Abstract

Background: The Kaban modification of the Pruzansky classification (KPC) is based on the severity of the mandibular deformity in Hemifacial Microsomia (HFM) as assessed by clinical exam and plain x-ray. Despite advances in medical imaging, the KPC continues to be the benchmark for the classification of HFM. Our goal is to examine HFM using current 3DCT technology to determine its relationship to traditional KPC as well as its reproducibility between individual raters.

Methods: An IRB approved retrospective review of all patients with a diagnosis of HFM and a preop 3DCT was performed. An in-house score based on consensus between surgeons at our institution stratified the population into Mild (0-1), Moderate (2A), and Severe (2B-3). Clinical KPC score was used as a second comparison. 3DCT scans were evaluated by surgeons and rated according to the KPC. Percent agreement was compared between these standards and the scores of our raters and average Cohen’s kappa was calculated. ANOVA was used for statistical significance. Fleiss’ Kappa was used to assess inter-rater reliability among all expert evaluators.

Results: Sixteen craniofacial surgeons with 248 yrs (avg 15.5 yrs) of experience from 11 institutions were surveyed. Forty-one patients met criteria including 38 patients with documented clinical scores. Fleiss’ kappa was calculated to be 0.238 for clinical KPC score and 0.438 for the in-house score. When comparing the raters’ 3DCT-based classification to the clinical KPC scores, the average agreement was 39.2% (range: 24.4% for Type 2A to 57.6% for Type 3) (average Cohen’s Kappa = 0.257 +/- 0.147). There was improved rater identification of Type 3 mandibles (p<.001), however, as a group all raters were equally unable to accurately identify mandibular severity as compared to clinical assessment (p=.90). When comparing the raters’ 3DCT-based classification to our in-house score, the average total agreement was 69.7% (range: 45.0% for Moderate to 84.8% for Severe) (average Cohen’s Kappa = 0.576 +/- 0.140) with improved identification of Mild and Severe mandibles (p<.001). As a group all raters were equally unable to accurately identify mandibular severity as compared to the gold standard (p=.99). (Figure 1)

Figure 1. Agreement among evaluators and two scoring standards.
Conclusions: The introduction of 3DCT into the diagnostic paradigm highlights the inaccuracy and variability of traditional classification systems. Our results question the accuracy and reproducibility of the current clinical paradigm suggesting the need to reexamine the classification of HFM.

References: