Accepted Manuscript

Study on the criteria for assessing skull-face correspondence in craniofacial superimposition

Oscar Ibá ñez, Andrea Valsecchi, Fabio Cavalli, María Isabel Huete, Blanca Rosario Campomanes-Alvarez, Carmen Campomanes-Alvarez, Ricardo Vicente, David Navega, Ann Ross, Caroline Wilkinson, Rimantas Jankauskas, Kazuhiko Imaizumi, Rita Hardiman, Paul Thomas Jayaprakash, Elena Ruiz, Francisco Molinero, Patricio Lestón, Elizaveta Veselovskaya, Alexey Abramov, Maryna Steyn, Joao Cardoso, Daniel Humpire, Luca Lusnig, Daniele Gibelli, Debora Mazzarelli, Daniel Gaudio, Federica Collini, Sergio Damas

PII: S1344-6223(16)30104-3
DOI: http://dx.doi.org/10.1016/j.legalmed.2016.09.009
Reference: LEGMED 1361

To appear in: *Legal Medicine*

Received Date: 14 July 2016
Revised Date: 23 September 2016
Accepted Date: 28 September 2016

Please cite this article as: Ibá ñez, O., Valsecchi, A., Cavalli, F., Isabel Huete, M., Rosario Campomanes-Alvarez, B., Campomanes-Alvarez, C., Vicente, R., Navega, D., Ross, A., Wilkinson, C., Jankauskas, R., Imaizumi, K., Hardiman, R., Thomas Jayaprakash, P., Ruiz, E., Molinero, F., Lestón, P., Veselovskaya, E., Abramov, A., Steyn, M., Cardoso, J., Humpire, D., Lusnig, L., Gibelli, D., Mazzarelli, D., Gaudio, D., Collini, F., Damas, S., Study on the criteria for assessing skull-face correspondence in craniofacial superimposition, *Legal Medicine* (2016), doi: http://dx.doi.org/10.1016/j.legalmed.2016.09.009

Article Title: Study on the criteria for assessing skull-face correspondence in craniofacial superimposition

Authors: Dr. Oscar Ibáñez[1,2], Dr. Andrea Valsecchi[2], Dr. Fabio Cavalli[3], MS María Isabel Huete[4], Dr. Blanca Rosario Campomanes-Alvarez[2], MS Carmen Campomanes-Alvarez[2], MS Ricardo Vicente5,6, MS David Navega[5,6], Prof. Ann Ross[7], Prof. Caroline Wilkinson[8], Prof. Rimantas Jankauskas[9], Dr. Kazuhiko Imaizumi[10], Prof. Rita Hardiman[11], Prof. Paul Thomas Jayaprakash[12], MS Elena Ruiz[13], MS Francisco Molinero[14], MS Patricio Lestón[14], Prof. Elizaveta Veselovskaya[15], Prof. Alexey Abramov[16], Prof. Maryna Steyn[17], MS Joao Cardoso[18], Prof. Daniel Humpire[19], MS Luca Lusnig[3], Dr. Daniele Gibelli[20], M Debora Mazzarelli[20], MS Daniel Gaudio[20], MS Federica Collini[20], Dr. Sergio Damas[2].

Affiliation:

[1]Department of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain

[2]European Centre for Soft Computing, Mieres, Spain

[3] Research Unit of Paleoradiology and Allied Sciences, Ospedali Riuniti di Trieste, Trieste, Italy

[4]Physical Anthropology Laboratory, University of Granada, Granada, Spain

[5]Forensic Sciences Centre (CENCIFOR), Coimbra, Portugal

[6]Department of Life Sciences, University of Coimbra, Coimbra, Portugal

[7]Department Biological Sciences, North Carolina State University, Raleigh, USA

[8]Face Lab, Liverpool John Moores University, Liverpool L3 5TF, UK

[9] Department of Anatomy, Histology and Anthropology, Vilnius University, Vilnius, Lithuania

[10] National Research Institute of Police Science, Japan

[11]Melbourne Dental School, University of Melbourne, Melbourne, Australia

[12]Forensic Science Program, School of Health Sciences, Universiti Sains Malaysia, Kubang Kerian, Malaysia

[13]Legal Medicine School, Complutense University of Madrid

[14] Department of Identification - Criminalist Service, Civil Guard, Madrid, Spain

[15]Institute of Ethnology and Anthropology, Russian Academy of Science, Moscow, Russia

[16]Main Department of Criminalistics Investigation Committee of Russia, Moscow, Russia

[17]School of Anatomical Sciences, Faculty of Health Sciences, University of the Witwatersrand, South Africa

[18] Crime scene investigation section, Forensic Laboratory, Portuguese Criminal Police, Lisbon, Portugal

[19]Public Ministry, Lima, Peru

[20]Laboratorio di Antropología e Odontologia Forense, University of Milan, Milan, Italy

Corresponding author: Oscar Ibáñez

E-mail address: Oscar.ibanez@decsai.ugr.es

Phone: +0034666917443

Postal Address: Department of Computer Science and Artificial Intelligence, University of Granada, Granada 18014, Spain

Title: Criteria for assessing skull-face correspondence in craniofacial superimposition

Abstract

Craniofacial superimposition has the potential to be used as an identification method when other traditional biological techniques are not applicable due to insufficient quality or absence of ante-mortem and post-mortem data. Despite having been used in many countries as a method of inclusion and exclusion for over a century it lacks standards. Thus, the purpose of this research is to provide forensic practitioners with standard criteria for analysing skull-face relationships. Thirty-seven experts from 16 different institutions participated in this study, which consisted of evaluating 65 criteria for assessing skull-face anatomical consistency on a sample of 24 different skull-face superimpositions. An unbiased statistical analysis established the most objective and discriminative criteria. Results did not show strong associations, however, important insights to address lack of standards were provided. In addition, a novel methodology for understanding and standardizing identification methods based on the observation of morphological patterns has been proposed.

1. Introduction

Craniofacial superimposition (CFS) [1] is one of the approaches used in craniofacial identification [2, 3]. It involves the superimposition of a skull (or a skull model) over a number of ante mortem images of an individual and the analysis of their morphological correspondence. Since the first documented use of CFS for identification purposes [4] the technique has been undergoing continuous improvement. Although the foundations of the CFS method were laid by the end of the nineteenth century [5, 6], the associated procedures evolved as new technologies became available. As a result, distinct approaches have developed: photographic, video, and computer-aided superimposition [1, 7, 8]. Regardless of the applied technology, some authors have recently described three different stages for the whole CFS process [8, 9]: i) the acquisition and processing of the skull (or skull 3D model) and the ante mortem facial images together with the anatomical landmarks; ii) the skull-face overlay (SFO), which focuses on achieving the best possible superimposition of the skull and a single ante mortem image of the missing person. This process is repeated for each available photograph, obtaining additional overlays. Skull-face overlay thus corresponds to what traditionally has been known as the adjustment of the skull size and its orientation with respect to the facial photograph [10, 11]; iii) decision making. The degree of support for being the same person or not (exclusion) is determined by studying the relationship between the skull and the face based on the superimpositions achieved in the latter SFO stage: the morphological correlation, the match between the corresponding landmarks according to the soft tissue depth, and the consistency between asymmetries.

Although its reliability is still open to discussion, CFS has been employed by both the forensic anthropology community and law enforcement in the identification of unknown persons. It is used together with other techniques or on its own when there is insufficient information available to apply other techniques. CFS has been used for almost a century [2], contributing to the process of many identification cases, especially in scenarios like mass disasters [12], terrorism [13], missing person's identification [14], common grave investigation [15], and historical cases [16]. There is lack of protocols and standards in the application of the technique and varying information concerning its reliability [17-19]. The 'New Methodologies and Protocols of Forensic Identification by Craniofacial Superimposition (MEPROCS) project [20] aims to develop "a common framework to allow the extensive application of the CFS technique in practical forensic identification scenarios commonly tackled by European scientific law enforcement, providing an objective evaluation of the forensic identification results achieved by CFS, avoiding particular assumptions that could bias the process".

To this aim the MEPROCS international consortium, composed of 26 institutions including research centres, universities, police forces and international associations, set out to produce a set of work packages, meetings and inter-lab experimental studies. The latter are intended to provide quantitative and objective data that could support discussions and facilitate decision making processes in an unbiased way. In the first study [21,22] 26 participants from 17 different institutions in 13 countries were asked to deal with 14 CFS identification scenarios, some of them involving the comparison of multiple candidates with multiple unknown skulls. In total, 60 SFO problems, divided into female and male sets, were analysed. Participants followed their own methodologies and employed their own particular technologies. The data obtained from this large study was a key result leading to an international agreement on the first standard in the field. It includes good and bad practices, sources of error and uncertainties, technological requirements and desirable features, and finally a common scale for the craniofacial matching evaluation [22].

However, that study and the subsequent conclusions mainly focused on the process of superimposing the skull over the facial photograph, the aforementioned SFO stage [10]. Although it also deals with the relation between the quantity and quality of the materials (skull, ante mortem photographs) and the degree of support for a given identification decision, it did not cover the analysis of the skull and face anatomical relationship. One of the main reasons for this limitation is due to the different SFOs achieved by the participants in each single case. A visual inspection of participants' results clearly shows a important variability in the superimpositions achieved which, biased the following skull-face relationship assessment stage. In addition, as participants were asked to follow their own methodology, the set of anatomical criteria was different for each participant.

