Title Page (with authors and addresses)

Title

The unfamiliar face effect on forensic craniofacial reconstruction and recognition.

Authors

Won-Joon Lee, [a] ; Caroline M. Wilkinson, [b] .

Affiliations

[a] National Forensic Service, 10 Ipchun-ro, Wonju-si, Gangwon-do, 26460, Republic of Korea.

Tel: +82 33 731 7685

E-mail: ttolong2@gmail.com

[b] Face Lab, Liverpool John Moores University, Liverpool Science Park IC1, 131 Mount Pleasant, Liverpool, L3 5TF, United Kingdom.

Tel: +44 151 904 1188

E-mail: C.M.Wilkinson@ljmu.ac.uk

Corresponding author

Won-Joon Lee

National Forensic Service, 10 Ipchun-ro, Wonju-si, Gangwon-do, 26460, Republic of Korea.

Tel: +82 33 731 7685

E-mail: ttolong2@gmail.com

Abstract

Previous research into the reliability of forensic craniofacial reconstruction (CFR) has focused primarily on the accuracy of reconstructed faces from European or African ancestry skulls. Moreover, the recognition of CFR in relation to the experience and ancestry of the practitioners and the assessors has not been previously considered. The cross-race effect is a recognised phenomenon in psychology studies, where familiar ancestry faces are recognised more readily than unfamiliar ancestry faces, but there is a paucity of research addressing the relationship between the accuracy of reconstructed faces and the familiarity with this ancestry by the practitioners/assessors. The aims of this research were to investigate whether 'unfamiliar-race effect' has any influence on the accuracy of CFR and to evaluate how much the correct recognition rate of CFR is affected by the cross-race effect. Eight CFRs from three ancestry groups were produced by experienced practitioners in order to explore the aims. The results demonstrated that practitioners produced more recognisable CFRs using skulls from a familiar ancestry than skulls from unfamiliar ancestries.

Keywords

forensic; craniofacial reconstruction; approximation; face; recognition; accuracy; familiar; unfamiliar; ancestry; cross-race effect; own-race bias

Introduction

Forensic craniofacial reconstruction (hereafter referred to as CFR) is a technique based on both scientific standards and artistic skill to rebuild a face onto a skull to recreate the ante-mortem appearance of the individual [1]. It is also known as forensic facial approximation or depiction [2,3]. The ultimate aim of CFR is to recreate an *in vivo* countenance (ante-mortem appearance) of an individual that sufficiently resembles the decedent in order to allow recognition and then identification of the individual [4]. CFR might be employed in forensic investigations where other means are not successful or available to identify human remains. It has also been used in archaeology to recreate the faces of paleontological and archaeological humans [5-7].

The accuracy of CFR has been of primary importance since the practice first emerged scientifically. Research has been directed toward producing not only more reliable reconstructed faces, but in establishing better means of estimating their accuracy. There are a number of influential factors on the accuracy of CFR: the CFR standards utilised, the practitioners, the assessors (recognisers) and mechanism of facial perception [8]. There is on-going research relating to refining and developing standards and guidelines for practice to increase the accuracy of these techniques and to provide a stronger base for public faith in the reproducibility of facial reconstruction to present reliable facial features of the targets [9]. CFR practitioners can also contribute the accuracy in relation to familiarity with the biological profile of the subjects and the degree of expertise, experience or training. In addition the assessors may also contribute to the efficacy in relation to familiarity with the target and experience and skills in facial recognition. The experience and skills of the assessor may be related to their own biological profile, such as sex, age and ancestry [8].

The prediction guidelines for rebuilding facial components include anthropometrical or morphological analysis of the skull and many are currently employed to predict the facial components: eyes [10-14], nose [15-19], mouth [20-24] and ears [15,20,25,26]. In addition, numerous sets of facial soft tissue depth measurements have been recorded from living subjects using a variety of clinical imaging techniques including; lateral cephalometric radiographs [27,28], computed tomography (CT) [29-32], magnetic resonance imaging (MRI) [33-35] and ultrasound [36-40]. Through these techniques, a large amount of facial tissue depth data has been collected from various geographical and ethnic group populations relating to sex, age and body mass index (BMI) for use in CFR. With advances in three-dimensional (3D) medical diagnosis technology, new equipment has been modified for the purpose of collecting more accurate tissue depth data. Of those, recently-developed cone-beam CT (CBCT) scanner has been introduced to the study of tissue depth measurements [41-43].

Since CFR was introduced as an identification tool in forensic science, there has been a great deal of research into the reliability and accuracy of the techniques. In an early publication, Wilder [44] stated that given the appropriate procedure and tables of facial tissue measurements it would be difficult not to produce a successful facial reconstruction. However, there have been a few studies that disagree with Wilder's theory. The first recorded accuracy study by Von Eggeling [45] using death masks reported no resemblance between reconstructed and target faces. Further research by Stadtmüller [46], using two faces reconstructed from the skulls of an elderly and a young man and facial soft tissue depth data from cadavers [47] and postmortem images for comparison, reported no similarity. Suk [48] insisted that the most common method of measuring facial soft tissue measurements was inaccurate, and concluded that a facial reconstruction from the skull must resort to fantasy. Most recently Stephan and Henneberg [49] investigated the accuracy and reliability of CFR using sixteen reconstructed faces and thirty seven assessors. They argued that 'facial approximation' should be considered highly inaccurate and unreliable as 403 incorrect identifications were made out of 592 identification scenarios.

However, there have been some contrasting studies that add weight to Wilder's theory. Gerasimov [50] asserted that all his CFRs produced from the skulls of cadavers were recognised with strong similarity when compared to the photographs from the deceased. He also reported that all of the 140 CFRs attempted in his laboratory were successfully identified. Snow and colleagues [51] carried out an accuracy study using ante-mortem photographs. They produced two male and female CFRs from the skulls of body donors, and participants were asked to match the reconstructed face to the target from a face pool. The results showed 26% correct identification for the female and 67% for the male. Gatliff and Snow [52] claimed 70% and 65% success rates for identification in two studies using reconstructed faces produced from their CFR technique. Vanezis and colleagues [53] attempted a comparison between the manual and computer generated facial reconstruction methods, and showed that both methods could be employed as a useful tool for identification. Wilkinson and Whittaker [54] reported more optimistic results on the accuracy of facial reconstruction. They produced 3D manual facial reconstructions from five juvenile female skulls, and then compared to a face pool by 50 volunteers. The results showed that the mean hit rate was 44% and all hit rates were above chance (10%). Subsequently, Wilkinson and colleagues [1] tested the accuracy of CFRs from two CT scanned skulls of a male and female generated by a computer modelling system. The results in face pool comparison tests demonstrated that the combined hit rate was 50% above the level recorded by chance (20%). More recently, Fernandes and colleagues [55] reported a unique accuracy study using three Brazilian CFRs from a skull employing three different tissue depth datasets. In the study, the volunteers who were familiar to the target face tested the accuracy of the CFRs in three different recognition tests. The results demonstrated a 27% hit rate for the CFR using the average tissue depth data from Brazilian live subjects, 23% for the CFR using the date from Brazilian cadaver study and 20% from the use of Rhine and Moore [56] date which was derived from American White population, and they concluded that the CFR method might be a useful to investigate forensic cases.

Although skepticism remains, a considerable number of researchers have proved that CFR can provide a reasonable resemblance to a face for which an identity is sought. Also a number of forensic cases have shown that the technique of CFR can be used to assist in the identification of individuals from unknown skulls, particularly when other forensic tools are unavailable [4,51,52,57-62].

Among the possible contributing factors to the accuracy of CFR, the ancestry of practitioners or assessors appears to be overlooked in the field of CFR study. The influence of ancestry on recognising human faces has been rigorously established by relevant cognitive psychology research. Since the early 20th century, cognitive psychologists have recognised a phenomenon where people show a tendency to recognise and memorise more accurately the faces of their own ancestry group than the faces from a different ancestry. This fascinating subject has been empirically established as a robust theory named ‘cross-race effect (CRE)’ (also referred to as the ‘own-race effect/bias’ or ‘other-race effect’) by the research related to eyewitness identification and facial recognition over the last four decades (for reviews, see [63-65]). In early years, the CRE had been empirically studied employing the faces from Black and White subjects. More recent studies have included different ancestry or national groups, such as British, German, Chinese, Japanese, Korean and South Asian, and demonstrated the CRE in face identification or recognition performance associated with investigating the mechanism of CRE [66-70]. For an instance, Tanaka and colleagues [66] investigated the differences in recognising same-race and other-race faces between Caucasian (White European) and Asian participants, and confirmed that Caucasian participants recognised same-race faces more holistically than Asian faces, while Asian participants demonstrated a holistic pattern when recognising both same (familiar)-race and other (unfamiliar)-race faces. From the results, they suggested that the same-race effect may arise from the holistic recognition of faces from highly familiar ancestry faces.

From early studies to recent research, the CRE has been found universally across various ancestry and ethnic groups with strong consistency. It is assumed that many theoretical mechanisms are involved in the CRE. This has been established by the studies showing the primary factors influencing the phenomenon: contact and attitudes with other-racial groups, encoding and representational processes, perceptual-memory expertise, and perceptual categorisation (for reviews, see [69,70]). According to the research into the theoretical underpinnings of the CRE, the differences in experience with the same (familiar)- and other (unfamiliar)-race faces (also known as the ‘contact hypothesis’) are considered the main contribution to the CRE [65,71-74].

In face processing, a recent study by Blais and colleagues [75] discovered a unique phenomenon between Western Caucasian and East Asian. They employed Western Caucasian young adults and East Asian participants who just arrived in the UK from East Asia to investigate possible differences in their eye tracking movements when viewing the target faces. In contrast to adults from Western cultures, the results revealed that individuals from Eastern cultures fixate centrally on the nose region and generally avoid eyes when learning, recognising, and categorising faces. These results demonstrated that face processing can no longer be considered as arising from a universal series of perceptual events that was presumed as a triangular pattern, but employed to extract visual information from the faces differs across cultural bases. As a new paradigm that people from different cultural background achieve human face processing by focusing on different facial information has emerged, it is essential to question what aspect of culture is contributing to the process. The most plausible explanation has been provided by Nisbett and colleagues [76,77] that Westerners tend to engage analytic perceptual strategies for processing face configuration, whereas East Asians prefer holistic perceptual strategies (*analytical* versus *holistic* cultural framework) [75,78].

In relation to CFR, the CRE could be implicated in two ways; Firstly, the experience of CFR practitioners may influence the resemblance/reliability of the reconstructed face due to unfamiliarity with the depiction of races from outside of their experience. It is possible that unconsciously they could sculpt the face from a limited memory database especially when partially trained; Secondly, the facial recognition ability of the assessors utilised for accuracy assessment may be affected by unfamiliarity with the ancestry of the reconstructed face. Even though psychological research into the CRE has primarily been in memory recall rather than comparison without memory load, the CRE has also been observed in performances with minimal required memory [79] and in the early stage of perceptual encoding face process [68,80]. These studies suggest that there is a possibility that the CRE may affect the accuracy of recognising or identifying CFRs by assessors from another ancestry to the reconstructed face.

There have been numerous studies into the reliability and accuracy of CFR techniques and the identification or recognition of reconstructed faces. The impact, however, of the cross-race effect on CFR and recognition has not been clearly established. Studies on comparisons between CFRs produced by practitioners who have different biological profiles to the faces they reconstruct are rare. Moreover, research on the accuracy of CFR has been weighted more towards Caucasoid (White European)- or Negroid (Black African)- type skulls rather than Mongoloid (East Asian)- or other ancestry types [50,54,81-83]. Case reports [57,60] aside, very little literature about CFR and accuracy on Asian ancestry groups exists [84]. CFR from skulls of various ancestry groups and its recognition are therefore interesting to investigate not only in terms of the reliability of CFR itself, but also in regard to facial recognition by different ancestry groups.

