Intra- and Inter-Examiner Reliability and Inter-Method Comparison in Physical Anthropometry and Photogrammetry

Nikolay Mollov
Marquette University, nikolay.mollov@marquette.edu

INTRA- AND INTER-EXAMINER RELIABILITY AND INTER-METHOD COMPARISON IN PHYSICAL ANTHROPOMETRY AND PHOTOGRAMMETRY

by

Nikolay D Mollov DDS

A Thesis submitted to the Faculty of the Graduate School,
Marquette University,
in partial fulfillment of the Requirements for
the Degree of Master of Science

Milwaukee, Wisconsin
May 2012

ABSTRACT

INTRA- AND INTER-EXAMINER RELIABILITY AND INTER-METHOD COMPARISON IN PHYSICAL ANTHROPOMETRY AND PHOTOGRAMMETRY

Nikolay D Mollov, DDS

Marquette University, May 2012

Objective: Orthodontic treatment can often effect changes in the facial complex. In order to assess what those changes are, precise and reliable methods for measuring facial structures need to be used. While the techniques used for measuring have become increasingly more sophisticated they have also become more expensive and cumbersome for daily use. This study investigated the reliability of two methods, physical anthropometry and photogrammetry, that were inexpensive and relatively easy to set up.

Materials and Methods: Ten examiners measured a sample comprised of 20 dental students (10 male, 10 female) twice over three weeks. Eighteen measurements were acquired directly using a digital caliper. The 18 measurements were comprised of 20 facial landmarks previously defined by Farkas (1981). In addition, standardized facial photographs were made of the 20 participants, and the examiners were asked to identify the same points. The images were then calibrated and the same facial measurements computed. The intra-class correlation coefficient was used to determine the intra- and inter-examiner reliabilities. The Bland-Altman method was used to compare the two methods.

Results: Anthropometric intra-examiner reliability was very high for all measurements, while inter-examiner reliability exhibited a wide range of values, Overall the reliabilities were higher for easily identifiable landmarks, such as landmarks around the mouth, eyes, the nose, while bony landmarks covered by soft-tissue produced less reliable measurements. With few exceptions, photogrammetric reliability was high for both inter- and intra-examiner reliabilities. The least reliable measurement was the interlabial gap probably due to placement error of the points which was compounded by the small distances measured. The Altman-Bland plots showed large variability around the average difference.

Conclusions: The study found that examiners demonstrate high intra-rater reliability regardless of which method was used. Inter-examiner reliability showed larger variability dependent on the method. When using a caliper the examiner was not as consistent as on a photograph when selecting the facial landmarks. While both methods allow for accurate intra-examiner measurements, this study found that the photogrammetric method had greater inter-examiner agreement. However, large variability was found when comparing the two methods.

ACKNOWLEDGEMENTS

Nikolay D Mollov

I would like to thank my parents for their love, guidance and support throughout my education and especially over the last two years. Without their help this project would not be where it is today.

Just as importantly, I would like to extend my gratitude and respect to Dr Bosio who spent a great deal of time with me discussing, working on and improving this project. His work ethic and dedication are inspiring.

I would like to thank Dr. Bradley, Dr. Liu and Dr. Lobb and the staff in the orthodontic clinic for their help over the last two years.

I thank Mr. Tom Wirtz and Ms. Jessica Pruszynski for the patience they had with me while answering all my questions and helping me correct the mistakes I made. Their knowledge in the fields of computer science and statistics was invaluable for the completion of this project.

Last, but not least, my gratitude also goes out to Dr Hefti for his thoughtful advice and time he spent with me developing the project during its beginning stages. Without his insight, our work would not have been completed the way it did.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

Facial esthetics has long been an area of study for many disciplines in the health care field. In medicine, a series of studies were performed (Farkas and Posnick, 1992) to determine the anthropometric dimensions of the developing head. A large sample of approximately 1600 patients had more than 140 soft tissue parameters measured longitudinally over time. Proportions of the developing head and facial landmarks norms, for different age groups, were developed from that data. In medicine, this data can be used to diagnose and treat a variety of abnormalities. In dentistry, particularly orthodontics, the soft tissue esthetics of the lower face are of great interest, as movement of teeth and bony changes associated with growth and treatment will influence the soft tissue profile.

Orthognathic surgery involving either jaw has profound effects on the facial appearance of an individual. The magnitude of change produced with such procedures, makes proper and highly accurate soft tissue diagnostic tools very important. Arnett has advocated developing proper measurement and analytical techniques (Arnett and Bergman 1993, 1993, Arnett et al 1999) in order to maximize the effects of surgical procedures.

Orthodontic treatment is also capable of affecting changes in the lower third of the face, albeit more subtly in comparison to orthognathic surgery. Obtaining proper occlusion was, for many years, the main goal of orthodontics. However, in contemporary orthodontic treatment planning more emphasis is placed on occlusion *and* facial esthetics. Numerous studies have been carried out (Peck and Peck 1970, Park and Burstone 1986, Ferring et al. 2008) to investigate what makes the face attractive and how teeth and jaw positions/relations are related to facial attractiveness.

The methods used to measure different soft tissue facial landmarks are varied. Over the years, methods such as craniometry, physical anthropometry, cephalometry, photogrammetry, stereophotgrammetry, laser imaging and Cone Beam Computed Tomography (CBCT) imaging have been utilized by the scientific community to assess a variety of facial features.

It is therefore important that the precision of these various methods is known. Many studies have been conducted examining the accuracy, validity and repeatability of the information obtained from these measuring techniques. Of all measurement methods, direct physical anthropometric and photogrammetric are the simplest ones. The advantages of these methods are– the cost is low and the ease with which studies can be designed and carried out. However, significant limitations are found; for example, can the facial landmarks that are to be studied be identified consistently?, is there consistency between investigators? And are they accurate?

CHAPTER II

REVIEW OF THE LITERATURE

Importance of Understanding the Face

Over the last century orthodontics has moved from a science primarily concerned with the alignment and proper occlusion of the dentition, to one that places greater importance on the best possible combination of occlusion and facial esthetics. Presently, understanding facial proportions, esthetics and attractiveness (Ricketts 1982,Edler 2001,Naini et al 2006) is an essential component of orthodontic diagnosis.

Technology has become more prevalent, with more sophisticated methods developed to analyze and quantify what precisely makes the human face attractive. Peck and Peck (1970) compared the lateral and frontal photographs of a number of individuals who were previously “acclaimed” to be “possessing those qualities of facial esthetics which are the most pleasing” to cephalometric measurements. Farkas et al. (1999) and Budai et al. (2003), compared certain cephalometric measurements to anthropometric measurements taken directly from the face to determine if there was any correlation between the two. Results were inconsistent with some measurements showing strong correlation while others were very weak. When looking only at a cephalogram, Arnett attempted to develop soft tissue standards for treatment planning for orthognathic surgery (Arnett and Bergman 1993a, 1993b, Arnett et al. 1999). They placed metallic markers on 46 patients, and different aspects of the patient’s soft tissues were measured using cephalograms. Orthognathic surgery aims to improve not just the occlusal scheme of the patient but also provide for a much more esthetic facial result. As such, a proper understanding/ diagnosis of what has made the face deviate from the norm are essential.

Conventional orthodontics can also alter certain soft tissue landmarks and considerable research is devoted to this area. For example, Park and Burstone (1986) examined the position of the lower incisors and how their position related to the facial soft tissue harmony. Similar to Arnett, cephalograms were used to measure soft tissues structures.

Radiographs are not the only means for measuring facial soft tissues. Ferring and Pancherz (2008) examined the “divine proportions of the growing face” by taking photographs from a pre-set distance and completing the measurements subsequently. The purpose of the study was to understand how the face develops and if there was any proportionality among the different elements. Ferrario et al. (1998, 1999) used three-dimensional facial morphometry to digitally recreate the face. Wireless markers were placed on the face and charge-couple device cameras working in an infrared field were used to detect the soft tissue facial landmarks.

As well as facial form, an analysis of the smile and the position of the incisors are diagnostic parameters orthodontists use in devising a problem list. Sarver (2004) discussed the “macro-esthetics (the four-dimensional facial analysis and treatment planning of the soft-tissue paradigm), the micro-esthetics (four-dimensional smile structure – frontal, oblique, sagittal views) and the mini-esthetics (cosmetic dentistry principles – tooth size, shape, color, applied to orthodontics in order to finish a case). The analysis of these three components was completed via careful examination of the object of interest. This examination was performed both via physical anthropometric techniques, i.e. measuring the length of the lip, the width of the smile, etc as well as photogrammetric techniques – taking several photographs in order to properly assess the dynamics of the static and animated smile.

With the advent of cone-beam CT, digital photography and increasingly more powerful computer systems it has become possible to use all three and study the soft tissues of the face (Maal et al. 2008). Laser scanning and image fusion are examples of other technologies that have been used in order to examine the face. The advent of new technologies has helped us better understand how an individual grows and develops and what are the dynamic/static relationship of his/her facial soft tissues performing a variety of different measurements.

Methodologies for Measuring the Face

Some of the first measurements performed in the head and neck region involved the science of craniometry. Craniometry involves the physical measurement of dry skulls. Such measurements can be traced back to the times of Ancient Greece, but it was not until the 17^{th}-18^{th} centuries when new measurements were developed to allow for the comparison of skulls (Finlay, 1980). Using skull measurements, Camper developed his “facial angle” (Finlay, 1980), which is the intersection of the line connecting the most prominent part of the frontal bone in the area of the glabella to the slight convexity anterior to the upper teeth and the line connecting the lower part of nasal aperture and the center of the external auditory meatus. This angle allowed scientists to distinguish between different ethnicities and viewed as an indicator of intelligence. While this measurement was useful in classifying individuals based on certain anthropologic characteristics it was a very crude tool to study humans. Craniometry was not only used to derive simple classifications of the different facial norms but it also allowed scientists to develop a general idea for how growth has occurred. However, each skull could only be measured at a single time-point in the individual's life and thus did not provide extensive information. (Proffit, 2007)

In order for longitudinal studies to be conducted one has to measure living individuals over a period of time. Anthropometry is the scientific method that allows us to do that. The Merriam-Webster dictionary defines anthropometry as "the study of human body measurements especially on a comparative basis". In particular physical anthropometry is the direct identification of points on the human body and the resulting distances between these landmarks. Some of the first anthropometric techniques utilizing calipers, rulers and tapes were described by Hrdlicka (1920), who is also considered the "father of medical anthropometry" and their applications and improvements are seen to this day in the works of Farkas (1981). Anthropometric measurements were adopted in orthodontics in the early 20th century. It was Milo Hellman who introduced physical anthropometric measurements to the orthodontic field (Hellman 1939) which were augmented and further developed by others (Gosman 1950). The availability of a rigorous protocol for the evaluation of the face and the possible changes effected upon it, allowed orthodontists to better assess different facial structures.

Radiographic cephalometrics, introduced by Broadbent (1931) became prominent in the mid-20th century. It allowed investigators to combine both aspects of craniometry and anthropometry. For instance, precise measurements of individual craniums could be performed over time. In addition, soft tissue contours of the profile of the face could be seen and subsequently used for analysis (Burstone, 1959) (Behrents, 1985)

Two major limitations can be found in cephalometric measurements: primarily, the additive radiation dosage of progressive films and, secondly, it is a two dimensional representation of the three dimensional craniofacial region. Unless long-term follow-up using cephalograms was indicated for the patient this method ethically prevented scientists from conducting longitudinal studies. (Profitt 2007)

Photogrammametry was the next step of the evolution of anthropometric measurements and could be regarded as a subdivision of anthropometry. Its more specific definition concerns the determination of the geometric properties of an object through a photograph. It is a non-invasive, inexpensive and frequently used way of taking pre- and post-operative records to assess the conditions/changes that have occurred (Ettorre et al., 2006). It was not until the 1940s when (Sheldon 1940) released his work on somatotyping that the photograph was used for anthropometric measurements. Photogrammetry quickly became part of orthodontics and has been used for quite some time (Peck and Peck 1970) to assess physical beauty and perform some simple measurements; however recent developments in digital photography and advancements in data storage have rendered photogrammetry more useful. Due to the ease with which photographs can be manipulated and the quick and direct display of the images, photogrammetry has not only become part of initial record taking and analysis, but also has served us in determining ethnical variations (Al-Khatib 2010). Unlike cephalometrics, an object can be measured from multiple angles, i.e. multiple photographs from different angles could be taken, and thus the soft tissue envelope could be studied as desired. However, just as in cephalometrics, photogrammetry is a two-dimensional representation of a three-dimensional object. While there were ways to correct for the error due to the different dimensions, precise measurements were hard to obtain.

Computers have allowed scientists to take anthropometric, and in particular, facial measurements one step further. Several new measurement techniques that pertain to three-dimensional surface imaging have emerged over the last decade or so. A subdivision of photogrammetry, called stereophotogrammetry, has recently become more prominent in the of soft tissue facial research. While this technique is not new per se, it was first discussed by Thalmann-Degan in 1944 as referenced by Burke and Beard (1967), the digital innovations of late have reintroduced it and taken it to a new level. Stereophogrammetry uses several cameras that provide converging views of a given object and reconstruct said object in three dimensions (Hajeer et al, 2004)

Another technique that is non-invasive and available for facial measurements is laser imaging. This technique involves the projections of a known-pattern of laser light (Al-Khatib 2010) onto an object and recreating a three-dimensional digital image from that. Computer tomography (CT) and cone-beam computer tomography (CBCT) (Maal et al 2008) could also be used in medicine and dentistry to record and assess soft tissues. While they do show some promise in quality of the image acquisition, the fact that patients are irradiated with each image acquisition makes them, much like cephalograms, largely unusable for investigating soft tissues in living patients.

Reliability in Research

The term reliability in research is defined as “the degree with which repeated measurements, or measurements taken under identical circumstances will yield the same result.” (Lewis 1999) This definition also assumes that while the measurements are being made, no change is being effected to the measured subject. In essence, reliability looks the “randomness of the measurement process itself.” (Golafshani 2003) Reliability can also be defined as ”the precision or internal consistency of a test, and does not require comparison with an external standard.” (Karras 1997a, Karras 1997b)

Another term in statistics is validity. While reliability looks at the likelihood that a certain measurement will be the same after several repetitions, validity looks at how close the recorded measurement is to the true value. In order for those two terms to be clearly distinguished we need to define the concept of a “gold standard.” The term “valid” in the context of research implies that the parameter of interest is compared to an external variable, the “gold standard” which has a known and universally accepted value (Lewis, 1999). In an attempt to define a “gold standard” for facial body structures Farkas (Farkas et al. 1981, Farkas and Posnick 1992) developed norms for the different gender, ethnicities, ages, and for different structures of head. While the norms show us what the average is, they do not represent the “true” value of a given facial/head structure as the large variation of height, weight, skin texture, thickness, etc among individuals produces measurements that could be significantly outside of the norms. These individuals, however, are not abnormal, because all the different structures of their head/body are, in most instances, proportional. In addition, individuals are growing at different rates, thus making the previously derived norms limited in their application

This, however, renders the reliability measurements of the head and face difficult to accurately assess. Since no gold standard and a large variation between individuals in the size of their facial structures exist, the only way to accurately measure a given facial structure is to correctly identify the bony/soft tissue points that comprise it. In the head and neck, this is often a difficult task as the majority of structures are identified by bony points that are covered by soft tissue, which makes a precise identification difficult.

Any method developed to assess facial structures needs to show high intra-rater and inter-rater reliability so that accurate measurements can be completed. Thus, an accurate identification of the facial points comprising the afore-mentioned structures is necessary.

Reliability of Different Measuring Methods

Craniometry

While craniometry is a highly accurate method (Profitt 2007, Gribel et al 2011) its use in soft tissue measurements is non-existent. Recent advancements in three-dimensional technologies have allowed for measurement of skull landmarks on living patients. Gribel et al (2001) showed that craniometric and CBCT measurements on dry skulls produced highly accurate (the error was within 0.1mm) and repeatable results (the reliability was r=0.99 for the CBCT, and r=0.98 for the craniometric measurements). Craniometry is a highly accurate method for hard tissue structures only.

Cephalometrics

Accuracy of the point identification in cephalometric analyses has been a subject of extensive debate. Kamoen et al. (2001) attempted to identify the source of error in cephalometric digitization/tracing. The study looked at fifty randomly selected cephalograms digitized and afterwards repeatedly traced, by hand and on a computer, by four examiners. No statistically significant errors due to the digitizer nor significant intra- and inter-rater differences were found. Significant differences however were found for both intra- and inter-rater (higher error than inter-rater) reliabilities when landmark identification was performed on the original cephalogram. In addition, it was determined that the highest error came from the landmark recording and the particular landmark was important in determining the magnitude of the error.

A more rigorous study that looked at more parameters was completed by Trpkova et al (1997). They conducted a meta-analysis, which included six articles concerned with the repeatability and reproducibility of points identified on the cephalogram. According to them, the errors seen in cephalometrics are usually due to orientation and geometry. In addition, each landmark has an “envelope of error” – some landmarks are reproducible in a vertical, while others are more reproducible in a horizontal direction. The study found that only several landmarks were identified consistently with minimal error: “menton (Me) posterior nasal spine (PNS) anterior nasal spine (ANS) sella (S) pterygomaxillary fissure (Ptm), point A deepest point on the anterior maxillary margin (A) and point B deepest point on the anterior symphysis region of the mandible(B).”

The literature shows that the reliability between cephalometric films is largely dependent on the particular landmarks that are to be studied. Only seven points consistently showed high reproducibility. In addition to the minimal number of landmarks that are reproducible between investigators, the fact that cephalograms only show the lateral side of the head and even more importantly, they only show the contours of the soft tissue, makes cephalometrics ineffective in studying the soft tissues of the head.

Physical Anthropometry

Physical anthropometry has long been considered the primary way of investigating the morphology of the body and the craniofacial complex. The craniofacial complex has been where some of the most significant research has been done.

In his 1996 review Farkas examined the different aspects of the reliability of anthropometric craniofacial measurements. The data showed that the reliability of anthropometric measurements is dependent on a multitude of factors. Those include: *the ethnic composition* of the race being measured (different ethnicities may have different body height, but rarely different craniofacial structures), the *representative sample* being measured, *the environmental factors* (severity of climate, mean annual temperature, humidity, etc. can all affect the appearance and condition of the soft tissues*), the socioeconomic factors* (for example, it was observed that children of "upper middle-class families are taller than those of the working classes"; in order for the anthropometric data to be valid all socioeconomic levels needs to be represented). In order for a given study to produce sound and reliable results all of the above-given factors should be accounted for so that the studied sample is truly representative of the segment of society being studied.

Anthropometric reliability is usually tested by performing the measurement twice in a relatively short period of time (Hdrlicka 1920). Intra-rater testing can reveal the two important components: the consistency of the surface measurements and how skillful the examiner is (Farkas, 1996). The latter should also be tested by inter-rater reliability tests, because in longitudinal tests spanning several years and often including new examiners, it is important to know that everyone on the investigative team can perform the measurements with a high degree of reliability. Farkas' (1996) in a literature review showed inconsistency in the previous 60 years reports regarding the number of examiners in cross-sectional or longitudinal studies or "interobserver testing of collaborators."

In order for measurements to be accurate, an important aspect is the training of the examiner. Hrdlicka (1920) states that even extensive training is sometimes insufficient for examiner to provide reliable and accurate measurements. He also states that the two main components for good accuracy are "the ability to locate the facial landmark and to have a set of high-quality measuring tools.' Farkas (1996) adds the cooperation of the examinee as a necessary factor for high accuracy.

In his response to the readers, Farkas (1996) tried to distinguish between the two components of reliability – accuracy and precision. In his description, the former pertains to the bias in measurement while the latter is related to the repeatability of a certain measurement. The author suggests that while mistakes are made when the two components are combined, minimizing the possible errors is a way of reducing "unreliable measurements." Accuracy errors are due to examiner's bias and Farkas suggests that can be minimized via training while precision errors are due to the patient and obtaining a large data set can minimize those.

Photogrammetry

Tanner and Weinerr (1949) were amongthe first investigators to perform a study comparing the reliabilities of full-body anthropometric and photogrammetric measurements. They recognized that three errors could affect the reliability of photogrammetric measurements – the measuring of the photograph, the posing of the subject and the differing observers and attempted to account for those when performing the statistical analysis of his study. His findings suggest that there was very little difference between the physical anthropometric measurements and those performed on a photograph. Seventy participants were measured twice by two investigators, with a number of parameters investigated. Therefore the sample size of this study was probably insufficient to give a definitive answer as to the reliability of photogrammetric measurements.

Photogrammetry was used in medicine (Miskin 1959) and orthodontics to evaluate facial attractiveness (Peck and Peck 1970). Further testing was done to determine the reliability of such measurements in the craniofacial region (Farkas 1980). Farkas found of the 106 direct craniofacial measurements only 62 could be reproduced on photographs (due to loss of depth in the photograph). In addition, of those 62 measurements, only 26 showed to be reliable. He deemed the measurements reliable if the photogrammetric results were within 1mm or 2 degrees of the direct measurements. The most reliable measurements were around the mouth and lips.

Recently, Edler et al. (2003) looked at photogrammetry as a way to assess mandibular asymmetry. The authors compared four different types of measurements obtained from facial photographs to those from a Posterior-Anterior cephalogram: area, perimeter of the outlines and shape. While they did not perform measurements similar to Farkas (1980) or Tanner (1949), but rather looked more at volumetric analysis, they found the best repeatability in the photogrammetric method.

Three-dimensional surface imaging

The reliability of several different three-dimensional surface imaging techniques was examined. The majority of the literature supported the fairly high accuracy of measurement of these methods, however they all also had drawbacks that limited their usefulness in anthropometric studies.

Stereophotogrammetry

Investigations into the reliability of stereophotgrammetry date back to the 1960s. Burke and Beard (1967) conducted a study examining the accuracy of this measurement method using two “multiplex” cameras, with the measured subject being oriented the same way as in a cephalostat except the Frankfort plane being vertical. They performed anthropometric measurements and volumetric measurements off of plaster models on the subjects and compared them to stereophotogrammetric measurements. They found very acceptable error levels using this technique.

Stereophotogrammetry has evolved significantly. There are several stereophotogrammetric systems and they all show a high level of reliability in measurement acquisition. The Glasgow University system was used in orthognathic cases (Hajeer et al., 2004) to precisely assess the magnitude of surgical change. Its accuracy has been reported to be even greater than the laser scanning systems and be within 0.5mm. (Ayoub et al., 2003).

Another stereophtogrammetric system is the 3dMD FACE, which uses three different cameras (one color and one infrared) to capture the desired image. The capture time for this system is much shorter (approximately 1.5-2 milliseconds) than in the Glasgow University system, which creates less distortion and is more useful for data capture. The error detected with this system (Weinberg et al. 2006) was found to be extremely low, yielding results of a technical error well under 1 mm, and intraclass correlation coefficients from r=0.98 to r=1 . This system can increase the number of cameras, which could raise the level of detail and improve the already excellent results. Its high accuracy has allowed it to be used in studies looking variations of facial morphology and facial anomalies. While being highly reliable and accurate, the drawbacks of this system are similar to those of the Glasgow system, and in addition the extra number of cameras makes the set-up cost prohibitive for most researchers.

Laser Imaging

Several studies test the accuracy and precision of laser scanning. Kau et al. (2005) determined the reliability of measuring morphology at two time points T1=3 minutes after initial measurement and T2=3 days after the initial measurement using laser scanning and found that 90% of the created images were within an error of 0.85mm. Similar studies show that laser scanning can produce a “noninvasive, accurate, and reproducible means for medical applications” (Hajeer et al. 2002) Despite the fairly high accuracy of laser scanning there are obvious limitation in its usability – the method is expensive, the data acquisition is slow and the patient’s eyes need to be closed and head stabilized.

Cone-Beam Computer Tomography (CBCT)

CBCT imaging can accurately reproduce the identification of soft-tissue facial landmarks (Medelnik et al.2011), (Fourie et al.2011) and facial tissue depth measurement (Fourie et al. 2010).