As a consequence, the MEPROCS consortium designed the current study which aims to analyse the subjectivity and discriminative power of the different criteria for assessing the skull-face correspondence either proposed in the literature or by any of the MEPROCS partners. The following four tables (Tables 1 to 4) group all the craniofacial assessment criteria MEPROCS partners considered relevant. The 65 criteria represent an exhaustive list of the criteria employed and described in the most important studies in the field [1, 17, 18, 23-25]. These criteria, to be used in the assessment of the consistency between the skull and the face, are organized in four different groups analyzing anatomical criteria such as lines, landmarks and the corresponding soft tissue thickness, the concordance between the outlines of the face and the cranium, and positional relationship of specific facial and cranial features.

Table 1

Table 2

Table 3

Table 4

This novel study is expected to provide important insights to better understand: i) which are the most and least discriminative criteria; ii) which criteria depended more on the expert and which criteria are more independent, i. e. less subjective. Those criteria that are determined to be more discriminatory could be included as a recommended standard for CFS.

2. Material and methods

The dataset used in this study consisted of 18 different CFS problems, some datasets included more than one image of the same subject (24 SFO in total). Three-dimensional skull models were obtained from patients whose head was scanned with a Cone Beam Computed Tomography (CBCT) in orthostatic position. The patients gave their consent to use their clinical data and their scans were anonymized for this study.

In addition, frontal and lateral photographs were taken of the same patients to create a set of "positive matches", while other individuals with similar (according to Dr. Cavalli's personal criteria) facial geometry were photographed in order to compose a set of "negative matches". Half of 18 cases were positive matches and the other half were negative matches. Twelve of the photographs were lateral views and 12 were frontal views, half assigned as positive matches and the other half as negative matches.

The participants were directed to focus on the criteria for analysing only the skull and face relationship. The procedure to obtain each type of SFO differed.

For positive cases, optimal SFOs were achieved using the following procedure. The DICOM images resulting from the CBCT machine were automatically processed to obtain the corresponding 3D face and 3D skull models. After positioning homologous anatomical points in both the 3D face model and the photograph, the former was automatically projected onto the latter to obtain an ideal match. Then, the parameters originating from that match between the 3D face model and the photograph were applied to the 3D skull model resulting in an objective and accurate SFO. The latter superimposition is considered a ground truth SFO. Figure 1 presents an overview of the ground truth data creation process. Detailed explanations of this process are given in [26].

For negative cases, the SFOs were assembled by an anthropologist with experience in CFS (Ms. Maribel Huete) using Face2Skull^{TM} software. She was asked to assign the best possible SFO and to judge the skull-face relationship without being informed of the actual negative relationship to avoid biasing the SFO process.

2.1. Methodology

The participants were provided with the same 24 SFOs as a single image with four different layers (see Figure 2): face photograph with and without landmarks, skull projection with and without landmarks. First, participants were asked to indicate which criteria they using for evaluating the skull-face relationships (see Tables 1 to 4 in the supplementary material). Then, they were asked to evaluate the skull-face correspondence following a systematic approach.

Convincing studies quantifying a match between two patterns in individualization methods have not been reported to date [27]. Stoney [28], referring to the Daubert-driven challenge, emphasizes that to obtain scientific endorsement of the method, the level and degree of correspondence needs to be met. Thus, a scale to evaluate the degree of physical match was used. Because a certain level of subjectivity is involved in the comparisons, it is not possible to define each level within this scale, and the determination of the matching degree will correspond to the examiner judgment. Thus, for each SFO the degree of consistency of all the criteria previously selected were indicated using the following values:

- 0-not evaluable
- 1-no match
- 2-poor match
- 3-doubtful match
- 4-good match
- 5-perfect match

In order to avoid subjective interpretations, MEPROCS partners assigned in advance (before giving the instructions to the participants) the value 0 to those criteria they considered unable to be visually checked due to the noisy nature of the image (e.g. absence of the bony part or the pose of the face in the photograph). This was carried out for each SFO case. Specifically, evaluation of soft tissue thickness between trichion - trichion (2.10), evaluation of skull and face height (3.2), and confirming if the length of the skull from menton to bregma fits within the face (3.4) were removed from the study since it is not possible to objectively evaluate these in any of the SFO cases. A different set of criteria (1.2, 2.3, 2.4, 2.7, 3.1, 3.10, 4.8, 4.11, 4.14, 4.17) were then identified as “not able to evaluate” in some of the SFO cases.

Finally, for each SFO case (and also for each CFS case when it applies more than one SFO), participants were asked to indicate the final identification decision according to the following scale [22]:

- -3-strong support of not being the same person
- -2-moderate support of not being the same person
- -1-limited support of not being the same person
- 0-undetermined
- +1-limited support of being the same person
- +2-moderate support of being the same person
- +3-strong support of being the same person

A total of 37 individuals (e.g. MSc and PhD students, post-doc researchers, professors, and practitioners) from 16 different institutions participated in the study.

2.2. Data Analysis

Three statistical analyses were conducted with the following characteristics:

1. Depending on the data employed
   a. With all the data provided by the 37 participants
   b. Filtering (removing) the participants with a proficiency score (i.e. the rate of cases in which the expert answered correctly) lower or equal than 0.5
   c. Filtering the scenarios (SFO cases) with highest standard deviation (fourth quartile)
2. According to the view of the photographs: frontal vs. lateral
3. According to the family of criteria: lines, landmarks-soft tissue, outlines, and positional relationship

The statistical analyses applied, which relied on several concepts are introduced below along with an example:

- Cases with decision: the cases in which the expert's decision is not "undetermined" (i.e. different from 0)
- Expert proficiency: the proportion of cases with decision in which the expert evaluated the status of the case correctly.

  $\mathrm{EP} = \frac{\mathrm{TP+TN}}{\#\text{ of cases with decision}}$

  where TP is the number of positive cases with a positive decision and TN is the number of negative cases with a negative decision.
- Correlation coefficient. Correlation is any statistical dependence between two variables that have a linear relationship. Among different measures of correlation, we used Spearman's rank correlation coefficient, which can also assess a non-linear correlation between two variables X, Y. It is calculated as the linear correlation coefficient between the ranks of X, Y (i.e. the position of each value after sorting); that is $\frac{\sum_i (r(x_i)-\overline{r(x)})(r(y_i)-\overline{r(y)})}{\sqrt{\sum_i (r(x_i)-\overline{r(x)})^2 \sum_i (r(y_i)-\overline{r(y)})^2}}$ where r is the function that assigns a rank to each variable value.

For example, for a set of four cases i (X,Y) the rank r and the Spearman rank correlation p was calculated as follows

i	X	Y	r(X)	r(Y)
1	1	1	2	2
2	2	3	3	4
3	3	-1	4	1
4	0	2	1	3

$$\rho = \frac{(2-2.5)(2-2.5)+(3-2.5)(4-2.5)+(4-2.5)(1-2.5)+(1-2.5)(3-2.5)}{\sqrt{((2-2.5)^2+(3-2.5)^2+(4-2.5)^2+(1-2.5)^2)\,((2-2.5)^2+(4-2.5)^2+(1-2.5)^2+(3-2.5)^2)}}$$

$$\rho = \frac{-2}{5} = -0.4$$

- Criterion correlation with status: for each criterion G, it measures the correlation between the value of a criterion and the status of a case. The values assess the tendency of a criterion to have higher values on positive cases and lower values on negative ones. It is defined as the Spearman's rank correlation coefficient between the criterion values assigned by the expert and the status of a case.
- Using the data from the previous example, consider criterion A. For each expert, we compute the Spearman's rank correlation coefficient between the criterion values and the status of the corresponding case. Therefore, for Smith we calculate Spearman ((1,4,3), (0,0,1)) = 0. For Doe, Spearman ((3,2,4), (0,0,1)) = 0.866 with an average value of 0.43.
- Weighting. In parts of the analysis, the different measures are computed taking into account the proficiency of the expert. This means the most reliable experts have a higher weight (and thus, a larger influence on the outcome) and vice versa. For instance, the weighted criterion correlation with status is measured as the correlation with status, except that the correlation coefficients associated with each expert are weighted according to their proficiency.
- Consider criterion A again in the example data. This time we compute a weighted average between 0 and 0.866, using the weights 0.5 and 0.86, resulting in a mean of 0.55.
- Criterion variability: was computed as the standard deviation of the criterion evaluation over the same case, which is then averaged over all cases. It aims to assess the subjectivity of a criterion.

3. Results

With the aim of providing a feasible forum for discussion this study focused on the analysis on one scenario, where all the data (participants and CFS cases) are considered simultaneously (see subsection Data Analysis, 1.a). In addition, in segments of this manuscript, the second scenario where the data were divided in two different sets according to the view of photograph (frontal and lateral) is addressed. The remainder scenarios provided similar results (scenarios 1.b and 1.c) or resulted of minor interest (scenario 3). The remaining scenarios provided similar results (scenarios 1.b and 1.c) or resulted of minor interest (scenario 3) for the anthropology community.

