bases, and consisted in a blend of the Goslon yardstick (Mars et al., 1987), the ABO scoring index (Casko et al., 1998), and the PAR index (Richmond et al., 1992). Being a scale, the ordinal number of alterations was set to be proportional to each score. Whenever possible, the literature was consulted to help to determine the values. For instance, the range for crowding in Figure 2 was based on the classification presented by the classic work of Little (1975) with slight modifications to turn decimal numbers into integers in order to facilitate for the user.

A visual method to estimate crowding and spacing eliminated the use of special rulers (Figure 1). Instead of measuring the amount of crowding, the amount of surface overlapping was visually estimated. For example, the raters were not expected to precisely differentiate between a crowding 7 mm or greater, after all they were estimating, not measuring. When it comes to crowding, precision was not considered so relevant, given that crowding is in the bottom of the list in terms of malocclusion in the studied population.

Lingual tipping/tilting and in-and-out misalignments, which are also forms of crowding, were not quantified in the absence of overlapping. In the posterior region, rotations might take up space without overlapping, and, in such situations, they also were not counted. Unless mesial and distal spaces are present, anterior rotations nearly always lead to overlapping and were counted as such. In order to prioritize the assessment of occlusal parameters, data on soft tissues, pathologies, cephalometric values, and interdisciplinary procedures were not contemplated.

Before submitting the final draft of the COR system for approval by the local orthodontic panel, 2 cleft orthodontists (F.P. and C.F.) met 8 times to randomly rate study models with the aim of refining the system's latest version. Approval was granted only by the end of the panel's second meeting, totalizing 10 meetings before conclusion of the final version (Figure 2).

For practicality, it was necessary to devise the system, irrespective of treatment planning options, such as canine substitution or implants, orthognathic surgery or camouflage, extraction versus nonextraction, and so forth. Regardless of patients' specific needs, orthodontic techniques, and clinician's preferences, the goal was to measure the amount of occlusal improvement based on the following objective parameters: (1) positive overjet and overbite, (2) cusp-to-embrasure or cusp-to-sulcus intercuspation, (3) normal transverse relationship, (4) reasonable tooth alignment, (5) absence of nonrestorative spaces, and (6) absence of root collisions. The choice of occlusal features and their weight was mostly based on the clinical experience of the panel.

The rationale behind each main decision was explained in Figure 1 by listing the teeth that were considered in the evaluation of each component or feature and by illustrating the most common variations and how they were handled. It was necessary to make a decision in face of each specific variation. For instance, as illustrated by the first picture in Figure 1, teeth that were partially in crossbite and partially in edge to edge were considered to be in edge to edge as this type of crossbite variation tends to be more easily corrected. In the anteroposterior assessment based on the upper incisors, the reader was instructed to consider the canines in substitution cases. Another example was the standardized way to assess interdigitation. Considering that settling can naturally occur following removal of orthodontic appliances, the explanation above the fourth picture in Figure 1 emphasizes that the mere impression of being in contact sufficed in terms of interdigitation. As skeletal discrepancies tend to involve a greater number of teeth, the number of teeth in crossbite or with a negative overbite was the most important parameter to determine the severity of transverse and vertical problems, respectively.

Due to the well-known limitations of panoramic films, information on bone level, bone graft, implants, and root resorption was not utilized for it can be influenced by factors unrelated to orthodontics. Instead, it was agreed that the panoramic radiograph would be useful to distinguish between agenesis/extraction and impaction/ectopic position as the latter can considerably increase the treatment complexity. In addition, with the aid of a panoramic radiograph, the rater would be able to differentiate between an ectopic unerupted tooth amenable to correction and another for which extraction would be inevitable. Although it is sometimes impossible, this type of radiograph could also help distinguish between fusion and gemination. Considering the degree of subjectivity when assessing root parallelism without a measuring tool, it was decided that, instead of root parallelism, root proximity, such as contact and/or overlapping of the roots, would be a more objective evaluation. Regions such as the canines', where distortion usually occurs, were disregarded. Also, considering that appropriate root position is not very often achieved adjacent to the cleft, this area was also excluded from the radiographic evaluation.

As subjectivity could introduce bias and increase execution complexity, the following was agreed to be disregarded: root dilacerations, full transpositions, extra cusps, Bolton discrepancy, and missing teeth.

The raters evaluated 62 pairs of study models and 62 panoramic radiographs, 31 pretreatment and 31 posttreatment. The newly developed COR system (Figure 2) was applied to calculate the amount of occlusal improvement due to orthodontic treatment. The general guideline was to always select the score based on the most severe trait, multiplying it in each row by its respective weight to find the row score. The sum of the row scores yields the total score. The total score at T2 (final model) must then be subtracted from the total score at T1 (initial model). The higher the COR improvement score (T1-T2), the greater the improvement.

