Published in final edited form as:

Am J Orthod Dentofacial Orthop. 2019 May ; 155(5): 693–701. doi:10.1016/j.ajodo.2018.06.018.

3D Stereophotogrammetry versus Traditional Craniofacial Anthropometry: Comparing Measurements from the 3D Facial Norms Database to Farkas's North American Norms.

Seth M. Weinberg, PhD[1,2,3]

[1]Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA 15219, USA

[2]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261, USA

[3]Department of Anthropology, University of Pittsburgh, Pittsburgh, PA 15261, USA

Abstract

Introduction: Datasets of soft-tissue craniofacial anthropometric norms collected using different methods are available, but there is little understanding of how the measurements compare. Here we compare a set of standard facial measurements between two large datasets: the 3D Facial Norms (3DFN) dataset collected using 3D stereophotogrammetry (N=2454) and the Farkas craniofacial norms collected using direct anthropometry (N=2326).

Methods: A common set of 24 craniofacial linear distances were compared by computing standardized effects sizes (Cohen's d) for each measurement, which describe the overall direction and magnitude of the difference between the two datasets.

Results: Variables with higher mean d-values (suggesting greater discrepancy across datasets) included measurements involving the ear landmark tragion, the landmark nasion, the width of nasolabial structures, the vermilion portion of the lips, and palpebral fissure length. Variables with lower mean d-values included smaller midline measurements involving the lips and lower face and horizontal distance measures between the eyes. Eight measurements showed a significant negative correlation ($p < 0.05$) between Cohen's d and age, indicating greater similarity across the two datasets as age increased.

Conclusions: There are considerable differences between the 3DFN and Farkas norms. In addition to the measurement method, other factors accounting for discrepancies may include secular trends in craniofacial morphology or differences in ethnic composition.

Corresponding Author details: 100 Technology Drive, Suite 500, Pittsburgh, PA 15219, 1-412-648-8928, smwst46@pitt.edu.

DATA AVAILABILITY STATEMENT

The individual-level measurements and raw 3D surface images for all participants in the 3DFN dataset are available through the controlled-access FaceBase repository (https://www.facebase.org/). In addition, genotypic markers for these individuals are available to the research community through the dbGaP controlled-access repository (https://www.ncbi.nlm.nih.gov/gap) at accession number: phs000949. v1.p1. The summary statistics for the Farkas dataset are published and publically available.[23]

INTRODUCTION

Craniofacial soft-tissue norms are typically comprised of a set of standard anthropometric measurements taken on the head and face from a healthy, population-based sample. Similar to cephalometric norms, soft-tissue anthropometric norms are used as an aid to syndrome delineation,[1–3] in pre-surgical planning and surgical outcome assessment,[4,5] and for making quantitative morphological comparisons.[6,7] To be maximally useful, craniofacial norms need to be sex-, age-, and ethnicity-specific. The traditional method used to collect normative craniofacial datasets has been manual anthropometry, using tools such as calipers and tape measurers.[8] This measurement approach, however, has numerous drawbacks. Some of the limitations include a reliance on distances, angles and indices, which can only provide crude measures of complex three-dimensional morphology, an inability to re-measure subjects or derive additional measures not collected at the time of participation, the extensive training required to learn proper anthropometric techniques, the relatively invasive nature of the measurement methods which is often poorly tolerated by young children and those with certain developmental disabilities, and the amount of time it takes to collect a battery of measurements. In addition, traditional normative databases have been limited to summary statistics, while the raw underlying individual-level data is generally not available to the research or clinical community. This fact severely limits the usefulness of the dataset for analysis and comparative purposes.

The 3D Facial Norms (3DFN) database was created in an effort to overcome many of the limitations present in existing craniofacial anthropometric datasets.[9] An image-based repository, 3DFN uses digital stereophotogrammetry to capture quantitative information about the face. Initiated in 2009, 3DFN was created as part of the FaceBase Consortium (https://www.facebase.org/), an NIH-funded effort designed to generate publically available data resources for the scientific community.[10] 3DFN is a crossectional dataset consisting of 2454 unrelated male and female individuals of self-reported European ancestry, ranging in age from 3-40 years. The dataset has an interactive web interface (https://www.facebase.org/facial_norms/) where users can perform custom searches, explore and download summary statistics for a variety of 3D measurements, and calculate Z-scores. Most importantly, however, for every individual in the dataset, the raw 3D facial surface images, derived measurements and demographic descriptors are available to the research and clinical community. In addition, through the dbGaP repository (https://www.ncbi.nlm.nih.gov/gap) full genomic data is now available for every individual in the dataset. Thus, with proper permissions, individual investigators can gain access to raw data from the entire 3DFN dataset. Owing to these features, 3DFN represents a truly unique data repository. To date, 3DFN data has been used in studies of facial dysmorphology[7], in the genomic analysis of human facial traits,[11–12] in anthropological studies,[6,13] and as a testing dataset for the development of novel image analysis methods.[14]

One of the principal rationales behind creating the 3DFN repository has been the increasing use of 3D facial surface imaging (particularly 3D stereophotogrammetry) in clinical and research environments. With the availability of relatively low cost and easy to use 3D camera systems on the market, traditional craniofacial anthropometry is rapidly being replaced. With this movement toward 3D technology, however, appropriate norms based on the new technology were still unavailable. Just as care must be taken to use norms appropriate for factors such as age, sex and ethnicity, so too must the technology used to generate the norms be taken into account. There are several reasons why facial measurements obtained through direct anthropometry may be systematically different from those collected indirectly on 3D images. For instance, direct anthropometry often requires contact with the skin which can deform the pliable tissues of the face during the course of measurement.[15] Even when such contact is not required for a particular measurement, soft-tissue deformation may occur inadvertently. This can disproportionally impact measures on more pliable structures like the nose, lips, and ears. Other measurements which are readily available through direct anthropometry are either difficult or impossible to collect with indirect 3D imaging. For example, regions with landmarks covered by hair interfere with the collection of virtually all cranial vault measurements. Also, some measures involve skeletal landmarks that are extremely difficult to accurately localize on 3D facial surface images, such as gonion or zygion. Thus, each method has its limitations.

Several previous studies have investigated the agreement between direct and indirect (3D image based) facial measurement. When measurements are taken on inanimate mannequin heads and compared across methods, a high degree of agreement (typically sub-millimeter) can be achieved.[16] In contrast, when measurements have been collected on live participants, differences across methods can vary from minimal to substantial.[17–21] This suggests that while direct and indirect image-based methods of facial measurement are capable of achieving good congruence in ideally controlled scenarios, typical data collection involving living participants presents challenges to both methods which can lead to discrepancies. Despite the potential for incongruence between measurement methods, researchers have occasionally analyzed 3D facial morphology using norms based on traditional anthropometry.[22] The indiscriminant use of inappropriate norms could lead to seriously biased results, potentially under- or over-estimating the magnitude of facial differences. However, there is currently no agreement on which facial measurements may be most problematic and subject to this type of method bias.

In the current study, an attempt was made to address this concern by comparing facial measurements from the 3DFN repository to previously published traditional anthropometric norms collected by Leslie Farkas in the 1970s and 1980s. The Farkas dataset[23] is perhaps the most comprehensive and widely used resource for traditional craniofacial anthropometric norms ever collected. The 3DFN and Farkas datasets have a number of measurements in common, both collected from North American White males and females. The age ranges of the two datasets do differ, but there is substantial overlap allowing for age-specific comparisons. Comparing a number of sex- and age-specific measurements across these two datasets will help demonstrate the degree of concordance between two sets of craniofacial norms generated with different techniques. The results of this study will hopefully provide clarity as to which facial measurements, if any, can be safely compared or combined across such datasets.

MATERIALS AND METHODS

The 3DFN dataset is comprised of 2454 unrelated individuals of self-reported European ancestry: 952 males and 1502 females. These participants were recruited from 2010 to 2013 from the general population at four US sites: Pittsburgh, PA; Seattle, WA; Houston, TX; and Iowa City, IA. The full dataset is comprised of males and females ranging in age from 3 to 40 years. As described in detail elsewhere,[9] all participants were screened for any personal or family history of medical conditions that might affect the structure of the craniofacial complex and a personal history of significant craniofacial trauma or surgery. 3D facial surface images were acquired on each participant using 3dMD camera systems and 24 well-defined 3D soft-tissue facial landmarks were collected.[8,23] A set of 29 standard linear distances, corresponding to traditional anthropometric measurements, were then calculated from the 3D landmark coordinates. These distances were then aggregated across demographic categories to produce age- and sex-specific means and standard deviations. Institutional ethics (IRB) approval was obtained and all participants in the 3DFN dataset provided their written informed consent prior to participation.

As described in Weinberg et al. [9], 3DFN facial images were captured following a standardized protocol: “In preparation for facial imaging, participants were asked to remove any jewelry or accessories that could interfere with the capture process. When necessary, the participant’s hair was pinned back to keep it from obscuring the ears and forehead. Selected landmarks were labeled directly on the participant’s face using skin-safe markers (e.g., tragion, gnathion, and pronasale) to facilitate later landmark identification from the resulting 3D surface images. Participants were positioned in front of the imaging system with their head facing forward and tilted slightly back to ensure coverage under the nose and chin. During capture, participants were instructed to keep their eyes open and their lips gently closed, to maintain a neutral facial expression, and to keep their face relaxed. Each capture was inspected on the spot to ensure 3D surface quality; additional captures were obtained as needed.”

The facial landmarking process and subsequent derivation of linear distance measurements was also subjected to extensive quality control. From Weinberg et al. [9] : “To ensure quality and consistency, each evaluator engaged in landmarking completed a three-phase training process prior to working with any 3DFN surfaces. In the first phase, presumptive evaluators were required to familiarize themselves with landmark definitions and identification strategies. In the second phase, evaluators were introduced to the landmarking software environment (3dMDvultus) and asked to identify all 24 landmarks on a test set of 10 different facial surfaces of varying age and sex. An independent expert then reviewed the placement of the landmarks and provided feedback to the evaluator regarding any problems. In the third phase, the evaluator was required to landmark an additional test set of 20 surfaces twice, with at least 48 hours between landmarking sessions. The degree of Intraobserver error was then assessed by comparing the x, y, and z components of each landmark across the two sessions with intraclass correlation coefficients. The threshold for acceptable intraobserver error for each landmark in each of the three principal axes was 0.90. Values below this threshold indicated that additional practice was required, and evaluators could not proceed to working on 3DFN data until they had successfully remediated. After collection, additional quality control measures were put in place to check the resulting landmark data and derived measurements. Using a semiautomated process, the 24 landmark coordinates collected from each 3D facial surface were screened for common errors such as incorrect order and left-right reversals. This was accomplished by visually inspecting the landmark configuration for each subject as a simple wireframe using a locally developed program and subjecting the landmark coordinate data to simple logic rules based on expected spatial patterns. Each of the automatically generated set of 29 inter-landmark distances for each participant was then screened for outliers by calculating sex- and age-appropriate z scores. Any z scores greater than or less than 3.0 were flagged, and the participant's 3D surface was checked manually for errors in landmark placement or potential problems with the participant's age."

The Farkas normative dataset[23,24] is comprised of 2326 individuals of self-reported European ancestry (1096 males and 1230 females), recruited from 1973 until 1986 in the Canadian provinces of Ontario, Alberta and Quebec. The individuals in the Farkas dataset range in age between birth and 25 years. The dataset is comprised of 132 craniofacial measurements all obtained using traditional direct anthropometry. Age- and sex-specific means and standard deviations are available for each measurement in the dataset.[23]

Because there are some key demographic differences between the 3DFN and Farkas datasets, steps were necessary to align both datasets prior to comparison. First, individuals over 25 years of age were excluded from the 3DFN dataset in order to match the upper age limit of the Farkas dataset. Likewise, individuals under the age of three years were excluded from the Farkas dataset in order to match the lower age limit of the 3DFN dataset. Furthermore, the 3DFN dataset originally divided three and four year olds into half year intervals, where as Farkas did not subdivide these ages. To match Farkas, these half years were collapsed in 3DFN into full year intervals. In the Farkas dataset individuals between 19 and 25 were collapsed into a single adult age category, whereas in 3DFN these are discrete. To match Farkas, these ages in the 3DFN dataset were collapsed to make a single 19-25 year old age category. Following these adjustments, each dataset contained both males and females at 17 age intervals: 16 full year intervals spanning between three and 18 years and one additional adult interval that included 19-25 year olds.

The Farkas dataset contains many more measurements than 3DFN. However, all of the 29 linear distance measurements in 3DFN are also contained in the Farkas dataset. For bilateral measurements, only the right side was focused on, effectively reducing the total number of measurements to 24. These 24 common measurements formed the basis for our statistical comparison (Figure 1). All of the linear distance measurements in 3DFN were defined using the same standard facial surface landmarks as Farkas.[23]

Because only published aggregated data (sex- and age-specific means and standard deviations) were available from the Farkas dataset, the types of statistical comparisons we could perform were limited. The primary metric used for comparison was the Cohen's d effect size.[25] Cohen's d is a widely used, unitless standardized measure of effect size for comparing means between two samples, describing the overall direction and magnitude of the difference. The larger the value of d, the greater the difference between samples. By convention, values of d are traditionally broken into categories of effect: very small (< 0.20), small (0.20-0.49), moderate (0.50-0.79), and large (≥ 0.80). d was calculated by comparing means and standard deviations across the two datasets at each age interval and sex separately for all 24 measurements. This resulted in 816 separate d values (17 age intervals × 2 sexes × 24 measurements). To aid interpretation, the individual d values were then averaged across the age intervals and sexes for each measurement, resulting 24 mean d values (with accompanying 95% confidence intervals). The direction and magnitude of these mean d values provides a simple descriptive statistic for determining the overall direction and magnitude of the differences across datasets for each of the 24 measurements. One-sample t-tests were also performed, comparing 3DFN values to the means from the Farkas dataset, which were considered the reference values for testing against.

In addition, a non-parametric Spearman correlation was used to examine the relationship between effect size and age for each of the 24 measurements. For these calculations, males and females were combined and the absolute values of the effect sizes were used. A significant negative correlation would indicate that effect sizes tend to get smaller (less difference between the means of each dataset) as age increases from 3 to 25 years. Statistical tests were performed in SPSS v22. The threshold for statistical significance was set at 0.05.

RESULTS

The observed mean d values varied considerably across the 24 measurements and are shown as a forest plot in Figure 2. For half of the measurements, the mean d value was positive (meaning the measurement was larger in 3DFN than in Farkas). Variables with large and moderate d values – suggesting greater discrepancy across datasets – included those measurements involving the ear landmark tragion (e.g. measures of facial depth), the landmark nasion (e.g., measures of facial height), the width of central nasolabial structures (e.g., labial fissure width and philtrum width), and the vermillion portion of the lips. The variables with the highest mean d values were the length of the palpebral fissure ($d = -1.41$) and nasal height ($d = 1.40$). Variables with small or very small mean d values – suggesting greater concordance across datasets – included smaller midline measurements involving the lips and lower face and horizontal distance measures between the eyes. The variables with the smallest mean d values were lower lip height and philtrum length ($d = -0.02$ for both measures). For these two measurements, the 95% confidence intervals for d included zero; for the other 22 measurements the confidence intervals excluded zero.

The raw d values and t-test results at each age interval by sex are provided for all 24 variables as supplemental files (Tables S1–S24). Not surprisingly, t-test results tended to be significant at most ages for variables with higher mean d values. As expected, the differences observed between datasets were very similar between males and females. Although left side measurements were not focused on here, the results corresponded well to the right side.

Eight measurements showed a significant negative correlation ($p < 0.05$) between d and age, indicating a greater similarity across the two datasets as age increased. These measurements include cranial base width, nasal bridge length, cutaneous lower lip height, upper lip height, upper face depth, nasal height, lower lip height, and nasal protrusion (Table 1). An additional nine measurements also showed negative correlations and seven showed positive correlations, although none of these were statistically significant.

DISCUSSION

The results indicate that, while some facial measurements showed reasonably good concordance between the 3DFN and Farkas normative datasets, many other measurements showed large discrepancies. All measurements, except perhaps five with very small mean effect sizes, showed differences great enough to warrant caution in how these craniofacial norms should be used. Of the 24 measurements compared, half were larger and half were smaller in the 3DFN dataset compared with the Farkas dataset. The types of measurements that were larger in 3DFN tended to be expansive (e.g., cranial base width), although this was not always the case (e.g., philtrum width). A number of measures with more pronounced discrepancies involved landmarks that can be difficult to locate either on 3D surface images or direct anthropometry. Tragion, the point just anterior to the external ear canal at the superior border of the tragus, is often used in measures of facial depth and cranial base width. In 3D surface images, the tragus can be difficult to locate accurately because it is often obscured by hair or may be distorted due to the fact that it is located at the edge of the image viewing field. In direct anthropometry, the tragion may be deformed by the tip of the calipers during measurement which could result in underestimation. Another example of a potentially difficult landmark is nasion, which is the midline point of the frontonasal suture. Its determination on soft-tissue can be a challenge[15,26] and in direct anthropometry typically relies on palpation of the suture. With indirect 3D measurement methods, palpation is not possible, so nasion must be approximated using information gleaned from surrounding structures.[8] Such differences in the localization of landmarks could result in measurement discrepancies. Indeed, nasal height (a measure from nasion to subnasale) was among the most discrepant measures reported, with the 3DFN dataset showing a 4.4mm increase ($d = 1.40$) on average from the Farkas dataset (averaged across ages and sex).

Measurements involving eye landmarks showed conflicting results. Intercanthal width, which involves the endocanthion landmarks, was very similar across datasets. The endocanthion is a measurement that is highly stable and easily to localize both in-person and on 3D images. The exocanthion points at the outer corner of the eyes can be more challenging. Outercanthal width, the distance between the left and right exocanthions, also showed reasonable concordance across datasets, suggesting that the point can be accurately localized using both methods. However, the palpebral fissure length, which is the distance between the endo- and exocanthion landmarks, was the most discrepant measurement observed here. The explanation for this may have to do with the nature of the measurement itself. Individuals are naturally wary of instruments being placed close to the eye – a necessity for direct anthropometry to obtain accurate palpebral fissure length measurements. When the lids are not completely open and relaxed, the proper location of the exocanthion cannot be easily determined and must be estimated.

Likewise, measurements involving lip landmarks were highly variable. Lower lip height and philtrum length were the most concordant measurements we observed. However, philtrum width, labial fissure width, and upper and lower vermillion heights were very discrepant. All of these measures involve landmarks on highly pliable tissues making direct measurement difficult. Labial fissure width, which involves the landmark chelion at the lateral commissure of the mouth, was found to be smaller in 3DFN compared to Farkas. This could be related to soft-tissue deformation or difficulty in determining the landmarks on 3D images. With the lips closed, it can be difficult to determine the exact edges of the labial fissure, particularly in individuals with thinner lips. Philtrum width, while similarly discrepant in magnitude, was found to be larger in 3DFN. This measurement involves the crista philtri landmarks, which can be very difficult to locate depending on how well the margins of the philtrum are defined. As such, this measurement is either being systematically overestimated with 3D photogrammetry, systematically underestimated with direct anthropometry, or a combination of the two.

Beyond technical issues such as difficultly with landmark localization and tissue-deformation during measurement, other potential reasons for some of the observed measurement discrepancies may be related to secular trends or differences in the geographic origin and ethnic composition of two samples. It must be noted that the datasets compared here were collected decades apart, and changes in body composition which can affect facial morphology (e.g. BMI) have occurred over this period.[27] Complicating this overall body size argument, however, is the fact that the discrepancies were not simply biased in one direction (e.g. all larger or smaller in the 3DFN dataset). There have been several recent reports showing changes in specific craniofacial dimensions over similar time periods,[28–30] suggesting that secular changes in craniofacial morphology may be much more nuanced. Unfortunately, we do not have time-separated measurement data from either sample used in the present study to confirm these types of trends. Minor differences in the ethnic composition of the 3DFN and Farkas dataset may also be present. Both datasets included individuals self-described as having European ancestry. Participants for 3DFN were recruited at four sites spread across the US, whereas the Farkas dataset was compiled at three Canadian sites. Some reports have shown that facial differences may be present even across geographically close European countries.[31,32] Detailed descriptions of the specific proportions of ethnic subpopulations for both datasets are not currently available to investigate this possibility.

For 70% (17/24) of the measurements included, the discrepancy between datasets tended to decrease as individuals increased in age. Among those measurements with significant correlations, were several with very large dataset differences (e.g., cranial base width, nasal height). This pattern of inverse correlation may stem from the difficulty collecting both caliper measurements and 3D images on children due to non-compliance. One advantage of 3D imaging, in this regard, is that repeated images can be taken with little effort until one suitable for measurement is obtained.

Our results have practical implications for how measurements from the 3DFN and Farkas datasets – and potentially other datasets similar to them – should be used. A large percentage of the measurements examined here showed discrepancies of great enough magnitude to be a cause for concern. This concern stems from the fact that use of inappropriate norms can lead to biased results and erroneous conclusions. To illustrate this point, consider a hypothetical 18 year old White male patient being evaluated for a series of eye measurements using 3D stereophotogrammetry (as in 3DFN). This patient is determined to have a palpebral fissure length of 29.5mm. Comparing this value against the sex- and age-specific average for this measurement from the 3DFN dataset, our hypothetical patient would be +0.23 standard deviations from the norm. If, however, this value was compared to the sex- and age-specific average from the Farkas dataset, our hypothetical patient would now be −1.30 standard deviations from the norm. Such discrepant outcomes demonstrate why researchers and clinicians need to exercise extreme caution when comparing or combining measures derived from 3D images to those collected with traditional manual anthropometry. Moreover, our results underscore the need for high quality craniofacial norms based on 3D surface imaging, as this technology is rapidly replacing more traditional facial measurement approaches in clinical and research settings. To our knowledge, 3DFN is the only publically available, large-scale 3D craniofacial normative resource of its kind – yet, its demographic limitations are apparent. Similar resources for very young children and other ethnic groups either do not exist, are inaccessible to the broader community, or are limited in the types of measurements and demographic categories they cover. Correcting this deficiency will require a concerted and coordinated effort.

CONCLUSIONS

Of the 24 facial measurements investigated, all but a handful showed meaningful differences between the 3DFN and Farkas normative datasets, with over half showing moderate to large effect sizes ($d \geq 0.50$). These differences were not systematically biased in any direction; half the measurements were larger and the other half smaller in the 3DFN dataset. While many of the differences noted here may be related to the method of measurement (3D image-based indirect versus caliper-based direct anthropometry), other possibilities include secular trends or differences in the ethnic composition of the datasets. These results suggest that caution is warranted when using published craniofacial norms, especially when different measurement methods were used to collect data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Thanks to Zachary D. Raffensperger and Raquel S. Sandoval for their assistance with data collection and to Michael L. Cunningham, Carrie L. Heike, Jacqueline T. Hecht, George L. Wehby, Lina M. Moreno, and Mary L. Marazita for their assistance with the 3DFN project. This work was funded by grants from the National Institute of Dental and Craniofacial Research (U01-DE020078; R01-DE016148) and the Centers for Disease Control and Prevention (R01-DD000295). The funders played no part in the design or execution of the work presented here and the content is solely the responsibility of the author. The author has no conflict of interest to declare.

REFERENCES

1. Allanson JE, O’Hara P, Farkas LG, Nair RC. Anthropometric craniofacial pattern profiles in Down syndrome. Am J Med Genet 1993;47:748–52. [PubMed: 8267006]

2. Ward RE, Jamison PL, Allanson JE. Quantitative approach to identifying abnormal variation in the human face exemplified by a study of 278 individuals with five craniofacial syndromes. Am J Med Genet 2000;91:8–17. [PubMed: 10751082]
3. Kolar JC, Salter EM, Weinberg SM. Preoperative craniofacial dysmorphology in isolated sagittal synostosis: a comprehensive anthropometric evaluation. J Craniofac Surg 2010;21:1404–10. [PubMed: 20856028]
4. Yamada T, Mori Y, Minami K, Mishima K, Tsukamoto Y. Three-dimensional analysis of facial morphology in normal Japanese children as control data for cleft surgery. Cleft Palate Craniofac J 2002;39:517–26. [PubMed: 12190340]
5. Brons S, van Beusichem ME, Bronkhorst EM, Draaisma JM, Bergé SJ, Schols JG, et al. Methods to quantify soft tissue-based cranial growth and treatment outcomes in children: a systematic review. PLoS One 2014;9:e89602. [PubMed: 24586904]
6. Kesterke MJ, Raffensperger ZD, Heike CL, Cunningham ML, Hecht JT, Kau CH, et al. Using the 3D Facial Norms Database to investigate craniofacial sexual dimorphism in healthy children, adolescents, and adults. Biol Sex Differ 2016;7:23. [PubMed: 27110347]
7. Weinberg SM, Leslie EJ, Hecht JT, Wehby JL, Deleyiannis FWB, Moreno LM, et al. Hypertelorism and orofacial clefting revisited: an anthropometric investigation. Cleft Palate Craniofac J 2017;54:631–8. [PubMed: 27505181]
8. Kolar JC, Salter EM. Craniofacial Anthropometry: Practical Measurement of the Head and Face for Clinical, Surgical and Research Use. Springfield: Charles C. Thomas; 1997.
9. Weinberg SM, Raffensperger ZD, Kesterke MJ, Heike CL, Cunningham ML, Hecht JT, et al. The 3D Facial Norms Database: Part 1. A web-based craniofacial anthropometric and image repository for the clinical and research community. Cleft Palate Craniofac J 2016;53:e185–97. [PubMed: 26492185]
10. Hochheiser H, Aronow BJ, Artinger K, Beaty TH, Brinkley JF, Chai Y, et al. The FaceBase Consortium: A comprehensive program to facilitate craniofacial research. Dev Biol 2011;355:175–82. [PubMed: 21458441]
11. Shaffer JR, Orlova E, Lee MK, Leslie EJ, Raffensperger ZD, Heike CL, et al. Genome-wide association study reveals multiple loci influencing normal human facial morphology. PLoS Genet 2016;12:e1006149. [PubMed: 27560520]
12. Lee MK, Shaffer JR, Leslie EJ, Orlova E, Carlson JC, Feingold E, et al. Genome-wide association study of facial morphology reveals novel associations with *FREM1* and *PARK2*. PLoS One 2017;12:e0176566. [PubMed: 28441456]
13. Larson JR, Manyama MF, Cole JB, Gonzalez PN, Percival CJ, Liberton DK, et al. Body size and allometric variation in facial shape in children. Am J Phys Anthropol 2018;165:327–42. [PubMed: 29178597]
14. Liang S, Wu J, Weinberg SM, Shapiro LG. Improved detection of landmarks on 3D human face data. Conf Proc IEEE Eng Med Biol Soc 2013:6482–5. [PubMed: 24111226]
15. Farkas LG. Sources of error in anthropometry and anthroposcopy In: Farkas L, editor. Anthropometry of the Head and Face. Second ed. New York: Raven Press; 1994 p 57–71.
16. Weinberg SM, Naidoo S, Govier DP, Martin RA, Kane AA, Marazita ML. Anthropometric precision and accuracy of digital three-dimensional photogrammetry: comparing the Genex and 3dMD imaging systems to one another and to direct anthropometry. J Craniofac Surg 2006;17:477–83. [PubMed: 16770184]
17. Aung SC, Ngim RCK, Lee ST. Evaluation of the laser surface scanner as a measuring tool and its accuracy compared with direct facial anthropometric measurements. Br J Plast Surg 1995;48:551–8. [PubMed: 8548155]
18. Weinberg SM, Scott NM, Neiswanger K, Brandon CA, Marazita ML. Digital three-dimensional photogrammetry: evaluation of anthropometric precision and accuracy using a Genex 3D camera system. Cleft Palate Craniofac J 2004;41:507–18. [PubMed: 15352857]
19. Sforza C, Dellavia C, Colombo A, Serrao G, Ferrario VF. Nasal dimensions in normal subjects: conventional anthropometry versus computerized anthropometry. Am J Med Genet Part A 2004;130A:228–33. [PubMed: 15378539]

20. Ghoddousi H, Edler R, Haers P, Wertheim D, Greenhill D. Comparison of three methods of facial measurement. Int J Oral Maxillofac Surg 2007;36:250–8. [PubMed: 17113754]
21. Dindaroğlu F, Kutlu P, Duran GS, Görgülü S, Aslan E. 2016 Accuracy and reliability of 3D stereophotogrammetry: A comparison to direct anthropometry and 2D photogrammetry. Angle Orthodont 86(3):487–494. [PubMed: 26267357]
22. Edler R, Abd Rahim M, Wertheim D, Greenhill D. The use of facial anthropometrics in aesthetic assessment. Cleft Palate Craniofac J 2010;47:48–57. [PubMed: 20078203]
23. Farkas LG. Anthropometry of the Head and Face. New York: Raven Press; 1994.
24. Farkas LG, Munro IR. Anthropometric Facial Proportions in Medicine. Springfield: Charles C. Thomas; 1987.
25. Olejnik S, Algina J. Measures of effect size for comparative studies: applications, interpretations, and limitations. Contemp Educ Psychol 2000;25:241–286. [PubMed: 10873373]
26. Ashley-Montagu MF. Location of nasion in the living. Am J Phys Anthropol 1935;20:81–93.
27. Flegal KM. Epidemiologic aspects of overweight and obesity in the United States. Physiol Behav 2005;86:599–602. [PubMed: 16242735]
28. Gyenis G Rapid change of head and face measurements in university students in Hungary. Anthropol Anz 1994;52:149–58. [PubMed: 8067725]
29. Jantz RL, Meadows Jantz L. Secular change in craniofacial morphology. Am J Hum Biol 2000;12:327–38. [PubMed: 11534023]
30. Hossain MG, Saw A, Ohtsuki F, Lestrel PE, Kamarul T. Change in facial shape in two cohorts of Japanese adult female students twenty years apart. Singapore Med J 2011;52:818–23. [PubMed: 22173252]
31. Hajnis K, Farkas LG, Ngim RCK, Lee ST, Venkatadri G. Racial and ethnic morphometric differences in the craniofacial complex In: Farkas L, editor. Anthropometry of the Head and Face. Second ed. New York: Raven Press; 1994 p 201–18.
32. Kau CH, Richmond S, Zhurov A, Ovsenik M, Tawfik W, Borbely P, et al. Use of 3-dimensional surface acquisition to study facial morphology in 5 populations. Am J Orthod Dentofacial Orthop 2010;137(Suppl 4):e51–9.

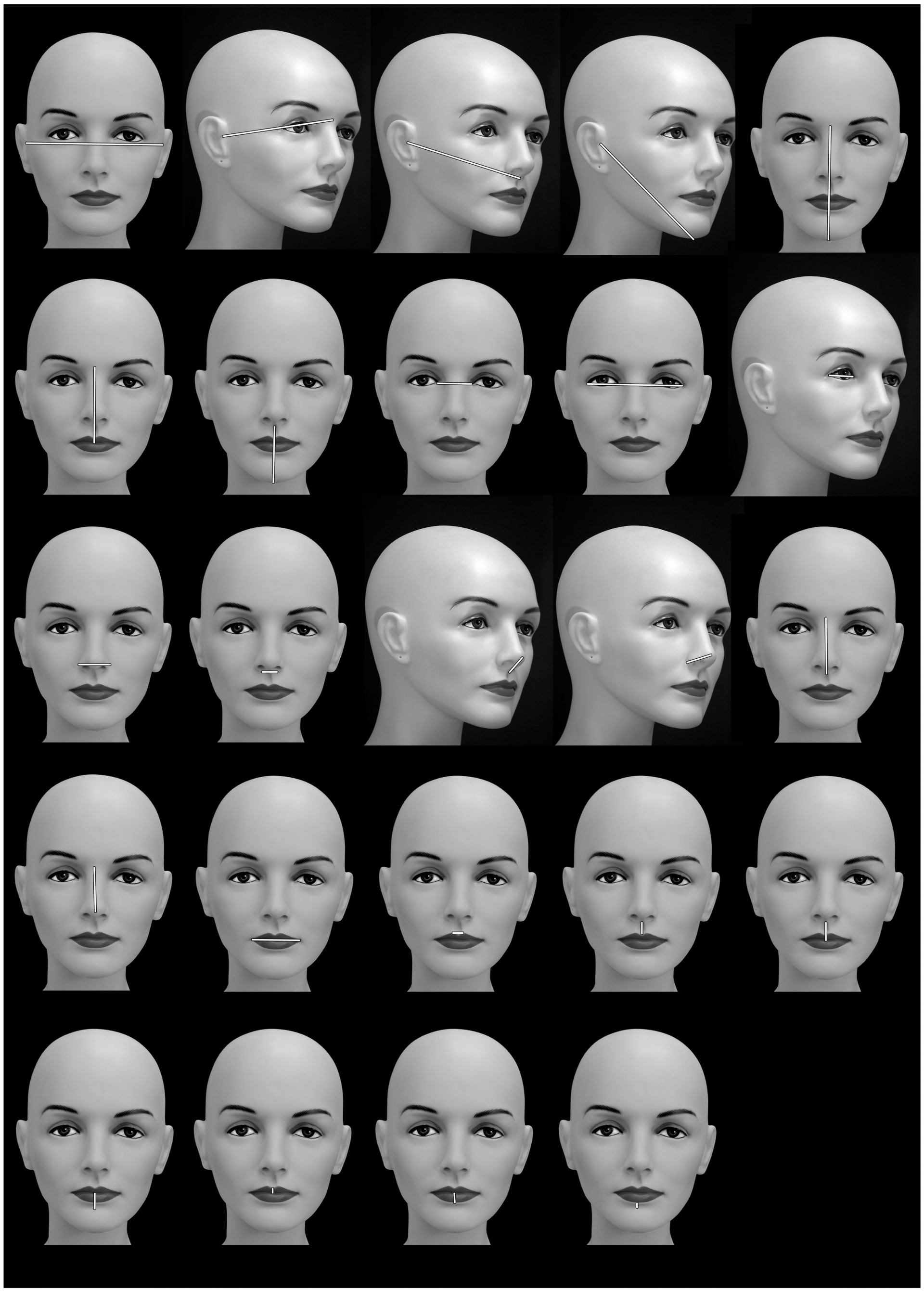

Fig 1.

The 24 measurements used in the current analysis (ordered from top left to bottom right): Cranial base width; Upper facial depth (right); Middle facial depth (right); Lower facial depth (right); Morphological facial height; Upper facial height; Lower facial height; Intercanthal width; Outercanthal width; Palpebral fissure length (right); nasal width; Subnasal width; Nasal protrusion; Nasal ala length (right); Nasal height; Nasal bridge length; Labial fissure width; Philtrum width; Philtrum length; Upper lip height; Lower lip height; Upper vermilion height; Lower vermilion height; Cutaneous lower lip height.

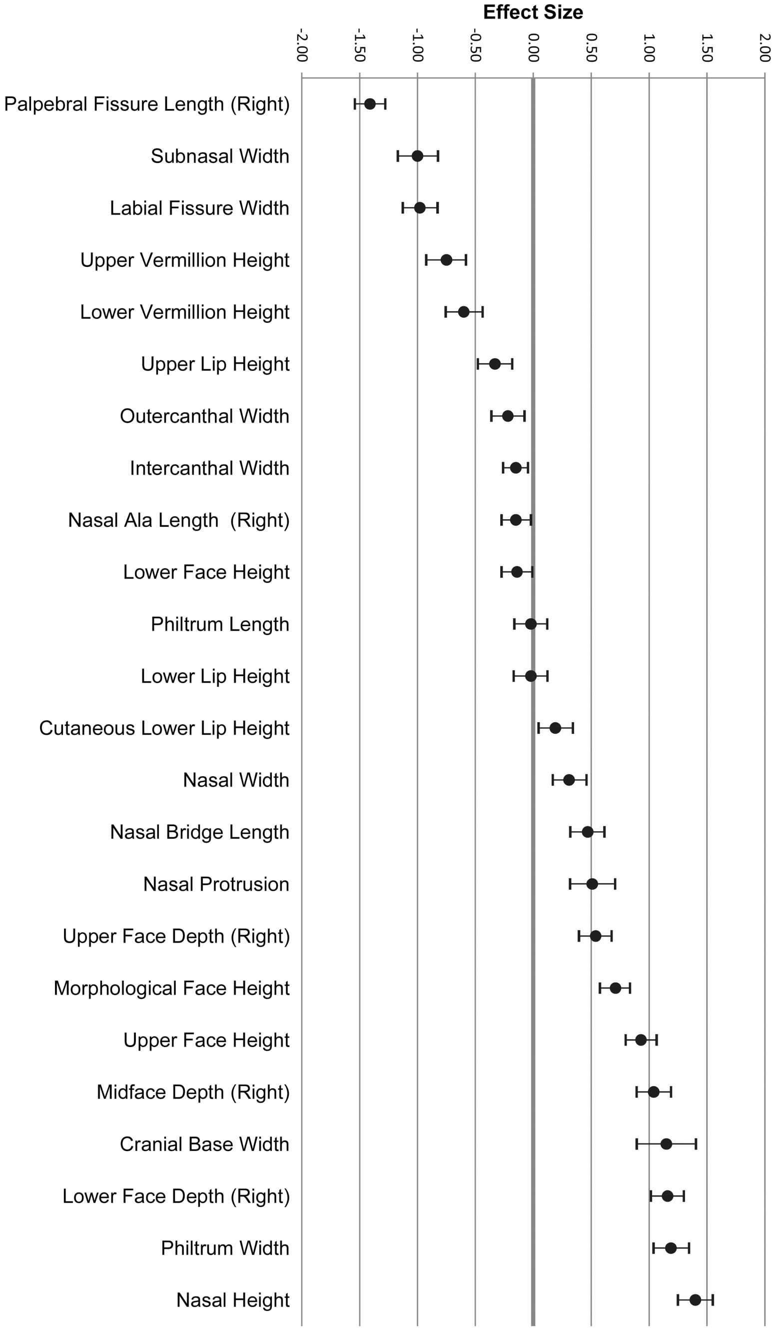

Fig 2.

Forrest plot showing the 24 measurements arranged by mean effect size (d) with associated 95% confidence intervals. For each measurement, the effect size shown here is averaged across the age and sex categories; the raw data used to compute the averaged d values are available in Supplemental Tables S1–S24. Positive d values indicate that a measurement is larger in 3DFN compared to the Farkas norms.

Table 1.

Relationship between effect size (d) and age

Measurement	**Correlation**
Cranial Base Width	−0.823 [***]
Nasal Bridge Length	−0.600 [***]
Cutaneous Lower Lip Height	−0.578 [***]
Upper Lip Height	−0.461 [**]
Upper Face Depth (Right)	−0.436 [**]
Nasal Height	−0.431 [*]
Lower Lip Height	−0.394 [*]
Nasal Protrusion	−0.345 [*]
Lower Face Height	−0.309
Palpebral Fissure Length (Right)	−0.283
Upper Vermillion Height	−0.205
Upper Face Height	−0.194
Lower Vermillion Height	−0.176
Intercanthal Width	−0.095
Nasal Ala Length (Right)	−0.054
Morphological Face Height	−0.046
Outercanthal Width	−0.036
Philtrum Length	0.031
Midface Depth (Right)	0.058
Labial Fissure Width	0.117
Philtrum Width	0.129
Nasal Width	0.227
Lower Face Depth (Right)	0.301
Subnasal Width	0.302

A significant negative correlation indicates that effect size decreases (difference between Farkas and 3DFN gets smaller) as age increases.

[*] Correlation significant at $p < 0.05$;

[**] Correlation significant at $p < 0.01$;

[***] Correlation significant at $p < 0.001$.

Correlations are non-parametric (Spearman method) and use the absolute value of d to aid with interpretation of the coefficients.