3.1. Criterion usage

Figure 3 (a and b) shows the number of times a criterion was evaluated over the total number of assessments (each participant for each SFO case, “Usage C”). It also shows the percentage of participants that employed a criterion at least once or “Usage P”. Statistics are presented for all the cases (“All”), only frontal cases (“Frontal”), and only lateral cases (“Lateral”). Criteria that were employed in less than 10% of the cases were not considered. Specifically, the criterion evaluating the soft tissue thickness consistency between mastoidale - subaurale (2.26) does not reach the required 10% of usage. Criteria that were employed by less than 30% of the participants were also filtered out. These were omitted in order to avoid potential lack of significance resulting from a reduced data set. It is also important to note that some criteria that were classified as able to evaluate in frontal view in the literature were evaluated by the participants in lateral view, which included the assessment of zygion to zygion soft tissue estimation (2.9), ectocanthion line (1.1), consistency of the oblique line of the mandible (3.18), and the positional relationship of nasion in comparison to the nasal root (4.10). Similarly, two criteria categorized *a priori* for being examined in lateral view, which included the consistency of gonial and jaw angle outlines (3.7), consistency between the outline of the frontal process of the zygomatic bone, and the process seen in the face (3.8).

3.2. Expert proficiency

Figure 4 shows the overall poor average performance for the expert proficiency achieved by each participant. Only two participants had a correct rate of 75%. However, 37% (14 participant experts) did not achieve 50% correct match. As a result, the 14 experts were removed from one sub-study (1.b) and from subsequent analyses.

A boxplot with the experts’ assessment across the scenarios is depicted in Figure 5. This boxplot shows the variability within each of the experts’ responses. In general, both negative (red bars) and positive (blue bars) cases have similar performance rates although a lower variability resulted in the evaluation of the positive cases. While there are only two negative cases (4-2 and 11-1) where most of the participants (≥ 75%) made a correct evaluation, there are four positive cases with a similar successful evaluation (3-1, 7-1, 13-1, and 18-1). The median values (black horizontal line inside the boxes) shows three negative cases that were incorrectly assessed by most of the participants (SFO cases 4-1, 15-1, and 16-1). The median values of the other three cases (8-1, 10-1, and 17-1) fall within the undetermined category (value 0). Similarly, there are three positive cases that were incorrectly evaluated by most of the participants (SFO cases 5-1, 5-2, and 12-1) with 75% of the participants not making the correct identification. Differences were not observed between the evaluations of lateral versus frontal views.

3.3. Criterion subjectivity

Subjectivity has been measured as the standard deviation of the evaluations. The standard deviation was computed for each case across the different experts. Then, the values were averaged across all the cases. Because the ranges of values for the criteria are within the intervals (1, 5), results show that there is a significant distribution in the evaluations by the different participants with standard deviations ranging from 0.857 to 1.315 (see Figure 6).

3.4. Criterion discriminative power

Figure 7 corresponds to Spearman's rank correlation coefficient between a criterion's value and the status of the case. According to this study, criteria 3.1 (the outline of the frontal bone follows the forehead outline) and 3.19 (dental matching) are the most correlated (in this order). Criteria 4.14 (the landmark porion aligns posterior to the tragus, slightly inferior to the crus of the helix) and 3.16 (the sagittal outline of the nasal cartilage is the mirror image of the contour of the pyriform aperture) have a much lower correlation. The discriminative power of the rest of criteria decrease linearly (as depicted in the Figure 7) from criteria that are still correlated, which include 1.8, 1.7, 2.27, 2.9, 4.12, 2.19, 2.20, 3.13, 2.4, 2.18, 3.12, 2.3, and 3.10, to criteria holding a low correlation with the final decision 1.2, 2.1, 1.6, 2.13.

3.5. Criterion subjectivity and discriminative power

The scatter plots (Figure 8 and 9) visualize, at the same time, criteria standard deviation (related to the ease of objective assessment) and correlation with status (related to the discriminatory power of a single criterion), both weighted (multiplied) using the expert’s proficiency. Figure 8 and 9 distinguish criteria in both frontal and lateral views, respectively. In both cases, criteria with similar characteristics were grouped (both subjectivity and those with discriminative power).

For the frontal cases, five groups were differentiated. At the top of scatter plot the criterion with the highest discriminative power, dental matching (3.19), is clearly observable. Below this criterion, criteria with a good trade-off between subjectivity and discriminative power were identified (4.20, 4.5, 1.7, 2.18, 1.8, 4.2, 4.7, 4.1, 2.22, and 2.23). On the left side of the scatter plot, an easy to evaluate criterion is depicted (3.10) with the least amount of variability. In addition, it is an important criterion in terms of discriminative power with high correlation values. In the center of the scatter plot the majority of the criteria that show the least amount of difference among them is presented. Finally, the right bottom region groups the least useful criteria with the highest subjectivity, which cannot be used to discriminate between face and skull (1.2, 2.1, 1.6, 4.18, 4.11, 2.3, 3.6, 1.1, 2.8, 3.18, 1.4, 2.6, and 4.10). Note that 4.8 (nasion is higher than the nasal root) and 2.2 (soft tissue thickness at gnathion – menton) refer to the same anatomical correspondence criteria as 4.11 and 2.1, respectively, for which are analysed in different image views.

For the lateral views, although they can be grouped into eight separate groups, the two groups in the center (between correlation values of 0 and 1.2) are considered as part of the same group of criteria with almost no discriminatory power. On the top left corner the best criteria are represented (i.e., outline of the frontal bone follows the forehead outline, 3.1, and the external auditory meatus opening lies medial to the tragus of the ear, 4.17). They also have the greatest discriminatory power and least variability. The red curve at the top right corner encloses a group with high discriminatory power but with high variability (4.14 and 3.19). Below the latter, a group was identified (criteria 2.4, 3.11, 2.9, 1.1, 4.10, 2.12, and 2.19) that showed important correlation with the identification decisions, although they showed significant variability. Similarly, criterion 3.18 holds important discriminatory power with significantly lower variability. As in the other two sections, the central part of the scatter plot contains the majority of the criteria, which do not hold significant correlation values. Finally, the bottom right part of the scatter plot contains the criteria with the greatest subjectivity and lowest discriminating power (4.15, 3.9, 2.13, 4.11, and 4.16). Note that 3.14 refers to the same anatomical correspondence criterion as 3.9 although it was analysed in different image views.

4. Discussion and conclusions

The purpose of this experimental study was to provide quantitative and objective data for discussion in order to attempt to reach an international agreement (among MEPROCS partners) of a set of recommended craniofacial assessment criteria.

Although we have examined all the material produced according to the three different statistical analyses and the different types of correlation, the results and discussion included in this manuscript focus only on the most relevant results. In particular, we have included results corresponding to all the data provided by the 37 participants and also separately analysing the correlation according to the view of the photographs, frontal vs. lateral. Among the three different ways of calculating the correlation, we focus on the "simple weighted correlation". We preferred it against "no weighed correlation" since we think that decision mechanisms followed by better performing participants have to somehow bias the study. We preferred "simple weight" against "correlation-based weight" since the scale employed for evaluating each CFS case not only considered the matching degree but also the quality and quantity of the materials, variables that were not introduced in the statistical analysis.

The most important differences, and at the same time the more practical ones, were found when subjectivity and discriminative power are considered together. This procedure applied to the data provided by all the participants lead to an initial set of recommended criteria, i.e., the one described together with Figures 8 and 9 for frontal and lateral view, respectively. However, although this set of criteria was directly and objectively obtained from the statistical study it reflects a few inconsistencies and errors as we detail as follows:

- Criterion 4.17, “The external auditory meatus opening lies medial to the tragus of the ear”, was identified as one of the best criteria for lateral view. However it is quite obvious that this can’t be determined in lateral view. When we provided the template to the participants we accidentally pre-assigned value “0” to this criteria in the frontal cases instead of in the lateral/oblique cases. That explains why less than 40% of the participants assessed 4.17. Thus, we have decided to remove it from the final set of recommended criteria.
- Criteria 2.9 “Evaluate soft tissue thickness at Zygion – zygion”, 4.10 “medial margin of orbit aligns and superimpose with the endocanthion”, 2.12 “Evaluate soft tissue thickness at ant lacrimal crest - medial canthus”, 2.19 “Evaluate soft tissue thickness at Porion – tragion”, and 1.1 “ Ectocanthion line”, showed an important discriminative power and a significant variability in lateral view in the statistical analysis. However, it is considered a criterion that can only be assessed in frontal view. The template should have constrained 2.9, 4.10, 2.12, 2.19, and 1.1 values to “0” in lateral/oblique views but it was not. Thus, we have decided to remove it from the final set of recommended criteria.
- Criteria 4.5 “The stomion lies at the central incisors (at the occusal line)” was identified as a criterion with a good trade-off between subjectivity and discriminative power in frontal view. However, it was removed since criterion 4.7 “The stomion lies at the central incisors (Incisal margin of the upper incisors)” refers to the same anatomical criterion and it was also included in the same category.

Thus, MEPROCS partners agreed on a reduced set of assessment criteria split into two different sets according to the view of the photograph (e.g. frontal and lateral). The recommended criteria for frontal view photographs were reduced to 10, and are shown in Table 5. The single best criterion was found to be dental information, but this is obviously unfortunately only usable if the teeth are visible on an antemortem photograph. Other criteria included features of the postero-lateral jaw, the fit of the soft tissues of the mouth and the eye.

Table 5

Similarly, the recommended criteria for lateral view photographs are shown in Table 6 and included the outline of the frontal bone as the best criteria. Other good criteria included once again the dental information, the positional relationship of the porion and tragus, as well as the outline of the mandible, the evaluation of the soft tissue thickness at glabella, and outline of the face. The previous tables were directly obtained from quantitative and objective data. This is the first time for almost a century of use of the CFS technique that such a study has been carried out. Thus, there is a great deal of room for improvement in relation to rigour and reliability. Although the aim of the study was to provide objective data to guide the selection of a recommended set of criteria, the design of the study, the quality and quantity of the data employed, the number of participants, and the statistical analysis could be improved in the future.

Table 6

The percentage of correct identification decisions was very low (see Figure 4) even if an ideal SFO was provided for the positive cases. The performance (identification rates) is significantly worse than previous studies that also involved the SFO stage [17-19, 21, 23-25]. Possible explanations for this low performance rates are:

1. The absence of a complete cranium (the maximum field of view of CBCT is optimized for the acquisition of the jaw or malar). Thus, the criteria recognized as having the highest discriminative power in this study may differ if a whole skull image would had been available
2. The quality of some 3D models, which in some cases present noisy parts and artefacts. The displayed grey levels in CBCT systems are arbitrary and do not allow for the assessment of bone quality as performed with Hounsfield Units (HU) in medical CT. In addition, the signal / noise ratio is lower than Multiple Detector Computed Tomography (MDCT)
3. The precision of the positive SFO provided, although demonstrated to be accurate, do not represent a perfect solution. Among the 37 participants, one of them claimed that 3 SFO cases are wrongly performed
4. The material given to the participants did not include the 3D skull models but just a projection on a 2D plane
5. The lack of facility to examine the SFOs in wipe mode
6. The isolation of the decision making stage given a SFO

While the negative influence of the first five issues outlined above is quite evident, the sixth is not clear at all. In contrast to recent computerized approaches that clearly differentiates between SFO and Decision making stages [8, 9, 10], the process of overlaying the skull over the face traditionally [2, 11, 25] involved a continuous comparison of the skull-face relationship. While, this task is avoided in the present study since the participants directly received the SFO, the approach followed guarantees a more objective framework where all the participants examine the list of criteria using the same SFO (that are objective and accurate in the case of positive cases). Thus, positional relationship criteria, consistency of the outlines/morphological curves, lines or soft tissues thickness are evaluated in exactly the same conditions by all the participants.

Thus, in order to have more reliable and significant data that could lead to a new set of recommended criteria we think that at least the first four issues have to be overcome. The first three problems are directly related to the use of CBCT scan models and we do not know how to solve them. We preferred this technology rather than MDCT for the following reasons: i) It obtains a scan in orthostatic position avoiding gravity deformations of the soft tissue [30]; ii) There is no systematic error comparing average homologous landmark coordinates in conventional digital cephalograms and CBCT-generated cephalograms [31]; iii) The low exposure dose of CBCT [32]. The fourth issue has a simple solution, share also the 3D models. Similarly, the fifth can be easily tackled providing a list of existing open source software that allows wipe mode examination of images.

Additional analysis of the data could have been carried out. In this sense, another proposal is to extract criteria by level or degree of importance. For that aim, we could bring concepts from multi-criteria optimization [33]. Two conflicting variables, correlation and deviation have to be maximized and minimized, respectively. The criteria for each level could be selected using the pareto-front concept [33] and removing those criteria from the pareto in order to select the criteria of the subsequent level.

Acknowledgements

We would like to thank all the participants that give us the permission to work with both their head scan and facial photographs and Drs. Luca Contardo and Domenico Dalessandri to the support provided during images acquisition and head scanning. We are also grateful towards the Universitary Hospital of Trieste and Ortoscan for supporting this research.

Bibliography

1. Yoshino M. Craniofacial Superimposition. In: Wilkinson C, Rynn C, editors. Craniofacial Identification. University Press, Cambridge, 2012: 238-253

2. Aulsebrook WA, Iscan MY, Slabbert JM, Beckert P. Superimposition and reconstruction in forensic facial identification: a survey. Forensic Sci Int 1995; 75(2-3):101-120

3. Stephan CN. Craniofacial identification: Techniques of facial approximation and craniofacial superimposition. In: Blau S, Ubelaker DH, editors. Handbook of Forensic Anthropology and Archaeology, Left Coast Press, Walnut Creek, 2009; 25:304-321

4. Glaister J, Brash JC. Medico-legal aspects of the Ruxton case. E and S Livingstone, Edinburgh, 1937

5. Bertillon A. The Bertillon System of Identification. McClaughry RW, Chicago, IL, 1896

6. Broca P. Instructions craniologiques et craniométriques de la Société d'Anthropologie de Paris [in French]. In: G Masson, editor, vii, Paris, 1875

7. Ubelaker DH, Bubniak E, Odonnell G. Computer-Assisted Photographic Superimposition. J Forensic Sci 1992; 37(3):750-762

8. Damas S, Cordón O, Ibáñez O, Santamaría J, Alemán I, Botella M. Forensic identification by computer-aided CFS: A survey. ACM Comput Surveys 2011; 43(4): article 27

9. Huete MI, Kahana T, Ibáñez O. Past, present, and future of craniofacial superimposition: Literature and international surveys. Legal Medicine 2015; 17(4):267-78

10. Campomanes-Álvarez BR, Ibáñez O, Navarro F, Alemán I, Botella M, Damas S, Cordón O. Computer Vision and Soft Computing for Automatic Skull-Face Overlay in Craniofacial Superimposition. Forensic Sci Int 2014; 245:77-86.

11. Fenton TW, Heard AN, Sauer NJ. Skull-photo superimposition and border deaths: identification through exclusion and the failure to exclude. J of Forensic Sci 2008; 53:34-40.

12. Al-Amad S, McCullough M, Graham J, Clement J, Hill A. Craniofacial identification by computer-mediated superimposition. J Forensic Odontostomatol 2006; 24(2):47-52

13. Indriati E. Historical perspectives on forensic anthropology in Indonesia. In: Blau S, Ubelaker D, editors. Handbook of Forensic Anthropology, Left Coast Press, 2009;115-125

14. Ross AH. Use of digital imaging in the identification of fragmentary human skeletal remains: A case from the Republic of Panama. Foren. Sci. Comm 2004; 6(4)

15. Jankauskas R. Forensic anthropology and mortuary archaeology in Lithuania. Anthropologischer Anzeiger 2009; 67(4):391-405.

16. Kolesnikov LL, Pashinyan GA, Abramov SS. Anatomical appraisal of the skulls and teeth associated with the family of Tsar Nicolay Romanov. Anat Rec 2001; 265(1):15-32

17. Austin-Smith D, Maples WR. The reliability of skull/photograph superimposition in individual identification. J Forensic Sci 1994; 39(2):446-55

18. Chai DS, Lan YW, Tao C, Gui RJ, Mu YC, Feng JH, et al. A Study on the standard for forensic anthropological identification of skull-image superimposition. J Forensic Sci 1989; 34(6):1343-56

19. Gordon GM, Steyn M. An investigation into the accuracy and reliability of skull-photo superimposition in a South African sample. Forensic Sci Int 2012; 216(198):1-6

20. http://www.meprocs.eu

21. Ibáñez O, Vicente R, Navega DS, Wilkinson C, Jayaprakash PT, Huete MI, et al. Study on the performance of different craniofacial superimposition approaches (I). Forensic Sci Int 2015; 257:496-503.

22. Damas S, Wilkinson C, Kahana T, Veselovskaya E, Abramov A, Jankauskas R, et al. Study on the performance of different craniofacial superimposition approaches (II): best practices proposal. Forensic Sci Int 2015; 257:504-8.

23. Yoshino M, Imaizumi K, Miyasaka S, Seta S. Evaluation of anatomical consistency in craniofacial superimposition images. Forensic Sci Int 1995; 74:125-134.

24. Lan Y. A study on national differences in identification standards for Chinese skull-image superimposition. Forensic Sci Int 1995; 74:135-153.

25. Jayaprakash PT, Srinivasan GJ, Amravaneswaran MG. Cranio-facial morphanalysis: a new method for enhancing reliability while identifying skulls by photo superimposition. Forensic Sci Int 2001; 117:121-143.

26. Ibáñez O, Cavalli F, Campomanes-Alvarez BR, Campomanes-Álvarez C, Valsecchi A, Huete MI. Ground truth data generation for skull-face overlay. Int J of Legal Med 2015; 129:569–581

27. Jayaprakash PT. Practical relevance of pattern uniqueness in forensic science. Forensic Sci Int 2013; 231(1-3):403.e1-16.

28. Stoney DA. Update, History and development of fingerprinting. In: Lee HC, Gaensslen RE, editors. Advances in Fingerprint Technology, CRC Press, 2001; 52–53.

29. Jacob B, Alt KW. Advances in Forensic Sci: Forensic odontology & anthropology / in cooperation with K. W. Alt ; P. Pieper. Köster 1995; Vol. 7.