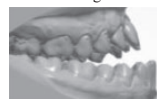| Feature	Teeth	Rationale	Evaluation Examples
A-P - Frontal view	Upper central and lateral incisors	-Upper canines usually not part of overjet assessment, but to be considered at T2 in canine substitution cases	Teeth #11 and #12 are in crossbite whereas tooth #21 is rotated and edge-to-edge
A-P – Buccal view	Upper premolars and first molars	-Upper canines not to be considered as they frequently serve as lateral incisor substitutes -Cusp-to-sulcus/embrasure intercuspation instead of conventional Class I, II and III molar classification	Tooth #24 A-P position is incorrect whereas teeth #25 and #26 have acceptable cusp-to-sulcus/embrasure relationship
Vertical – Frontal view	Upper central and lateral incisors as well as canines	- Upper canines to be considered as they frequently serve as lateral incisor substitutes	Despite the negative overjet, upper incisors and canines have a positive overbite
Vertical – Buccal view	Upper premolars and first molars	-Upper canines not to be considered as they frequently serve as lateral incisor substitutes -One single apparent point of contact counts -Partially erupted teeth count as lack of contact due to unpredictability	Despite incorrect posterior A-P relationship, teeth #24 - #26 appear to be in contact
Transverse	Upper canines, premolars and first molars	-Upper canines not to be considered when substituting for a missing lateral -Edge-to-edge cleft-side canine to score higher due to increased difficulty -Molars considered acceptable even when one cusp is edge-to-edge and the other is normal	As the upper canines are not substituting for lateral incisors, they were taken into account during the transverse assessment
Crowding ("overlapping")	From first molar to first molar	-Not to consider 2nd molars as they rarely overlap 1st molars and usually are found to be partially erupted at T1 -No actual measurement, but just visual estimation of the amount of dental surface overlapping	The labial surface of tooth #12 overlaps approximately 2 mm the palatal surface of tooth #13 and approximately 1 mm the palatal surface of tooth #11.
Spacing	From first molar to first molar	-Not to consider 2nd molars as they rarely have excess space and usually are partially erupted at T1 -Not to consider spaces due to unerupted/ectopic teeth at T1 -Not to consider spaces left at T2 for restorative purposes	Space between teeth #21 and #24 was not counted as tooth #23 was severely ectopic at T1
Tooth shape/size	From first molar to first molar	-Not to consider 2nd molars as they are usually partially erupted at T1 and not essential for function -To only consider taurodontia, peg-shaped teeth, and fusion/germination	Although narrow, teeth #11 and #21 were not peg shaped and were considered as normal
Root angulation	From first molar to first molar	-Not to consider canines due to frequent image distortion in the area -Not to consider cleft-side teeth as they usually lack final root parallelism at T2 -To be considered non-parallel when a root apical third collides or overlaps another	Imperfect, yet acceptable root parallelism
Supernumeraries and ectopic/impacted teeth	From second molar to second molar	-To only consider those that are amenable to correction -Complexity increases if in close proximity to adjacent teeth	Tooth #15 is ectopic but must be disregarded as it is not amenable to correction and will be extracted
Variations	Primary teeth	-Same rules apply to primary teeth as they are expected to serve both function and esthetics if present at T2	This upper right "e" will be counted as tooth #15 and will be considered in contact despite the unideal A-P relationship
Weight assignment	Weights for A-P and transverse relationship	-The highest weight was attributed to A-P-Frontal view as an incorrect overjet can impact periodontal health, speech, and esthetics -A weight was also attributed to transverse discrepancies as they can cause occlusal instability and mandibular shift -No weight was assigned to vertical discrepancies as it is the number of affected teeth that will determine their severity	Despite the severe open bite, weights were applied to A-P and transverse only but the increased number of non-contacting teeth will uptick the vertical score

Figure 1. Summary of main rationales behind the panel decisions.

Instructions: Always select the score based on the most severe trait. In each row, multiply the selected score by its respective weight* to find the row score. Sum the row scores to find the total score.	Teeth	Score 0	Score 1	Score 2	Score 3	Score 4	Row Score
Anteroposterior	(Mx 2-2) *Weight: 3 NB₁: Consider 3's in canine substitution case	Positive O.J.	1 tooth edge-to-edge	>1 tooth edge-to-edge	1 tooth in x-bite	>1 tooth in x-bite	
	(Mx 4-6) *Weight: 1 NB₁: Cusp-to-sulcus/ embrasure intercuspation	Ideal	Unideal: 1 side, 1-2 teeth	Unideal: 1 side, >2 teeth	Unideal: both sides, 1-2 teeth per side	Unideal: both sides, >2 teeth on at least 1 side	
Vertical NB₁: Contact on at least 1 point	(Mx 3-3) *Weight: 1	Positive O.B.	≥1 tooth, 0mm O.B.	Open bite, 1 tooth	Open bite, 2-3 teeth	Open bite, > 3 teeth	
	(Mx 4-6) *Weight: 1	All teeth contact	1 side, 1 tooth out of contact	1 side, 2 teeth out of contact	One side, 3 teeth out of contact	Both sides, 1 or more teeth out of contact	
Transverse NB₁: If sides differ, select score based on the most severe side	(Mx 3-6) *Weight: 2	Normal	1 side, edge-to-edge	Both sides, edge-to-edge	1 side, buccal/lingual x-bite and/or cleft-side canine is edge-to-edge	Both sides, buccal/lingual x-bite	
Crowding ("overlapping") NB₁: Select the most crowded arch and sum the amount of overlapping	(6-6) *Weight: 1	No crowding	Very mild: between 0-2mm	Mild: between 2-4mm	Moderate: 4-7mm	Severe: >7mm	
Spacing NB₁: Select the most spaced arch. At T1: spaces due to extractions and agenesis must be counted; spaces that were restored at T2 or that resulted from treatable ectopic, impacted, unerupted or partially erupted teeth must not be counted	(6-6) *Weight: 1	No spacing	Very mild: between 0-2mm	Mild: between 2-4mm	Moderate: between 4-7mm	Severe: >7mm	
Tooth shape/size NB₁: Only taurodontia, peg-shaped, and fused/germinated teeth	(6-6) *Weight: 1	Normal	1 peg-shaped tooth	2 peg-shaped teeth	1 taurodont and/or 1 fused/germinated tooth	>1 taurodont and/or >1 fused/germinated tooth	
P A N **Root angulation** NB₁: Normal if no contact and/or overlapping NB₂: Disregard canines and cleft-adjacent teeth	(6-6) *Weight: 1	Normal 0-1 root(s) in contact and/or overlapping	2-3 root(s) in contact and/or overlapping	4-6 roots in contact and/or overlapping	7-9 roots in contact and/or overlapping	>9 roots in contact and/or overlapping	
Super# and ectopic/ impacted teeth NB₁: Only consider if amenable to correction	(7-7) *Weight: 1	None	Any unerupted super# not in contact with adjacent teeth	Super# mesiodens and/or any super# in contact with adjacent teeth	1 ectopic/impacted tooth	>1 ectopic/impacted tooth	
					TOTAL SCORE (SUM OF ROW SCORES) =		

Mx = maxillary, O.J. = overjet, O.B. = overbite, x-bite = crossbite, PAN = panoramic radiograph, Super# = supernumerary, NB = *Nota bene* (Latin expression for "observation"), T1 = initial model, and T2 = final model.

