

Manuscript Number: JCMS-D-16-00895R1

Title: Accuracy of measurements used to quantify cranial asymmetry in deformational plagiocephaly

Article Type: Original Paper

Keywords: positional plagiocephaly; skull deformation; anthropometrics; cephalometry; stereophotogrammetry; 3D imaging

Corresponding Author: Mr. Henri Aarnivala, MD

Corresponding Author's Institution: Uni Oulu, Oulu University Hospital

First Author: Henri Aarnivala, MD

Order of Authors: Henri Aarnivala, MD; Ville Vuollo, MSc; Tuomo Heikkinen, DDS, PhD; Virpi Harila, DDS, PhD; Lasse Holmström, MSc, PhD; Pertti Pirttiniemi, DDS, PhD; A Marita Valkama, MD, PhD

Abstract: Objective

Various measurements are used to quantify cranial asymmetry in deformational plagiocephaly (DP), but studies validating cut-off values and comparing the accuracy of such measurements are lacking. In this study, we compared the accuracy of four different measurements in classifying children with and without DP diagnosed by visual assessment, and sought to determine their optimal cut-off values.

Study Design

Two experts rated 407 3D craniofacial images of children aged between 3 and 36 months using the Argenta classification. We then measured the following asymmetry-related variables from the images: Oblique Cranial Length Ratio (OCLR), Diagonal Difference (DD), Posterior Cranial Asymmetry Index (PCAI), and weighted Asymmetry Score (wAS). We created receiver operating characteristic curves to evaluate the accuracy of these variables.

Results

All variables performed well, but OCLR consistently provided the best discrimination in terms of area under the curve values. Subject's age had no clear effect on the cut-off values for OCLR, PCAI, and wAS; however, the cut-off for DD increased monotonically with age. When subjects with discrepant expert ratings were excluded, the optimal cut-off values for DP (Argenta class  $\geq 1$ ) across all age-groups were 104.0% for OCLR (83% sensitivity, 97% specificity), 10.5% for PCAI (90% sensitivity, 90% specificity), and 24.5 (unitless) for wAS (88% sensitivity, 90% specificity).

Conclusion

We recommend using OCLR as the primary measurement, although PCAI and wAS may also be useful in monitoring cranial shape. Subject's age did not impact the threshold of relative asymmetry required for making the diagnosis of DP, and thus we discourage using measurements not relative to head size as a primary measurement of cranial asymmetry or as a definition for DP.

# Accuracy of measurements used to quantify cranial asymmetry in deformational plagiocephaly

Henri Aarnivala<sup>1,2,3</sup>, MD; Ville Vuollo<sup>2,4,5</sup>, MSc; Tuomo Heikkinen<sup>2,4</sup>, DDS, PhD; Virpi Harila<sup>2,4</sup>, DDS, PhD; Lasse Holmström<sup>2,5</sup>, MSc, PhD; Pertti Pirttiniemi<sup>2,4</sup>, DDS, PhD; A. Marita Valkama<sup>1,2,3</sup>, MD, PhD

<sup>1</sup>*Department of Children and Adolescence, Oulu University Hospital, Oulu, Finland*

<sup>2</sup>*Medical Research Center Oulu, University of Oulu, Oulu, Finland*

<sup>3</sup>*PEDEGO Research Group, University of Oulu, Oulu, Finland*

<sup>4</sup>*Department of Orthodontics, Oral Health Sciences, University of Oulu, Oulu, Finland*

<sup>5</sup>*Department of Mathematical Sciences, University of Oulu, Oulu, Finland*

**Corresponding author:** Henri Aarnivala, Dpt of Children and Adolescence, Oulu University Hospital, Oulu, Finland: Tel. +358 45 136 4342, e-mail [henri.aarnivala@student oulu.fi](mailto:henri.aarnivala@student oulu.fi)

The authors declare no conflicts of interest. The research was funded with a grant from the Oulu University Scholarship Foundation. The foundation was not involved in the execution of the study in any way. The first draft of the manuscript was written by H. Aarnivala. No payments were given for producing the manuscript.

## Abstract

### Objective

Various measurements are used to quantify cranial asymmetry in deformational plagiocephaly (DP), but studies validating cut-off values and comparing the accuracy of such measurements are lacking. In this study, we compared the accuracy of four different measurements in classifying children with and without DP diagnosed by visual assessment, and sought to determine their optimal cut-off values.

### Study Design

Two experts rated 407 3D craniofacial images of children aged between 3 and 36 months using the Argenta classification. We then measured the following asymmetry-related variables from the images: Oblique Cranial Length Ratio (OCLR), Diagonal Difference (DD), Posterior Cranial Asymmetry Index (PCAI), and weighted Asymmetry Score (wAS). We created receiver operating characteristic curves to evaluate the accuracy of these variables.

### Results

All variables performed well, but OCLR consistently provided the best discrimination in terms of area under the curve values. Subject's age had no clear effect on the cut-off values for OCLR, PCAI, and wAS; however, the cut-off for DD increased monotonically with age. When subjects with discrepant expert ratings were excluded, the optimal cut-off values for DP (Argenta class  $\geq 1$ ) across all age-groups were 104.0% for OCLR (83% sensitivity, 97% specificity), 10.5% for PCAI (90% sensitivity, 90% specificity), and 24.5 (unitless) for wAS (88% sensitivity, 90% specificity).

### Conclusion

We recommend using OCLR as the primary measurement, although PCAI and wAS may also be useful in monitoring cranial shape. Subject's age did not impact the threshold of relative asymmetry

required for making the diagnosis of DP, and thus we discourage **using** measurements not relative to head size **as a primary measurement of cranial asymmetry or as a definition for DP.**

**Keywords:** positional plagiocephaly; skull deformation; anthropometrics; cephalometry; stereophotogrammetry; 3D imaging

## INTRODUCTION

The increasing prevalence of Deformational Plagiocephaly (DP) has generated a need for reliable methods to diagnose and classify cranial deformities, and several different approaches have been proposed. (*American Academy of Pediatrics, 1992, Hutchison et al., 2004*) Traditionally, the diagnosis has been made via clinical assessment, sometimes relying on a set of reference images as described by Argenta, to determine the type and degree of deformation. (*Argenta et al., 2004, Robinson and Proctor, 2009, Rogers, 2011*) However, these methods reportedly have only moderate inter-rater reliability and repeatability, and may be biased by the clinician's knowledge of the infants' referral status, or whether the assessment is made pre- or post-treatment. (*Mortenson and Steinbok, 2006, Spermon et al., 2008, Atmosukarto et al., 2010*) Therefore, various objective measurements for quantifying cranial asymmetry have been developed. The most commonly used measurement is the ratio between two transcranial diagonals measured in a particular angle—usually 40°—to the anteroposterior midline. These diagonals can be measured directly with calipers or a flexicurve, from a 2D photograph, from a thermoplastic band, or from a 3D image. (*Hutchison et al., 2005, van Vlimmeren et al., 2006, Atmosukarto et al., 2010, Schaaf et al., 2010a, Aarnivala et al., 2014, Skolnick et al., 2015*) This index has also been termed the Cranial Vault Asymmetry Index (CVAI) and Oblique Diagonal Difference Index (ODDI), but in this study we use the term Oblique Cranial Length Ratio (OCLR). (*Hutchison et al., 2005, van Vlimmeren et al., 2006, Schaaf et al., 2010b*)