30. Gwen R, Swennen J, Schutyser F. Three-dimensional cephalometry: Spiral multi-slice vs cone-beam computed tomography. Am J Orthod Dentofacial Orthop 2006; 130:410-416

31. Grauer D, Cevidanes LSH, Styner MA, Heulfe I, Harmon ET, Zhu H, et al. Accuracy and Landmark Error Calculation Using Cone-Beam Computed Tomography–Generated Cephalograms. Angle Orthod 2010; 80 (2):286–294

32. Loubele M, Bogaerts R, Van Dijck E, Pauwels R, Vanheusden S, Suetens P, Marchal G, Sanderink G, Jacobs R. Comparison between effective radiation dose of CBCT and MSCT scanners for dentomaxillofacial applications. Eur J Radiol 2009; 71 (3):461-468

33. Steuer RE. Multiple Criteria Optimization: Theory, Computation and Application. John Wiley, New York, 1986.

Fig 1 Overview of the ground truth data creation process

Fig 2 Example of the four layers image provided to the participants. From left to right: face photographs with and without landmarks, skull model projections with and without landmarks.

Fig 3 Percentage of criterion usage by all participants (ordered by Usage C-all).

Fig 4 Expert proficiency (correct identification decision rate) of each participant

Fig 5 Statistical representation of the expert's assessment for each (negative and positive) SFO case. Expert decisions (between -3 and +3) on the y-axis and SFO cases on the x-axis. F and L, in brackets after the number of the case, indicate frontal and lateral view cases, respectively.

Fig 6 Criterion subjectivity according to the standard deviation values. Sorted from lowest to highest

Fig 7 Criterion discriminative power according to the Spearman's rank correlation coefficient between a criterion's value and the status of the case. Sorted from highest to lowest

Fig 8 Scatter plot including criteria for frontal view assessment under study spatially distributed according to their subjectivity (x-axis) and discriminative power (y-axis)

Fig 9 Scatter plot including criteria for lateral view assessment under study spatially distributed according to their subjectivity (x-axis) and discriminative power (y-axis)

Fig 10 Marking line used to analyze anatomical consistency.

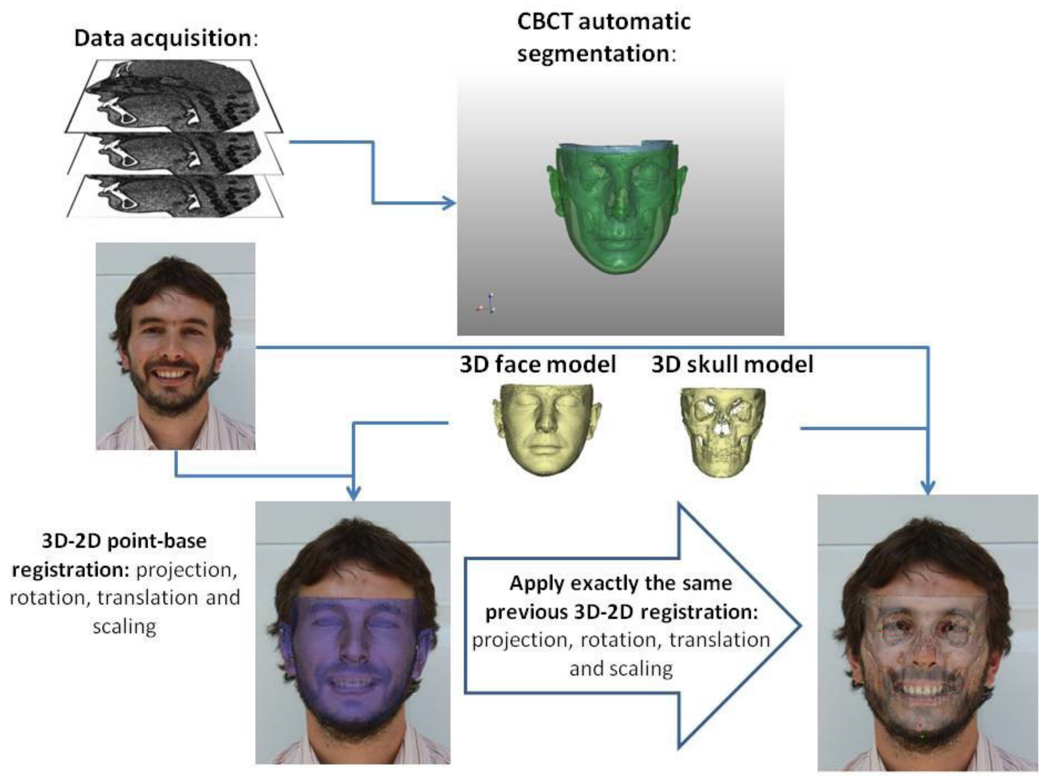

Data acquisition:
CBCT automatic segmentation:
3D face model
3D skull model
3D-2D point-base registration: projection, rotation, translation and scaling
Apply exactly the same previous 3D-2D registration: projection, rotation, translation and scaling

Fig. 1

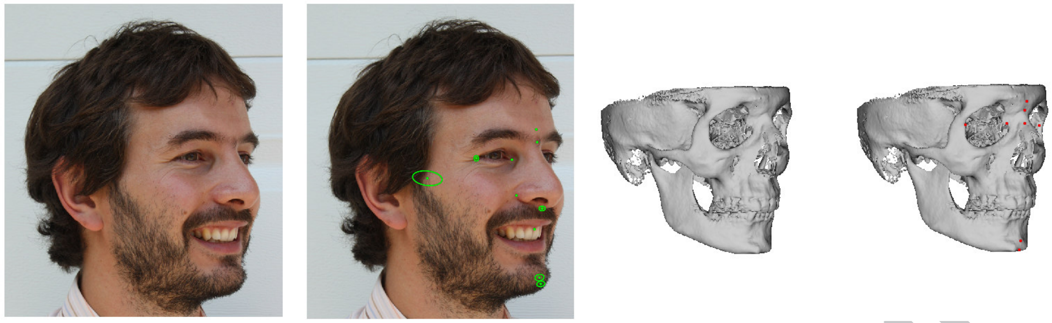
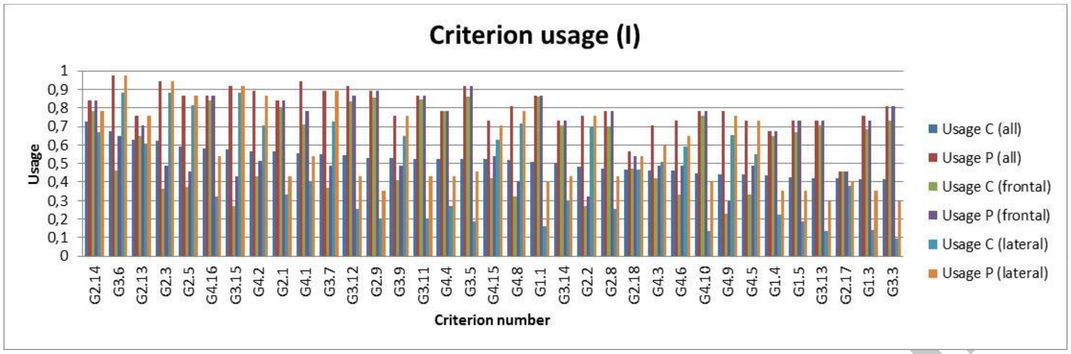
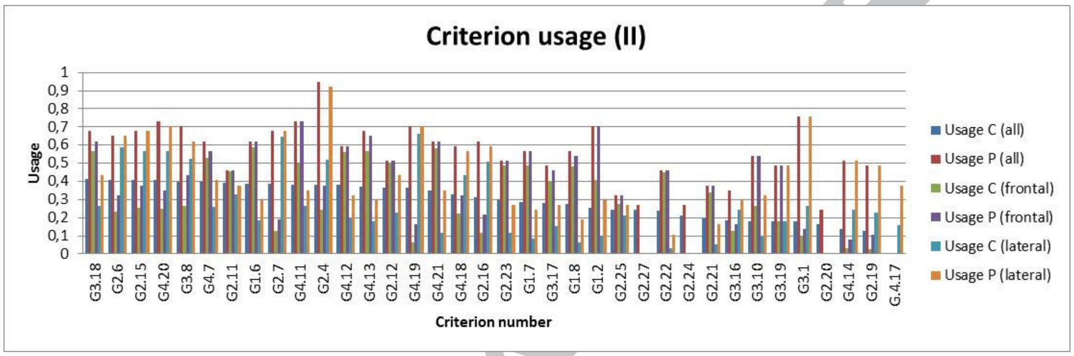