This research examines whether the accuracy of CFRs are affected by the degree of practitioner experience in relation to the ancestry of the faces reconstructed. The possibility that the ancestry origin of the assessor may influence a correct recognition of reconstructed faces is also examined. Thus the objectives of this research are to; analyse whether the 'unfamiliar-race effect' influences practitioners of CFR; and investigate whether the CRE has an impact on the accuracy rate for the recognition of CFR performed by assessors from various ancestry groups.

Materials and methods

General design

To produce CFRs for this research, two independent teams composed of experienced practitioners were organised in the UK (the Centre for Anatomy and Human Identification at the University of Dundee) and South Africa (SA) (the Department of Anatomy at the University of Pretoria). Each team provided two skull models with the related CFRs from previous forensic investigations (Caucasoid-type for the UK team and Negroid-type for the SA team) (Tab. 1). In addition, two Mongoloid-type skulls were provided from China (the Tieling 213 Research Institute). Each team then produced two CFRs using skull models from unfamiliar-race (ancestry) groups and the CFRs produced in the forensic cases were also utilized in this study as familiar-race cases. The accuracy of the CFRs was then tested by employing a face pool comparison method and voluntary assessors. Ethical permissions were not required to utilise the skull models due to the validity of previous permissions from each institute for the research or forensic investigation into the skulls used.

Skull models & production of CFR

Six skull models were utilized; two Caucasoid (C1 and C2), two Negroid (N1 and N2)-, and two Mongoloid (M1 and M2) type skull models (Fig. 1). Previously, the UK team produced two familiar-race CFRs from the skull models C1 and C2. The SA team also produced familiar-race CFRs from the skull models N1 and N2. Each team then received two skull models of unfamiliar ancestry – the UK team received one Negroid-type (N2) and one Mongoloid-type (M2) skull model and the SA team received one Caucasoid-type (C1) and one Mongoloid-type (M1) skull model. The models were provided as either plaster copies or 3D laser scanned digital model depending on which method would be employed by the team. In order to standardise the application of facial soft tissue depth data to the forensic casework, the UK and South African teams provided facial soft tissue depth measurements to the other team and the same tissue depth dataset for the Mongoloid-type skulls was provided to both teams by the researcher [20,85]. All information about the six skulls is described in Table 1. The teams applied the same method to the CFRs; the combination method [4]. All ante-mortem photographs were provided by the teams and collected by the researcher for the accuracy tests. The ante-mortem facial photographs were not exposed to the practitioners until the CFRs had been completed to ensure blind testing.

Practitioners

This study was based on the assumption that CFRs would be generated by experienced practitioners who have fulfilled qualification in producing reliable CFRs. Two practitioners from the UK team took part in this study; one practitioner had five years' experience in the field of CFR and the other had twenty years' experience. The two British practitioners had experience in forensic casework for majority White European and Indian Sub-continent faces. The South African practitioner had two years' experience in CFR and had majority experience of reconstructing Black African faces. Each practitioner decided which method they would employ to reproduce the facial reconstructions depending on the applicability and circumstances of their institute (Tab. 1).

Completion of CFRs

The completed CFRs were collected from the teams as 3D scanned image files or photographic images (Fig. 2). In total eight CFRs were completed and collected. N1-SA and N2-SA were produced by SA team; C1-UK and C2-UK were produced by UK team as familiar-race CFRs. N2-UK and M2-

UK were produced by UK team; M1-SA and C1-SA were produced by SA team as unfamiliar-race CFRs. These CFRs were compared to the corresponding target faces in order to estimate their accuracy using face pool comparison method.

Accuracy assessment

Preparation of photo arrays

Facial images of the CFRs were utilised in photo arrays. Images of the CFRs generated by the 3D manual method were taken using a digital camera (EOS 400D, Canon®, Japan) with 50 mm zoom (AF 17-50 mm Lens, Tamron®, Japan) three metres away from the CFR under adequate illumination. The images of the CFRs generated by the computerised 3D method were captured using the 'print screen' tool in FreeForm Modelling Plus™. When processing the images using the camera or capture tool on computer screen, the facial angles of the CFRs were aligned in a frontal view for the images used in the photo arrays. Facial images for possible foils (distractors) used in the photo arrays were obtained from various databases founded in the website of the Facial Recognition Homepage [86] which is maintained by the Department of Wireless Communications at the Faculty of Electrical Engineering and Computing in the University of Zagreb, Croatia. All databases linked in the website were open to the other researchers for the purpose of facial recognition studies only. Therefore, additional ethical permissions were not required to utilise the photographic database for this research.

There are a number of empirical standards currently being employed in the studies of facial recognition and identification in the field of face cognitive psychology. Unlike the studies in psychological experiments, however, it is almost impossible to standardise the format of facial photographs in terms of a unified facial expression, angle of head posture, hair style, illumination, resolution and background, since applicable ante-mortem facial photographs in real forensic scenarios were taken mostly without any considerations of forensic investigation or empirical study. However, it was determined that the selection of foils for this research followed the recommended guideline [87] as correctly as possible. Consequently nine foils were established to be utilised for each photographic line-up of eight face pool comparison trials. The image quality of the foil photographs was altered to make them a similar looking quality to the target utilising Adobe® Photoshop® (CS5, Verson 12.0.1, USA). The photo array consisted of a photographic line-up of ten subjects (including the target) directly below the CFR image. Thus the targets were present in all arrays. The target facial image and nine foils were randomly allocated. The prepared eight trials were arranged randomly to be posted in an on-line survey tool.

Assessors

Since this study involves cases of a forensic nature, the assessors understood the confidentiality requirement for such a survey. There were twenty two White European assessors from the UK (UK assessors) comprised of staff from the Centre for Anatomy and Human Identification at the University of Dundee, or police officers at the Department of Tayside Police in Dundee; nineteen assessors from South Africa (SA assessors) who were police officers at the Department of Facial Identification in the South African Police Service from a mixed race background; and twenty one Northeast Asian assessors from the Republic of Korea (KOR assessors), who were doctors of forensic medicine/dentistry or researchers working at the National Forensic Service or the Departments of Forensic Medicine in various medical colleges in Republic of Korea. No attempt was made to record any personal data from the assessors, except for their ancestry origin and duration of residence in the country.

Accuracy survey

Due to geographical and occupational difficulties of the participants being based in three different countries and secured working places, the survey for assessing the accuracy of CFRs was carried out via an on-line survey tool (Surveymonkey.com™). Using design and edit options in the website, the prepared eight trial sheets were uploaded one by one into the web pages in order to complete a series of questionnaires. For the participants whose first language is Korean, a Korean language version was prepared.

At the beginning of the survey via the web site, assessors were asked to read and follow the instructions and provided consent enabling them to take part in the survey. Subsequently, they were instructed to compare the CFR to the ten facial photographs in the array, and decide which one of the faces most closely resembled the CFR, and to mark their answer with a tick in the bracket assigned for that particular numbered photograph. The assessors were informed that the target face of the corresponding CFR may or may not be present in the ten photographs and that if none of the faces in the photographs appeared to resemble the reconstruction, then they were to tick the bracket assigned to "no similar faces". Assessors were advised to make their own decision without being influenced by, or asking the advice of, another person. No time limit was imposed for any trial.

The results were analysed in order to establish possible cross-race effects between the CFRs in relation to the experienced practitioner groups. The hit rates for the eight CFRs by the experienced practitioners were statistically compared in terms of the three ancestry groups of the CFRs; the forensic experience of the practitioners in the UK and South Africa in relation to ancestry; the three ancestry groups of the assessors from the UK, South Africa and Korea; familiar/unfamiliar-race of the CFRs in relation to the practitioners and the assessors. Since this study aims to investigate the effect of familiar ancestry rather than the accuracy of CFR, only correct hit rates (when assessors correctly chose the target) were analysed. A series of paired t-test, One-way ANOVA and Kruskal-Wallis analyses were conducted to assess the correlations between variables using Microsoft® Excel® (2010, USA) and IBM® SPSS® Statistics (version 20, USA).

Results

Most of targets did not receive the majority hit rate, with the exception of C1-UK and C2-UK as recognised by the South African assessors (Tab. 2). The assessors had 50% chance of deciding whether the target would exist in the face pool or not, and 10% chance of choosing correct face when they decided that the target was present in the face pool. Therefore, it was expected that each target had 5% chance rate (50% × 10%) as all the face pool comparisons were target-present trials. With the exception of N2-UK and M2-UK (4.8% and 0.0% respectively), all mean hit rates were above chance rate suggesting that these CFRs were recognised at a rate greater than guessing. Mean hit rates were 12.5% for UK assessors, 13.1% for Korean assessors and 12.5% for South African assessors. One way

ANOVA showed that there was no significant difference between the mean hit rates from the three assessor groups ($p > 0.05$) (Tab. 2). This proposes that on average the assessor groups had equal levels of face recognition ability.

The mean hit rates for each CFR varied from 0.0% (for M2-UK) to 24.2% (for C1-UK and C2-UK). M2-UK from Mongoloid (East Asian)-type skulls received very low hit rates (0.0%). One way ANOVA demonstrated that there was a significant difference in the mean hit rates between the eight CFRs ($p = 0.015$) implying that there was a considerable variation in the recognisability of individual CFRs. To investigate overall differences in each face pool trial between the three assessor groups, distributions of response rates across eleven options were examined using Kruskal-Wallis test. Also the assessor groups were split into two categories depending on if they were the familiar or unfamiliar ancestry as the CFR in order to examine the difference in the distributions of response rates across eleven options between the two groups. There were no significant differences in the response rate distributions across eleven options for each trial either between the three assessor groups (Asymptotic significance > 0.05) or between the familiar and unfamiliar ancestry assessor groups as the skull (Asymptotic significance > 0.05). These results suggest that the three assessor groups demonstrated statistically similar distributions when choosing a possible face from the face pool in all trials, and this pattern was maintained when the assessor groups were split into two groups according to familiar or unfamiliar ancestry as the CFRs.

The cases were divided into three groups in relation to the ancestry groups of CFRs in order to investigate whether the ancestry of CFR and assessor influenced hit rates. One-way ANOVA presented that the assessors from the UK and South Africa demonstrated significant differences in the hit rates between the ancestries of the CFRs ($p = 0.036$ and 0.005 respectively) (Fig. 3). The UK assessors revealed higher hit rates when recognising Caucasian (White European/Indian Sub-continent) and Negroid (Black African) CFRs (16.7% and 15.2% respectively) than for Mongoloid (East Asian)-type CFRs (2.3%). The South African assessors exhibited higher hit rates when recognising Caucasian-type CFRs (28.1%) than the other two ancestries (5.3% for Negroid and 0.0% for Mongoloid CFRs). There appeared to be a tendency for Korean assessors to recognise Caucasian

CFRs (15.9%) better than the other ancestry types (12.7% for Negroid and 9.5% for Mongoloid), but this difference was not statistically significant (p = 0.587). Mean hit rates from all the assessors demonstrated a significant difference between the recognitions of the CFR ancestries (p = 0.039). Overall, the assessors revealed a higher hit rate when recognising Caucasian CFRs (20.2%) than Negroid (11.0%) or Mongoloid (3.9%) types. These results can be interpreted that the CFRs from Caucasoid skulls were recognised at a higher rate than CFRs from the other two ancestries, regardless of the ancestry of the assessors or the practitioners.