Figure 2. Final version of the Cleft Occlusal Rating system for patients with unilateral cleft lip and palate.

The COR system grid (Figure 2) also contains specific information in regard to possible variations related to each component and their attributed weigh. Higher weights were attributed to the sagittal and transverse components as inspired by the Goslon yardstick wherein these 2 features were crucial to differentiate one score from another. The sagittal component was considered to be more associated with the facial appearance, which can sometimes result in psychosocial distress.

The transverse component was ranked in second in face of the increased challenge to correct and retain. Furthermore, a review of the literature seemed to agree that the vertical maxillary growth appears overall to be fairly normal in untreated patients, hence implying that vertical alterations due to surgical interventions will tend to be less severe (Will, 2000). Also, in the Americleft study (Daskalogiannakis et al., 2011), despite significant differences in terms of maxillary prominence

among the centers, which also corresponded to the differences in soft-tissue morphology, there was no difference in regard to the vertical dimension, implying that center- and surgery-related variables might play a less important role in the vertical aspect.

The system can still be used despite of fractured or extracted teeth. For example, if a fractured tooth is not in contact, the case is automatically downgraded due to the lack of occlusal contacts. Even with badly fractured teeth, it is still possible to estimate the anteroposterior relationship, given that the system uses the cusp-to-embrasure relationship as reference. When a tooth is extracted, it must be counted as a space in the same manner as for agenesis. In cases in which it is absent due to being ectopic, impacted, unerupted, or partially erupted, it must not be counted as space. Refer to the note under "spacing" in Figure 2.

Quantitative Rating Session

The raters consisted of 4 certified orthodontists (F.P., C.F., R.K., E.K.) with 5 to 18 years of clinical experience. For calibration purposes, they attended a group review of the panel decisions (Figure 1) and a mock rating session with 11 random models. After this, each rater worked independently using the newly developed COR system (Figure 2).

To be able to classify the level of occlusal improvement based on a COR score, the sample mean $\pm$ SD was used. For instance, scores lower than the mean minus SD represented a mild improvement, whereas those between the mean minus SD and the mean plus SD represented a moderate improvement. Scores higher than the mean plus SD represented a remarkable improvement. Scores 0 and below represented either no improvement or worsening. Although this was a systematic way to categorize the scores, it was still necessary to compare the mathematically determined score ranges with the local panel's classification of improvement.

Qualitative Rating Session and Development of an Interpretation Scale

The local orthodontic panel was requested to watch a slow-paced PowerPoint presentation containing static side-by-side initial and final views of each patient. The slides were presented once again for the panel members to independently rate the improvement perceived in each case using the following subjective classification categories: (1) no improvement/worsening, (2) mild improvement, (3) moderate improvement, or (4) remarkable improvement. Being this an incredibly subjective exercise without the aid of a specific method, the goal was purely to record what the majority of the members of the panel thought, thus putting together the overall identity of the clinical team. This was one of the reasons why the median was calculated. The ultimate goal was to correlate the prevailing improvement perception by the local orthodontic panel with the mathematically determined COR level of improvement (T1-T2) to create an interpretation scale. All data were stored

in an Excel spreadsheet and submitted for statistical analysis using SigmaPLOT 14.0 for Windows version 7.0. In all occasions, the significance level was set at 5%.

Statistical Analysis

The sample size calculation was based on a 5% α, 80% test power, and correlation coefficient of 0.5, indicating the need of a minimum of 29 individuals. To decrease the probability of type II error, the Shapiro-Wilk test was used and detected normal distribution ($P = .693$). As recommended by Houston (1983), a subset of 25 randomly selected models (80.64%) were reassessed 2 weeks apart. The intraclass correlation coefficient (ICC) was used to test the inter- and intra-rater repeatability since it is suitable for independent and normally distributed quantitative variables, taking into account the differences in ratings or orderings. The agreement level was interpreted as suggested by Fleiss (1986): <0.40 (poor), 0.40 to 0.75 (fair to good), and >0.75 (excellent). Despite the sample having normal distribution, the relationship between the local orthodontic panel's perception and the COR improvement score (T1-T2) was evaluated using the Spearman correlation test since it is indicated to analyze the correlation between qualitative and quantitative variables that are ranked whose relationship may or may not be linear.

Results

The pretreatment and post-treatment mean ages were 15.19 ($\pm$2.87) and 21.16 ($\pm$3.19) years, respectively.

Table 1 contains all sample statistic parameters such as mean COR score and SD for each case as assessed by the raters before (T1) and after (T2) treatment. The subtraction of T2 mean score from T1 mean score represents the COR improvement score for each case. Only 16% of the sample (cases number 9, 11, 15, 24, and 26) had an SD equal to half or more than half of the value of the mean scores. This was limited to the T2 ratings. On the other hand, 84% of the sample at T2 had an SD that was lower than half of the mean, showing that, for the most part, disagreements were not excessive.

