Reviewer's report

Title: Agreement between cranial and facial classification through clinical observation and anthropometric measurement among Envigado school children

Version: Date: 21 February 2014
Reviewer: Curtis Deutsch

Reviewer's report:

Major questions and possible revisions

Efforts to relate clinical impression using visual inspection to anthropometric measurement are important. The authors have done an excellent job providing the background and significance of these metric issues.

There is a statistical concern that the authors might consider. There may well be poor association for craniofacial variation in this study, as they found in their analysis. However, this result may in part be due to diminished statistical power.

The anthropometric measurements reported here are continuously distributed variables, and reducing them to merely three levels diminishes their statistical power. Thus, correlations of measurements with visual impression might be modest in magnitude.

One means of increasing power with the present data set would be to correlate the continuously-distributed anthropometric values with the categorical visual impressions.

A word about the cutoff points used in the study, which are unavoidably arbitrary. Would a method more current than Retzius', based on measurements corresponding to the population distribution, be indicated?

There is another question about the interpretation of the magnitude Cohen's Kappa, which is not always cut-and-dry. Some statisticians, e.g., Landis and Koch (1977) employ other criteria for adequacy of agreement [The measurement of observer agreement for categorical data". Biometrics 33 (1): 159–174]. Would it be worth considering another classification scheme? For example, Landis and Koch would classify a Kappa of 0.388 as fair and approaching "good".

Minor Revisions

"Multiple sample sizes were calculated for agreement studies, both for the cranial index and the facial index" (page 8). Please clarify.

"Negative values were interpreted as equal to 0.0" (page 9). Please give an example.
“Kappa’s coefficient to assess intra-observer agreement... for cranial index visual measurement was 0.917 ± 0.057, and for facial index it was 1.0, indicating an almost perfect match for cranial index visual measurement” (page 11). Were these very high values expected for subjective ratings?

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.