The hit rates were examined in relation to whether the practitioner was the familiar- or unfamiliar-race as the skull. For example, if a practitioner from the UK or South Africa reconstructed a Mongoloid-type CFR, then it was classified as 'unfamiliar-race'. The familiar-race CFRs were recognised more frequently than the unfamiliar-race CFRs (18.2% versus 6.8% in the UK assessors; 20.2% versus 6.0% in Korean assessors; 19.7% versus 5.3% in South African assessors) (Fig. 4). These differences were statistically significant at the $p < 0.05$ level in One-way ANOVA analysis. When the results from all the assessor groups were combined there was a significant difference (p = 0.014) in the hit rates between the familiar-race CFRs (19.4%) than the unfamiliar-race CFRs (6.0%). These results imply that CFRs from familiar-race skulls as the practitioner are recognised more accurately than those from unfamiliar-race skulls.

Hit rates for the CFRs were subdivided further depending on the two groups of practitioners (UK and South Africa) in relation to assessor groups to examine differences in the hit rates between familiar- and unfamiliar-race CFRs as practitioners. Familiar-race CFRs for the UK practitioners demonstrated higher hit rates than unfamiliar-race CFRs recorded by UK (15.9% versus 2.3%), Korean (23.8% versus 2.4%) and South African (34.2% versus 2.6%) assessors and with the total assessors (24.2% versus 2.4%) (Fig. 5). With the exception of the UK assessors (p = 0.051), these differences in Korean and South African assessors were statistically significant at p = 0.012 and 0.014 respectively. In contrast, when we look at the results for the South African practitioners the differences are not so clear. Although the hit rates from the two groups of assessors (UK and Korean) and the total assessors were higher for the familiar-race CFRs than the unfamiliar-race CFRs, the differences were smaller and were not significant statistically ($p > 0.05$). These results suggest that the cross-race effect can be demonstrated with the UK practitioners, but was not evident with the South African practitioners.

The familiar-race assessors recorded a slightly lower hit rate (9.3%) than the unfamiliar-race assessors (15.7%) for the CFRs by the UK practitioners, but this difference was not statistically significant at $p > 0.05$ level (Fig. 6). The familiar- and unfamiliar- race assessors recorded similar hit rates (11.9% and 11.8% respectively) for the CFRs by the South Africa practitioners. When all the practitioner groups were combined, there was no difference between the familiar- and unfamiliar-race assessor hit rates (10.6% and 13.8% respectively). These results suggest that the CFRs were recognised independently from whether the assessors were the familiar- or unfamiliar-race to the CFR.

In summary for the CFRs produced by the experienced practitioners; the ancestry of assessors did not affect recognition; CFRs produced by familiar-race practitioners were recognised more reliably than those by unfamiliar-race practitioners; CFRs from Caucasoid-type skulls were recognised more frequently than those from Negroid- and Mongoloid-type skulls; Familiar-race CFRs produced by the UK practitioners were recognised the most and the cross-race effect can be demonstrated for UK practitioners but not with the South African practitioners.

Discussion

The overall mean hit rate of 12.7% can be compared to other previous studies used face pool comparison tests: 11% [49]; 70% [51]; 44% [54]; 23% [55]; 21% [88]; 21% [89]. Although each study employed slightly different methodologies and performance procedures, the mean hit rate for the CFRs was generally lower than those from the previous studies. Among the hit rates for the eight CFRs, the two CFRs from the UK practitioners (C1-UK and C2-UK) recorded the highest hit rates at 24%. This hit rate is close to the average hit rate from the above six previous studies at 28%. Previous research using Caucasoid-type skulls [49,51,54,55,88,89] are comparable. A research [89] has suggested that the hit should exceed 18% in order to be meaningful recognition. According to this suggestion [89], the two Caucasoid-type CFRs by the UK practitioners can be considered as a standard to judge the accuracy of other CFRs and the existence of ‘unfamiliar-race effect’ within this study.

The hit rates for each CFR varied from 0 to 24%. The two CFRs recording the lowest hit rates were from Mongoloid-type skulls with unfamiliar-race practitioners (UK and SA). This result suggests that the eight CFRs recorded fluctuating degrees of reproducibility in relation to resemblance to the targets, and that the degree of familiarity to the ancestry of CFRs might be responsible for this variation.

Overall, the CFRs from Caucasoid-type skulls recorded the highest hit rate (20%) compared to the other two skull types (11% for Negroid and 4% for Mongoloid) (Fig. 3). This result implies that the Caucasoid-type CFRs were more successfully recognised than those from Negroid- and Mongoloid-type skulls, regardless of the ancestry of the assessors or the practitioners. This higher hit rate was mainly caused by the UK practitioners. The Mongoloid-type CFRs recorded the lowest hit rate at 4% and this may be due to the unfamiliar-race effect.

The CFRs produced by familiar-race practitioners recorded overwhelmingly higher hit rates than those of unfamiliar-race practitioners (3.2 times more correctly recognized) (Fig. 4). This result is one of the most significant finding in this study suggesting that CFRs produced by a familiar-race practitioner were more accurate than those from unfamiliar-race practitioners. This compares to other cross-race effect psychology studies [70,71] and suggests that the degree of familiarity to a certain ancestry type (including the skull) can be transmitted to the ability of producing more accurate CFR.

The familiar-race CFRs by the UK practitioners scored higher hit rates than unfamiliar-race CFRs by all assessor groups (Fig. 5). However, the correlation between skull and practitioner familiarity with ancestry was less obvious for the CFRs produced by the South African practitioner. These results suggest that the unfamiliar-race effect was more obvious for the UK practitioners than the South African practitioner.

It was found that the CFRs were recognised independently from whether the assessors were the familiar- or unfamiliar-race as the skull. The cross-race effect relating to the assessors was also not observed where the CFRs were classified by the ancestry of the skulls. These results suggest that the ancestry of the assessors relative to the ancestry of the CFRs did not influence their recognisability. None of the previous studies examined the recognition rates regarding the ancestry of the assessors [49,51,54,55,88,89]. Even though the hit rates from this study were not directly comparable to the results from earlier research, we cautiously suggest that all assessor groups were presumed to employ universally categorised processing for the matching performance between the CFR and possible faces in the photo array. This presumption does not appears to be consistent with the results of previous studies which established that different ancestry people, especially between White European and East Asian, demonstrated a different processing strategy when recognising the human face – White Europeans recognise familiar-race faces more holistically than Asian faces, whereas East Asians demonstrate holistic strategies for processing both familiar- and unfamiliar-race facial configuration [65,74,77]. It therefore hypothesised that the assessors may alter the perceptual strategy (holistic processing) when configuring CFRs, even from the familiar-race skulls, because the process of recognising CFR requires more analytic approaches than recognising human faces.

General Discussion

Accuracy assessment is crucially important to estimate the reliability of the CFR in recognition or identification of unknown human remains. Previous research has reported diverse results regarding the accuracy of CFR employing various empirical methods [84,89,90]. For accuracy surveys, researchers have to present an image of the CFR alongside an ante-mortem facial photograph of the identified individual, utilising blind tests. However, there are fundamental problems on this general protocol that the studies are practically and ethically difficult to represent a forensic scenario as based on familiar face recognition [8,90]. More significantly ante-mortem photographs were typically provided with inferior resolution and diverse facial expressions, which might create difficulties to unify the photographs with distractors allowing standardised objective studies [83]. Thus, these blind studies inevitably must rely on unfamiliar face recognition using variable quality photographs (even though the pictures were closely aligned to the target).

It is well established by extensive studies that people are poor at recognising unfamiliar faces [91-93]. Cognitive psychological research, employing face pool comparison tasks, demonstrated that the true positive hit rates were ranged below 80%. The hit rates dramatically dropped when facial textures, hairs, skin colours, light or posture were removed or changed from the actual features [8, 94-97]. The overall lower hit rates in this research maybe caused by these limitations of difficulties in unifying the photographs and unfamiliar face recognition for the accuracy survey. In addition, this lower hit rates trend could be raised by the difficulties in choosing the correct targets from the photographic line-ups of this face pool comparison tasks. The majority of assessors indeed left comments after the experiment that the task was extremely hard as the most of the faces in the photo array looked very similar to the corresponding CFR. The selection of distractor faces followed the recommendation of the Police and Criminal Evidence Act [87] and perhaps this is not appropriate when studying CFR accuracy. Perhaps more distinctive faces should be included in the photo array so that face type is one of the distinguishing characteristics.

The hit rate for the familiar-race CFRs by the South African practitioner was lower than the result from the UK practitioners, whereas the hit rate for unfamiliar-race skull was higher than those from the UK practitioners. Hence, the South African practitioner demonstrated lesser discrepancy in the hit rates between the ancestry groups of the CFRs than those of the UK practitioners. Some explanations may be possible; firstly, the experience of the South African practitioner in CFR was two years which is shorter than those of the UK practitioners (five and fifteen years for the two practitioners). It can be estimated that the highest hit rates of the two CFRs from the UK practitioners were contributed by not only the unfamiliar-race effect but also the longer experience periods of the practitioners in CFR. In addition, the South African practitioner was considered to have the higher contact experience for both Black and White faces. From these results, we could establish that training and experience in unfamiliar-race faces and skulls may promote the recognisability of CFRs.

The cross race effect was not observed for the assessors and this may be due to the recognition of CFRs being performed ‘analytically’ (as similar to perceive unfamiliar faces) rather than ‘holistically’ (as similar to perceive familiar faces) by universal assessors regardless of their ancestry or cultural bases. Namely, even if there was concealed cross-race effect on recognising the CFRs, the influence was not revealed since the reproducibility (recognisability) of CFRs were biased already by the practitioners; additionally, it is hypothesised that observers tend to look at CFR as like an object that needs to be analysed. However, the result presenting the non-statistical difference in the hit rates for the familiar-race skull to the Korean assessors suggests a possibility that the cross-race effect might be combined in recognising CFR. Therefore, further research is required into this issue employing more rigorously categorised assessor groups in term of the race.

It is suggested that further research employing more strictly standardised methodology, skulls, practitioners and assessors with abundant number of the subject would be required to establish more rigorous results. Consequently, in order to investigate more robust cross-race effects on CFRs, the above possible interference factors ought to be minimised as much as possible when designing further studies and exploring the cross-race effect on CFRs.

Conclusions

The practitioners demonstrated the unfamiliar-race effect when producing CFRs and produced more recognisable familiar-race CFRs.

The UK practitioners produced the most recognisable familiar-race CFRs, but demonstrated the greatest unfamiliar-race effect.

The assessors were not affected by ancestry.

This study established that the ancestry experience of practitioners can influence the recognisability and accuracy of CFRs.

References

[1] C.M. Wilkinson, C. Rynn, H. Peters, M. Taister, C.H. Kau, S. Richmond, A blind accuracy assessment of computer-modeled forensic facial reconstruction using computed tomography data from live subjects, Forensic Sci. Med. Pathol. 2 (3) (2006) 179-188.

[2] R.M. George, The lateral craniographic method of facial reconstruction, J. Forensic Sci. 32 (5) (1987) 1305-1330.

[3] C.N. Stephan, Anthropological facial 'reconstruction'– recognizing the fallacies, 'unembracing' the errors, and realizing method limits, Sci. Justice 43 (4) (2003) 193-200.

[4] C.M. Wilkinson, Forensic facial reconstruction, Cambridge University Press, Cambridge, 2004.

[5] C.M. Wilkinson, Facial analysis of preserved bodies, Yearbook of Mummy Studies 1 (2011) 171-180.

[6] C.N. Stephan, Facial approximation: a review of the current state of play for archaeologists, Int. J. Osteoarchaeol. 15 (4) (2005) 298-302.

[7] W.J. Lee, A.Y. Yoon, M.K. Song, C.M. Wilkinson, D.H. Shin, The archaeological contribution of forensic craniofacial reconstruction to a portrait drawing of a Korean historical figure, J. Archaeol. Sci. 49 (1) (2014) 228-236.

[8] S. Davy-Jow, The devil is in the details: a synthesis of psychology of facial perception and its applications in forensic facial reconstruction, Sci. Justice 53 (2) (2013) 230-235.

[9] W.J. Lee, S. Mackenzie, C.M. Wilkinson, Facial identification of the dead, in: S. Black, E. Ferguson (Eds.), Forensic anthropology 2000-2010, CRC Press, 2011, pp. 363-386.

[10] W.M. Krogman, M.Y. İşcan, The Human Skeleton in Forensic Medicine, Charles C. Thomas, Illinois, 1986.

[11] S. Tian, Y. Nishida, B. Isberg, G. Lennerstrand, MRI measurements of normal extraocular muscles and other orbital structures, Graefes. Arch. Clin. Exp. Ophthalmol. 238 (5) (2000) 393-404.

[12] B.P. Gatliff, Facial sculpture on the skull for identification, Am. J. Forensic Med. Pathol. 5 (4) (1984) 327-332.

[13] C.M. Wilkinson, S.A. Mautner, Measurement of eyeball protrusion and its application in facial reconstruction, J. Forensic Sci. 48 (1) (2003) 12-16.

[14] J.L. Angel, Restoration of head and face for identification, in Proceedings of the 30th annual meeting of the American Academy of Forensic Sciences, St. Louis, Missouri, 1987.

[15] M.M. Gerasimov, The reconstruction of the face from the basic structure of the skull, Translator, W. Tshernezky, Publishers unknown, Russia, 1955.

[16] C. Rynn, C.M. Wilkinson, Appraisal of traditional and recently proposed relationships between the hard and soft dimensions of the nose in profile, Am. J. Phys. Anthropol. 130 (3) (2006) 364-373.

[17] C. Rynn, C.M. Wilkinson, H.L. Peters, Prediction of nasal morphology from the skull, Forensic Sci. Med. Pathol. 6 (1) (2009) 1-15.

[18] C.N. Stephan, M. Henneberg, W. Sampson, Predicting nose projection and pronasale position in facial approximation: a test of published methods and proposal of new guidelines, Am. J. Phys. Anthropol. 122 (3) (2003) 240-250.

[19] K.M. Lee, W.J. Lee, J.H. Cho, H.S. Hwang, Three-dimensional prediction of the nose for facial reconstruction using cone-beam computed tomography, Forensic Sci. Int. 236 (2014) 194.e1–e5.

[20] G.V. Lebedinskaya, T.S. Balueva, E.V. Veselovaskaya, Principles of facial reconstruction, in: M.Y. İşcan, R.P. Helmer (Eds.), Forensic analysis of the skull, Wiley-Liss, New York, 1993, pp. 183-198.

[21] R.M. George, Anatomical and artistic guidelines for forensic facial reconstruction, in: M.Y. İşcan, R.P. Helmer (Eds.), Forensic analysis of the skull, Wiley-Liss, New York, 1993, pp. 215-127.

[22] C.M. Wilkinson, M. Motwani, E. Chiang, The relationship between the soft tissues and the skeletal detail of the mouth, J. Forensic Sci. 48 (4) (2003) 728-732.

[23] C.N. Stephan, M. Henneberg, Predicting mouth width from inter-canine width - a 75% rule, J. Forensic Sci. 48 (4) (2003) 725-727.

[24] C.N. Stephan, S.J. Murphy, Mouth width prediction in craniofacial identification: cadaver tests of four recent methods, including two techniques for edentulous skulls, J. Forensic Odontostomatol. 26 (1) (2008) 2-7.

[25] B.A. Fedosyutkin, J.V. Nainys, The relationship of skull morphology to facial features, in: M.Y. İşcan, R.P. Helmer (Eds.), Forensic analysis of the skull, Wiley-Liss, New York, 1993, pp. 199-213.

[26] P. Guyomarc'h, C.N. Stephan, The validity of ear prediction guidelines used in facial approximation, J. Forensic Sci. 57 (6) (2012) 1427-1441.

[27] R.M. George, The lateral craniographic method of facial reconstruction, J. Forensic Sci. 32 (5) (1987) 1305-1330.

[28] H. Utsuno, T. Kageyama, K. Uchida, K. Kibayashi, Facial soft tissue thickness differences among three skeletal classes in Japanese population. Forensic Sci. Int. 236 (2014) 175-180.

[29] D. Vandermeulen, P. Claes, D. Loeckx, S. De Greef, G. Willems, P. Suetens, Computerized craniofacial reconstruction using CT-derived implicit surface representations, Forensic Sci. Int. 159 (1) (2006) 164-174.

[30] J.P. Moss, A.D. Linney, S.R. Grindrod, C.A. Mosse, A laser scanning system for the measurement of facial surface morphology, Opt. Lasers Eng. 10 (3-4) (1989) 179-190.

[31] K.D. Kim, A. Ruprecht, G. Wang, J.B. Lee, D.V. Dawson, M.W. Vannier, Accuracy of facial soft tissue thickness measurements in personal computer-based multiplanar reconstructed computed tomographic images, Forensic Sci. Int. 155 (1) (2005) 28-34.

[32] D. Cavanagh, M. Steyn, Facial reconstruction: soft tissue thickness values for South African black females, Forensic Sci. Int. 206 (1-3) (2011) 215.e1-215.e7.

[33] D. Sahni, I. Jit, M. Gupta, P. Singh, S. Suri, Sanjeev, H. Kaur, Preliminary study on facial soft tissue thickness by magnetic resonance imaging in northwest Indians. Forensic Sci. Commun. 4

(2002) 1-7.

[34] D. Sahni, Sanjeev, G. Singh, I. Jit, P. Singh, Facial soft tissue thickness in northwest Indian adults, Forensic Sci. Int. 176 (2-3) (2008) 137-146.

[35] S. Sipahioğlu, H. Ulubay, H.B. Diren, Midline facial soft tissue thickness database of Turkish population: MRI study, Forensic Sci. Int. 219 (1-3) (2011) 282.e1-282.e8.

[36] S. De Greef, P. Claes, D. Vandermeulen, W. Mollemans, P. Suetens, G. Willems, Large-scale in-vivo Caucasian facial soft tissue thickness database for craniofacial reconstruction, Forensic Sci. Int. 159 (1) (2006) S126-146.

[37] M.H. Manhein, G.A. Listi, R.E. Barsley, R. Musselman, N.E. Barrow, D.H. Ubelaker, In vivo facial tissue depth measurements for children and adults, J. Forensic Sci. 45 (1) (2000) 48-60.

[38] C.M. Wilkinson, In vivo facial tissue depth measurements for white British children, J. Forensic Sci. 47 (3) (2002) 459-465.

[39] S.L. Smith, G.S. Throckmorton, A new technique for three-dimensional ultrasound scanning of facial tissues. J. Forensic Sci. 49 (3) (2004) 451-457.

[40] W.N.J. Chan, G.A. Listi, M.H. Manhein, In vivo facial tissue depth study of chinese-american adults in New York city. J. Forensic Sci. 56 (2) (2011) 350-358.

[41] Z. Fourie, J. Damstra, P.O. Gerrits, Y. Ren, Accuracy and reliability of facial soft tissue depth measurements using cone beam computer tomography, Forensic Sci. Int. 199 (1-3) (2010) 9-14.

[42] H.S. Hwang, K. Kim, D.N. Moon, J.H. Kim, C.M. Wilkinson, Reproducibility of facial soft tissue thicknesses for craniofacial reconstruction using Cone-Beam CT images, J. Forensic Sci. 57 (2) (2012) 443-448.

[43] H.S. Hwang, M.K. Park, W.J. Lee, J.H. Cho, B.K. Kim, C.M. Wilkinson, Facial soft tissue thickness database for craniofacial reconstruction in Korean adults. J. Forensic Sci. 57 (6) (2012) 1442-1447.

[44] H.H. Wilder, The physiognomy of the Indians of Southern New England. Am. Anthropol. New Series 14 (3) (1912) 415-436.

[45] H. Von Eggeling, Die leistungsfahigkeit physiognomischerrekon-struktionsversuche auf grundlagc des schädels, Arch. Anthropol. 12 (1913) 44-47.

[46] F. Stadtmüller, Zur beurteilung der plastischen rekonstruktionsmethode der physiognomie auf dem schädel, J. Morph. Anthropol. 22 (1922) 337-372.

[47] J. Kollmann, W. Büchly, Die Persistenz der rassen und die reconstruction der physiognomie prahistorischer schädels, Arch. Anthropol. 25 (1898) 329-359.

[48] V. Suk, Fallacies of anthropological identifications, Publications de la Facultae des sciences de l'Universitae Masaryk, 207 (1935) 3-18.

[49] C.N. Stephan, M. Henneberg, Building faces from dry skulls: Are they recognized above chance rates? J. Forensic Sci. 46 (3) (2001) 432-440.

[50] M.M. Gerasimov, The face finder, Lippincott, New York, 1971.

[51] C.C. Snow, B.P. Gatliff, K.R. McWilliams, Reconstruction of facial features from the skull: an evaluation of its usefulness in forensic anthropology, Am. J. Phys. Anthropol. 33 (2) (1970) 221-228.

[52] B.P. Gatliff, C.C. Snow, From skull to visage, J. Biocommun. 6 (2) (1979) 27-30.

[53] P. Vanezis, R.W. Blowes, A.D. Linney, A.C. Tan, R. Richards, R. Neave, Application of 3-D computer graphics for facial reconstruction and comparison with sculpting techniques. Forensic Sci. Int. 42 (1-2) (1989) 69-84.

[54] C.M. Wilkinson, D.K. Whittaker, Juvenile forensic facial reconstruction: a detailed accuracy study, in: Proc. of the 10th conference of the International Association of Craniofacial Identification, Bari, Italy, 2002.

[55] C.M.S. Fernandes, M. da Costa Serra, J.V.L. da Silva, P. Yoshito Noritomi, F.D.A. de Sena Pereira, R.F.H. Melani, Tests of one Brazilian facial reconstruction method using three soft tissue depth sets and familiar assessors, Forensic Sci. Int. 214 (1-3) (2012) 211.e1-211.e7.

[56] J.S. Rhine, C.E. Moore, Tables of facial tissue thickness of American Caucasoids in forensic anthropology, Maxwell Mus. Tech. Ser. 1, 1984, cited in: K.T. Taylor, Forensic Art and Illustration, Boca Raton, CRC Press, Florida, 2001.

[57] T. Suzuki, Reconstitution of a skull, Int. Crime Police Rev. 264 (1973) 76-80.

[58] W.D. Haglund, D.T. Reay, Use of facial approximation techniques in identification of Green River Serial Murder victims, Am. J. Forensic Med. Pathol. 12 (2) (1991) 132-142.

[59] J. Prag, R.A.H. Neave, Making faces, British Museum Press, London, 1997.

[60] K.T. Taylor, Forensic art and illustration, Boca Raton, CRC Press, Florida, 2001.

[61] C.M. Wilkinson, R.A.H. Neave, The reconstruction of a face showing a healed wound, J. Archaeol. Sci. 30 (10) (2003) 1343-1348.

[62] BBC, Corstorphine Hill murder: Reconstruction of murder victim's face released. 2013. Available at: < http://www.bbc.com/news/uk-scotland-edinburgh-east-fife-23131524>, [Accessed on 12 January 2016]

[63] P.N. Shapiro, S. Penrod, Meta-Analysis of Facial Identification Studies, Psychol. B. 100 (2) (1986), 139-156.

[64] R.K. Bothwell, J.C. Brigham, R.S. Malpass, Cross-racial identification, Pers. Soc. Psychol. B. 15 (1) (1989) 19-25.

[65] C.A. Meissner, J.C. Brigham, Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review, Psychol. Public Pol. Law 7 (1) (2001) 3-35.

[66] J.W. Tanaka, M. Kiefer, C.M. Bukach, A holistic account of the own-race effect in face recognition: Evidence from a cross-cultural study, Cognition, 93 (1) (2004) B1-B9.

[67] S. Sangrigoli, C. Pallier, A.-M. Argenti, V.A.G. Ventureyra, S. De Schonen, Reversibility of the other-race effect in face recognition during childhood, Psychol. Sci. 16 (6) (2005) 440-444.

[68] P.M. Walker, M. Hewstone, A perceptual discrimination investigation of the own-race effect and intergroup experience, Appl. Cognitive Psych. 20 (4) (2006) 461-475.

[69] K.J. Hancock, G. Rhodes, Contact, configural coding and the other-race effect in face recognition, Brit. J. Psychol. 99 (1) (2008) 45-56.

[70] J.L. Marcon, C.A. Meissner, R.S Malpass, Cross-race effect, 2008. Available at: <http://digitalcommons.utep.edu/cgi/viewcontent.cgi?article=1034&context=christian_meissner>, [Accessed on 12 January 2016].

[71] G. Rhodes, W.G. Hayward, C. Winkler, Expert face coding: configural and component coding of own-race and other-race faces, Psychon. B. Rev. 13 (3) (2006) 499-505.

[72] R.S. Malpass, J. Kravitz, Recognition for faces of own and other race, J. Pers. Soc. Psychol. 13 (4) (1969) 330-334.

[73] J.C. Brigham, R.S. Malpass, The role of experience and contact in the recognition of faces of own- and other- race persons, J. Soc. 41 (3) (1985) 139-155.

[74] P. Chiroro, T. Valentine, An investigation of the contact hypothesis of the own-race advantage in face recognition, Q. J. Exp. Psychol. 48 (4) (1995) 879-894.

[75] C. Blais, R.E. Jack, C. Scheepers, D. Fiset, R. Caldara, Culture shapes how we look at faces, PLoS ONE 3 (8) (2008) art. no. e3022.

[76] R.E. Nisbett, T. Masuda, Culture and point of view, Proc. Nat. Acad. Sci. USA 100 (19) (2003) 11163-111170.

[77] R.E. Nisbett, Y. Miyamoto, The influence of culture: holistic versus analytic perception, Trends Cogn. Sci. 9 (10) (2005) 467-473.

[78] D.J. Kelly, S. Liu, H. Rodger, S. Miellet, L. Ge, R. Caldara, Developing cultural differences in face processing, Dev. Sci. 14 (5) (2011) 1176-1184.

[79] D.J. Turk, T.C. Handy, M.S. Gazzaniga, Can perceptual expertise account for the own-race bias in face recognition? a split-brain study. Cogni. Neuropsychol. 22 (7) (2005) 877-883.

[80] P.M. Walker, J.W. Tanaka, An encoding advantage for own-race versus other-race faces, Perception 32 (9) (2003) 1117-11125.

[81] W.M. Krogman, The reconstruction of the living head from the skull, FBI Law Enforcement Bulletin (1946) 11-18.

[82] M.Y. İşcan, R.P. Helmer, Forensic analysis of the skull (Eds.): craniofacial analysis, reconstruction, and identification, Wiley-Liss, New York, 1993.

[83] C.N. Stephan, M. Henneberg, Recognition by forensic facial approximation: case specific examples and empirical tests, Forensic Sci. Int. 156 (2-3) (2005) 182-191.

[84] W.J. Lee, C.M. Wilkinson, H.S. Hwang, An accuracy assessment of forensic computerized facial reconstruction employing Cone-Beam Computed Tomography from live subjects, J. Forensic Sci. 57 (2) (2012) 318-327.

[85] F. Birkner, Beiträge zur Rassenanatomie der Chinesen, Archiv. Anthropol. n.s. IV (1905) 1-40.

[86] Face recognition homepage, Department of Wireless Communications at the Faculty of Electrical Engineering and Computing in the University of Zagreb, Croatia, 2011, Available at: <http://www.face-rec.org/databases/>, [Accessed on 12 January 2016].

[87] Home Office Police and criminal evidence act 1984: code D, 2008, Available at: <http://www.homeoffice.gov.uk/publications/police/operational-policing/pace-codes/pace-code-d?view=Binary>, [Accessed on 12 January 2016].

[88] C.N. Stephan, R. Arthur, Assessing facial approximation accuracy: how do resemblance ratings of disparate faces compare to recognition tests? Forensic Sci. Int. 159 (suppl. 1) (2006) 159-163.

[89] C.N. Stephan, J. Cicolini, Measuring the accuracy of facial approximations: a comparative study of resemblance rating and face array methods, J. Forensic Sci. 53 (1) (2008) 58-64.

[90] C.M. Wilkinson, Facial reconstruction - anatomical art or artistic anatomy? J. Anat. 216 (2) (2010) 235-250.

[91] V. Bruce, Z. Henderson, K. Greenwood, P.J.B. Hancock, A.M. Burton, P. Miller, Verification of face identities from images captured on video, J. Exp. Psychol. Appl. 5 (4) (1999) 339-360.

[92] P.J. Hancock, V.V. Bruce, A.M. Burton, Recognition of unfamiliar faces, Trends Cogn. Sci. 4 (9) (2000) 330-337.

[93] R. Kemp, N. Towell, G. Pike, When seeing should not be believing: photographs, credit cards and fraud, Appl. Cogn. Psych. 11 (3) (1997) 211-222.

[94] A.W. Young, D.C. Hay, K.H. McWeeny, B.M. Flude, A.W. Ellis, Matching familiar and unfamiliar faces on internal and external features, Perception 14 (6) (1985) 737-774.

[95] M. Bindemann, A.M. Burton, The role of colour in human face detection, Cogn. Sci. 33 (6) (2009) 1144-1156.

[96] V. Bruce, P. Healey, M. Burton, T. Doyle, A. Coombes, A. Linney, Recognising facial surfaces, Perception 20 (6) (1991) 755-769.

[97] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face recognition by humans: nineteen results all computer vision researchers should know about, Proc. IEEE 94 (11) (2006) 1948-1961.

TABLE 1. Biological profiles of the six skulls, status of exchanging skull models between the two teams, applied average facial soft tissue depth data and methods for the CFRs.

Skull	Biological profiles			Skull exchange		Applied tissue depth date	Method for CFR	
	Sex	Ancestry	Age (years)	Provided from	Distributed to		by unfamiliar-race team	by familiar-race team
N1	Female	Negroid	20 - 30	South Africa	South Africa	Cavanagh and Steyn [32]	N/A	3D manual (N1-SA)
N2	Female	Negroid	20 - 30	South Africa	UK South Africa	Cavanagh and Steyn [32]	3D manual (N2-UK)	3D manual (N2-SA)
M1	Female	Mongoloid	20 - 30	China	South Africa	Created from Birkner [85] & Lebedinskaya [4]	3D manual (M1-SA)	N/A
M2	Male	Mongoloid	30 - 40	China	UK	Created from Birkner [85] & Lebedinskaya [4]	3D computerized (M2-UK)	N/A
C1	Male	Caucasoid	20 - 30	UK	South Africa UK	Sahni et al. [34]	3D manual (C1-SA)	3D computerized (C1-UK)
C2	Female	Caucasoid	15 - 21	UK	UK	Helmer [4]	N/A	3D manual (C2-UK)

N/A = Not applicable.

TABLE 2. Overall hit rates for the eight CFRs.

CFR	**Hit rate (%)** Assessor group UK (n = 22)	Korea (n = 21)	South Africa (n = 19)	Total assessors	**Significance (*p*-value)**
N1-SA	22.7	19.0	5.3	***16.1***	**0.015***
N2-UK	4.5	4.8	5.3	***4.8***	
N2-SA	18.2	14.3	5.3	***12.9***	
M1-SA	4.5	19.0	0.0	***8.1***	
M2-UK	0.0	0.0	0.0	***0.0***	
C1-UK	18.2	23.8	31.6	***24.2***	
C1-SA	18.2	0.0	15.8	***11.3***	
C2-UK	13.6	23.8	36.8	***24.2***	
Mean hit rate	***12.5***	***13.1***	***12.5***	**Overall mean hit rate (%)**	***12.7***
Significance (*p*-value)		0.993			

*, $p < 0.05$.

FIGURE 1. Laser scanned images of the six skull models. N1 and N2 from Negroid type; M1 and M2 from Mongoloid type; C1 and C2 from Caucasoid type.

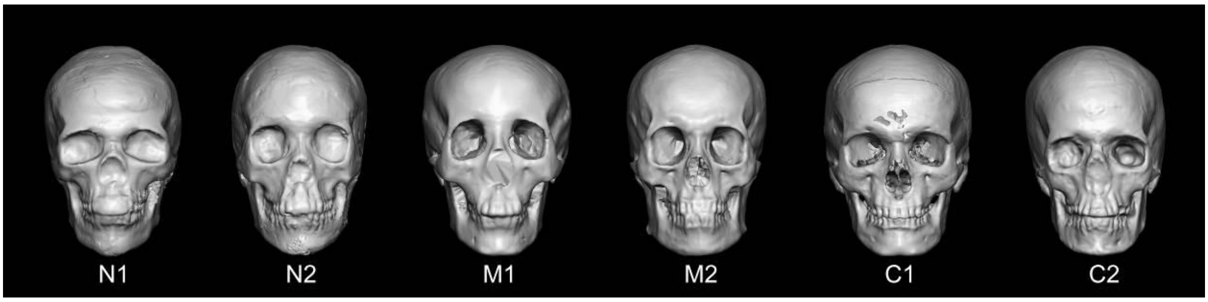


FIGURE 2. Eight CFRs produced by the two teams. N1-SA and N2-SA were produced by SA team; C1-UK and C2-UK were produced by UK team as familiar CFRs. N2-UK and M2-UK were produced by UK team; M1-SA and C1-SA were produced by SA team as unfamiliar-race CFRs.