Table 2 contains the data to estimate the error of the method with and without the utilization of a panoramic radiograph. When considering the method with a panoramic radiograph, the inter-rater reliability ICC varied between 0.762 and 0.814 at T1, meaning that excellent agreement was achieved, and between 0.489 and 0.698 at T2, a level considered fair to good (Table 2). The intra-rater agreement ICC varied between 0.851 and 0.936 at T1, which is considered excellent, and between 0.578 and 0.906 at T2, which ranges from fair/good to excellent (Table 2). The level of agreement for the improvement scores (T1-T2) was excellent overall (Table 2). There was a slight increase in the ICC levels of agreement without the panoramic radiograph, but such increase did not result in changes to the classification of the ICC values (Table 2).

The graph in Figure 3 confirms that the score classification that had been mathematically determined based on mean $\pm$ SD

Table 1. COR Score Sample Mean ($\overline{x}$), Sample SD (S) Before (1) and After (2) Orthodontic Treatment, COR Improvement Score Sample Mean ($\overline{x}_1 - \overline{x}_2$), its SD Differences (S1-S2), and Confidence Interval (CI).

Cases	$\overline{x}_1$	S1	$\overline{x}_2$	S2	$\overline{x}_1 - \overline{x}_2$	S1-S2
1	28.50	1.29	6.00	1.15	22.50	0.58
2	32.75	3.59	5.50	2.38	27.25	1.50
3	30.25	0.96	4.25	1.71	26.00	1.15
4	36.00	1.41	6.50	1.91	29.50	2.38
5	34.50	2.38	11.25	0.96	23.25	2.87
6	31.50	2.89	9.00	1.83	22.50	2.38
7	25.75	2.50	6.25	1.50	19.50	3.32
8	36.25	2.63	4.00	1.83	32.25	3.59
9	14.50	3.32	1.50	1.29	13.00	2.58
10	35.25	3.30	7.50	3.11	27.75	4.78
11	17.00	3.74	4.75	3.09	12.25	4.35
12	29.50	2.08	5.50	1.91	24.00	1.41
13	24.75	2.99	8.25	2.63	16.5	3.70
14	30.50	1.73	6.50	1.29	24.00	0.82
15	19.75	1.50	3.75	3.30	16.00	4.08
16	23.75	5.19	11.75	2.75	12.00	6.22
17	32.25	2.50	4.75	1.26	27.50	3.11
18	23.50	3.69	5.75	1.71	17.75	4.65
19	27.00	1.82	4.75	0.96	22.25	1.26
20	31.75	1.26	7.25	2.99	24.50	2.38
21	30.50	5.45	9.25	1.26	21.25	4.65
22	13.75	1.26	5.75	1.50	8.00	1.41
23	28.75	2.06	5.50	1.29	23.25	1.71
24	21.25	1.71	3.50	3.11	17.75	3.59
25	23.50	5.80	4.75	1.50	18.75	5.74
26	27.25	2.63	3.50	2.38	23.75	2.63
27	17.75	3.30	5.00	1.41	12.75	3.30
28	27.50	1.91	7.25	1.50	20.25	1.71
29	31.50	1.91	11.50	1.91	20.00	1.63
30	28.75	3.30	2.00	0.82	26.75	2.63
31	27.75	3.30	5.50	1.00	22.25	3.86
Study sample	27.20	6.53	6.07	3.03	21.13	6.29
CI 95%	24.90-29.50		5.01-7.14		18.91-23.34	

Abbreviation: COR, Cleft-Customized Occlusal Rating.

Table 2. Intra- and Inter-Examiner Agreements With and Without a Panoramic Radiograph.

	Time	Comparison	ICC with PAN	ICC without PAN
Intraexaminer	T1	1	0.875	0.875
		2	0.851	0.850
		3	0.936	0.922
		4	0.906	0.925
	T2	1	0.783	0.790
		2	0.578	0.596
		3	0.906	0.909
		4	0.780	0.837
	T1-T2	1	0.823	0.831
		2	0.703	0.693
		3	0.936	0.939
		4	0.785	0.861
Interexaminer	T1	1 vs 2	0.811	0.808
		1 vs 3	0.814	0.802
		1 vs 4	0.768	0.816
		2 vs 3	0.804	0.837
		2 vs 4	0.762	0.799
		3 vs 4	0.810	0.828
	T2	1 vs 2	0.698	0.717
		1 vs 3	0.513	0.543
		1 vs 4	0.546	0.633
		2 vs 3	0.489	0.543
		2 vs 4	0.582	0.675
		3 vs 4	0.689	0.728
	T1-T2	1 vs 2	0.778	0.781
		1 vs 3	0.771	0.757
		1 vs 4	0.615	0.711
		2 vs 3	0.784	0.819
		2 vs 4	0.661	0.700
		3 vs 4	0.762	0.795

Abbreviation: PAN, panoramic radiograph.

coincided with the local panel's subjective classification. Note the linear predisposition of the correlation between the medians of improvement perception of the local orthodontic panel ("qualitative rating") and the COR improvement scores ("quantitative rating") generated by the 4 participating raters. The correlation coefficient was 0.69, which, together with a strong positive relationship ($P < 0.001$) and confirmation of normal distribution, confirmed that the mathematically determined score range classification matched the level of perception by the local panel of experts. None/worsening referred to scores less or equal to 0, mild to scores 0.1 to 14.83, moderate to 14.84 to 27.42, and remarkable to scores greater than 27.42 (Table 3). The graph in Figure 3, which represents the correlation between the subjective (qualitative) and COR score range (quantitative) classifications, depicts great similarity for each level of agreement. As it can be seen in the graph (Figure 3), most cases considered mild had scores lower than 13 whereas those considered moderate varied from 13 to approximately 27.

The majority of cases whose outcome was considered remarkable had scores greater than 27. Note that, instead of 31, there are 29 clearly identifiable dots on the graph in Figure 3. This happened because 2 pairs of cases had the same score (22.5 and 22.25) and the same qualitative classification, thereby leading to perfect superimposition of their dots when plotted.