Image fusion is a technique that allows us to superimpose a 3D photograph on a CBCT image. While there are some errors associated with this method (Maal et al. 2008), it is a promising development of anthropometric measurements in the digital world and the hope is that eventually we will be able to carry out accurate examinations of the soft tissue of the face. However, an anthropometric measurement on a CBCT reproduced image is still hindered by the software’s rendering of the patient’s skin texture, color, facial line angles, light reflection and other factors. In addition, just like in cephalometrics, patients are exposed to radiation every time an image is taken.

Method Comparison

Studies (Ghoddousi et al 2006, Guyot et al 2003, Aksu et al.2010) have been conducted examining the inter-method reliability of the different measurement techniques. While the number is sizable, the heterogeneity of the design of the different studies makes it very difficult to be able to generalize their comparability.

Current State of the Problem

The majority of the methods used to perform facial soft tissue measurements are extremely resource intensive and impractical. Although a variety of measuring methods are available, direct clinical and photographic measurements provides a simpler way of investigating soft tissue facial landmarks. The cost is low, and the methods are simple to implement. The limitations are related to the landmark identification and different investigators measurement acquiring consistency. If the orthodontically produced changes to the facial soft tissue are to be measured before and after orthodontic treatment using direct clinical measurements, the reliability of the investigators needed to be reported.

There were three objectives of this study:

a) To determine the intra-rater and inter-rater reliability of direct facial caliper measurements in a large group of examiners

b) To determine the intra-rater and inter-rater reliability of a simple photogrammetric systems in the same group

c) To compare the two methods

Our working hypothesis was that there would be no difference in the intra- and inter-rater reliabilities of both measuring method, but that there would be a significant difference when the two methods were compared.

CHAPTER III

MATERIALS AND METHODS

This study was approved by the Institutional Review Board at Marquette University. (Protocol # HR-2083)

A total of twenty dental students were recruited and randomly selected to participate as subject of measurement in the project. The group was comprised of 10 male and 10 female students of the Marquette University School of Dentistry (MUSoD). The exclusion criteria for participants were:

a. Congenital facial abnormalities
b. Having any medical/pharmacological treatment that could produce distortion of normal facial landmarks.
c. Age was not considered as an exclusion criteria

Ten examiners were selected from the postgraduate orthodontic program and from the undergraduate dental students at the MUSoD. One examiner was a full time faculty member (JB). The examiner population was comprised of 5 females and 5 males. Due to the number of examiners standard calibration was not feasible. Instead, the examiners were provided with a detailed write-up (Addendum A) and a Power-Point (Microsoft, Seattle, WA) presentation on how to identify the facial landmarks (Fig 1, Fig 2; note original figures were published in Arnett and McLaughln 2004; permission obtained to use and distribute figures is attached in Addendum B). The examiners practiced identifying the facial landmark points and took measurements on each other until they felt comfortable with performing the measurements on the study participants.

The points used in the study were described by Farkas (1981). However, left (L) and right (R) were identified for the appropriate landmarks in order to make it easier for the examiners to identify the different points. The points used in the study are described in Addendum C.

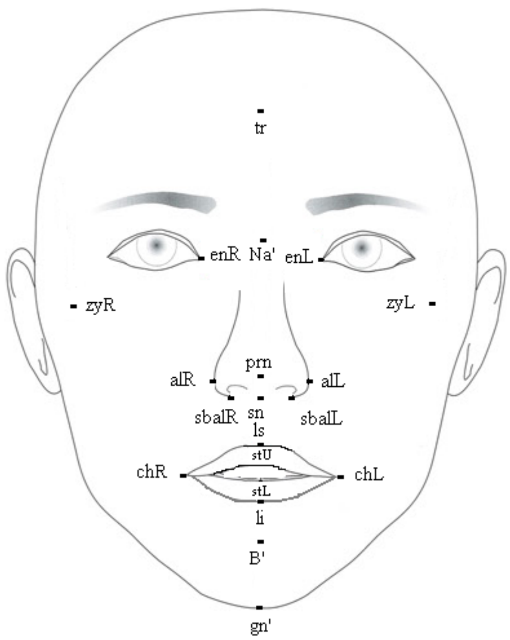


Figure 1. Points identified on the face by the examiners.

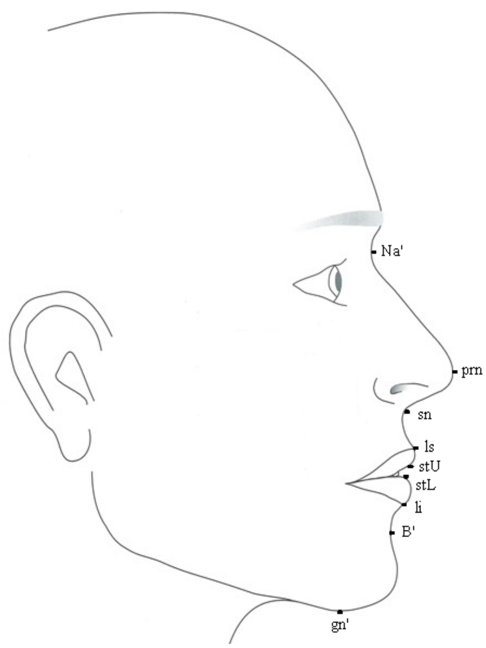


Figure 2. Profile of the points identified by the examiners

The facial measurements were taken in the graduate orthodontic clinic using an 8mm sliding digital Mitutoyo calipers (Aurora, IL). The measurement error for all Mitutoyo calipers (0.1 mm precision rate) was identical per the company's description. The examiners were paired in teams – one examiner recorded the measurements while the other recorded the data. The participants were seated in the dental chair with their head relaxed and in an upright position. In order to establish a repeatable position of the mandible, the study participants were guided into mandibular rest position and asked to remain with their lips relaxed. The examinees were sitting upright in the chairs while the different measurements were being taken. (Fig 3) The measurements were recorded in the standard form for all participant subjects (Addendum D). The study participants were recalled approximately a month later and the whole procedure was repeated.

Figure 3. Examples of the different measurements taken. In clockwise direction: Na'-sn, alR-alL, zyR-zyL, sbalR-sbalL, prn-ls, chR-chL

Each participant had photographs taken at the beginning of the project. The examinees were placed in the cephalometric machine and digital photographs were taken five feet away using a Nikon D40 camera with a 60mm, 1/ 2.8f lens. The images were calibrated using an object of known size – for that purpose a circle of diameter 20 mm was taped to one of the earholders of the cephalometric machine. (Fig 4) The examinees were positioned so that their Frankfort horizontal was approximately paralell to the floor and the camera was placed at approximately the level of the examinee's Frankfort horizontal plane. The examinees were asked to pronounce the word "Emma" and relax their lips in order to mimic the mandibular rest position attained during the caliper measurements.

Figure 4. Examinee in cephalostat; calibration device is on the left side of the examinee.

The digital photographs were reviewed by the investigators and anywhere facial muscle strain was evident were eliminated. One photograph was randomly selected from a pool of photographs for each examinee and uploaded in a software package designed to assist in obtaining the digitized landmarks. (Fig 5) The test performed consisted of two separate trial runs. The order of the photographs in each trial was randomized. Each examiner was trained how to use the software on a sample photograph of the lead investigator. The examiners were then allowed to carry out the point identification at their own discretion with the one condition being to allow at least a few days between the two trials. The points were labeled one through twenty and a legend was given to each examiner so that they knew what number corresponds to the specific point to be digitized.

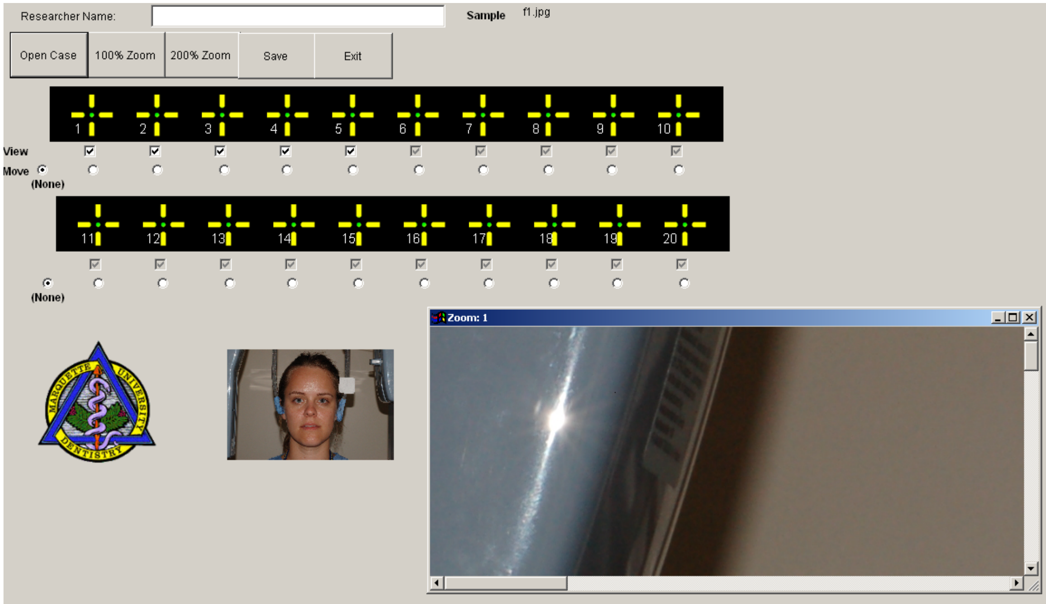


Figure 5. The software package for digitizing the proper landmarks. The different points were identified by a number (1-20) and the examiners were assigned a legend. The software allowed the examiners to zoom in, as well as save their work without having completed the particular measurement.

Statistical Analysis

The intra-class correlation coefficient, and in particular the Shrout-Fleiss (Shrout and Fleiss, 1979) method, was used to determine both the intra- and inter-investigator reliabilities. This correlation coefficient is a general measure of agreement between two or more raters. The Bland-Altman method was used to compare the two methods.

CHAPTER IV

RESULTS

Physical Anthropometric Measurements

The reliability coefficients for the 18 facial soft tissue measurements and the intra-examiner and inter-examiner differences with a 95% confidence interval are shown in Table 1. The first five measurements were considered horizontal, whereas the last 13 were considered vertical.

Intra-examiner differences

All 10 examiners showed consistently high intra-examiner reliability between T1 and T2. None of the calculated reliabilities fell below R=0.934. The least reliable measurements were nasal width at base of the nose, soft tissue B point to gnathion and mouth height. Even for those 3 measurements, the average reliabilities varied betweenR=0.934 to R=0.943. The 18 measurements exhibited very high reliabilities with nasal width (al-al, R=0.992), middle third of the face (Na'- sn, R= 0.989), and upper lip length (sn – ls, R=0.992) showing the highest reliabilities.

Inter-examiner differences

When comparing the measurements among the 10 examiners, a larger reliability distribution was found. The reliabilities for the 18 measurements can be placed in three distinct groups. Group one is made up of a few measurements showing consistently high reliabilities. Those include alR-alL (r=0.922) and sn-ls (r=0.926). As noted before, those same two measurements also showed very high intra-examiner reliability.

Significant reliability measurements differences are seen in the second group with a larger number of measurements showing poor reliability. Most notable are nasal width at base of nose (r=0.590), mouth height (r=0.585) and B' – gn (r=0.623). The first two measurements also showed the lowest intra-examiner reliability.

Most of the remaining measurements can be placed in group three which showed reliabilities that fell somewhere in between the extremes with mouth width (chR-chL,r=0.863), the third of the face (tr – Na', r=0.827; Na' – sn, r=0.899; sn – gn',r=0.867), measurements around the mouth (stL – li, r = 0.865; stU-stL=0.882) being the most consistent. Measurements between the left and right commissures differed greatly (sn – chL, r = 0.758; sn – chR, r = 0.837).

No significant differences were found between horizontal and vertical measurements. Both categories feature some reliable and some unreliable measurements.

Measurement	Caliper Interexaminer Reliability	Caliper Intraexaminer Reliability
Zygomatic Width (zyR-zyL)	0.696(0.55-0.837)	0.958(0.924-0.981)
Mouth Width(chR-chL)	0.863(0.774-0.932)	0.984(0.972-0.993)
Nasal Width at Widest Nostrils(alR-alL)	0.922(0.866-0.963)	0.992(0.985-0.996)
Nasal Width at Base of Nose(sbalR-sbalL)	0.590(0.428-0.765)	0.935(0.882-0.970)
Intraorbital Width(enL-enR)	0.775(0.65-0.884)	0.972(0.949-0.987)
Hairline – Nasion(tr-Na’)	0.827(0.723-0.914)	0.980(0.963-0.991)
Nasion – SubNasale(Na’-sn)	0.899(0.83-0.951)	0.989(0.98-0.995)
SubNasale – Gnathion(sn-gn’)	0.867(0.78-0.935)	0.985(0.973-0.993)
Nasion - Tip of Nose(Na’-prn)	0.763(0.635-0.877)	0.97(0.946-0.986)
Stomion Lower – Soft Tissue B Point(stL-B’)	0.706(0.562-0.843)	0.96(0.928-0.982)
Soft Tissue B point – Gnathion(B’-gn’)	0.623(0.465-0.788)	0.943(0.897-0.974)
SubNasale – Right Commissure(sn-chR)	0.837(0.736-0.919)	0.981(0.965-0.991)
SubNasale – Left Commissure(sn-chL)	0.758(0.628-0.874)	0.969(0.944-0.986)
Tip of Nose – Upper Lip(prn-ls)	0.850(0.755-0.926)	0.983(0.969-0.992)
Mouth Height(ls-li)	0.585(0.423-0.762)	0.934(0.88-0.97)
SubNasale to Upper Lip(sn-ls)	0.926(0.872-0.965)	0.992(0.986-0.996)
Lower Lip Thickness(stL-li)	0.865(0.778-0.934)	0.985(0.972-0.993)
Interlabial Gap(stU-stL)	0.882(0.803-0.942)	0.987(0.976-0.994)

Table 1. Reliabilities of physical anthropometric (caliper) measurement

Photogrammetric Measurements

The majority of photogrammetric reliabilities exhibited extremely high values. The examiners had virtually an unlimited amount of time to complete the project. They were told to inform the investigator when they had completed the point identification and returned the data sets to the investigators.

It is worth noting that the initial data contained errors, which yielded very low reliabilities. Upon review, it was found that numerous points had not been placed on the photograph. The problem occurred predominantly with zy L/R and stU/L. The investigators were unable to ascertain if the problem was due to faulty software or because the examiners had forgotten to place the points.

Just like with the caliper measurements, the intra-examiner and inter-examiner reliability coefficients for the same measurements with a 95% confidence interval are shown in Table 2.

Intra-Examiner Differences

The majority of the measurements exhibited reliability values higher than r=0.99. The most reliable values were nasal width (alR-alL), length of nose measured to the tip (Na’-prn), certain measurements around the mouth (prn-chR, ls-li, sn-ls) with all of them exhibiting reliabilities r=0.999. While all measurements exhibited very reliable results, the ones that showed the lowest values were interlabial gab (stU-stL, r=0.952), and soft tissue measurements in the lower third of the face (stL-B’, r=0.981; B’-gn’,r=0.987).

Inter-Examiner Differences

Unlike the caliper measurements, the reliabilities for almost all inter-examiner measurements were high. The highest ones paralleled the highest intra-examiner reliabilities (alR-alL, Na'-prn, prn-chR, ls-li, sn-ls), albeit with an insignificantly lower level of reliability.

There was a slightly larger drop-off in terms of the least reliable measurements however the least reliable measurements mirror exactly the least reliable intra-examiner measurements. The overall reliability however, drops off to an r=0.832-0.882 range.

Measurement	Computer Inter-examiner Reliability	Computer Intra-examiner Reliability
Zygomatic Width (zyR-zyL)	0.919 (0.862-0.962)	0.991(0.984-0.996)
Mouth Width (chR-chL)	0.844 (0.746-0.922)	0.982 (0.967-0.992)
Nasal Width at Widest Nostrils (alR-alL)	0.997 (0.995-0.999)	0.999 (0.999-1)
Nasal Width at Base of Nose (sbalR-sbalL)	0.865 (0.777-0.934)	0.985 (0.972-0.993)
Intraorbital Width (enL-enR)	0.987 (0.976-0.994)	0.999 (0.998-1)
Hairline – Nasion (tr-Na’)	0.926 (0.873-0.965)	0.992 (0.986-0.996)
Nasion – SubNasale (Na’-sn)	0.936 (0.889-0.970)	0.993 (0.988-0.997)
SubNasale – Gnathion (sn-gn’)	0.925 (0.871-0.964)	0.992 (0.985-0.996)
Nasion - Tip of Nose (Na’-prn)	0.992 (0.985-0.996)	0.999 (0.998-1)
Stomion Lower – Soft Tissue B Point (stL-B’)	0.835 (0.734-0.918)	0.981 (0.965-0.991)
Soft Tissue B point – Gnathion (B’-gn’)	0.882 (0.804-0.943)	0.987 (0.976-0.994)
SubNasale – Right Commissure (sn-chR)	0.995(0.991-0.998)	0.999 (0.998-1)
SubNasale – Left Commissure (sn-chL)	0.981 (0.966-0.991)	0.998 (0.997-0.999)
Tip of Nose – Upper Lip (prn-ls)	0.999 (0.998-1)	0.999 (0.999-1)
Mouth Height (ls-li)	0.999 (0.999-1)	0.999 (0.999-1)
SubNasale to Upper Lip (sn-ls)	0.987 (0.977-0.994)	0.999 (0.998-0.999)
Lower Lip Thickness (stL-li)	0.966 (0.940-0.984)	0.997 (0.994-0.998)
Interlabial Gap (stU-stL)	0.663 (0.511-0.816)	0.952 (0.912-0.978)

Table 2. Reliabilities of physical photogrammetric measurements

Inter-Method Differences

The summary of the method comparison is given in Table 3. Addendum E shows the scatter plots of the Bland-Analysis of the two methods.

The average difference between the two methods for each measurement was minimal for the majority of measurements. Only three measurements showed large variability (*sn – chL, prn-ls and ls-li*) in the average value of the difference. While most of the averages seem to be close to a particular range, that range appears to be larger than desired so that these results have no clinical value.

Measurement	Average Difference Between the Methods (Caliper – Computer)	95% Confidence Interval of the Difference
Zygomatic Width	-6.38	(-21.00, 8.24)
Mouth Width	-1.16	(-7.29, 4.98)
Nasal Width at Widest Nostrils	-3.49	(-8.04, 1.05)
Nasal Width at Base of Nose	-0.74	(-4.33, 2.84)
Intraorbital Width	-3.71	(-7.81, 0.38)
Hairline – Nasion	-3.18	(-13.97, 7.62)
Nasion – SubNasale	5.30	(-3.93, 14.53)
SubNasale – Gnathion	-1.81	(-11.39, 7.78)
Nasion – Tip of Nose	9.23	(-4.92, 23.38)
Stomion Lower – Soft Tissue B Point	0.59	(-3.28, 4.45)
Soft Tissue B Point – Gnathion	-2.32	(-8.69, 4.05)
SubNasale – Right Commissure	7.16	(-36.00, 50.32)
SubNasale – Left Commissure	-37.25	(-77.69, 3.18)
Tip of Nose – Upper Lip	-38.20	(-81.82, 5.40)
Mouth Height	-40.79	(-86.14, 4.57)
SubNasale to Upper Lip	-0.28	(-4.55, 3.99)
Lower Lip Thickness	-0.34	(-2.51, 1.84)
Interlabial Gap	-0.91	(-3.91, 2.09)

Table 3. Average difference of the clinical and photogrammetric measurements with a 95% confidence interval.

CHAPTER V

DISCUSSION

In order to evaluate the changes in the soft tissue contour before and after orthodontic treatment within a large sample, a strong reliability test is necessary. This study was designed to evaluate the reliability of soft tissue measurements performed on a sample of dental student volunteers.

The time it took to acquire the measurements was not recorded. However, we made a general observation where most of the 18 measurements were collected in less than 4 minutes. If we were to only acquire the reliable measurements in future studies, this time can be greatly reduced.

Some particular measurements were different from those performed in previous studies. However, the majority of facial landmarks used in the study (Addendum C) were developed, similarly to the points used by Farkas(1981). The one exception was stomion upper (stU) and stomion lower (stL). In Farkas' description, stomion was a point described by the intersection of the facial midline with the "horizontal labial fissure of the gently closed lips." In our study, the participants were requested to relax their mandible and, consequently, their lips were also relaxed, which often resulted in an interlabial gap. Thus, the lower-most point of the upper lip and the upper-most point of the lower lip (both crossing the imaginary facial midline) were defined as stomion upper and stomion lower.

Burstone (1959) used cephalometric headplates in lieu of measurements taken live. He believed that those measurements would diminish accuracy associated with soft-tissue flexibility. He also stated that the time factor was relevant, since the operator could not be as leisurely with that method and the patient could not be expected to hold a given pose for a long period of time. However, transverse measurements are not seen on those cephalometric headplates and tracing errors would also have to be investigated.

Farkas(1981) identified three particular sources of error – improper measuring technique, problems with the measuring instruments and improper identification of the facial landmarks. We attempted to eliminate the first two by training all ten examiners well prior to the study and by having the ten examiners use the exact same caliper model, as well as measuring the sample on the same day and in the same clinical setting. Thus, the only variable that could produce error among the different examiners was the facial landmark identification.

The examiners in this study exhibited high intra-investigator reliability for essentially all measurements. The findings of this study agree with previous studies where a minimal number of investigators were used. Shaner et al. (1998) used two examiners to measure similar anthropometric facial measurements and found the majority of the measurements were in good agreement. Farkas (1981) also found minimal differences in measurements when looking at one examiner over different time points. The present findings showed that the examiners consistently pick the same points.

However, without a gold standard for identifying some of these points, those overlying a bony structure (zygion, gnathion) or those that require several different angles for precise identification (pronasale), the precise determination of the points becomes difficult. Thus, while we can say that the examiners consistently picked the same point we cannot state with certainty if those points were the correct ones or if they were what the examiner believed was the correct point.

Inter-examiner reliability showed a much larger variation. This was confirmed by previous studies. Mommaerts et al (2008) investigated several distances similar to those measured in this project and found the majority of those to be highly unreliable. The measurements that showed the highest reliability involved points that were very easy to identify - in his study, the pupils in the interpupillary distance measurement, supraorbitale, gnathion. The distance between the two zygomatic points (right and left) was found not to be reliable similar to the results of this study.

Geerts et al. (2004) attempted to evaluate the reliability of measuring the vertical dimension of rest by essentially measuring the distance between pronasale and gnathion with a caliper. They used an examiner sample of N=20 (1 patient, 1 measurement, 10 times) and found good inter-examiner reliability for those two points. This was confirmed in another study (Sakar et al. 2011) that attempted to evaluate the measurement of the vertical dimension of rest using pronasale and an additional point on the chin. In our particular study, the measurements involving pronasale fell in the second group – while the reliability was acceptable, it was not ideal. This again was dependent on the points that comprise the particular measurement – those that involve clearly identifiable points produce, as expected, a more reliable measurement. The least reliable measurement was the mouth height (ls-li). This result was possibly generated due to subject difficulties to maintain their lips relaxed during measurement acquiring.

Unlike Shaner's(1998), this study did not attempt to mark the landmarks on the participant subject faces for two reasons. First, we wanted to allow all the examiners to identify the points themselves, and second, we did not want to spend an excessive amount of time acquiring the measurements. Landmark identification relationship between different examiners needed to be proven strong, as well as how successfully could these examiners reproduce that landmark identification from T1 to T2.

Lastly, while observing the reliability between the set of horizontal versus the set of vertical measurements no differences were found. Both groups had some very reliable measurements and some very poor ones. This was probably due to the reliabilities being dependent on how easy it was to define the facial landmarks as opposed to how the measuring device was being held.

Photogrammetry is being used in a variety of different fields and its reliability has being looked at for different purposes. For example, Naylor et al. (2011) used photogrammetry in order to perform goniometric measurements and determine knee range motion. Photogrammetry proved to be highly reliable yielding intra-rater and inter-rater reliabilities higher than r=0.9. Tanner and Weinerr (1949) also looked at how photogrammetric measurements compared to direct measurements of the "living body" and found the measurements to be as reliable as anthropometric measurements. He found that 2/3 of the errors in photogrammetric measurements came from "posing differences, measuring error [accounted for] one-fifth and observer differences the rest.” Thus, the author concluded that photogrammetry is very reliable and the errors occur predominantly because of external factors and not the method itself. In particular, Tanner (1949) stated that the *dimension* measured was the most important determinant in the reliability of that particular measurement.

While the anthropometric measurements were taken from essentially two views: profile and frontal views, an attempt was made with the photogrammetric measurements to simply use a frontal view as Farkas (1980) reported that measurements taken from a photograph of landmarks around the eyes and lips are more accurate from frontal photographs. The majority of the measurements of interest are located around those areas. The investigators wanted to assess the reliability of certain photogrammetric measurements just from a frontal photograph and thus decide to record the lateral anthropometric measurements on the frontal photograph.

Farkas (1980) found that the most reliable photogrammetric measurements were found around the lips and mouth. He stated that the reason for that are the clearly defined facial landmarks. In addition, in his study, a large number of reliable measurements were inclinations. This agrees somewhat with the results of our study. The highest reliabilities for both the intra- and inter-examiner measurements were around the mouth, but they did not necessarily involve any of the landmarks associated with it. For example, alR-alL, Na’-prn, sn-chR were reliable for both the intra- and the inter-examiner comparisons and only one of them involved a landmark associated with the mouth (chR). Our study indicated that measurements involving pronasale yielded highly reliable photogrammetric results. In addition, only two of the measurements performed here were inclinations (sn-chR/L) and they also showed strong reliability. Farkas' study was conducted over 40 years ago, when digital photography was non-existent. The recent ability to manipulate a digital photograph allows the examiner to study the subject in great detail and more accurately pick the desired landmark. Landmarks around the mouth are still easy to identify, leading to high reliabilities, but now we can also select other landmarks fairly easily which yield an increase in the reliabilities.

It is interesting to note that while the measurements around the mouth in this study are fairly reliable, interlabial gab has the lowest reliability for both intra- and inter-examiner measurements. This could be attributed to the size of the points placed on the photograph and the distance between stU and stL. While the examiners had the ability to zoom in and place the point precisely where they thought it should go, there was no way of determining how many of them did exactly that. In addition, because the two points were so close any imprecision in placing them on the photograph could lead to large distortions.

Franke-Gromberg et al. (2010) looked at the validity of photogrammetric measurements and measurements taken directly off the face. The authors found that direct facial anthropometric measurements appeared to be approximately 7.6% shorter than the measurements obtained with the photograph. The authors believe that both methods are valid but do not assess how reliable they are among examiners.

The Altman-Bland inter-method comparison plots show that while the majority of the measurements are located around an average for each measurement, the 95% confidence interval is almost always too large for us to be able to compare these measurements effectively. The reason for this large variability probably stems from two sources – one, the magnification factor was not correct due to the loss of depth in the photograph and two, photographic measurements, even if accurately calibrated, only represent the distance on a two-dimensional plane. Caliper measurements, made directly on the face include all in three dimensions. While the latter may not have contributed much in the overall large variability of the results, it is something that cannot be ignored when comparing measurements like this.

Similarly to this study, Aksu et al (2010) compared direct caliper measurements to photogrammetric measurements. Unlike this study, however, the photographs were calibrated using five different reference distances on the head. The distance had been measured with a caliper after which it was measured on the photograph. A ratio was developed between the two numbers and that was used to develop as magnification factor for the other measurements of interest. The benefits of this method are that that the tool used to calibrate the image is on the face and thus, there is no loss of depth when calibrating the image. The authors found that only three reference lines were reliable and they were only reliable for a total of three measurements. Thus, their calibration method was not very effective in standardizing direct and photogrammetric measurements.

Other studies (Ghoddousi et al 2006 and Guyot et al 2003) found a higher level of agreement between the two methods. Ghoddousi et al attempted a calibration method that combined the previous two – they placed a 2x2cm square on the cheeks and forehead of the examinee. While this method appears to yield better results it also covers areas of the face that are of interest to the researcher.

CHAPTER VI

CONCLUSIONS

Based on the results of this study intra-examiner reliability of facial soft-tissue measurements tends to be very high for either one of the two methods used. While photogrammetric reliability is very high, physical anthropometric reliability exhibits sufficiently high reliability as well. There are a few exceptions, where the reliability is low, and they usually occur when the landmark in question is defined by an underlying bony structure. Landmarks and measurements associated with the mouth tend to show higher reliabilities. We can thus accept our hypothesis that intra-rater reliability is that same for either one of the two methods.

Inter-examiner reliability was, as expected, overall lower than intra-examiner and was similarly distributed. Landmarks associated with structures like the zygomatic prominence, soft tissue B point, gnathion showed lower reliabilities than structures associated with clearly definable soft-tissue landmarks located around the mouth and nose. In addition, photogrammetric reliability was again higher than the clinical measurements. That data obtained in this study, suggests that while we can accept the hypothesis that photogrammetric inter-rater reliabilities are the same, the same is not true for direct anthropometric measurements. Thus, we have to reject that hypothesis that those measurements are the same.

While both methods showed acceptable intra-examiner reliabilities, the photogrammetric method appeared to be much more useful if conducting research featuring multiple investigators. Future research can be directed toward improving the photogrammetric method developed here. In particular, comparing the points identified on the photographs and the true location of those facial landmarks should be examined. If a correlation can be established between those two, the photogrammetric method used in this study can provide a very cheap, accurate and effective alternative for facial measurements.

It is possible that the two methods compare effectively, and can be used interchangeably, however, a better calibration method would be required. In their current form, the two methods are significantly different to be of any clinical use and we can accept our working hypothesis regarding the inter-method comparison.

The data shows that one examiner can consistently measure a given parameter on a subject. However, a study involving multiple examiners will most likely produce unreliable results and minimize its significance.

BIBLIOGRAPHY

"Antrhopometry." *Merriam-Webster.com*. Merriam-Webster, 2012.Web. 15 March 2012.

Aksu M, Kaya D, Kocadereli I. Reliability of reference distances used in photogrammetry. Angle Orthod. 2010, 80(4):482-9.

Al-Khatib AR Facial three dimensional surface imaging: an overview. Archives of Orofacial Sciences 2010, 5(1): 1-8.

Arnett GW, Bergman RT. Facial keys to orthodontic diagnosis and treatment planning. Part I. Am J Orthod Dentofacial Orthop. 1993a,103(4):299-312.

Arnett GW, Bergman RT. Facial keys to orthodontic diagnosis and treatment planning. Part II. Am J Orthod Dentofacial Orthop. 1993b,103(5):395-411.

Arnett GW, Jelic JS, Kim J, Cummings DR, Beress A, Worley CM Jr, Chung B, Bergman R. Soft tissue cephalometric analysis: diagnosis and treatment planning of dentofacial deformity. Am J Orthod Dentofacial Orthop. 1999,116(3):239-53.

Arnett GW, McLaughlin RP. Facial and Dental Planning for Orthodontists and Oral Surgeons. Elsevier Mosby, London. 2004. Print.

Ayoub A, Garrahy A, Hood C, White J, Bock M, Siebert JP, Spencer R, Ray A. Validation of a vision-based, three-dimensional facial imaging system. Cleft Palate Craniofac J. 2003, 40(5):523-9.

Behrents R. Growth of the Aging, Craniofacial Skeleton Craniofacial Growth Series. Ann Arbor: Center for Human Growth and Development: The University of Michigan; 1985. Print.

Broadbent BH: A new x ray technique and its application to orthodontics. Angle Orthod: 1931:45-66.

Budai M, Farkas LG, Tompson B, Katic M, Forrest CR. "Relation between anthropometric and cephalometric measurements and proportions of the face of healthy young white adult men and women." *J Craniofac Surg* 2003 14(2):154-61; discussion 162-3.

Burke PH, Beard FH. Stereophotogrammetry of the face. A preliminary investigation into the accuracy of a simplified system evolved for contour mapping by photography. Am J Orthod. 1967, 53(10):769-82.

Burstone C. Integumental contour and extension patterns. Angle Orthod. 1959, 29(2): 93-104.

Edler R, Wertheim D, Greenhill D. Comparison of radiographic and photographic measurement of mandibular asymmetry. Am J Orthod Dentofacial Orthop. 2003,123(2):167–174.

Edler RJ. Background considerations to facial aesthetics. J Orthod. 2001, 28(2):159-68.

Ettorre G, Weber M, Schaaf H, Lowry JC, Mommaerts MY and Howaldt HP . Standards for digital photography in cranio-maxillo-facial surgery - Part I: Basic views and guidelines. J Craniomaxillofac Surg, 2006, 34(2): 65-73.

Farkas LG. Examination, Photogrametry of the face, craniofacial anthropometry in clinical genetics. In: Farkas LG (ed.), Anthropometry of the Head and Face. New York: Raven Press Ltd, 1994. Print.

Farkas LG. Accuracy of anthropometric measurements: past, present, and future. Cleft Palate Craniofac J, 1996, 33(1): 10-18; discussion 19-22.

Farkas LG, Anthropometry of the Head and Face in Medicine, Elesvier, New York. 1981. Print.

Farkas LG, Bryson W, Klotz J. Is photogrammetry of the face reliable? Plast Reconstr Surg. 1980, 66(3):346-55.

Farkas LG, Posnick JC. Growth and development of regional units in the head and face based on anthropometric measurements. Cleft Palate-Craniofacial Journal. 1992. 29:301-329.

Farkas LG, Tompson B, Phillips JH, Katic MJ, Cornfoot ML. Comparison of anthropometric and cephalometric measurements of the adult face. J Craniofac Surg. 1999, 10(1):18-25; discussion 26.

Ferrario VF, Sforza C, Schmitz JH, Miani A Jr, Serrao G. A three-dimensional computerized mesh diagram analysis and its application in soft tissue facial morphometry. Am J Orthod Dentofacial Orthop. 1998, 114(4):404-13.

Ferrario VF, Sforza C, Serrao G, Colombo A, Ciusa V. Soft tissue facial growth and development as assessed by the three-dimensional computerized mesh diagram analysis. Am J Orthod Dentofacial Orthop. 1999, 116(2):215-28.

Ferring V, Pancherz H. Divine proportions in the growing face. Am J Orthod Dentofacial Orthop. 2008, 134(4):472-9.

Finlay LM, Craniometry and Cephalometry: A History Prior to the Advent of Radiography. Angle Orthod. 1980, 50(4):312-21.

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

Fourie Z, Damstra J, Gerrits PO, Ren Y. Evaluation of anthropometric accuracy and reliability using different three-dimensional scanning systems. Forensic Sci Int. 2011, 207(1-3):127-34

Franke-Gromberg C, Schüler G, Hermanussen M, Scheffler C. Digital 2D-photogrammetry and direct anthropometry--a comparing study on test accomplishment and measurement data. Anthropol Anz. 2010, 68(1):11-20.

Geerts GA, Stuhlinger ME, Nel DG. A comparison of the accuracy of two methods used by pre-doctoral students to measure vertical dimension. *J Prosthet Dent* 2004, 91(1):59-66.

Ghoddousi H, Edler R, Haers P, Wertheim D, Greenhill D. Comparison of three methods of facial measurement. Int J Oral Maxillofac Surg. 2007, 36(3):250-8.

Golafshani N, Understanding Reliability and Validity in Qualitative Research *The Qualitative Report* 2003, 8(4):597-607.

Gosman, S D. Anthropometric method of facial analysis in orthodontics. *Am J Orthod* 1950, 36(10):749-62.

Gribel BF, Gribel MN, Frazäo DC, McNamara JA Jr, Manzi FR. Accuracy and reliability of craniometric measurements on lateral cephalometry and 3D measurements on CBCT scans. Angle Orthod. 2011, 81(1):26-35.

Guyot L, Dubuc M, Richard O, Philip N, Dutour O. Comparison between direct clinical and digital photogrammetric measurements in patients with 22q11 microdeletion. Int J Oral Maxillofac Surg. 2003, 32(3):246-52.

Hajeer MY, Ayoub AF, Millett DT, Bock M, Siebert JP. Three-dimensional imaging in orthognathic surgery: the clinical application of a new method. Int J Adult Orthodon Orthognath Surg. 2002;17(4):318-30.

Hajeer MY, Millett DT, Ayoub AF, Siebert JP. Applications of 3D imaging in orthodontics: part I. J Orthod. 2004, 31(1):62-70.

Hellman, M. Some facial features and their orthodontic implication. *Am J Orthod Oral Surgery* 1939, 25(10): 927-951.

Hrdlicka A. Practical Anthropometry, Wistar Institute of Anatomy and Biology. Philadelphia, 1920. Print.

Kamoen A, Dermaut L, Verbeeck R. The clinical significance of error measurement in the interpretation of treatment results. Eur J Orthod. 2001, 23(5):569-78.

Karras DJ. Statistical methodology: II. Reliability and variability assessment in study design, Part A. Acad Emerg Med. 1997, 4(1):64-71.

Karras DJ. Statistical methodology: II. Reliability and variability assessment in study design, Part B. Acad Emerg Med. 1997, 4(2):144-7.

Maal TJ, Plooij JM, Rangel FA, Mollemans W, Schutyser FA, Bergé SJ. The accuracy of matching three-dimensional photographs with skin surfaces derived from cone-beam computed tomography. Int J Oral Maxillofac Surg. 2008, 37(7):641-6

Medelnik J, Hertrich K, Steinhäuser-Andresen S, Hirschfelder U, Hofmann E. Accuracy of anatomical landmark identification using different CBCT- and MSCT-based 3D images: an in vitro study. *J Orofac Orthop,* 2011, 72(4) : 261-78.

Miskin EA. The application of photogrammetric techniques to medical problems. The Photogrammetric Record, 1956, 2(8): 92–110.

Mommaerts MY, Moerenhout BA. "Reliability of clinical measurements used in the determination of facial indices." *J Craniomaxillofac Surg,* 2008, 36(5): 279-84.

Lewis R. Reliability and Validity: Meaning and Measurement. Annual Meeting of the Society for Academic Emergency Medicine, Boston MA, 1999. Print.

Naini FB, Moss JP, Gill DS. The enigma of facial beauty: esthetics, proportions, deformity and controversy. *Am J Orthod Dentofacial Orthop* 2006, 130(3): 277–282.

Naylor JM, Ko V, Adie S, Gaskin C, Walker R, Harris IA, Mittal R. Validity and reliability of using photography for measuring knee range of motion: a methodological study. BMC Musculoskelet Disord. 2011, 12:77.

Park YC, Burstone CJ. Soft-tissue profile--fallacies of hard-tissue standards in treatment planning. Am J Orthod Dentofacial Orthop. 1986, 90(1):52-62.

Paul SM, Chamberlin AP, Hatt C, Nayak AV, Danoff JV. Reliability, validity, and precision of an active stereophotogrammetry system for three-dimensional evaluation of the human torso. Med Eng Phys. 2009, 31(10):1337-42

Peck H, Peck S. A concept of facial esthetics. Angle Orthod. 1970, 40(4):284-318.

Proffit WR, Fields HW. Contemporary orthodontics, ed 4. St Louis:Mosby. 2007.Print.

Ricketts, RM. "The biological significance of the divine proportion and Fibonacci Series." *Am J Orthod,* 1982, 81(5): 351–370.

Sakar O, Sülün T, Kurt H, Gençel B. "Reliability and comparison of two facial measurements to detect changes of occlusal vertical dimension in complete denture wearers." *Gerodontology* 2011,28(3): 205-8.

Sarver DM. Principles of cosmetic dentistry in orthodontics: Part 1. Shape and proportionality of anterior teeth. Am J Orthod Dentofacial Orthop. 2004, 126(6):749-53.

Shaner DJ, Bamforth JS, Peterson AE, Beattie OB. Technical Note: Different techniques, different results – a comparison of photogrammetric and caliper-derived measurements. Am J Phys Antropol. 1998, 106 (4): 547-52.

Sheldon WH, StevensSH, and Tucker RH. The Varieties of Human Physique, Harper & Brothers, New York, 1940. Print.

Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979, 86(2):420-8.

Tanner RJM, Weinerr JS. The reliability of the photogrammetric method of anthropometry, with a description of a miniature camera technique. Am J Phys Anthropol. 1949, 7(2):145-86.

Trpkova B, Major P, Prasad N. Cephalometric landmarks identification and reproducibility: a meta-analysis. Am J Orthod Dentofacial Orthop. 1997, 112(2):165-70.

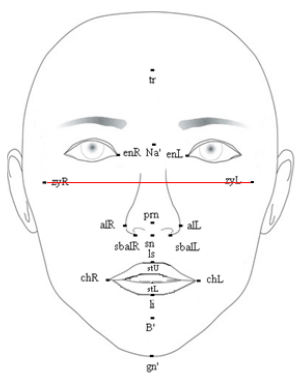
Weinberg SM, Naidoo S, Govier DP, Martin RA, Kane AA, Marazita ML. Anthropometric precision and accuracy of digital three-dimensional photogrammetry: comparing the Genex and 3dMD imaging systems with one another and with direct anthropometry. J Craniofac Surg. 2006, 17(3):477-83.

ADDENDUM A

Thank you for participating in the project entitled “Intra- and Inter Examiner Reliability of Clinical Anthropometry and Photogrammetry”. This document will provide with a detailed description of each measurements is to be taken. It is your responsibility to be familiar with all the points and the ways they are to be measured. In order to ensure consistency of measurement, the identification of each point and the way the measurements are to be taken will be described in greater detail below. Please refer to Figures 1 and 2 for a large-scale representation of the points. Smaller versions with the corresponding measurement are given in the write-up below.

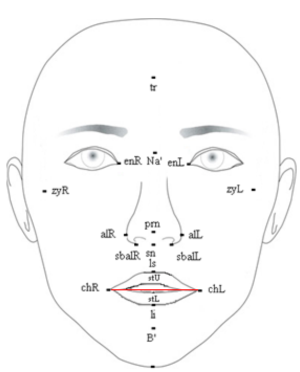
The Mitutoyo digital caliper consists of two measuring sides. The larger side is found below the digital screen. The measurement that it records is found between the two **inner** edges of the caliper. The smaller side found above the digital screen measures a distance located between the two **outer** edges. The study participant is to be seated upright with the lips relaxed. In order for you not to change the study participant’s position you may have to use one of the two sides. We have provided indications for when it is necessary to use the short side. Please, do no mark the points on the face.
Let’s being:

1. **Zygomatic Width.** Defined as the straight-line distance between zyR and zyL.



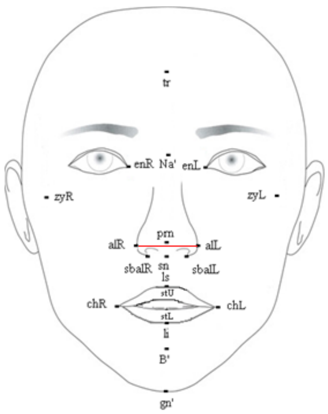
zyR/zyL are defined as the most prominent points of the cheekbone (zygoma) on either the right or left side. Palpate the area in order to select the point. Hold the caliper horizontally. Use the longer side to measure the distance.

2. **Mouth Width.** Defined as the straight line distance between chR and chL



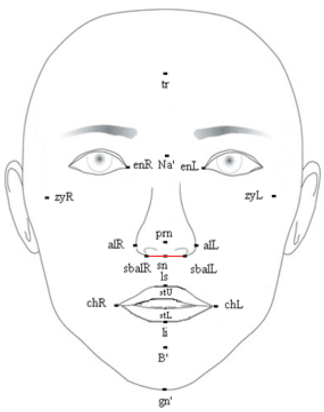
chR / chL are the commissures of the mouth, or the end points of the mouth in the transverse plane. Locate the points by determining where the upper and lower lip vermillions intersect with the skin of the face.
Hold the caliper horizontally. Use the longer side to measure the distance.

3. **Alar Width.** Defined as widest portion of the nose in the nostril area. (alR-alL)



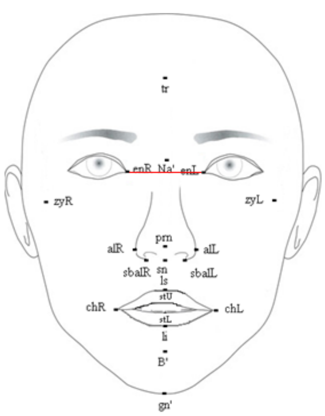
alR/alL are the points of the alae that yield the widest portion of the nose in the nostril area. Locate the points by determining the most lateral points of the nostrils. Connect the two points to obtain the measurement.
Hold the caliper horizontally.
Use the longer side to measure the distance.

4. **Alar Base Width.** Defined as the straight-line distance between the base of the two alae.(sbalR-sbalL)



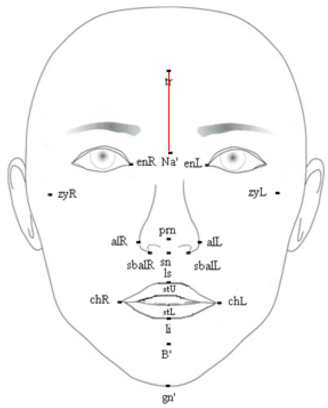
sbalR/sbalL are the points where the nostrils connect with the skin of the upper lip. Locate the points by determining the intersection of the nostrils with the upper lip.
Hold the caliper horizontally.
Use the longer side to measure the distance.

5. **Intra-orbital width.** Defined as the distance between the two innermost points of the orbits (enR-enL)



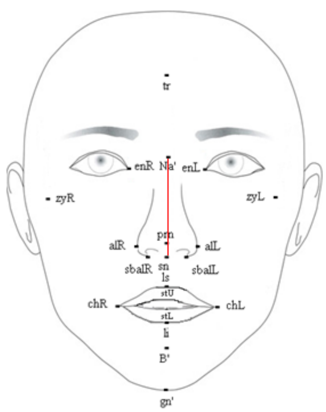
enR/enL are the innermost points of the right and left orbits. Locate the points by determining the intersection of the orbits with skin of the face.
Hold the caliper horizontally.
Use the longer side to measure the distance. The tips of the longer side should be pointing upward and NOT toward the face. Do not touch the points directly. Get as close to the points as possible and project the location of enR/enL.

6. **Hairline (tr) – Soft Tissue Nasion (Na').** Defined as the distance between the hairline and Nasion.



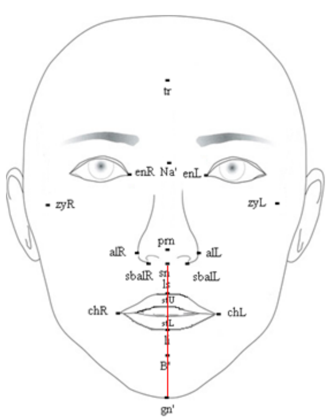
tr is located at the intersection of the hair and the skin of the forehead. Na' is the soft tissue point representing the bony intersection between the frontal and nasal bones. Locate Na' by palpating the innermost point between the forehead and nose. Hold the caliper vertically. Use the short side to measure the distance.

7. **Nasion (Na')– SubNasale (sn).** Defined as the distance between Nasion and SubNasale.



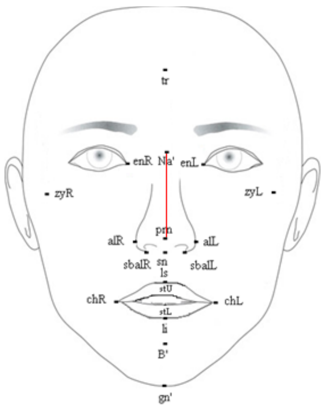
Na' is defined as in **(6)**. sn is defined as the intersection of the columnella with the philtrum. Hold the caliper vertically. Use the longer side to measure the distance.

8. **SubNasale (sn) – Soft Tissue Gnathion (gn').** Defined as the distance between SubNasale and menton.



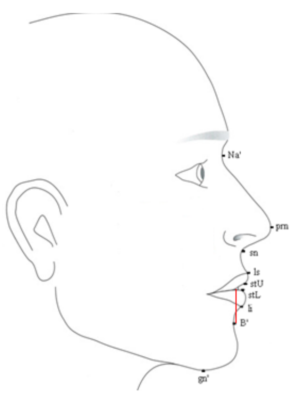
sn is defined as in (7). Me is defined as the most inferior point of the mandible in the midline. In order to determine gn' locate the intersection of the most inferior point of the chin/mandible and midline. Use the philtrum as an indicator for the midline.
Hold the caliper vertically. Use the short side to measure the distance.

9. **Nasion (Na')– Tip of Nose (prn).** Defined as the distance between Na' and ToN.



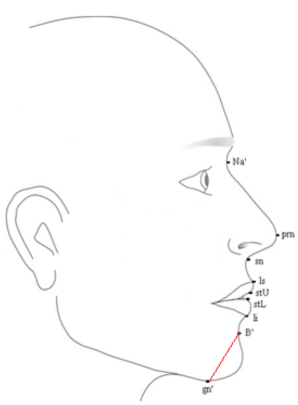
Na' is defined as in **(6)**. prn is the most anterior point of the nose in the alar area. Palpate the nose in order to determine prn. Hold the caliper vertically. Use the short side to measure the distance.

10. **Stomion Lower (StL) – Soft tissue B point (B').** Defined as the distance between the uppermost point of the lower lip and the innermost point between the lower lip and the chin.



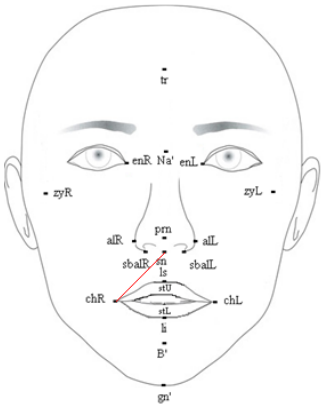
StL is the uppermost point of the lower lip that you can locate. Locate B' by examining the area below the lower lip and the chin and determining the innermost point. Measure the distance between the two
Hold the caliper vertically. Have the tips of the caliper point toward the participant. Use the short side to measure the distance. NOTE: in order to not touch the participant's chest have the tail of the caliper point upward.

11. **Soft Tissue B Point (B') – Menton (Me).** Defined as the distance between B' and Me.



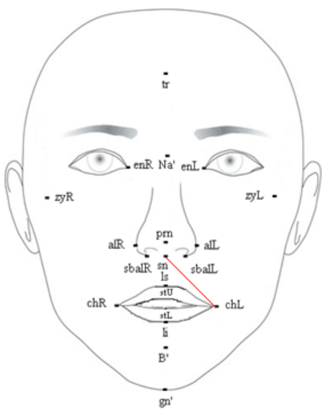
B' is defined as in (10). Me is defined as in (8). Measure the distance between the two points. Hold the caliper vertically. Have the tips of the caliper point toward the participant. Use the short side to measure the distance. NOTE: in order to not touch the participant's chest have the tail of the caliper point upward.

12. **SubNasale – Right Commisure (chR).** Defined as the distance between SubNasale and the right outermost point of the mouth



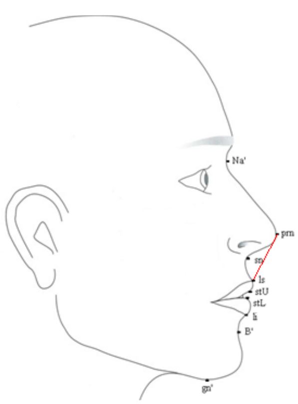
sn is defined as in (7). chR is defined as the intersection of the upper lip vermillion, lower lip vermillion and right side of the skin of the face. Measure the distance between the two points. Hold the caliper at an angle/diagonally. Use the short side to measure the distance.

13. **Subnasale – Left Commissure (ch L).** Defined as the distance between SubNasale and the left outermost point of the mouth



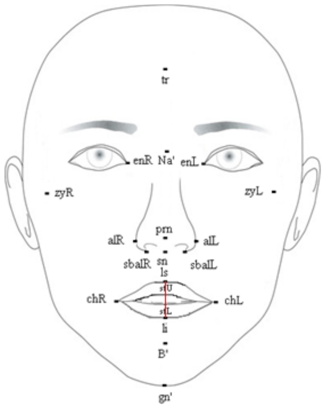
sn is defined as in (7). chL is defined as the intersection of the upper lip vermillion, lower lip vermillion and left side of the skin of the face. Measure the distance between the two points. Hold the caliper at an angle/diagonally. Use the short side to measure the distance.

14. **Tip of Nose (prn) – Upper Vermillion Border (ls).** Defined as the distance between the tip of the nose and the line passing through the intersection points of the philtrum with the upper vermillion border



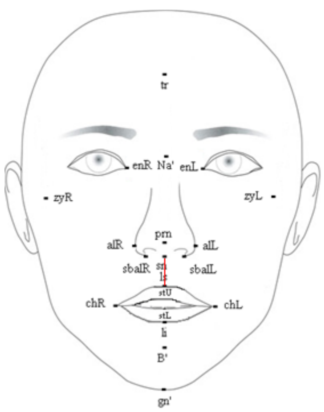
prn is defined as in (9). ls is defined as the imaginary line connecting the intersection of the philtrum columns with the upper vermillion border.
Hold the caliper vertically. Use the short side to measure the distance. Have the tail of the caliper point upward.

15. **Mouth Height.** Defined as the distance between the upper vermillion border (ls) and the lower vermillion border (li).



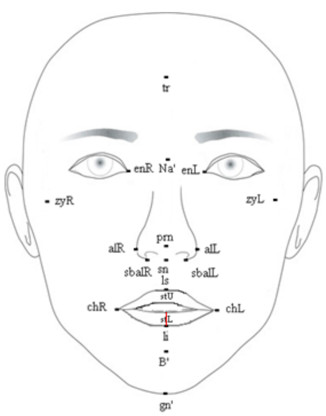
ls is defined as in (14). li is the horizontal lowermost line that passes through the intersection of the lower lip with the skin of the face. Measure the distance between the two
Hold the caliper vertically. Use the longer side to measure the distance. Have the tail of the caliper point upward.

16. **Upper lip length.** Defined as the distance between subnasale and ls.



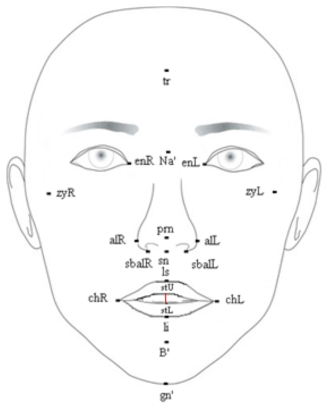
sn is defined as in (7). ls is defined as in (14). Measure the distance between the two.
Hold the caliper vertically. Use the longer side to measure the distance. Have the tail of the caliper point upward.

17. **Lower Lip Thickness .** Defined as the distance between stomion lower and labius inferius.



stL is defined as in (10). li is defined as in (15). Measure the distance between the two points.
Hold the caliper vertically. Use the longer side to measure the distance. Have the tail of the caliper point upward.

18. Interlabial gap. Defined as any space present between Stomion Upper (stU) and Stomion Lower (stL) when the participant is in repose.



StL is defined as in (10). StU is the lowermost point the you can locate on the upper lip. Measure the distance between the two. Hold the caliper vertically. Use the longer side to measure the distance. Have the tail of the caliper point upward.

ADDENDUM B

From: Pritchard, Laura (ELS-OXF) [L.Pritchard@elsevier.com]
Sent: Friday, January 27, 2012 8:29 AM
To: Mollov, Nikolay
Subject: Permission Request

Dear Dr Mollov,

We hereby grant you permission to reproduce the material detailed below in **print and electronic format** at no charge subject to the following conditions:

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"This article was published in Publication title, Vol number, Author(s), Title of article, Page Nos, Copyright Elsevier (or appropriate Society name) (Year)."

3. This permission is granted for non-exclusive world rights in all languages.

4. Reproduction of this material is granted for the purpose for which permission is hereby given, and includes use in any future editions.

Kind regards

Laura

Laura Pritchard

Senior Rights Associate | ELSEVIER | The Boulevard| Langford Lane | Kidlington | Oxford OX5 1GB |

Tel: +44 1865 843517 Fax: +44 1865 853333

l.pritchard@elsevier.com

ADDENDUM C

trichion (tr)	The point on the hairline in the midline of the forehead. Note: for this project, not participants with visible hair loss or abnormally high hairline were selected for participation
soft tissue nasion (Na')	The soft tissue covering the point located in the midline of both the nasal root and the nasofrontal suture
endocanthion (en)(Left or Right)	The point at the inner commissure of the eye fissure
zygion (zy)(L or R)	The most lateral point of each zygomatic arch; identified by trial measurements. **Note**: in this project left and right are identified, when applicable, in order to help the investigators in communicating with the study examiners
pronasale(prn)	The most protruded point of the apex nasi
alare (al)(L or R)	The most lateral point on each alar contour
subnasale (sn) (L or R)	The midpoint of the columnella base at the apex of the angle where the lower border of nasal septum and the surface of the upper lip meet
subalare (sbal)(L or R)	The point at the lower limit of each alar base, where the alar base disappears into the skin of the upper lip
labiale superius (ls)	The midpoint of the upper vermillion line
labiale inferius (li)	The midpoint of the lower vermillion line
cheilion (ch)(L or R)	The point located at each labial commissure
stomion (sto) (Upper and Lower)	The imaginary point at the crossing of the vertical facial midline and the horizontal labial fissure between the upper/lower lip and the oral cavity as seen from a frontal view. Note: in this project the study participants were asked to relax their lips, hence the visible border of each lip was used as the horizontal landmark
soft tissue B point (B')	The deepest curvature of the soft tissue between the lower lip and the chin point
gnathion (gn')	The lowest median landmark of the lower border of the mandible