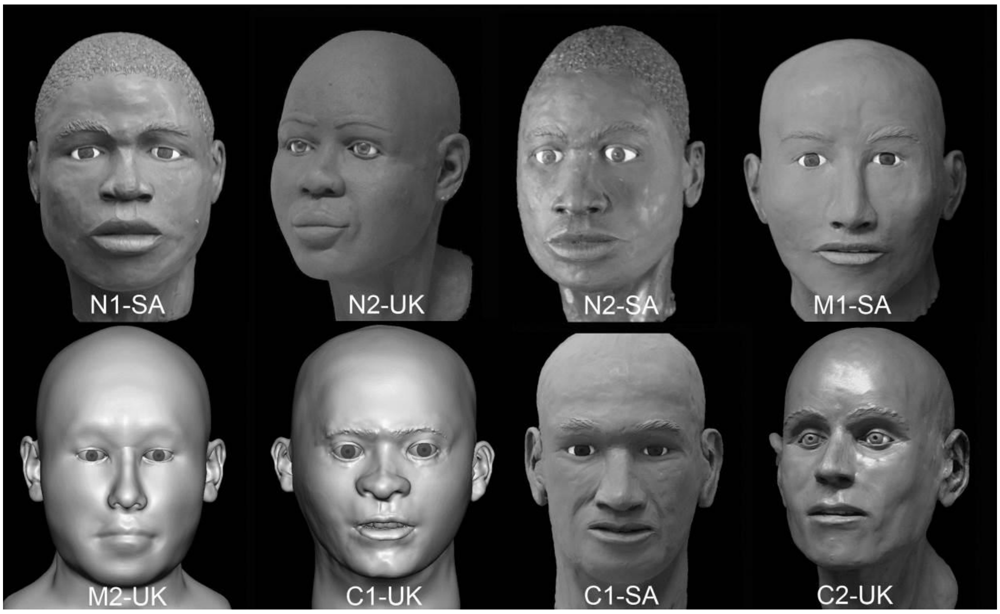
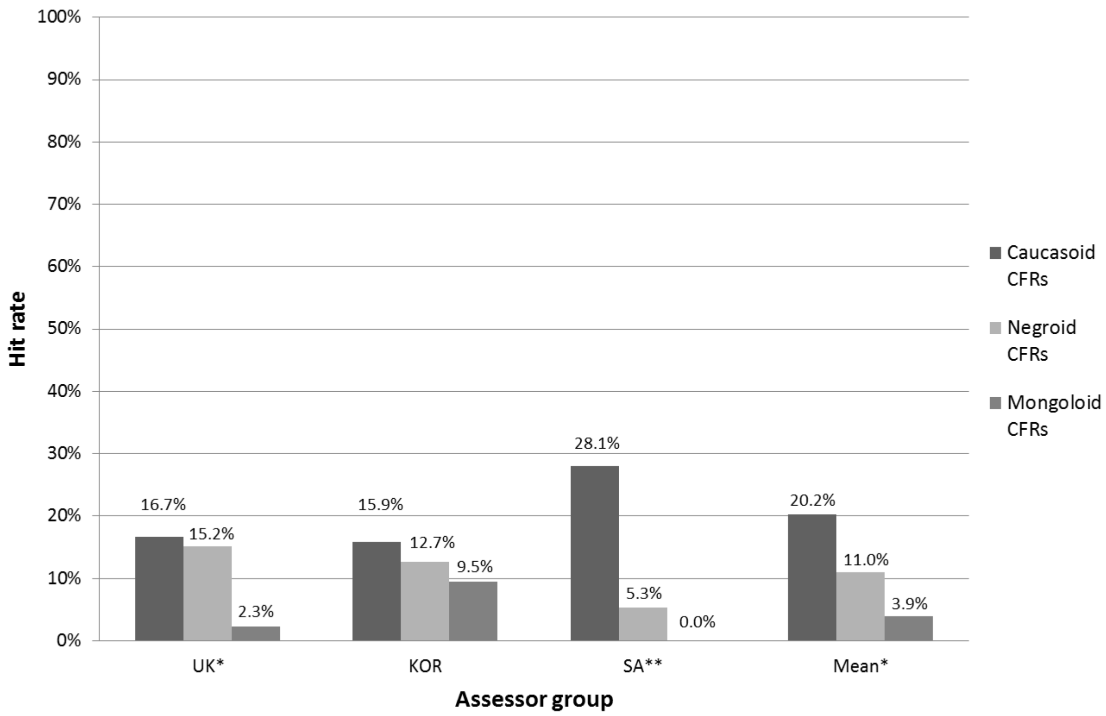


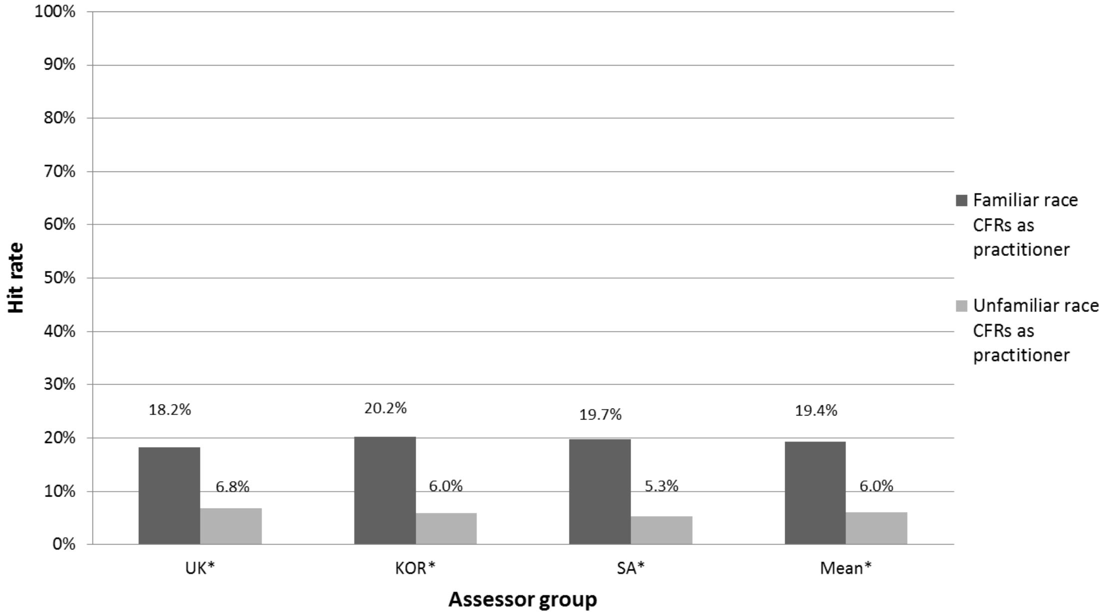
FIGURE 3. Graphs showing the hit rates from three assessor groups relating to the familiarity with the ancestry of the CFRs as the practitioners.



*, $p < 0.05$; **, $p < 0.01$.

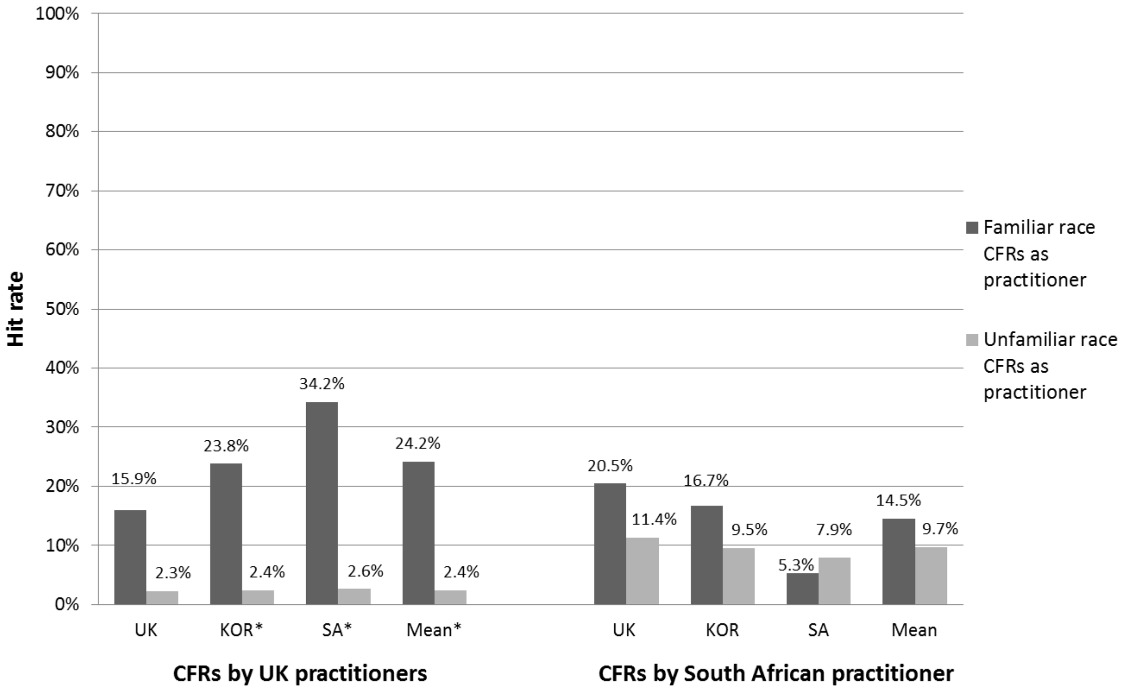
"UK" = UK assessors; "KOR" = Korean assessors; "SA" = South African assessors.

FIGURE 4. Graphs showing the hit rates from three assessor groups relating to the familiarity with the ancestry of the CFRs as the practitioners.



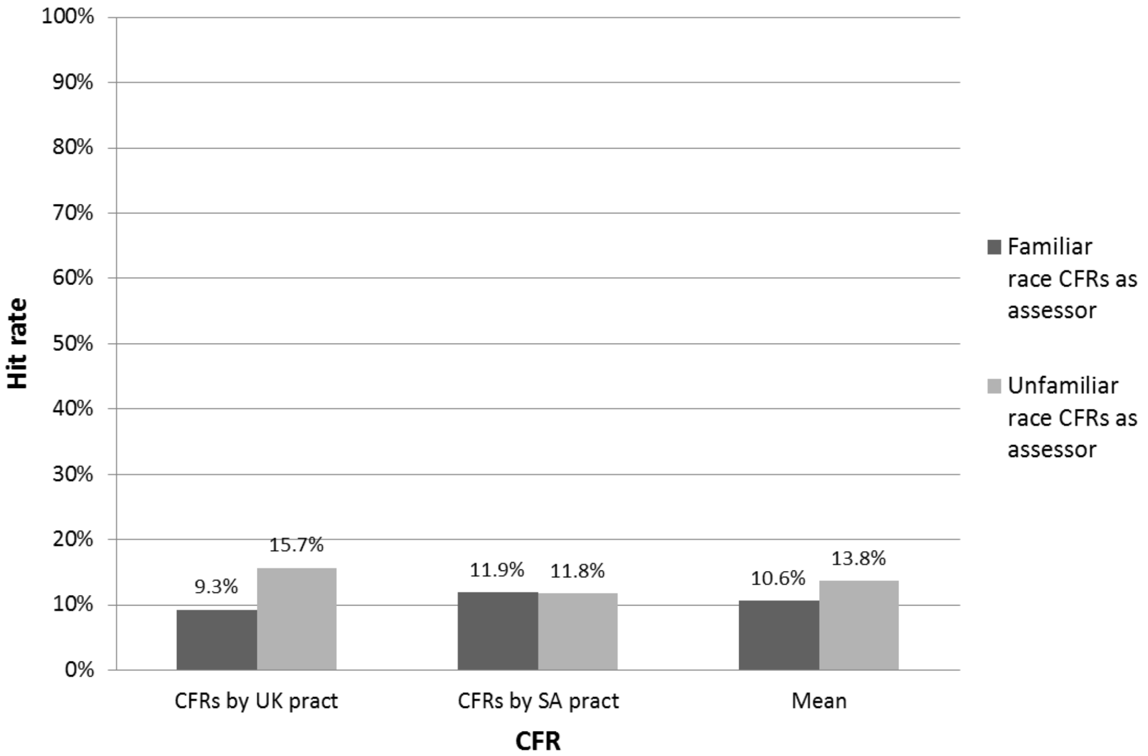
*, $p < 0.05$.
"UK" = UK assessors; "KOR" = Korean assessors; "SA" = South African assessors.

FIGURE 5. Graphs showing the hit rates from three assessor groups relating to the familiarity with the ancestry of CFRs as the UK and South African practitioners.



*, $p < 0.05$.
"UK" = from UK assessors; "KOR" = from Korean assessors; "SA" = from South African assessors.

FIGURE 6. Graphs showing the hit rates between the practitioner groups relating to the familiarity with the ancestry of the CFRs as the assessors.



"UK pract" = UK practitioner.