Fig. 2


Criterion usage (I)
Usage
Criterion number
G2.14 G3.6 G2.13 G2.3 G2.5 G4.16 G3.15 G4.2 G2.1 G4.1 G3.7 G3.12 G2.9 G3.9 G3.11 G4.4 G3.5 G4.15 G4.8 G1.1 G3.14 G2.2 G2.8 G2.18 G4.3 G4.6 G4.10 G4.9 G4.5 G1.4 G1.5 G3.13 G2.17 G1.3 G3.3
Usage C (all)
Usage P (all)
Usage C (frontal)
Usage P (frontal)
Usage C (lateral)
Usage P (lateral)

a


Criterion usage (II)
Usage
Criterion number
G3.18 G2.6 G2.15 G4.20 G3.8 G4.7 G2.11 G1.6 G2.7 G4.11 G2.4 G4.12 G4.13 G2.12 G4.19 G4.21 G4.18 G2.16 G2.23 G1.7 G3.17 G1.8 G1.2 G2.25 G2.27 G2.22 G2.24 G2.21 G3.16 G3.10 G3.19 G3.1 G2.20 G4.14 G2.19 G.4.17
Usage C (all)
Usage P (all)
Usage C (frontal)
Usage P (frontal)
Usage C (lateral)
Usage P (lateral)

b

Fig. 3

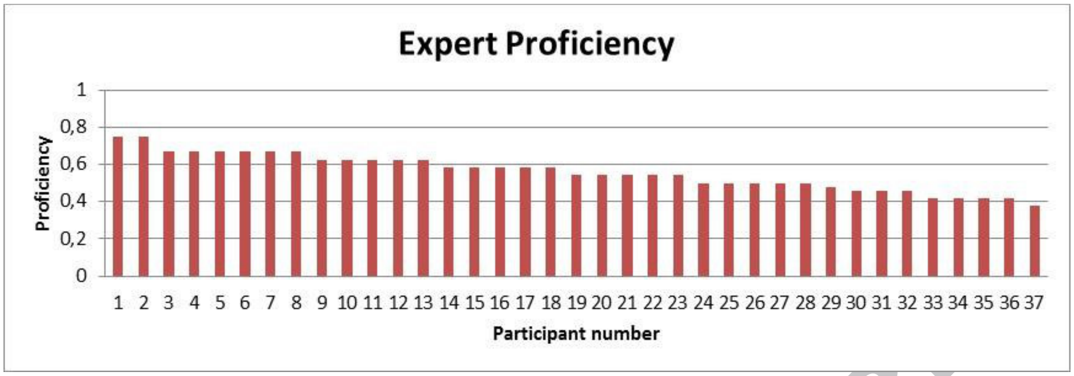
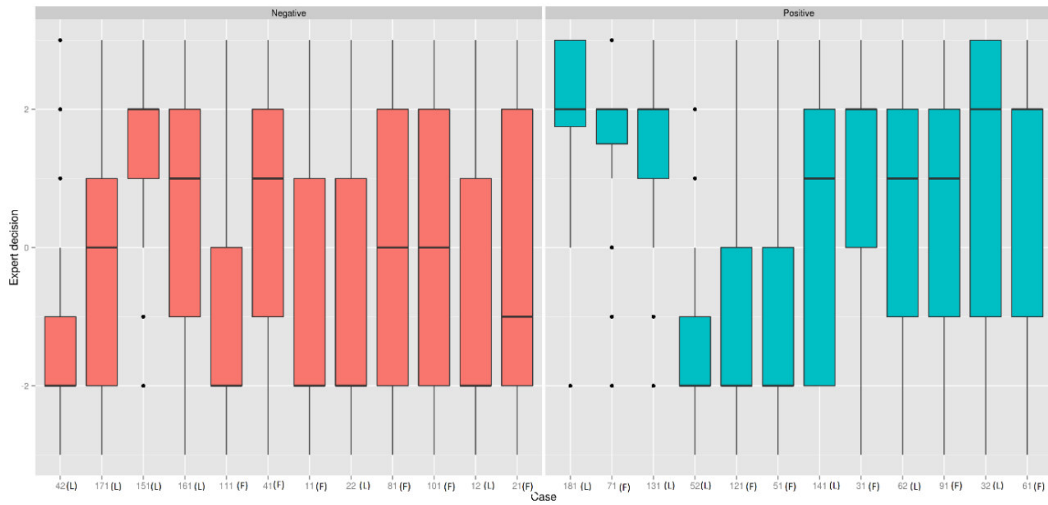

Expert Proficiency
Proficiency
1
0,8
0,6
0,4
0,2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37
Participant number

Fig. 4



Negative
Positive
Expert decision
2
0
-2
42(L)
171(L)
151(L)
161(L)
111(F)
41(F)
11(F)
22(L)
81(F)
101(F)
12(L)
21(F)
181(L)
71(F)
131(L)
52(L)
121(F)
51(F)
141(L)
31(F)
62(L)
91(F)
32(L)
61(F)
Case

Fig. 5

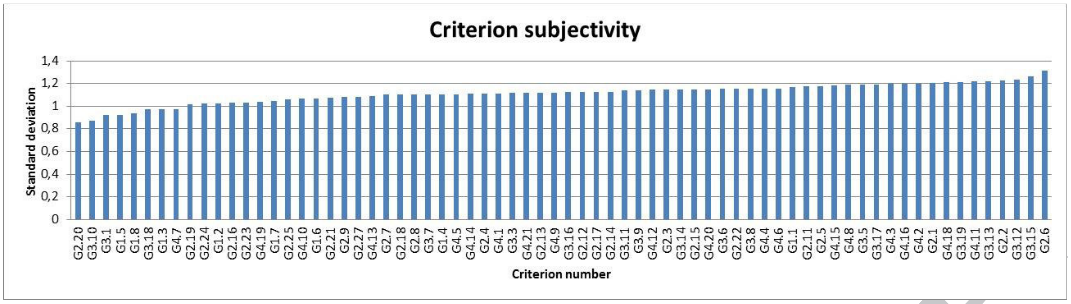

Criterion subjectivity
Standard deviation
1,4
1,2
1
0,8
0,6
0,4
0,2
0
G2.20 G3.10 G3.1 G1.5 G1.8 G3.18 G1.3 G4.7 G2.19 G2.24 G1.2 G2.16 G2.23 G4.19 G1.7 G2.25 G4.10 G1.6 G2.21 G2.9 G2.27 G4.13 G2.7 G2.18 G2.8 G3.7 G1.4 G4.5 G4.14 G2.4 G4.1 G3.3 G4.21 G2.13 G4.9 G3.16 G2.12 G2.17 G2.14 G3.11 G3.9 G4.12 G2.3 G3.14 G2.15 G4.20 G3.6 G2.22 G3.8 G4.4 G4.6 G1.1 G2.11 G2.5 G4.15 G4.8 G3.5 G3.17 G4.3 G4.16 G4.2 G2.1 G4.18 G3.19 G4.11 G3.13 G2.2 G3.12 G3.15 G2.6
Criterion number

Fig. 6

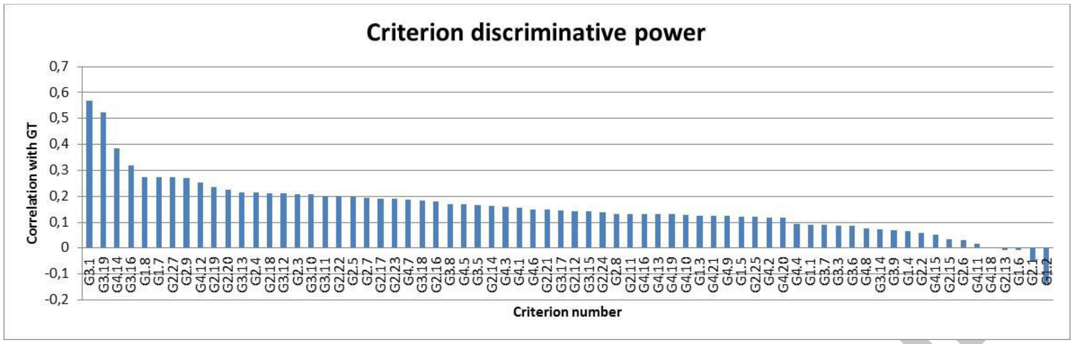

Criterion discriminative power
Correlation with GT
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
-0,2
G3.1 G3.19 G4.14 G3.16 G1.8 G1.7 G2.27 G2.9 G4.12 G2.19 G2.20 G3.13 G2.4 G2.18 G3.12 G2.3 G3.10 G3.11 G2.22 G2.5 G2.7 G2.17 G2.23 G4.7 G3.18 G2.16 G3.8 G4.5 G3.5 G2.14 G4.3 G4.1 G4.6 G2.21 G3.17 G2.12 G3.15 G2.24 G2.8 G2.11 G4.16 G4.13 G4.19 G4.10 G1.3 G4.21 G4.9 G1.5 G2.25 G4.2 G4.20 G4.4 G1.1 G3.7 G3.3 G3.6 G4.8 G3.14 G3.9 G1.4 G2.2 G4.15 G2.15 G2.6 G4.11 G4.18 G2.13 G1.6 G2.1 G1.2
Criterion number