The often-cited problem with diagonal measurements is that they only quantify cranial symmetry on a single plane. (*van Vlimmeren et al., 2006, Atmosukarto et al., 2010, Meyer-Marcotty et al., 2013, Aarnivala H et al., 2015*) By contrast, as 3D images capture surface data of the whole craniofacial complex, using 3D imaging enables the calculation of various different volume- and surface-related variables, in addition to the previously mentioned point-to-point variables. (*Aarnivala et al., 2016*) Ratios of volumes within different quadrants of the cranium have been used to quantify anterior and

posterior asymmetry, while variables based on the distribution of cranial surface normal vector directions have been used to measure posterior asymmetry and flatness.(*Atmosukarto et al.*, 2010, *Meyer-Marcotty et al.*, 2013) A recent study demonstrated that a better correlation with the clinical diagnosis of DP can be reached by using a smooth Kernel Density Estimate (KDE) of the directional data defined by the normal vectors instead of fixed-bin histograms defined on the spherical coordinate plane.(*Vuollo et al.*, 2016)

Although many of these variables are commonly used in clinical practice, no cut-off values for DP have been established for any of the variables, with the exception of OCLR. For the latter, there is wide variation in the suggested cut-off values found in literature (from OCLR 103.5% to 106.0%), most of which are based solely on expert opinion.(*Hutchison et al.*, 2005, *van Vlimmeren et al.*, 2006, *Glasgow et al.*, 2007, *van Adrichem et al.*, 2008, *Wilbrand et al.*, 2012) Even the perhaps most commonly used cut-off of 104.0% is deemed "premature" by the authors in the paper in which it was established.(*van Vlimmeren et al.*, 2006) No study has investigated the accuracy of these cut-off values in discriminating DP and non-DP infants in an unselected population, and studies comparing the accuracy of different variables are scarce.(*Atmosukarto et al.*, 2010, *Vuollo et al.*, 2016)

In this study, we tested and compared the accuracy of different variables in differentiating subjects with DP from subjects without DP in an unselected population. We performed exploratory analyses and created Receiver Operating Characteristic (ROC) curves for each of the variables in order to both compare their accuracy and to determine the optimal cut-off values with the best specificity and sensitivity. We also investigated whether a subject's age impacts the threshold of cranial asymmetry required for making the clinical diagnosis of DP.

## **MATERIALS & METHODS**

The study material consisted of 407 3D craniofacial images, with the subjects being either 3, 6, 12, or 36 months of age. We acquired the images from a database generated over previous studies on DP. (Aarnivala *et al.*, 2016) We had originally recruited the subjects in the images at birth at Oulu University Hospital on pre-selected dates, so as to represent the normal general population. However, we had excluded infants with craniosynostosis or cleft lip and/or palate, as well as syndromic infants.

We obtained the 3D images using a 3dMDhead 5-pod camera system (3dMD, Atlanta, Georgia), with a tight, soft nylon cap fitted on the subject's head. The process of removing excess data from the images and standardizing the image positions is described in detail here. (Vuollo *et al.*, 2016) In brief, we first identified the craniofacial landmarks: endocanthion left, endocanthion right, tracion left, and tracion right on the 3D image. Then, using these landmarks and the mirror face method, (Zhurov *et al.*, 2010) a custom-written macro automatically standardized the image position and defined the sagittal, coronal, and transverse reference planes.

After the image positions were standardized, an independent investigator (VV) created files with the images de-identified and randomized into different orders. All of the images were then visually assessed by two pediatricians with experience in craniofacial dysmorphology (HA and MV). Both assessors performed this assessment twice with two-week intervals, with the images in a different order for each assessment. The assessors rated the heads as symmetrical or deformed using the Argenta classification to quantify any deformity on a five-stage scale. (Argenta *et al.*, 2004) On the Argenta scale, type I deformation consists of only occipital asymmetry, type II adds malposition of the affected ear, type III adds forehead deformity, type IV adds malar deformity, and type V adds either occipital lift or temporal bulging (**Appendix 1**).

We calculated the asymmetry-related variables OCLR, Diagonal Difference (DD), Posterior Cranial Asymmetry Index (PCAI), and weighted Asymmetry Score (wAS) from the 3D images using a custom-written macro. We also calculated the Cephalic Index (CI), which is related to cranial

proportions. As seen in **Figure 1**, OCLR is the ratio between the longer and shorter oblique transcranial diagonal (measured in a 40° angle to the midline) x 100%, and DD is the difference between the same diagonals in millimeters. CI is the ratio between maximum cranial width and maximum cranial length x 100%. PCAI is calculated from the volumes within the posterior quadrants using the following formula: (larger cuboid volume – smaller cuboid volume) / smaller cuboid volume x 100 %. WAS is calculated from the surface normal vector directions in the area within the posterior quadrants by integrating the square of difference between the Kernel Density Estimate (KDE) function and its reflection against the sagittal plane, which is then multiplied by a weigh coefficient. The weigh coefficient is the distance of the outermost point of the occiput from the sagittal plane divided by maximum cranial width.(*Vuollo et al.*, 2016) OCLR, DD, and CI were measured on a plane at the level of the maximum posterior curvature in the occipital region (i.e., level of maximum head circumference) parallel to the transverse plane. PCAI and WAS were calculated from the surface area above a plane immediately above the highest part of the helix of the higher set ear, also parallel to the transverse plane.

Finally, we created ROC curves, from which we calculated the Area Under the Curve (AUC) values and Youden's indices to assess and compare the accuracy of OCLR, DD, PCAI, and WAS in differentiating children with DP from children without DP, and to determine cut-off values for different classes of DP.(*Zweig and Campbell*, 1993) We compared the ROC curves visually and tested the significance of difference between the AUC values according to Hanley and McNeil.(*Hanley and McNeil*, 1982) We used the Independent Samples t Test to assess the significance of differences between mean CI values, and Cohen's Kappa coefficient to measure inter- and intra-rater reliability.