"SA pract" = South African practitioner.

fig 1.jpg
Click here to download high resolution image

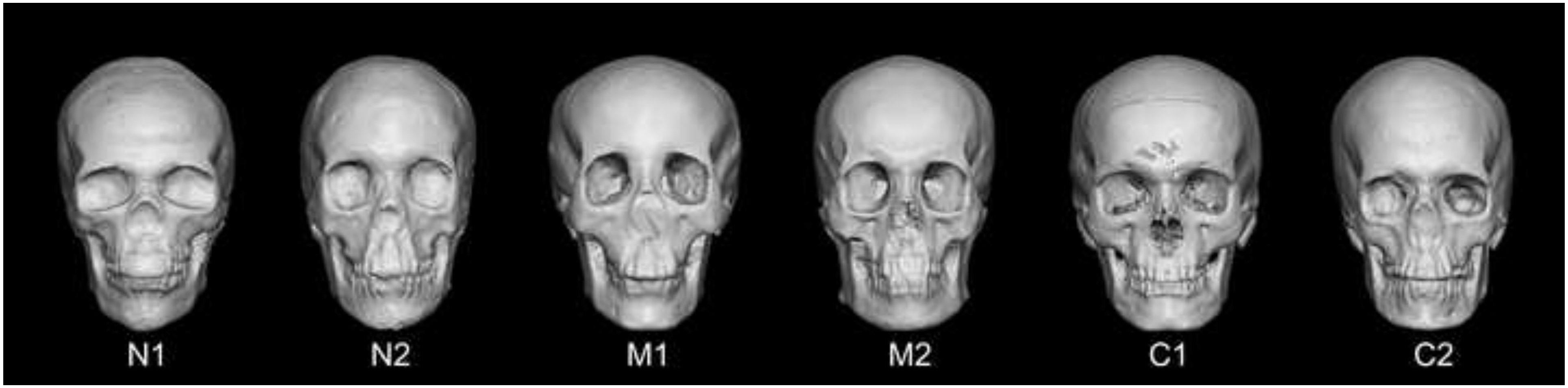


fig 2.jpg
Click here to download high resolution image

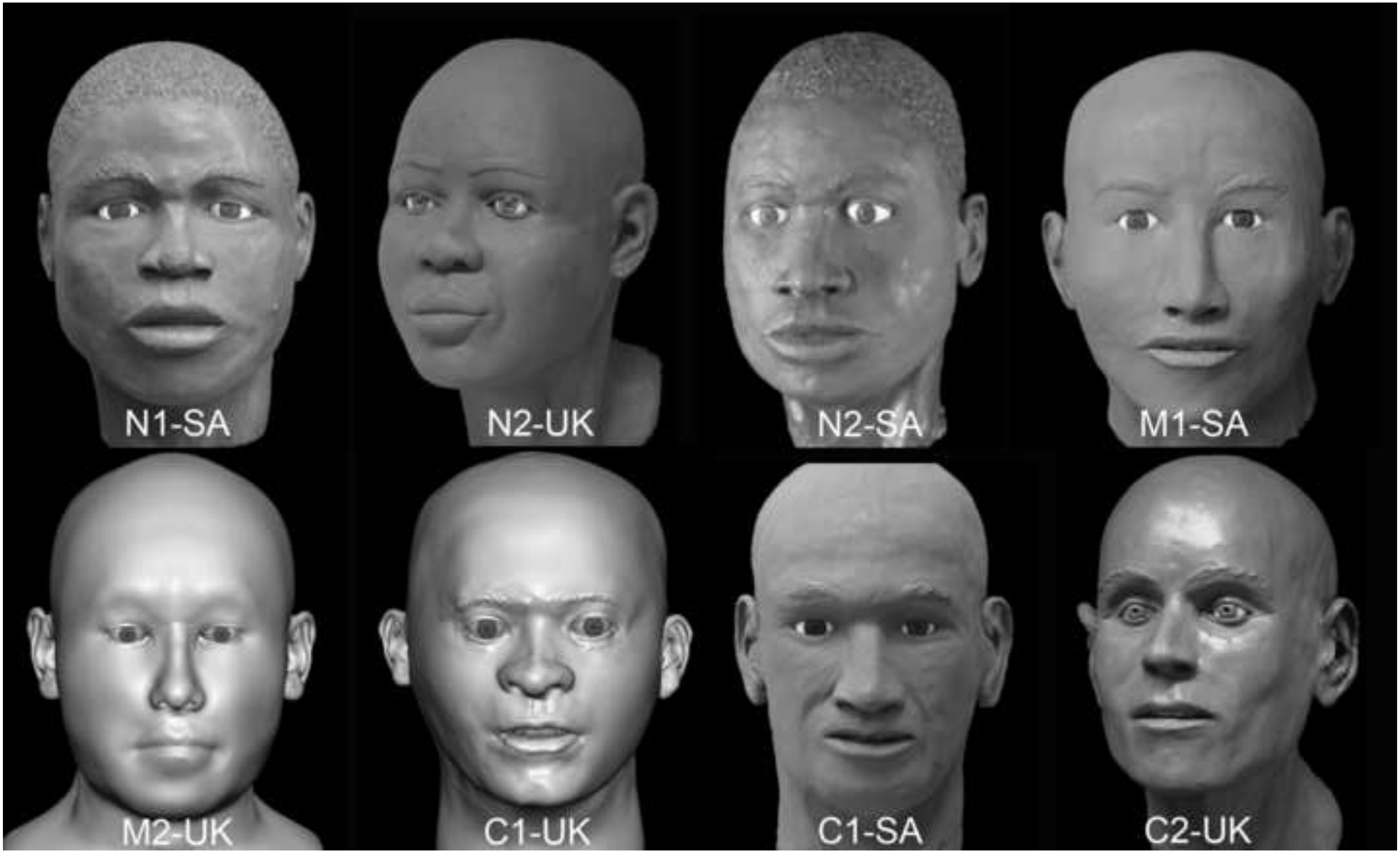


fig 3.jpg
Click here to download high resolution image

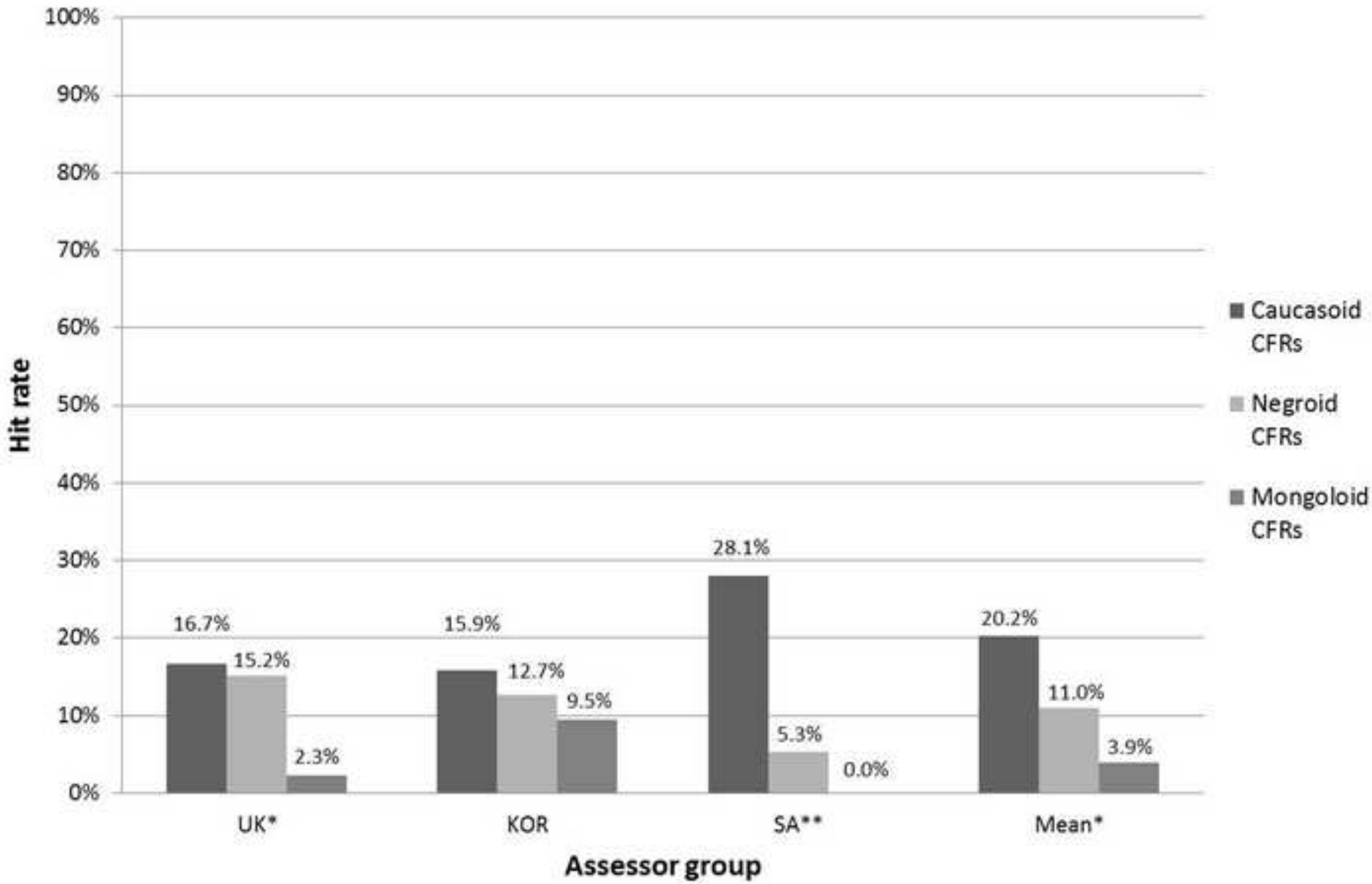


fig 4.jpg
Click here to download high resolution image

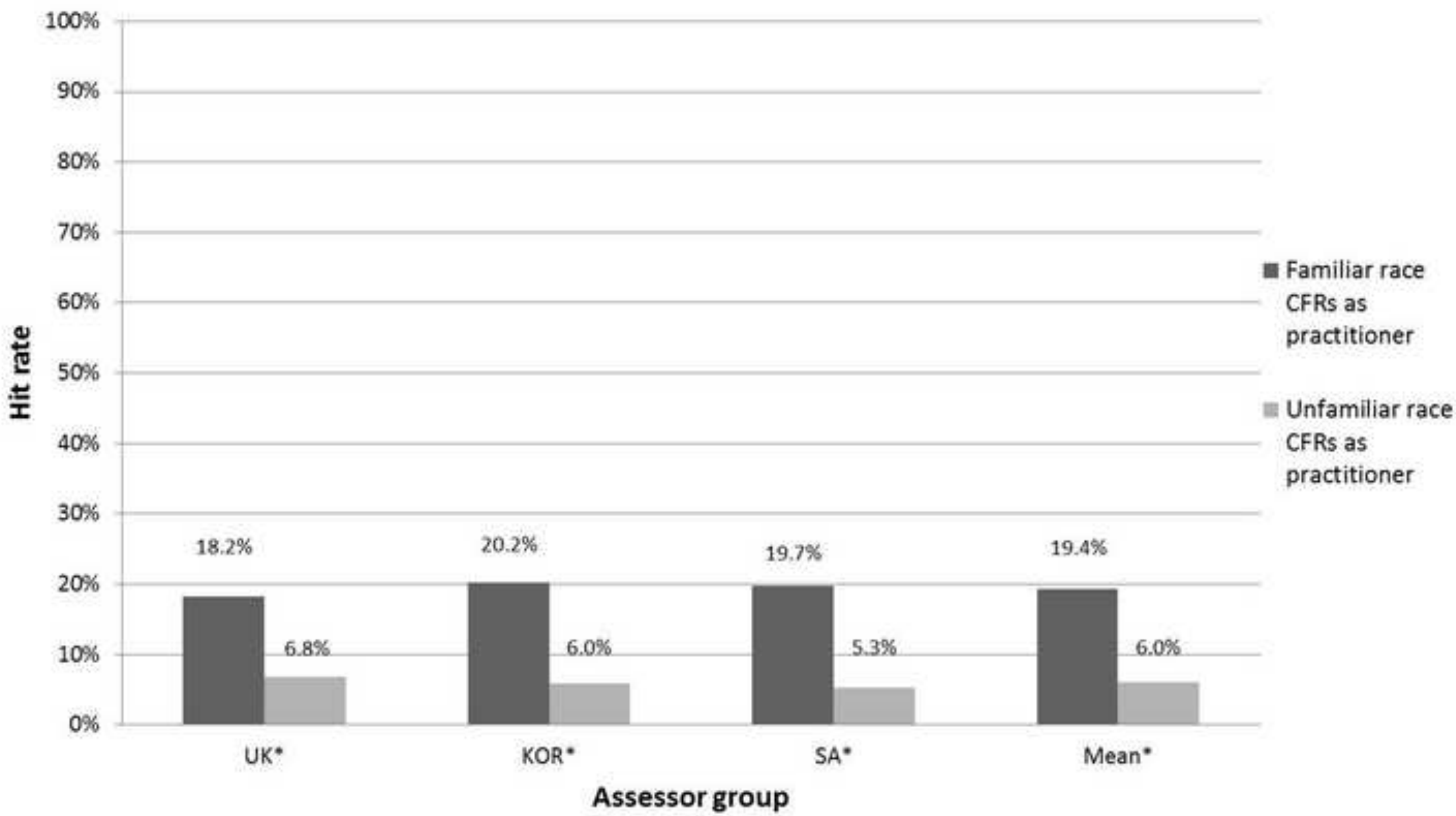


fig 5.jpg
Click here to download high resolution image

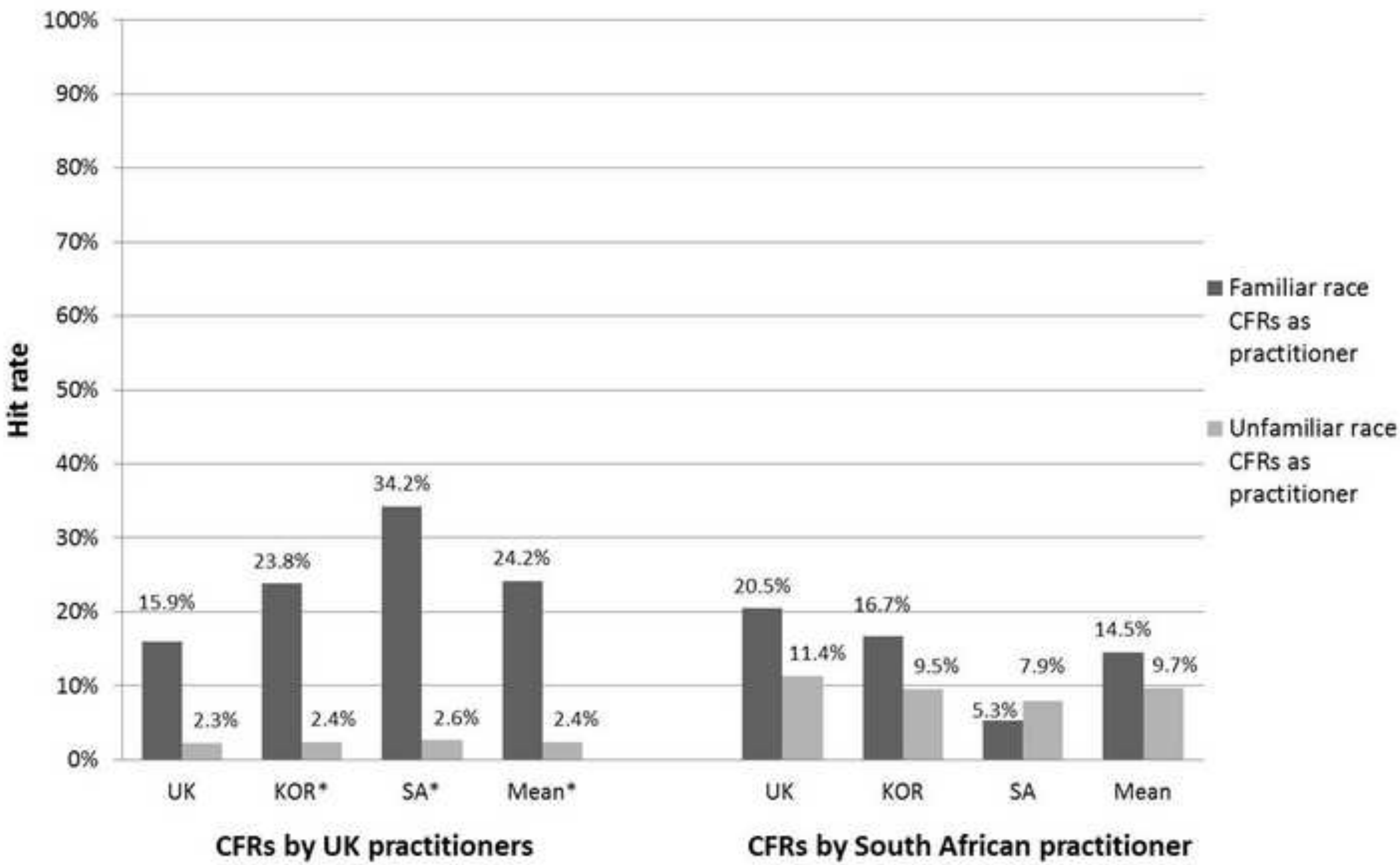
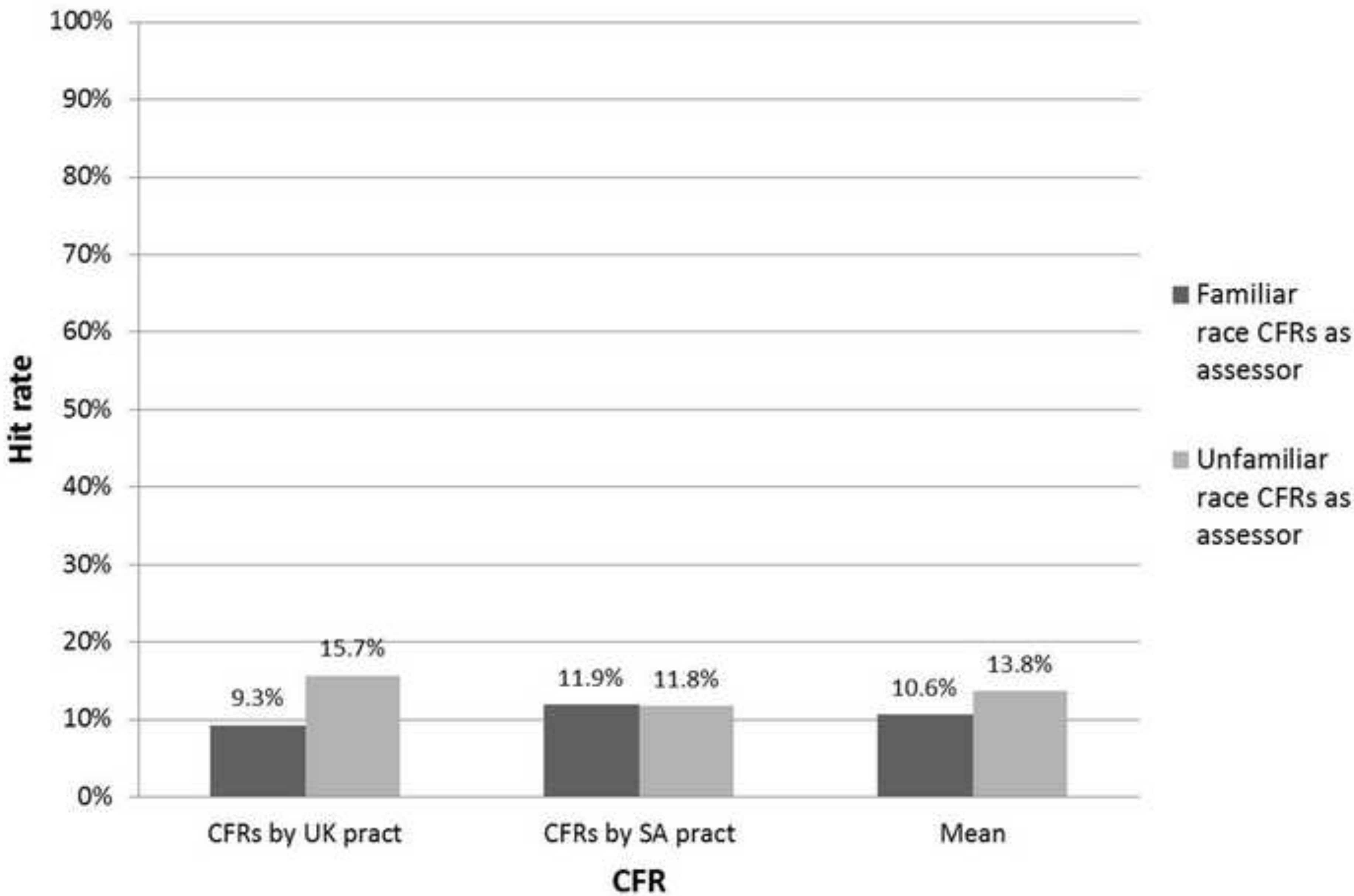


fig 6.jpg
Click here to download high resolution image



Tables

TABLE 1. Biological profiles of the six skulls, status of exchanging skull models between the two teams, applied average facial soft tissue depth data and methods for the CFRs.

Skull	Biological profiles			Skull exchange		Applied tissue depth date	Method for CFR	
	Sex	Ancestry	Age (years)	Provided from	Distributed to		by unfamiliar-race team	by familiar-race team
N1	Female	Negroid	20 - 30	South Africa	South Africa	Cavanagh and Steyn [32]	N/A	3D manual (N1-SA)
N2	Female	Negroid	20 - 30	South Africa	UK South Africa	Cavanagh and Steyn [32]	3D manual (N2-UK)	3D manual (N2-SA)
M1	Female	Mongoloid	20 - 30	China	South Africa	Created from Birkner [85] & Lebedinskaya [4]	3D manual (M1-SA)	N/A
M2	Male	Mongoloid	30 - 40	China	UK	Created from Birkner [85] & Lebedinskaya [4]	3D computerized (M2-UK)	N/A
C1	Male	Caucasoid	20 - 30	UK	South Africa UK	Sahni et al. [34]	3D manual (C1-SA)	3D computerized (C1-UK)