Fig. 7

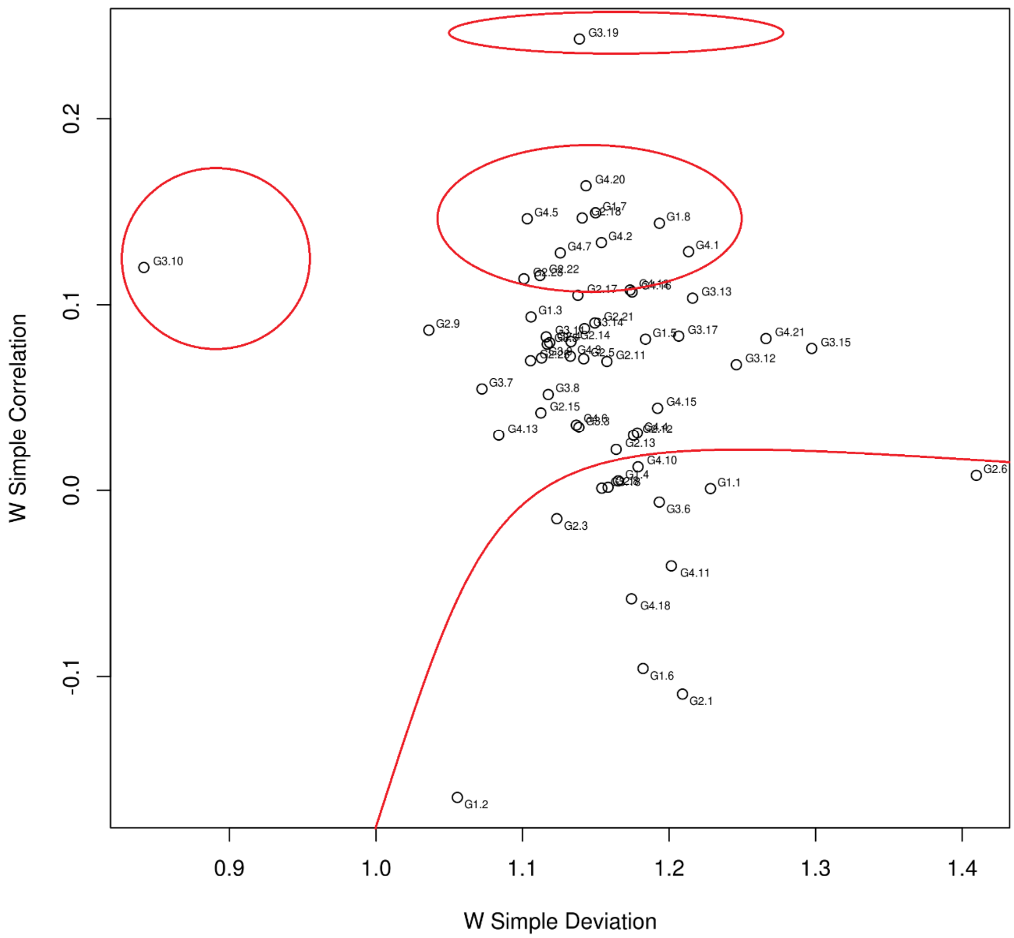

W Simple Correlation
W Simple Deviation
0.2
0.1
0.0
-0.1
0.9
1.0
1.1
1.2
1.3
1.4
G3.19
G4.20
G4.5
G1.7
G1.8
G4.2
G4.7
G4.1
G3.10
G2.17
G3.13
G1.3
G2.21
G2.9
G2.14
G1.5
G3.17
G4.21
G3.15
G2.11
G3.12
G3.7
G3.8
G2.15
G4.15
G4.13
G2.13
G4.10
G2.6
G1.4
G1.1
G3.6
G2.3
G4.11
G4.18
G1.6
G2.1
G1.2

Fig.8

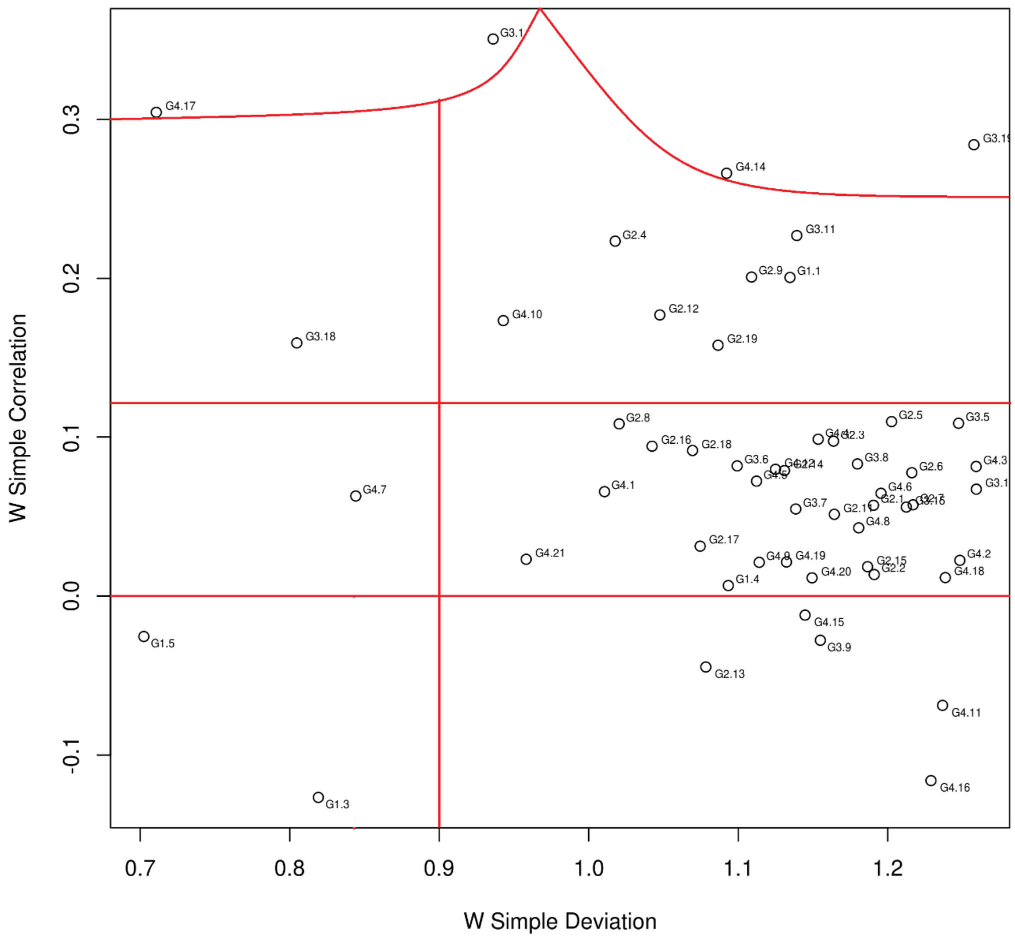

W Simple Correlation
W Simple Deviation
0.7
0.8
0.9
1.0
1.1
1.2
-0.1
0.0
0.1
0.2
0.3

Fig. 9

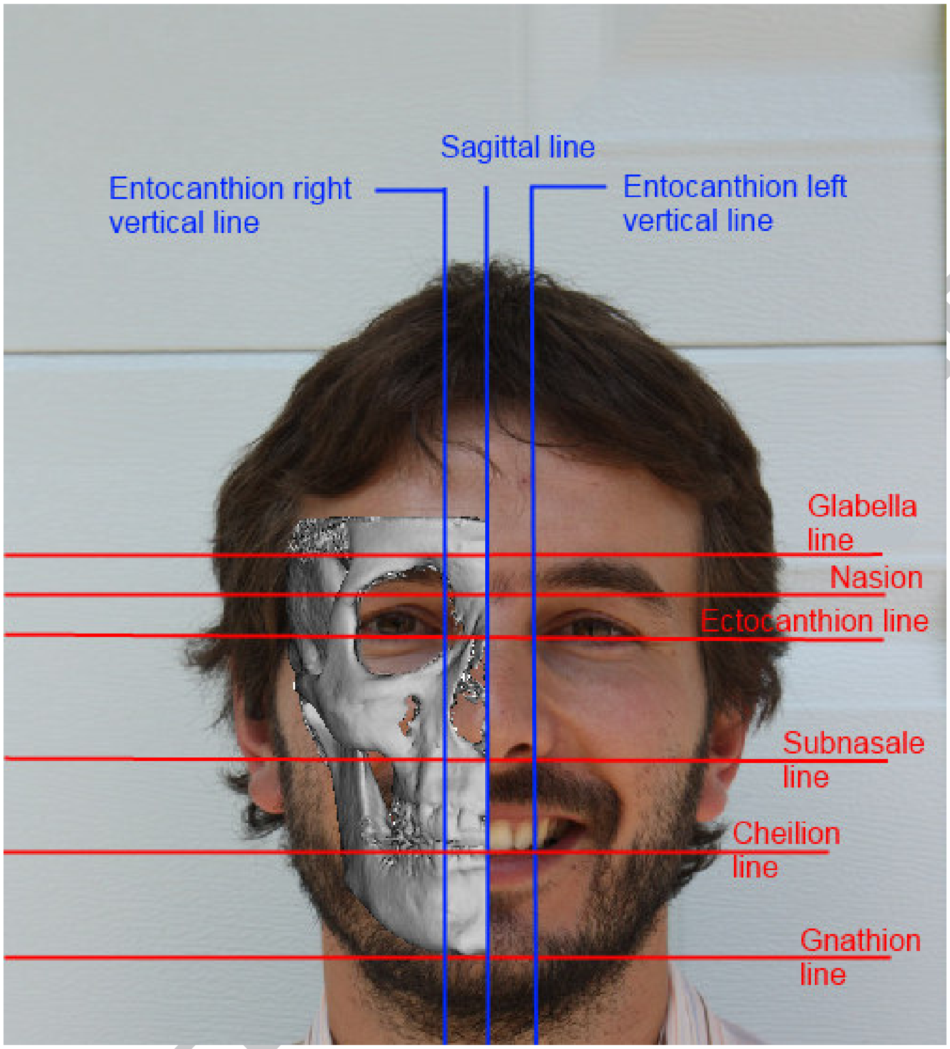

Sagittal line
Entocanthion right vertical line
Entocanthion left vertical line
Glabella line
Nasion
Ectocanthion line
Subnasale line
Cheilion line
Gnathion line

Fig. 10

Table 1. Marking lines used to analyze anatomical consistency.