The 3D images were processed and analysed with Rapidform2006 (INUS Technology, Inc., Seoul, South Korea) 3D software, and the custom macros were written with Visual Basic for Applications (VBA). More complex mathematical analyses were performed with Matlab R2014b (MathWorks,



Natick, Massachusetts). Statistical work was carried out with SPSS v. 22.0 (IBM, Armonk, New York).

The study was approved by the Ethics Committee of the Northern Ostrobothnia Hospital District and was carried out according to the Declaration of Helsinki. Written, informed consent was obtained from the parents of the participants, prospectively.

## RESULTS

There were 123 images (30%) of 3-month-olds, 117 (29%) of 6-month-olds, 110 (27%) of 12-month-olds, and 57 (14%) of 36-month-olds. 84 (21%) of the heads were rated as asymmetrical and 277 (68%) as symmetrical, respectively, by both assessors. Ratings for 46 (11%) heads were discrepant. Out of the heads rated as asymmetrical by both assessors, 1 head was type IV, 10 were type III, 29 were type II, and 44 were type I deformation on the Argenta scale. Out of the heads with discrepant ratings, 1 head was rated as type III, 9 were type II, and 36 were type I by the other assessor. The assessors achieved decent intra-rater reliability for the visual classification ( $\kappa = 0.680 - 0.714$ ), and the inter-rater reliability was also good ( $\kappa = 0.677$ ).

Means for all of the measured variables were roughly equal across all age groups in the no deformation group. Means in other groups showed more fluctuation, especially in the groups with Argenta class  $> 1$ , but both visually perceived and anthropometrically measured asymmetry was milder in older subjects. Descriptive statistics for the measured variables are presented in **Table 1**.

ROC curves for all of the measured variables were first created separately in each age group to determine and compare the optimal thresholds for differentiating children with DP (Argenta class  $\geq 1$ ) from children without DP (no deformation). The differences between the AUC values of the curves for different age groups were not statistically significant. Moreover, the Youden's indices and visual interpretation of the curves showed that the optimal thresholds for OCLR, PCAI, and

wAS were similar in all age groups, whereas the threshold for DD increased monotonically with age (see **Table 2**). Hence, the final ROC curves for OCLR, PCAI, and wAS were created with all of the images pooled together (**Figure 2**), while further analysis of DD was omitted. The cut-off values for OCLR, PCAI, and wAS that produced the most accurate classification, along with their respective Youden's J values, are shown in **Table 3**.

OCLR and DD showed the highest bivariate correlation (Pearson's  $r = 0.99$ ), which was expected because they are calculated from the same transcranial diagonals. Between OCLR and PCAI, the Pearson's  $r$  was a slightly lower 0.89, and between wAS and OCLR a clearly lower 0.69. **Figure 3** shows that, although neither OCLR or wAS provided perfect classification, they may complement each other, as only three heads (4%) with consistent expert ratings simultaneously fell under the suggested cut-off values for both outcome variables.

Comparing the groups that received asymmetrical and symmetrical expert ratings, there was a significant difference between the mean CI:s (79.5% vs. 76.9%,  $p < 0.001$ ). It is noteworthy that with all outcome variables, the subjects receiving false negative results had a significantly higher mean CI than the subjects receiving false positive results when using the cut-off values reported in **Table 3**. The mean CI:s were 82.3% vs. 75.7% ( $p < 0.01$ ) using OCLR 103.5%; 80.4% vs. 76.4% ( $p < 0.05$ ) using PCAI 10.5%; and 79.4% vs. 76.2% ( $p < 0.05$ ) using wAS 24.5 (unitless).

## **DISCUSSION**

In this study, we assessed and compared the diagnostic accuracy of several asymmetry-related indices. To our knowledge, this is the first study to investigate the impact of age on the threshold of cranial asymmetry required for making the clinical diagnosis of DP, as it is the first study to produce candidate cut-off values for different types of DP by analysing ROC curves based on a random sample of the normal general population.

The variables chosen to be investigated in this study have reportedly shown to have good correlations with the clinical evaluation of asymmetry, and all of them are used in clinical practice. (Hutchison *et al.*, 2005, Glasgow *et al.*, 2007, Meyer-Marcotty *et al.*, 2013, Vuollo *et al.*, 2016) Although all of the studied variables performed well, OCLR consistently produced the most accurate classification, as indicated by the AUC and Youden's J values. The only subgroups where wAS performed better were the 3-month-old group and the Argenta class  $\geq 3$  groups in terms of Youden's J values, while PCAI did not outperform both OCLR and wAS in any subgroup. Further remarks in favor of OCLR are the simplicity of the measurement itself compared to PCAI and wAS, and the consistent suggested cut-off values that proved to be unaffected by age, as opposed to DD. Expectedly, the suggested cut-off for DD increased with age, as DD is an absolute measurement and not relative to the size of the head. This makes DD a relatively poor instrument for diagnosing and monitoring DP, as there is not only a need for different age- and size-specific cut-off values, but such unadjusted indices can simply be confusing, especially for non-professionals. For example, a similar plagiocephalic OCLR value of 105.0% at both 3 and 36 months would indicate no improvement, but, due to head growth, a similar DD value of 5.0 mm at both 3 and 36 months would indicate significant improvement in perceived asymmetry, and in the light of our results, recovery from DP. This can be confusing even for professionals, and some researchers have concluded that cranial deformation does not improve spontaneously due to an observed lack of improvement in the unadjusted measurements of untreated infants. (Wilbrand *et al.*, 2016) However, based on our results, infant age (i.e. head size) does not impact the threshold of relative asymmetry (i.e. geometric shape of the head) required for deformation to be visually perceived by professionals; it is the *asymmetry* that matters, not absolute measurement values. Therefore, we discourage solely using measurements not relative to head size in diagnosing and monitoring DP. The cut-off values for OCLR with the highest sensitivity and specificity, as determined by Youden's J, were in the range of those used in literature. Overall, the best discrimination (affected from

unaffected) was provided by OCLR 103.5%, which is close to the cut-offs proposed by Wilbrand et al. using percentile curves. (Wilbrand et al., 2012) However, contrary to their proposed age-specific cut-offs, our results suggest that the usage of the same cut-offs for OCLR, PCAI, and wAS regardless of subject age could be justified, making them easier to interpret when monitoring cranial asymmetry.