C2	Female	Caucasoid	15 - 21	UK	UK	Helmer [4]	N/A	3D manual (C2-UK)

N/A = Not applicable.

TABLE 2. Overall hit rates for the eight CFRs.

CFR	**Hit rate (%)** Assessor group				**Significance (*p*-value)**
	UK (n = 22)	Korea (n = 21)	South Africa (n = 19)	Total assessors	
N1-SA	22.7	19.0	5.3	***16.1***	**0.015***
N2-UK	4.5	4.8	5.3	***4.8***	
N2-SA	18.2	14.3	5.3	***12.9***	
M1-SA	4.5	19.0	0.0	***8.1***	
M2-UK	0.0	0.0	0.0	***0.0***	
C1-UK	18.2	23.8	31.6	***24.2***	
C1-SA	18.2	0.0	15.8	***11.3***	
C2-UK	13.6	23.8	36.8	***24.2***	
Mean hit rate	***12.5***	***13.1***	***12.5***	**Overall mean hit rate (%)**	***12.7***
Significance (*p*-value)		0.993			

*, p < 0.05.

Response to reviewers:

Reviewer #1: The article provides an interesting point of view on the importance of familiarity in facial reconstruction: it will be worth being published once minor revisions will be corrected.

-highlights missing: please use the conclusive points as highlights

→ The word file for the highlights have been uploaded on the submission system.

-references do not follow guidelines for authors: please correct them

→ The reference style has been changed following the journal guidelines.

-perhaps the description of materials and methods may be reworded in order to render it more readable: it is difficult sometimes to follow the number of CFR and the type of model in the different groups of research.

→ The overall materials and methods were amended as the reviewer's suggestions, especially about the numbering and naming of the skulls and CFRs, for a more readable manuscript.