In this study, most individuals (n = 22, 70.97%) were found to have experienced a moderate occlusal improvement after orthodontic treatment. The improvement was remarkable in 4 (12.90%) and mild in 5 (16.13%) individuals. None experienced lack of improvement or worsening.

Discussion

This article introduced a new method to assess occlusal improvement after orthodontic treatment in patients with UCLP. Due to sample availability and proper consent from the patients, the number of rated cases was slightly higher than that estimated by the sample size calculation. Statistically, this higher number of cases can either make no difference or be advantageous.
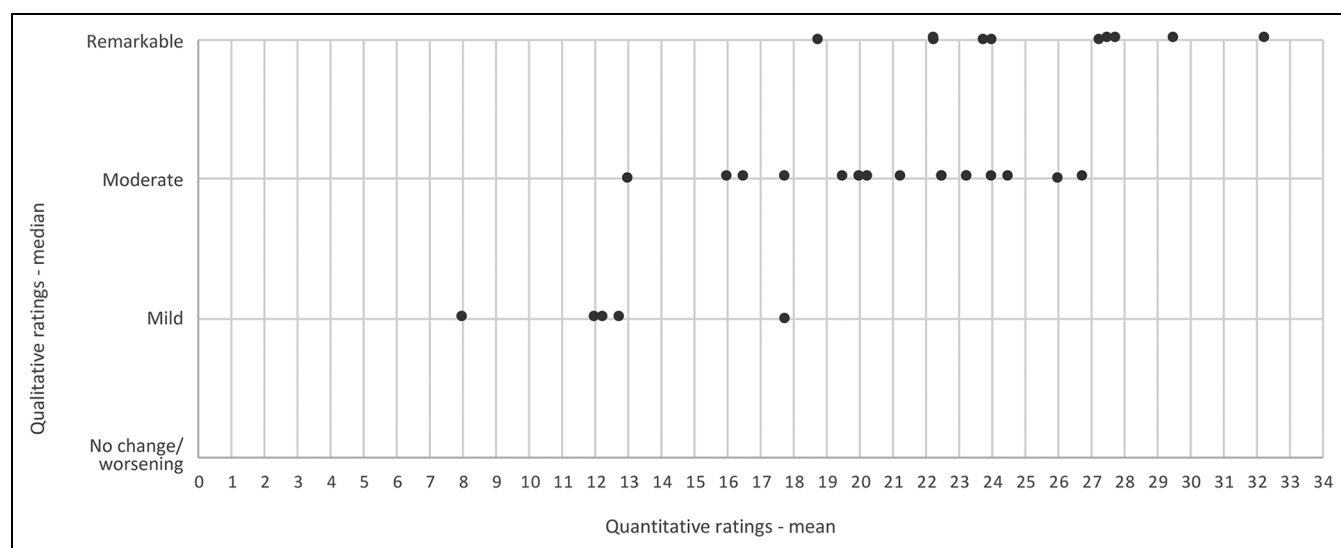
Figure 3. Graph of the correlation between the medians of improvement perception by the local orthodontic panel ("qualitative rating") and the Cleft-Customized Occlusal Rating improvement scores ("quantitative rating").

Table 3. Interpretation of COR Improvement Scores.

COR score range	Level of improvement
≤0	None or worsening
0.1-14.83	Mild
14.84-27.42	Moderate
>27.42	Remarkable

Deciding the weights and the scores for each of the anatomical features was an arduous task that required several meetings with exciting discussions. The highest weight assigned to the anteroposterior relationship lies on its correlation with skeletal discrepancy (Lombardo et al., 2012) and speech (Laine et al., 1985). Correction in the transverse plane was also assigned a weight, given the difficulty in achieving proper transverse dimensions and long-term retention (Nicholson & Plint, 1989). Whether a lower score should have been assigned to an untreatable bilateral crossbite and a higher score to a unilateral posterior crossbite with a shift incited conflicting arguments. The difficulty to reach a conclusion arose from the surprising scarce literature on the long-term negative effects of an untreated bilateral crossbite. For this reason, it was preferred to follow the instructions contained in the Goslon Yardstick, in which the authors stated that "marked narrowing of the upper arch with bilateral crossbite could indicate a more severe category . . ." (Mars et al., 1987). Although unilateral posterior crossbites with mandibular shifts are common in a population without cleft, unilateral posterior crossbites in patients with cleft tends to develop very early in life due to the collapse of the lesser segment (Mazaheri et al., 1993). Even though sometimes leaving a crossbite untreated seems to be the most correct thing to do, attempts to correct bilateral crossbites can facilitate the sagittal correction and induce possible improvement in airway (Reiser et al., 2010). Therefore, correction of bilateral crossbites should preferably be in the list of orthodontic treatment objectives. When indeed very severe (untreatable), this will be reflected on the initial high (T1) score. So, if a surgeon or orthodontist decides to leave a bilateral crossbite untreated, the case will not be dramatically "punished," given that, in this newly devised scoring system, the final score is the result from T1 to T2, and the higher the score, the greater the improvement. In other words, the score will not increase, but at least it will not decrease. Obviously, a case with a similar bilateral crossbite successfully treated will yield higher scores, but chances are that in such a case the bilateral crossbite would be anatomically less severe.

The higher SD observed in 16% of the sample at T2 implied that the raters tended to disagree more when assessing the T2 models (Table 1). This possibly occurred either because the occlusal shortcomings at T2 were much less noticeable than at T1 or because they were deemed as negligible by 1 or more raters. Analyzing this finding through another angle, the SDs remained more acceptable in comparison to their means in 84% of the sample, which seemed reasonable for a system devoid of measuring tools. In reality, for containing just the descriptive statistics data, Table 1 allows for very limited statistical analysis as it is impossible to distinguish whether a high SD derived from all raters disagreeing with one another or 2 raters disagreeing with each other. Also, it neither shows whether one rater is more inclined to disagree with the rest of the raters nor what was the extent of such disagreement. The ICC, as presented in Table 2, is a more acceptable tool for this type of analysis.