On the other hand, the cut-offs suggested by Youden's J indices in our study cannot be directly assumed to be the ideal cut-offs for clinical use. A closer inspection of the ROC curves shows that, with discrepant expert ratings excluded, using 103.5% as the cut-off for OCLR would lead to a relatively high false positive rate of 8% (92% specificity), whereas raising the cut-off to 104.0% would produce almost equally good discrimination, but also reduce the false positive rate to 3% (97% specificity), while reducing the sensitivity from 90% to 83%. As especially mild DP is primarily a cosmetic problem, the tolerance for false positive results should be low. Hence, the specificity should be set as high as possible while simultaneously retaining sufficient sensitivity; therefore, we propose OCLR 104.0% as the optimal cut-off for DP in clinical use. This is also supported by the fact that half of the subjects with discrepant expert ratings had an OCLR score less than 104.0%. For PCAI and wAS, increasing the specificity steeply decreases the sensitivity, so the optimal cut-offs are those shown in **Table 3**. Because wAS shows a relatively low bivariate correlation with OCLR and PCAI, but still performs almost as well—and in a few subgroups, better—than OCLR, it likely captures a different aspect of cranial shape than OCLR and PCAI. Therefore, using both OCLR and wAS likely provides more information compared to using only OCLR, and calculating both indices might be especially useful in monitoring cranial shape over time. However, it cannot yet be concluded that wAS is more accurate in identifying more severe forms of DP (with frontal asymmetry), as the number of subjects with Argenta class  $\geq 3$  was low in the present study.

It is noteworthy that the OCLR measurement performed much better in our study compared to a previous paper by Atmosukarto et al. (Atmosukarto et al., 2010). In this particular paper, the performance of the OCLR measurement and an absolute asymmetry score measurement (AAS, based on 2D histograms of surface normal vector angles) was evaluated with ROC curves in a setting quite similar to ours, except that their study population constituted of two highly contrasting groups of DP cases and controls. Even though subjects with discrepant expert ratings were excluded, they reported an AUC of only 0.79 for OCLR, and 0.91 for their AAS, both of which fall far below our results. As the WAS measurement used in our study is basically an improvement over AAS, and it has been previously shown to outperform AAS, the difference in the results likely stems from two things impairing their OCLR scores. First, Atmosukarto et al. measured the OCLR from a 2D snapshot of the 3D image manually, and in a 45° angle to the midline as opposed to the **more** commonly used 40° angle. Second, they apparently did not automatically adjust the image position for their measurements, and even if the 2D snapshots were captured automatically from a standardized angle, the orientation of the 2D images would have been lost afterwards, as they used no craniofacial landmarks. The importance of consistent and reliable image positioning must therefore be stressed, and correspondingly, it cannot be directly assumed that other methods for measuring the OCLR, such as direct measurements, plagiocephalometry, or HeadsUp, would yield results equivalent to ours. (Hutchison et al., 2005, van Vlimmeren et al., 2006) Then again, our method for positioning the 3D images aims to set the measurement plane for OCLR at the level of the maximum head circumference, so values measured with the aforementioned methods would likely be similar to ours.

By using the Argenta classification, we were able to test and compare cut-off values for DP subtypes with different clinical features. Then again, the Argenta classification does not directly serve as a "severity score" because it is possible to have quite severe occipital asymmetry without any associated frontal asymmetry, which leads to high values in asymmetry scores but only

classifies as type 1 or 2 on the Argenta scale (see **Figure 3**). Therefore, we could not study or produce cut-off values for different degrees of DP, which is a limitation of this study. Furthermore, the inter- and intra-rater agreements were far from perfect, albeit the inter-rater agreement was considerably higher compared to previous studies using visual classification. (*Mortenson and Steinbok, 2006, Spermon et al., 2008, Atmosukarto et al., 2010*) Nevertheless, as many as 46 (11%) of the 407 subjects received discrepant expert ratings, which reflects the difficulty and subjectivity of drawing the line between symmetrical and asymmetrical cranial shape, especially in an unselected population such as ours. Hence, it was essential to perform the statistical analyses with the discrepantly rating subjects included; when used in clinical practice, the accuracy of the cut-off values should be at least what is reported with the discrepant ratings included.

There is one final limitation to the continuous measurements used in this study. The general assumption has been that all of the asymmetry indicated by the numbers originates from the asymmetry of the calvarium. However, it is quite possible that a small proportion of subjects exceed the cut-off values for DP because of, for example, having isolated facial asymmetry that causes the glabella-nasion area to be slightly off center, resulting in an increased asymmetry score in the absence of visually perceived calvarial asymmetry, despite the use of the mirror face method to minimize potential error caused by facial asymmetry. (*Zhurov et al., 2010*) On the other hand, because the CI values of the heads receiving false negative results were significantly higher (i.e. more brachycephalic) than those receiving false positive results in terms of the outcome measurements, the overall proportions of the cranium likely impact the **visual perception of** asymmetry as well. Although DP and deformational brachycephaly both result from prolonged external forces on the infant's cranium and they frequently co-occur, it is possible that the oft-reported finding of infants with cranial asymmetry having higher CI values partly stems **s** from different cranial proportions altering the visual perception of cranial symmetry. (*Hutchison et al., 2004, McKinney et al., 2008, Collett et al., 2012*)

The subjectivity and inconsistency of visual assessment raises the question of whether measurable indices, such as the OCLR, should instead be considered the golden standard for the diagnostics and classification of DP. The potential sources for error with 3D stereophotogrammetry are restricted to the subjective identification of craniofacial landmarks, which seems to have a negligibly small effect as the measured parameters have previously shown excellent repeatability and reproducibility. (Zhurov *et al.*, 2010, Meyer-Marcotty *et al.*, 2013) Furthermore, 3D imaging is already considered the method of choice for monitoring the cranial shape over time, owing to its ability to document and store information regarding the whole craniofacial complex. However, these methods are still more time consuming than other methods and the required hardware remains quite expensive. 2D measurements, such as plagiocephalometry, are more widely accessible and they too have shown better reproducibility and inter-rater agreement than visual assessment. (van Vlimmeren *et al.*, 2006, Wilbrand *et al.*, 2011) On the other hand, DP is primarily regarded as a cosmetic condition, so it is necessary for the head to look asymmetrical to be considered abnormal, which has also been used as an argument against the necessity of measurable indices. (Feijen *et al.*, 2012) However, the potential sources of bias and low reproducibility of a clinical diagnosis may result in a relatively high number of false positive diagnoses, which again may cause unfounded interventions **resulting** in both unnecessary concern and costs for the family and **possible** discomfort for the infant. (van Wijk *et al.*, 2014) Despite the specificity of the outcome variables in this study not being perfect, using these variables in a clinical setting would not necessarily produce many false positive results, because the initial need for assessing cranial shape usually stems from either parental or expert concern—in other words, visual assessment—of the child's head shape. Therefore the function of the measurements would primarily be to either confirm or refute the "working diagnosis." Considering the accuracy of the outcome variables measured in the present study, using these objective measurements when making the initial diagnosis of DP would likely improve the diagnostic accuracy.

Future research should aim to develop objective measurements that would provide even better discrimination, possibly by taking the potential effect of cranial proportions (CI) into account. Validation of the cut-off values used in clinical practice to categorize different degrees of DP (i.e. moderate, severe, very severe) is also warranted. The software used for 3D stereophotogrammetry and image analysis should also be developed further to facilitate routine clinical use.

## **CONCLUSIONS**

OCLR produces the most accurate classification of cranial asymmetry compared to clinical assessment. While PCAI and wAS produce slightly less accurate classification, they may also be useful in monitoring cranial asymmetry. Age does not impact the threshold of relative cranial asymmetry required for making the clinical diagnosis of DP, and hence, only using measurements not relative to head size—such as DD—is discouraged.

## **ACKNOWLEDGMENTS**

The authors are grateful to all the families who participated in the study. The research was funded with a grant from the Oulu University Scholarship Foundation.

## **CONFLICTS OF INTEREST**

The authors declare no conflicts of interest.



## REFERENCES

- American Academy of Pediatrics Task Force on Infant Positioning and SIDS: Positioning and SIDS: *Pediatrics* 89: 1120 1126, 1992.
- Aarnivala H, Vuollo V, Harila V, Heikkinen T, Pirttiniemi P, Valkama AM: Preventing deformational plagiocephaly through parent guidance: a randomized, controlled trial. *Eur J Pediatr* 174: 1197 1208, 2015.
- Aarnivala H, Valkama AM, Pirttiniemi P: Cranial shape, size and cervical motion in normal newborns. *Early Hum Dev* 90: 425 430, 2014.
- Aarnivala H, Vuollo V, Harila V, Heikkinen T, Pirttiniemi P, Holmström L, Valkama AM. The course of positional cranial deformation from 3 to 12 months of age and associated risk factors: a follow-up with 3D imaging: *Eur J Pediatr* Epub ahead of print, 2016.
- Argenta L, David L, Thompson J. Clinical classification of positional plagiocephaly: *J Craniofac Surg* 15: 368 372, 2004.
- Atmosukarto I, Shapiro LG, Starr JR, Heike CL, Collett B, Cunningham ML, Speltz ML: Three-dimensional head shape quantification for infants with and without deformational plagiocephaly. *Cleft Palate Craniofac J* 47: 368 377, 2010
- Collett BR, Heike CL, Atmosukarto I, Starr JR, Cunningham ML, Speltz ML: Longitudinal, three-dimensional analysis of head shape in children with and without deformational plagiocephaly or brachycephaly. *J Pediatr* 160: 673 678, 2012
- Feijen M, Schuckman M, Habets E, van der Hulst R: Positional plagiocephaly and brachycephaly: is there a correlation between subjective and objective assessment of cranial shape? *J Craniofac Surg* 23: 998 1001, 2012.
- Glasgow TS, Siddiqi F, Hoff C, Young PC: Deformational plagiocephaly: development of an objective measure and determination of its prevalence in primary care. *J Craniofac Surg* 18: 85 92, 2007.
- Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29 36, 1982.
- Hutchison BL, Hutchison LA, Thompson JM, Mitchell EA: Plagiocephaly and brachycephaly in the first two years of life: a prospective cohort study. *Pediatrics* 114: 970 980, 2004.
- Hutchison BL, Hutchison LA, Thompson JM, Mitchell EA: Quantification of plagiocephaly and brachycephaly in infants using a digital photographic technique. *Cleft Palate Craniofac J* 42: 539 547. 2005.

- McKinney CM, Cunningham ML, Holt VL, Leroux B, Starr JR: Characteristics of 2733 cases diagnosed with deformational plagiocephaly and changes in risk factors over time. *Cleft Palate Craniofac J* 45: 208 216, 2008.
- Meyer-Marcotty P, Böhm H, Linz C, Kochel J, Stellzig-Eisenhauer A, Schweitzer T: Three-dimensional analysis of cranial growth from 6 to 12 months of age. *Eur J Orthodont* 36: 489 496, 2013
- Mortenson PA, Steinbok P: Quantifying positional plagiocephaly: reliability and validity of anthropometric measurements. *J Craniofac Surg* 17: 413 419, 2006.
- Robinson S, Proctor M: Diagnosis and management of deformational plagiocephaly. *J Neurosurg Pediatrics* 3: 284 295, 2009.
- Rogers GF: Deformational plagiocephaly, brachycephaly, and scaphocephaly. Part I: terminology, diagnosis, and etiopathogenesis. *J Craniofac Surg* 22: 9 16, 2011.
- Schaaf H, Pons-Kuehnemann J, Malik CY, Streckbein P, Preuss M, Howaldt HP, Wilbrand JF: Accuracy of three-dimensional photogrammetric images in non-synostotic cranial deformities. *Neuropediatrics* 41: 24 29, 2010a.
- Schaaf H, Wilbrand JF, Boedeker RH, Howaldt HP: Accuracy of photographic assessment compared with standard anthropometric measurements in nonsynostotic cranial deformities. *Cleft Palate Craniofac J* 47: 447 453, 2010b.
- Skolnick GB, Naidoo SD, Nguyen DC, Patel KB, Woo AS: Comparison of Direct and Digital Measures of Cranial Vault Asymmetry for Assessment of Plagiocephaly. *J Craniofac Surg* 26: 1900 1903, 2015.
- Spermon J, Spermon-Marijnen R, Scholten-Peeters W: Clinical classification of deformational plagiocephaly according to Argenta: a reliability study. *J Craniofac Surg* 19: 664 668, 2008.
- van Adrichem LN, van Vlimmeren LA, Cadanova D, Helders PJ, Engelbert RH, van Neck HJ, Koning AH: Validation of a simple method for measuring cranial deformities (plagiocephalometry). *J Craniofac Surg* 19: 15 21, 2008.
- van Vlimmeren LA, Takken T, van Adrichem LN, van der Graaf Y, Helders PJ, Engelbert RH: Plagiocephalometry: a non-invasive method to quantify asymmetry of the skull; a reliability study. *Eur J Pediatr* 165: 149 157, 2006.
- van Wijk RM, van Vlimmeren LA, Groothuis-Oudshoorn CG, Van der Ploeg CP, Ijzerman MJ, Boere-Boonekamp MM: Helmet therapy in infants with positional skull deformation: randomised controlled trial. *BMJ* 348: g2741, 2014.
- Vuollo V, Holmström L, Aarnivala H, Harila V, Heikkinen T, Pirttiniemi P, Valkama AM: Analyzing infant head flatness and asymmetry using kernel density estimation of directional surface data from a craniofacial 3D model. *Stat Med* 35: 4891 4904, 2016.