No.	Criteria	View
Group 1	**Superimposition of the following marking lines (Face – Skull)*:**	
1.1	excanthion – excanthion (A) - ectochconchion-ectoconchion (A') Ectocanthion line	F
1.2	glabella-gnathion (B) - glabella-gnathion (B') Frontal central line	F
1.3	superciliary-superciliary (C) - superciliary-superciliary (C') Supraciliary line	F
1.4	horizontal line at subnasal (D) – horizontal line at nasospinal (D') Subnasal line	F
1.5	cheilion-cheilion (E) – occusal line/horizontal line at stomion (E') Cheilion line	F
1.6	horizontal line at gnathion (H) – horizontal line at gnathion (H') Gnathion line	F
1.7	endocanthion-cheilion (F) – entocanthion –caninion (F') [right] Entocanthion vertical line	F
1.8	endocanthion-cheilion (F) – entocanthion –caninion (F') [left] Entocanthion vertical line	F

Table 2. Landmarks used to evaluate soft tissue thickness.

No.	Criteria	View
Group 2	**Overall consistency of the facial outline and facial soft tissue thickness at the following pair of homologous points (Skull – Face)*:**	
2.1	gnathion – menton	F
2.2	gnathion - menton	L-O
2.3	nasion - nasion	L-O
2.4	glabella – glabella	L-O
2.5	subespinale - subnasale	L-O
2.6	pogonion - pogonion	L-O
2.7	rhinion - rhinion	L-O
2.8	gonion – gonion	F
2.9	zygion – zygion	F
2.10	trichion - trichion	L-O
2.11	The minimal tissue thickness all along the contour is considered from the point of view of its symmetry by side for evaluating the match as acceptable	
2.20	Ant lacrimal crest - medial canthus	F - O
2.13	Prosthion - Supra-labiale	F – L - O
2.14	Alare – Alare	F – L - O
2.15	Gonion-gonion	L
2.16	Zygomaxilare - malare	L
2.17	Whitnall's tubercle - lateral canthus	F – L - O
2.18	occlusion mid-incisors - stomion	F – L - O
2.19	porion – tragion	L - O
2.20	crista conchalis - supra-alare	F – L - O
2.21	intercanine distance (75%) - chelion	F

2.22	eyeball position – pupilare	F
2.23	supraorbitale - sag eyebrow	F
2.24	two tangents nasal - pronasale	F – L - O
2.25	1st premolar / canine radiating line – chelion	F – L - O
2.26	mastoidale - subaurale	L - O
2.27	infraorbital foramen – chelion	F – L - O

Table 3. Consistency of the bony and facial outlines/morphological curves.

No.	Criteria	View
Group 3	**Overall consistency of the bony and facial outlines/morphological curves:**	
3.1	The outline of the frontal bone follows the forehead outline.	L-O
3.2	The skull and head height is similar (account for variation in soft tissue and distortion in the perception created by presence of hair).	L-O
3.3	The width of the cranium fills forehead area of the face.	F
3.4	The length of the skull from menton to bregma fits within the face.	F
3.5	The lateral line of the zygomatic bone matches the outline of cheek.	F
3.6	The chin outline is consistent with the mental outline.	L-O
3.7	The gonial outline follows the outline of jaw angle.	L-O
3.8	The outline of frontal process of the zygomatic bone can be aligned with the process seen in the face. The outline of the zygomatic arch can be fitted between the skull and the face. (This criteria are more easily appreciated in individuals with minimal soft tissue thickness).	L-O
3.9	The arcus supraciliariaris follows supraorbital margin.	L-O
3.10	The temporal line is consistent with the outline of the forehead (Sometimes the temporal line cannot be distinguished).	F
3.11	The outline of the face and the outline of the skull all along the contour follow each other maintaining symmetrical flow by side.	F
3.12	The outline of the nose in the face represented by shade distribution follows the outline of the nasal bone in the skull maintaining symmetrical flow by side.	F
3.13	The asymmetries in the facial organs especially the nose reveal consistency with the asymmetries in the organs of the skull including the nasal structures. These include: 1. Asymmetries in the nasal area including the nasal bone, piriform aperture and nasal spine. 2. Asymmetries in the zygomatic area, especially the extent of protrusion of the arch. 3. Asymmetries in the occlusal line caused by protruding or overriding anterior dentition reflected as asymmetries in the corresponding part of the lip closure line. 4. Asymmetries in the gonia.	F
3.14	The arcus supraciliariaris follows supraorbital margin.	F
3.15	The outline of the of the nasal bones follows the outline of the nose in the skull with minimal tissue thickness allowance	L
3.16	The sagittal outline of the nasal cartilage is the mirror image of the contour of the pyriform aperture, relative to Line №1 passing through the rhinion point (1) and parallel to Line №2 joining the nasion (2) and the prostion (3) anthropometric points. Line №1 splits the entire nasal cartilage into two symmetric mirrored halves: the protruding part of the nose cartilage is the mirror image of the cartilage filling the pyriform aperture of the cranium.	
3.17	Lateral nasal bulges	F
3.18	Oblique line of the mandible	F
3.19	Dental information (bony to bony consistency)	F – L - O

Table 4. Positional relationship analyzed to assess anatomical consistency.

No.	Criteria	View
Group 4	**Overall consistency positional relationships between the skull and face:**	
4.1	The prosthion lies posterior to the anterior edge of the upper lip. The occlusal and the lip closure line are consistent.	F

4.2	The lateral angle of the eye lies within the lateral wall of the orbit.	L-O
4.3	The lateral orbital margin at the Whinall's turbercle matches or approximates the position of the ectocathion	L-O
4.4	The piriform aperture width and height lies within the borders of the nose.	F
4.5	The stomion lies at the central incisors (at the occusal line)	L-O
4.6	The lateral margin of the piriform aperture matches or approximates the alare	L-O
4.7	The stomion lies at the central incisors (Incisal margin of the upper incisors).	F
4.8	The nasion is higher than the nasal root.	L-O
4.9	The prosthion lies posterior to the anterior edge of the upper lip. The occlusal and the lip closure line are consistent.	L-O
4.10	The medial margin of orbit aligns and superimpose with the endocanthion	F
4.11	The nasion is higher than the nasal root.	F
4.12	The Whitnall's tubercle aligns with the ectochantus on the horizontal plane and vertically the ectochantus lies medial to the tubercle. The orbital width is consistent with the eye-slit width.	F
4.13	The chelion lies between the canine and the first premolar (at the occusal line)	F
4.14	The porion aligns just posterior to the tragus, slightly inferior to the crus of the helix	L-O
4.15	The lower margin of piriform aperture matches the subnasale	L-O
4.16	The eyebrow generally follows the upper edge of the orbit over the medial two-thirds. At lateral superior one-third of the orbit the eyebrow continues horizontally as the orbital rim begins to curve inferiorly.	F
4.17	The external auditory meatus opening lies medial to the tragus of the ear. (Place a projecting marker at the ear canal to assess this criterion more easily).	F
4.18	The chelion lies between the canine and the first premolar (at the occusal line)	L-O
4.19	The anterior nasal spine lies posterior to the base of the nose near the most posterior portion of the lateral septal cartilage.	L-O
4.20	Gonial flare in the skull and the postero-lateral jaw angle outline in the face	L-O
4.21	Gonial flare in the skull and the postero-lateral jaw prominence in the face	F

Table 5. Set of assessment criteria recommended by MEPROCS partners in the case of frontal view photographs

Group properties	Criteria for frontal view
Highest discriminative power	3.19 Dental information (hard tissue to hard tissue consistency)
Good trade-off between subjectivity and discriminative power	4.20 Gonial flare in the mandible and the postero-lateral jaw angle outline in the face 4.2 The lateral angle of the eye lies within the lateral wall of the orbit. 4.7 The stomion lies at the central incisors (Incisal margin of the upper incisors) 4.1 The occlusal and the lip closure line are consistent 2.18 Evaluate soft tissue thickness at Occlusion mid-incisors – stomion 2.22 Evaluate consistency positional relationship between the expected

	position of the Eye ball in the skull and pupilare in the photographs 2.23 The soft tissue position just beneath the eyebrow should be more anterior than the orbital rim 1.7 and 1.8 Marking line used to analyze anatomical consistency: Entocanthion vertical line. Endocanthion-cheilion – entocanthion – caninion [left and right]. It is from entocanthion line to cheilion line, parallel with the front central line, used to mark the relationship of entocanthion and maxillary teeth (29). See Figure 10.
Easily to evaluate and important discriminative power	3.10 The temporal line is consistent with the outline of the forehead (Sometimes the temporal line cannot be distinguished)

Table 6.. Set of assessment criteria recommended by MEPROCS partners in the case of lateral view photographs

Group properties	Criteria for lateral/oblique view
Best criteria	3.1 The outline of the frontal bone follows the forehead outline
Highest discriminative power but also high variance	4.14 The porion aligns just posterior to the tragus, slightly inferior to the crus of the helix 3.19 Dental information (hard tissue to hard tissue consistency)
Easy to evaluate and important discriminative power	3.18 Consistency of the bony and facial outlines/morphological curves at the lower part of the face: Oblique contour of the mandible follows the outline of the jaw
Important discriminative power and a significant variability	2.4 Evaluate soft tissue thickness at glabella – glabella 3.11 The outline of the face and the outline of the skull all along the contour follow each other

Highlights:

Provide quantitative and objective data for discussion of a set of recommended craniofacial assessment criteria

Novel methodology for understanding and standardizing identification methods based on the observation of morphological patterns

MEPROCS partners agreed on a reduced set of assessment criteria