The graph in Figure 3 shows a reasonable sample variation despite the lack of cases whose occlusion remained unaltered or got worse. This was expected to occur as patients were treated in a specialized center.

Although the method was found to be user-friendly, some of the agreements varied from fair to good at T2, somewhat lower

than in the Goslon method (Mars et al., 1987), but slightly higher than the interclass random examiner reliability seen with the model-based ABO system (Casko et al., 1998). Whenever coarse discrete categories are used, such as in the Goslon yardstick (Mars et al., 1987), a higher level of agreement tends to be more easily attainable. This confirms that fine discrimination also adds to complexity and increases the frequency of fair inter-rater agreements. It is, however, important to point out that fair to good agreement is not poor agreement. The authors of the PAR index (Richmond et al., 1992) informed their agreement levels without distinguishing between T1 and T2, thereby making a comparison of the current data with theirs unfeasible. Besides the expected different levels of criticism among orthodontists, T2 represents the end of orthodontic treatment when discrepancies are much less evident. In addition, the ICC test becomes more sensitive as scores decrease. The fact that the raters at T2 tended to disagree more between one another, than with themselves, strengthens the assumption that the problem was rater related than method related, therefore calling the attention for better calibration. Another way is to surmise that, in comparison to novice raters, experienced raters tend to be more tolerant to the shortcomings of complex cases. This happens because they are more acquainted with the limitations imposed by complex malocclusions and the importance to reduce the burden of care. As a method developed for visual estimation, the "fair to good" agreements provide the user with a more realistic notion about what to expect in the event of basic calibration, yet maintaining the advantage of being easy and fast to use.

It is daring to leave rulers behind and rely on estimations, but minor imprecisions are a common ground among yardsticks, indexes, and scoring systems, being usually sorted out by increasing the number of raters and taking the mean/median. The authors believe that this is a controllable risk worth taking in favor of eliminating measuring tools.

The method to estimate crowding was unconventional. Instead of contact point displacement, crowding was herein defined as an overlapping of dental surfaces. However, there can be occasions in which a severely rotated, or displaced tooth, does not overlap adjacent teeth. For instance, lingual tipping/tilting might generate no overlapping, yet it is considered a form of crowding. As this usually can be easily corrected by extraction when severe, or by arch development and alignment, it was considered to be of less importance for the system. Posterior rotations also were overlooked when in the absence of overlapping, and being them usually corrected during alignment with no need of space creation, they were also considered less relevant. It is important to point out that this newly devised method was not tailored to precisely quantify crowding, nor it considers crowding to be of great importance in the studied population. Having to make a trade-off between the use of a measurement tool and visual estimation, it was decided that the latter sufficed to keep the method less stringent, as originally envisioned.

By not including a cephalometric radiograph, the same simplicity that characterized the creation of the Goslon yardstick and the PAR index was pursued. Another reason was the fact

that severe anteroposterior and vertical skeletal discrepancies are frequently treated with orthognathic surgery in patients with cleft, making the severity of the malocclusion be a relative one. In addition, whether the orthodontist was able to correctly interpret the cephalometric measurements and make decisions accordingly, this was expected to be reflected on the final occlusal outcome. For this reason, the proper use of a cephalometric radiograph was decided to be judged indirectly, rather focusing on the comparison between the initial and the final static occlusions, as planned initially.

On the other hand, including a panoramic radiograph was found to be essential, especially to avoid missing the frequent ectopic teeth in this studied population. While the general population's average for ectopic teeth ranges from 1.5% to 2% for the permanent canines (Bondemark & Tsiopa, 2007), 18.9% of a sample of patients with cleft showed an ectopic tooth in the anterior region of the cleft area (Rullo et al., 2015). Correction of ectopic upper canines, for instance, can increase treatment complexity and duration, hence requiring better finishing and settling skills. One might argue that some impacted and/or ectopic teeth could simply be extracted, thus drastically decreasing the degree of difficulty. To account for this, it was decided to differentiate between teeth amenable to correction and those whose correction would be unfeasible. In other words, how to distinguish between agenesis/extracted teeth or nonsalvable teeth from unerupted ectopic teeth that could be brought to the arch? Without the information provided by the panoramic radiograph, one could consider as comparable 2 cases that were finished with the same occlusal characteristics, yet neglecting how much more difficult one of them was. The last row in the COR system grid (Figure 2) contains instructions to only consider teeth amenable to correction.

The panoramic radiograph also contains information that sometimes can be useful to differentiate between fused and geminated teeth. Although not the most accurate method (Bouwens et al., 2011), panoramic radiographs can also be used to assess root proximity, something that has been linked with bone defects in the lower anterior region (Kim et al., 2008). As required for implants, one could argue that assessment of root parallelism would be important in the cleft area. Besides being an area prone to distortion, the necessary degree of root parallelism would depend on how the multidisciplinary team decided to close the space, for example, implant, canine substitution, bridge, or removable prosthesis. Moreover, how to determine the minimum amount of root parallelism necessary for an implant since this seems to depend on so many factors? On top of all that, the use of measuring methods to decrease subjectivity would be required. Therefore, root proximity was considered a more feasible form of assessment. Is this a limitation? Likely not because the developed system rather focus on how the teeth occlude in the end, regardless of how the spaces were or will be closed.

It is important to point out that the ultimate goal of this scoring system was to assess the orthodontist's skills in improving the static occlusion. To achieve this goal, it would not be enough to only evaluate the post-treatment occlusion, but also estimate the level of difficulty at T1. Certainly, for

being totally reliant on the occlusal/radiographic features presented, the method is unsuitable to assess the entire complexity of a case, especially in terms of dynamic and/or functional parameters, but still can offer information on how difficult or time-consuming a case is in comparison to another, based on the available information seen on models and panoramic radiographs.