- Wilbrand J, Lautenbacher N, Pons-Kuhnemann J, Streckbein P, Kahling C, Reinges MHT, Howaldt H, Wilbrand M: Treated Versus Untreated Positional Head Deformity. *J Craniofac Surg* 27: 13-18, 2016.
- Wilbrand J, Schmidtberg K, Bierther U, Streckbein P, Pons-Kuehnemann J, Christophis P, Hahn A, Schaaf H & Howaldt H (2012) Clinical classification of infant nonsynostotic cranial deformity. *J Pediatr* 161(6): 1120-1125.
- Wilbrand J, Wilbrand M, Pons-Kuehnemann J, Blecher J, Christophis P, Howaldt H & Schaaf H (2011) Value and reliability of anthropometric measurements of cranial deformity in early childhood. *J Craniomaxillofac Surg* 39(1): 24-29.
- Zhurov A, Richmond S, Kau CH & Toma A (2010) Averaging Facial Images. In: Anonymous *Three-Dimensional Imaging for Orthodontics and Maxillofacial Surgery*. , John Wiley & Sons, Ltd.: 126-144.
- Zweig MH & Campbell G (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39(4): 561-577.

**Table 1** Means  $\pm$  SDs for the measured variables grouped by age

	OCLR [%]	DD [mm]	PCAI [%]	wAS	CI [%]
3-month-olds	103,0 $\pm$ 2,7	3,9 $\pm$ 3,4	9,0 $\pm$ 8,3	25,9 $\pm$ 28,3	76,9 $\pm$ 4,3
6-month-olds	102,5 $\pm$ 2,7	3,5 $\pm$ 3,1	8,0 $\pm$ 7,0	20,3 $\pm$ 27,2	78,3 $\pm$ 4,4
12-month-olds	102,3 $\pm$ 1,9	3,4 $\pm$ 2,8	7,3 $\pm$ 6,4	19,5 $\pm$ 24,8	77,5 $\pm$ 4,1
3-year-olds	102,2 $\pm$ 1,5	3,5 $\pm$ 2,4	6,5 $\pm$ 5,6	24,0 $\pm$ 26,1	77,0 $\pm$ 3,3
Overall	102,6 $\pm$ 2,3	3,6 $\pm$ 3,0	7,9 $\pm$ 7,1	22,3 $\pm$ 32,5	77,5 $\pm$ 4,2

Means  $\pm$  SDs for the measured variables grouped by Argenta score

	OCLR [%]	DD [mm]	PCAI [%]	wAS	CI [%]
No deformation	101,6 $\pm$ 1,2	2,3 $\pm$ 1,7	5,1 $\pm$ 3,8	11,2 $\pm$ 11,2	77,0 $\pm$ 3,6
Any deformation	105,2 $\pm$ 2,3	7,1 $\pm$ 3,0	15,6 $\pm$ 8,2	52,3 $\pm$ 48,4	78,9 $\pm$ 5,1
Type I	104,2 $\pm$ 1,8	5,8 $\pm$ 2,5	12,8 $\pm$ 6,6	37,2 $\pm$ 27,1	77,9 $\pm$ 5,3
Type II	105,9 $\pm$ 1,8	8,1 $\pm$ 2,4	17,6 $\pm$ 6,6	60,4 $\pm$ 42,1	80,8 $\pm$ 4,8
Type III	108,1 $\pm$ 1,6	10,6 $\pm$ 1,9	23,7 $\pm$ 10,2	88,6 $\pm$ 7,9	79,4 $\pm$ 3,2
Type IV	113,6	16,6	44,9	120	80

**Table 2** Cut-off values for DP (Argenta score  $\geq 1$ ) and their respective (Youden's *J* values) in different age-groups

	OCLR [%]	DD [mm]	PCAI [%]	wAS
3-month-olds	103,5 (0,71)	4,3 (0,71)	10,5 (0,66)	24,1 (0,76)
6-month-olds	103,4 (0,75)	4,5 (0,75)	11,1 (0,67)	28,5 (0,67)
12-month-olds	103,6 (0,85)	5,1 (0,84)	10,4 (0,82)	28,0 (0,79)
3-year-olds	103,7 (0,99)	6,2 (0,99)	10,8 (0,87)	30,7 (0,94)

**Table 3** Cut-off values for DP and respective (*Youden's J values*) from the final ROC curves

<b>Excluding discrepant expert ratings</b>			
<i>Argenta score</i>	OCLR [%]	PCAI [%]	wAS
$\geq 1$	103,5 (0,82)	10,5 (0,80)	24,5 (0,78)
$\geq 2$	104,5 (0,87)	13,0 (0,82)	38,9 (0,79)
$\geq 3$	105,7 (0,93)	15,2 (0,83)	58,7 (0,96)
<b>Including discrepant expert ratings</b>			
<i>Argenta score</i>	OCLR [%]	PCAI [%]	wAS
$\geq 1$	103,5 (0,71)	10,5 (0,65)	24,5 (0,68)
$\geq 2$	104,5 (0,78)	13,0 (0,73)	38,9 (0,73)
$\geq 3$	105,7 (0,91)	15,2 (0,78)	58,7 (0,95)

Figure 1 part 1  
[Click here to download high resolution image](#)