In the definition of the scores and in the instructions provided in the rating grid (Figure 2), spaces at T1 that resulted from extractions and agenesis should be counted as "real" spaces whereas spaces that were restored at T2 or that resulted from treatable ectopic, impacted, unerupted, or partially erupted teeth should not. This happened because the panel understood that spaces that would likely have to be closed by the orthodontist alone, without the help of the other members of the dental team, should be treated differently than those that could be later restored. Likewise, spaces that would later be occupied by an unerupted or partially erupted tooth would not pose as much difficulty as a space that would require greater tooth translation. All these aforementioned details would be impossible to appraise without a panoramic radiograph. Therefore, it was concluded that the aforementioned advantages of including a panoramic radiograph outweighed the negligible decrease in ICC levels (Table 2).

Recognizing that conciseness is key to encourage use, a reasonable balance had to be found, leaving room for future modifications in order to customize the system for specific needs. For example, the score ranges and their specific interpretation indicated in Table 3 matched well the expectations of this local panel (Figure 3). However, they may not reflect the standards of a different panel, hence requiring a re-set to reflect higher or lower standards.

This newly devised rating system can play an important role in assessing the skills of orthodontic residents enrolled in programs with a strong focus on the treatment of patients with cleft. It was not supposed to be a robust form of assessment, but rather a fast and easy method to evaluate the initial and final static occlusions, what makes it attractive for multicenter research and governance initiatives. Given the absence of a measuring tool and a radiograph other than a panoramic, it seems to be ideal for public health systems in triages, quick audits, and resource planning.

In the area of cleft lip and palate, this system attempted to give more emphasis to orthodontic finishing. It seems prudent to recommend it to be used on patients with UCLP, given that it was developed based on the findings commonly encountered in this specific population. Having said that, it may not be incorrect to also use it to evaluate the occlusion of patients with other types of cleft. However, caution should be exercised as further studies are needed to confirm whether any refinement is required to cover peculiarities of other cleft types. It is neither better nor worse than the previous methods, but just specific. It definitely has its own limitations. For example, it consists in a method to evaluate the static occlusion with no allusion to functional aspects and mandibular shifts. The introduction of a dynamic analysis contemplating the assessment of mandibular shifts would increase subjectivity and could considerably reduce reproducibility. This would also limit the applicability of the system to prospective studies, given that it could become unfeasible or even unreliable to count on chart information. A prospective data collection would also be required to ensure that the presence or absence of a shift was correctly assessed and recorded. Prospective studies tend to be more costly and could lead to bias related to the inclusion/exclusion of patients. To keep transparency when data are retrospectively collected, it is rather advisable to rely on information present on standardized records that will be examined by all the raters than on information extracted from charts. Also, chart-based information might be intentionally altered or omitted in a multicenter study in order to favor a center's outcome. Scoring systems and yardsticks are supposed to be as simple as possible. Being particularly useful in the public health system and multicenter studies, the aforementioned issues would create obstacle to implementation.

The panel concluded that it is probably unfeasible to devise a system that accurately covers all aspects of orthodontic finishing without increasing complexity. The more thorough the method, the less practical and nonspecific it becomes, thus being necessary to balance and compromise.

Conclusion

It was possible to draw the following conclusions in regard to the newly devised COR system:

- The method had acceptable intra- and inter-rater reproducibility but requires good calibration.
- Clinical validation was observed based on a strong correlation between the system and the perception of a local orthodontic panel.
- The method of score classification is flexible and can be adapted to reflect local standards of practice.

ORCID iDs

Carolina Martins Frota, DDS, MSc ⬤ https://orcid.org/0000-0003-4820-1795
Renata Sathler, DDS, MSc, PhD ⬤ https://orcid.org/0000-0001-8543-6441

Renata Mayumi Kato, DDS, MSc https://orcid.org/0000-0002-7806-1167

References

American Dental Association. *Accreditation Standards for Clinical Fellowship Training Programs in Craniofacial and Special Care Orthodontics*. Commission on Dental Accreditation; 2018.

Ayoub A, Bell A, Simmons D, Bowman A, Brown D, Lo TW, Xiao Y. 3D assessment of lip scarring and residual dysmorphology following surgical repair of cleft lip and palate: a preliminary study. *Cleft Palate Craniofac J*. 2011;48(4):379-387.

Berg R. Post-retention analysis of treatment problems and failures in 264 consecutively treated cases. *Eur J Orthod*. 1979;1(1):55-68.

Bondemark L, Tsiopa J. Prevalence of ectopic eruption, impaction, retention and agenesis of the permanent second molar. *Angle Orthod*. 2007;77(5):773-778.

Bouwens DG, Cevidanes L, Ludlow JB, Phillips C. Comparison of mesiodistal root angulation with posttreatment panoramic radiographs and cone-beam computed tomography. *Am J Orthod Dentofacial Orthop*. 2011;139(1):126-132.

Casko JS, Vaden JL, Kokich VG, Damone J, James RD, Cangialosi TJ, Riolo ML, Owens SE Jr, Bills ED. Objective grading system for dental casts and panoramic radiographs. *Am J Orthod Dentofacial Orthop*. 1998;114(5):589-599.

Daskalogiannakis J, Mercado A, Russell K, Hathaway R, Dugas G, Long RE Jr, Cohen M, Semb G, Shaw W. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate. Part 3. analysis of craniofacial form. *Cleft Palate Craniofac J*. 2011;48(3):252-258.