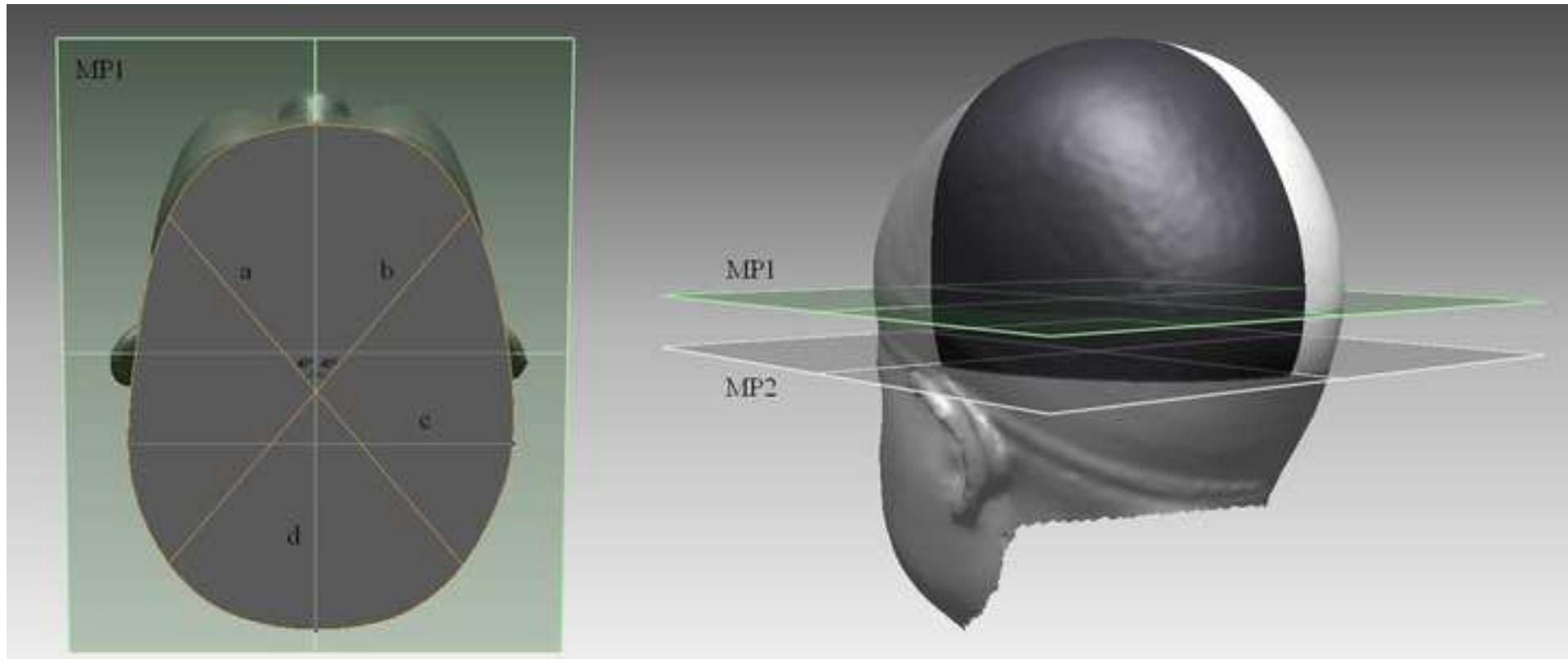


Figure 1 part 2

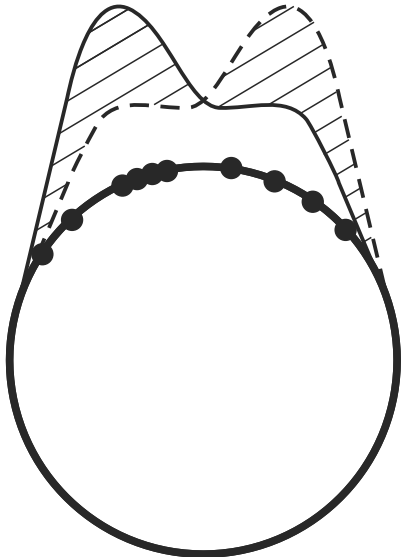
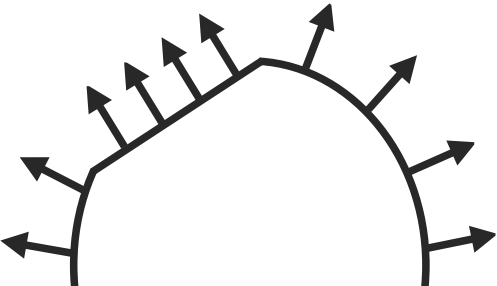




Figure 2

[Click here to download high resolution image](#)

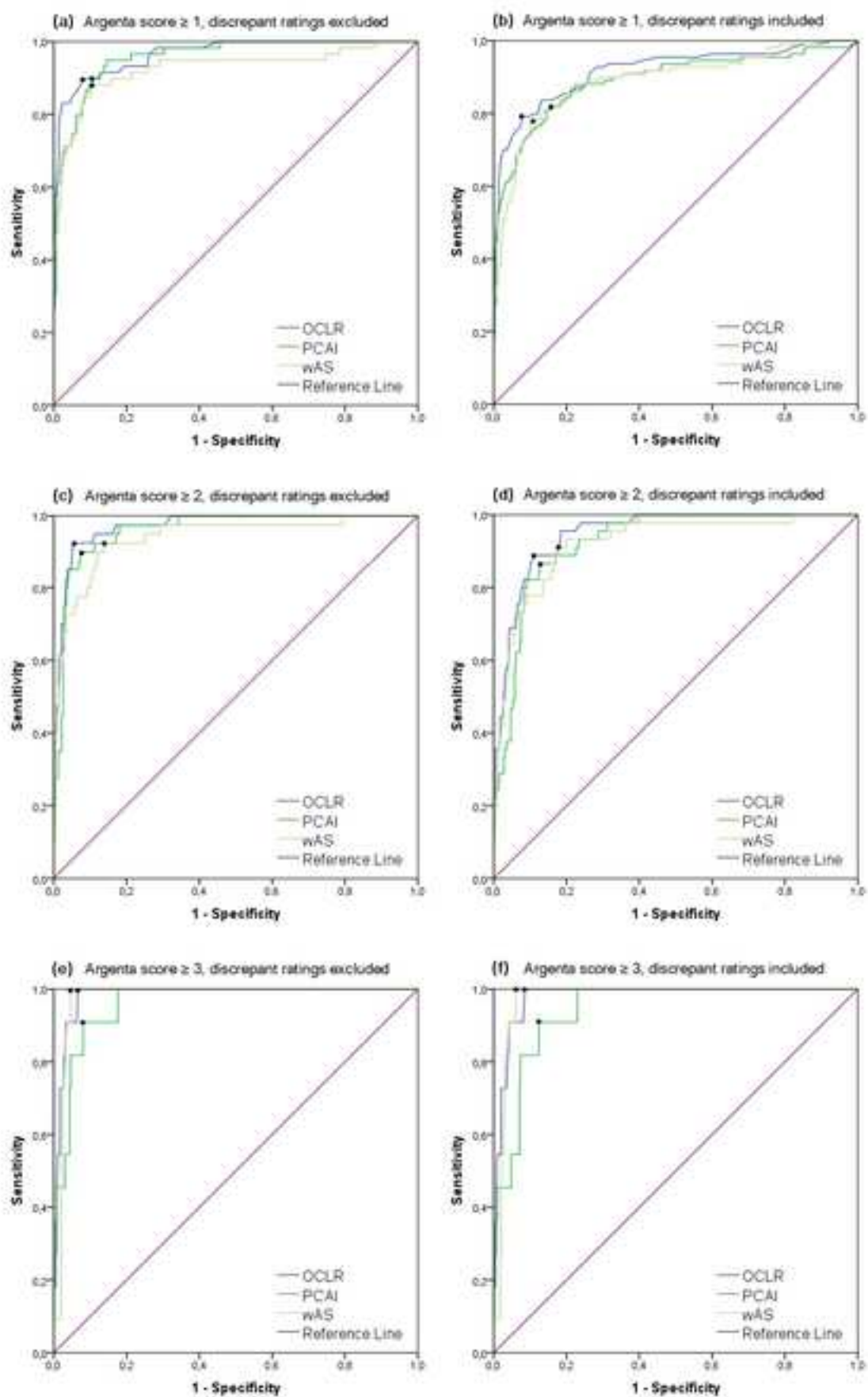
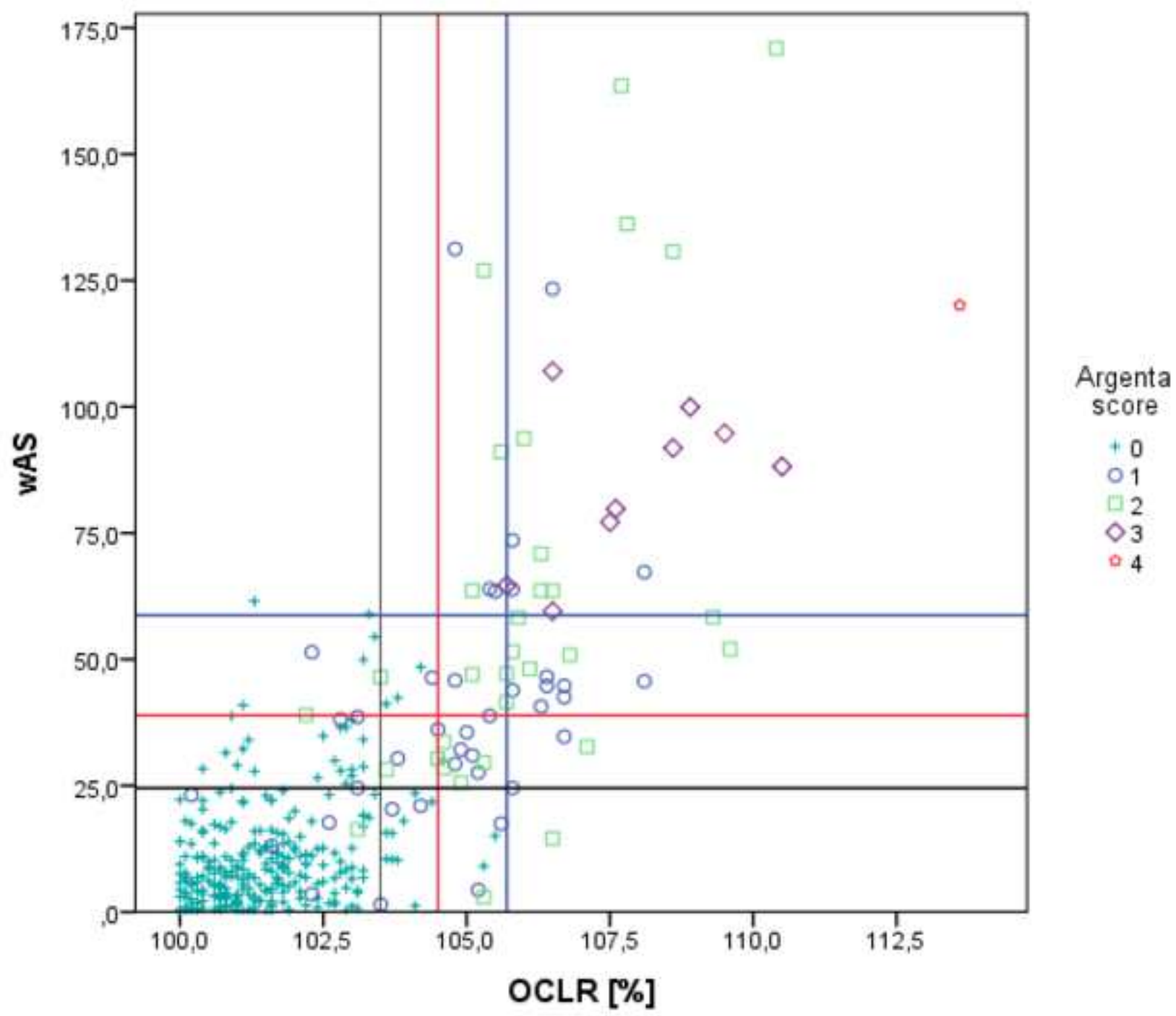


Figure 3  
[Click here to download high resolution image](#)



**Figure 1** The measurement plane for OCLR, DD, and CI is indicated as MP1, and the plane for PCAI and wAS as MP2.  $OCLR = a / b \times 100 \%$ , when  $a \geq b$ .  $DD = |a - b|$ .  $CI = c / d \times 100 \%$ . The black and white areas indicate the posterior quadrants, from where PCAI is calculated as (larger cuboid volume – smaller cuboid volume) / smaller cuboid volume  $\times 100 \%$ . wAS is calculated from the surface normal vector directions: on a flat area the surface normal vectors have similar directions, leading to a local maximum in the density function of direction angles, resulting in higher wAS scores when the flat area(s) are located asymmetrically within the posterior quadrants.

**Figure 2** ROC curves for the outcome variables OCLR, PCAI, and wAS. Separate curves are provided for Argenta classes 1-3, both excluding and including images with discrepant expert ratings. Youden's J are marked with a black dot, and the corresponding sensitivity (sens.) and specificity (spec.) values and the AUC values for each curve are as follows:

(a) OCLR = sens. 90%, spec. 92%, AUC 0.97; PCAI = sens. 90%, spec. 90%, AUC 0.96; wAS = sens. 88%, spec. 90%, AUC 0.93

(b) OCLR = sens. 79%, spec. 91%, AUC 0.92; PCAI = sens. 76%, spec. 90%, AUC 0.89; wAS = sens. 78%, spec. 90%, AUC 0.89

(c) OCLR = sens. 93%, spec. 95%, AUC 0.97; PCAI = sens. 90%, spec. 93%, AUC 0.96; wAS = sens. 85%, spec. 93%, AUC 0.94

(d) OCLR = sens. 89%, spec. 89%, AUC 0.95; PCAI = sens. 87%, spec. 87%, AUC 0.92; wAS = sens. 82%, spec. 91%, AUC 0.92

(e) OCLR = sens. 100%, spec. 93%, AUC 0.98; PCAI = sens. 91%, spec. 92%, AUC 0.96; wAS = sens. 100%, spec. 95%, AUC 0.98

(f) OCLR = sens. 100%, spec. 91%, AUC 0.98; PCAI = sens. 91%, spec. 88%, AUC 0.94; wAS = sens. 100%, spec. 94%, AUC 0.97

**Figure 3** Argenta scores for the participants, with wAS plotted against OCLR. Discrepant expert ratings not shown. The three colored lines represent the suggested cut-off values from the ROC curves for wAS and OCLR: black for Argenta score  $\geq 1$ , red for  $\geq 2$ , and blue for  $\geq 3$ .

**Appendix 1** Originally published in A Collection of Images and Illustrations by the Department of Plastic Surgery, Wake Forest University of Medicine. Used with permission.