Dobbyn L, Gillgrass T, McIntyre G, Macfarlane T, Mossey P. Validating the clinical use of the modified Huddart and Bodenham scoring system for outcome in cleft lip and/or palate. *Cleft Palate Craniofac J*. 2015;52(6):671-675.

Eismann D. A method of evaluating the efficiency of orthodontic treatment. *Trans Eur Orthod Soc*. 1974:223-232.

Eismann D. Reliable assessment of morphological changes resulting from orthodontic treatment. *Eur J Orthod*. 1980;2(1):19-25.

Enemark H, Krantz-Simonsen E, Schramm JE. Secondary bone grafting in unilateral cleft lip palate patients: indications and treatment procedure. *Int J Oral Surg*. 1985;14(1):2-10.

Fleiss JL. *The Design and Analysis of Clinical Experiments*. Wiley; 1986.

Genisca AE, Frias JL, Broussard CS, Honein MA, Lammer EJ, Moore CA, Shaw GM, Murray JC, Yang W, Rasmussen SA. Orofacial clefts in the national birth defects prevention study, 1997-2004. *Am J Med Genet Part A*. 2009;149(6):1149-1158.

Gottlieb EL. Grading your orthodontic treatment results. *J Clin Orthod*. 1975;9(3):155-161.

Hathaway R, Daskalogiannakis J, Mercado A, Russell K, Long RE Jr, Cohen M, Semb G, Shaw W. The Americleft study: an inter-center study of treatment outcomes for patients with unilateral cleft lip and palate. Part 2. dental arch relationships. *Cleft Palate Craniofac J*. 2011;48(3):244-251.

Heliovaara A, Kuseler A, Skaare P, Shaw W, Molsted K, Karsten A, Brinck E, Rizell S, Marcusson A, Saele P, et al. Scandcleft randomised trials of primary surgery for unilateral cleft lip and palate: 6. dental arch relationships in 5 year-olds. *J Plast Surg Hand Surg*. 2017;51(1):52-57.

Hinman C. The Dental Practice Board. Orthodontics – the current status. *Br J Orthod*. 1995;22(3):287-290.

Houston WJ. The analysis of errors in orthodontic measurements. *Am J Orthod Dentofacial Orthop*. 1983;83(5):382-390.

Kim T, Miyamoto T, Nunn ME, Garcia RI, Dietrich T. Root proximity as a risk factor for progression of alveolar bone loss: the Veterans Affairs Dental Longitudinal study. *J Periodontol*. 2008;79(4):654-659.

Laine T, Jaroma M, Linnasalo AL. Articulatory disorders in speech as related to the position of the incisors. *Eur J Orthod*. 1985;7(4):260-266.

Little RM. The irregularity index: a quantitative score of mandibular anterior alignment. *Am J Orthod Dentofacial Orthop*. 1975;68(5):554-563.

Lombardo L, Sgarbanti C, Guarneri A, Siciliani G. Evaluating the correlation between overjet and skeletal parameters using DVT. *Int J Dent*. 2012;2012(3):921-942.

Mars M, Plint DA, Houston WJ, Bergland O Semb G. The Goslon yardstick: a new system of assessing dental arch relationships in children with unilateral clefts of the lip and palate. *Cleft Palate Craniofac J*. 1987;24(4):314-322.

Mazaheri M, Athanasiou AE, Long RE Jr, Kolokitha OG. Evaluation of maxillary dental arch form in unilateral clefts of lip, alveolus, and palate from one month to four years. *Cleft Palate Craniofac J*. 1993;30(1):90-93.

Molsted K, Brattstrom V, Prahl-Andersen B, Shaw WC, Semb G. The Eurocleft study: intercenter study of treatment outcome in patients with complete cleft lip and palate. Part 3: dental arch relationships. *Cleft Palate Craniofac J*. 2005;42(1):78-82.

Nicholson P, Plint D. A long-term study of rapid maxillary expansion and bone grafting in cleft lip and palate patients. *Eur J Orthod*. 1989;11(2):186-192.

Ozawa TO, Shaw WC, Katsaros C, Kuijpers-Jagtman AM, Hagberg C, Ronning E, Semb G. A new yardstick for rating dental arch relationship in patients with complete bilateral cleft lip and palate. *Cleft Palate Craniofac J*. 2011;48(2):167-172.

Reiser E, Skoog V, Gerdin B, Andlin-Sobocki A. Association between cleft size and crossbite in children with cleft palate and unilateral cleft lip and palate. *Cleft Palate Craniofac J*. 2010;47(2):175-181.

Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD, Roberts CT, Andrews M. The development of the PAR index (Peer Assessment Rating): reliability and validity. *Eur J Orthod*. 1992;14(2):125-139.

Rullo R, Festa VM, Rullo R, Addabbo F, Chiodini P, Vitale M, Perillo L. Prevalence of dental anomalies in children with cleft lip and unilateral and bilateral cleft lip and palate. *Eur J Paediatr Dent*. 2015;16(3):229-232.

Semb G. A study of facial growth in patients with unilateral cleft lip and palate treated by the Oslo CLP Team. *Cleft Palate Craniofac J*. 1991;28(1):22-39.

Shapira Y, Lubit E, Kuftinec MM, Borell G. The distribution of clefts of the primary and secondary palates by sex, type, and location. *Angle Orthod*. 1999;69(6):523-528.

Summers CJ. The occlusal index: a system for identifying and scoring occlusal disorders. *Am J Orthod Dentofacial Orthop*. 1971;59(6): 552-567.

Tan ELY, Kuek MC, Wong HC, Ong SAC, Yow M. Secondary dentition characteristics in children with nonsyndromic unilateral cleft lip and palate: a retrospective study. *Cleft Palate Craniofac J*. 2018;55(4):582-589.

Will LA. Growth and development in patients with untreated clefts. *Cleft Palate Craniofac J*. 2000;37(6):523-526.