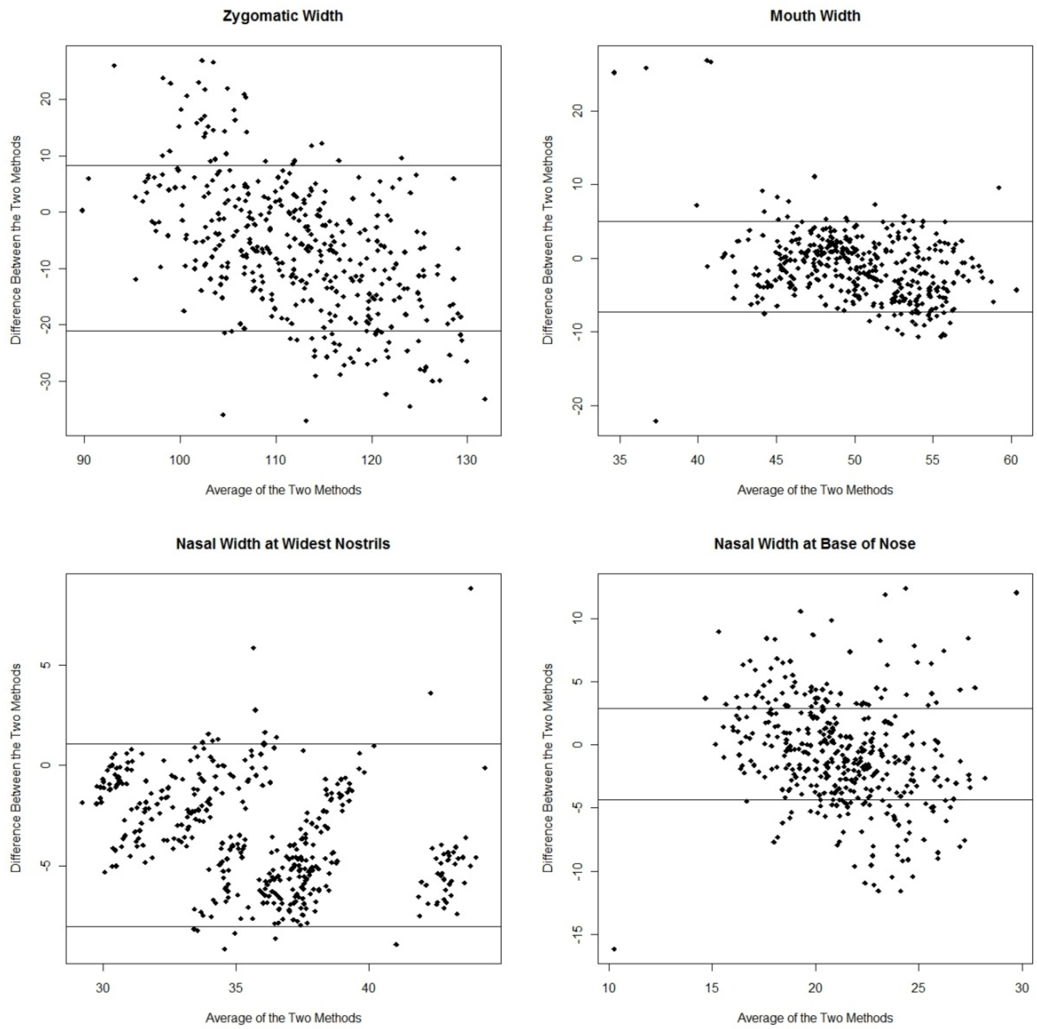
ADDENDUM D

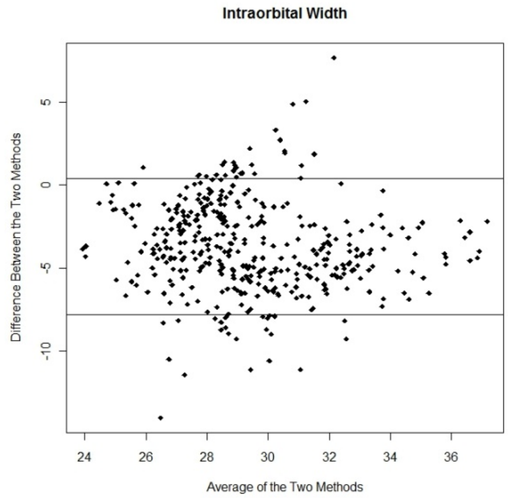
Chart used to record caliper measurements.

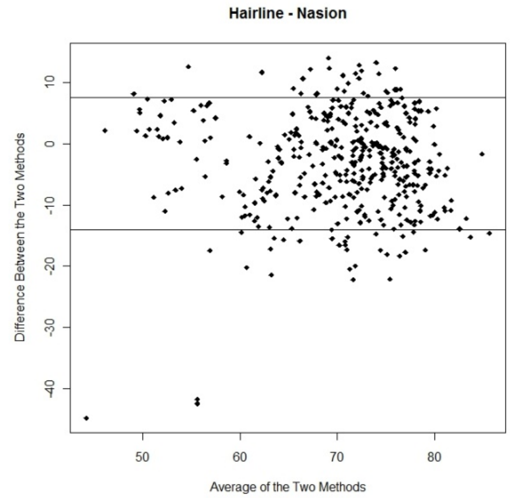
Dimension	Facial Landmark	Measurement (mm)
Horizontal Measurements	1. Zygomatic Width (zyR – zyL)	
	2. Mouth Width (chR - chL)	
	3. Nasal Width at widest nostrils (alR – alL)	
	4. Nasal Width at Base of Nose (sbalR – sbalL)	
	5. Intraorbital Width (enR – enL)	
Vertical Mesurements	6. Hairline-Nasion (tr – Na’)	
	7. Nasion – SubNasale (Na’ – sn)	
	8. SubNasale – Gnathion (sn – gn’)	
	9. Nasion – Tip of Nose (Na’ – prn)	
	10. Stomion Lower – Soft Tissue B point (li – B’)	
	11. Soft Tissue B point – Gnathion (B’ – gn’)	
	12. SubNasale – Right commissure (sn – chR)	
	13. SubNasale – Left commissure (sn – chL)	
	14. Tip of Nose - upper lip (prn - ls)	
	15. Mouth height (ls - stU)	
	16. SubNasale to Upper Lip (sn - ls)	
	17. Lower Lip Thickness (stL - li)	
	18. Interlabial Gap (stU – stL) - if lips are incompetent	

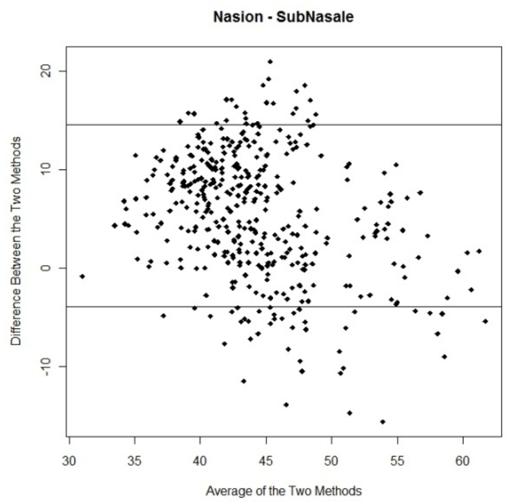
ADDENDUM E

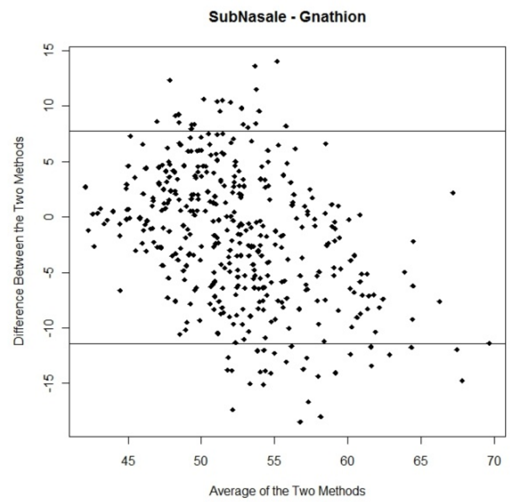
Graphical representation of the Bland-Altman inter-method comparison plots for the different measurements

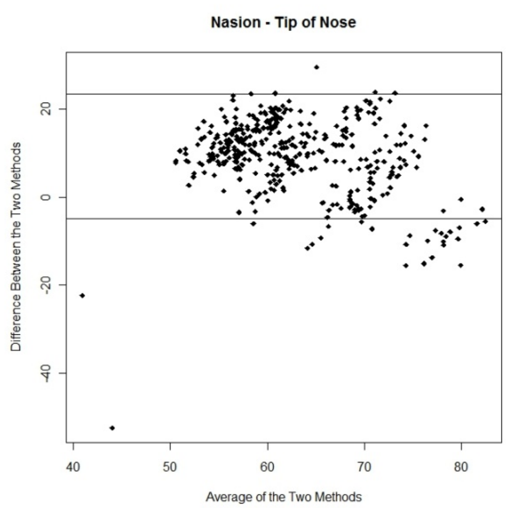


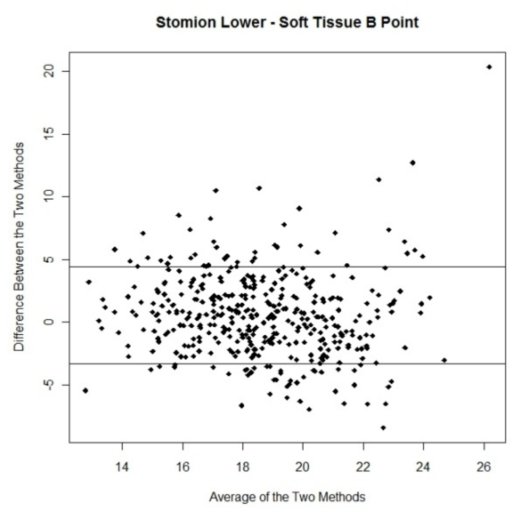

Intraorbital Width
Difference Between the Two Methods
Average of the Two Methods

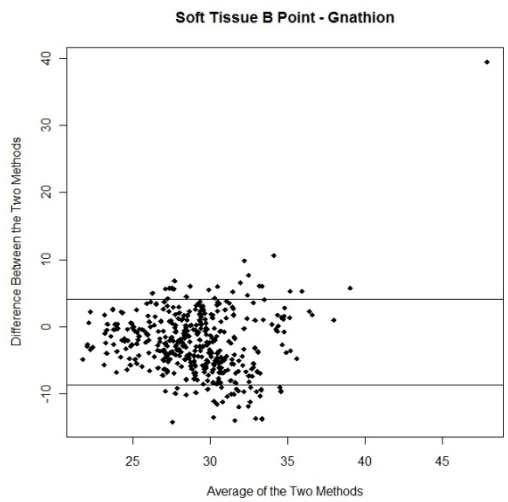

Hairline - Nasion
Difference Between the Two Methods
Average of the Two Methods

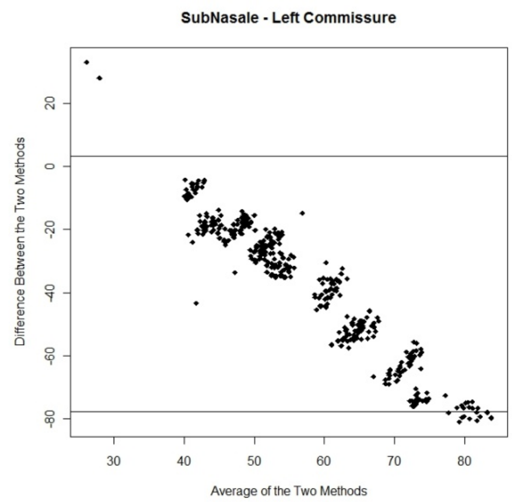

Nasion - SubNasale
Difference Between the Two Methods
Average of the Two Methods

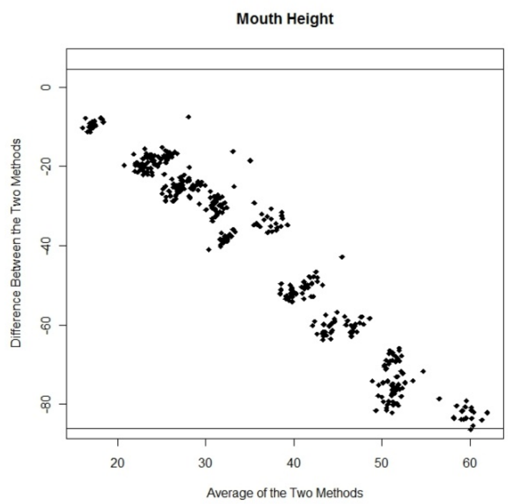

SubNasale - Gnathion
Difference Between the Two Methods
Average of the Two Methods

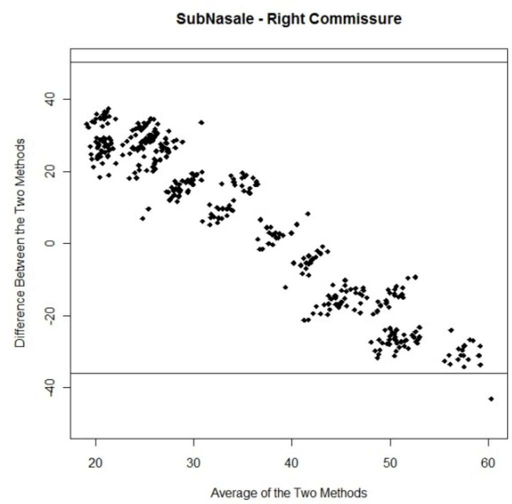

Nasion - Tip of Nose
Difference Between the Two Methods
Average of the Two Methods

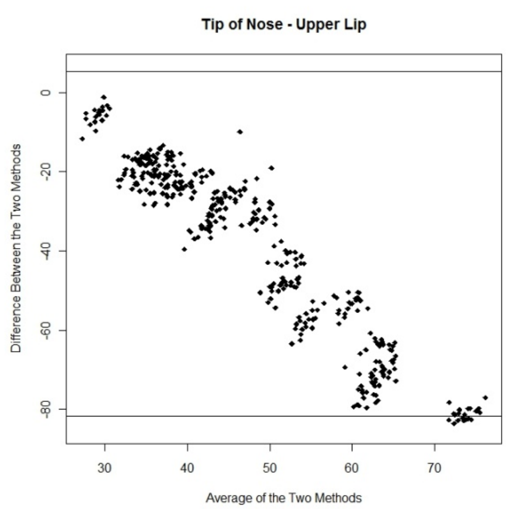

Stomion Lower - Soft Tissue B Point
Difference Between the Two Methods
Average of the Two Methods

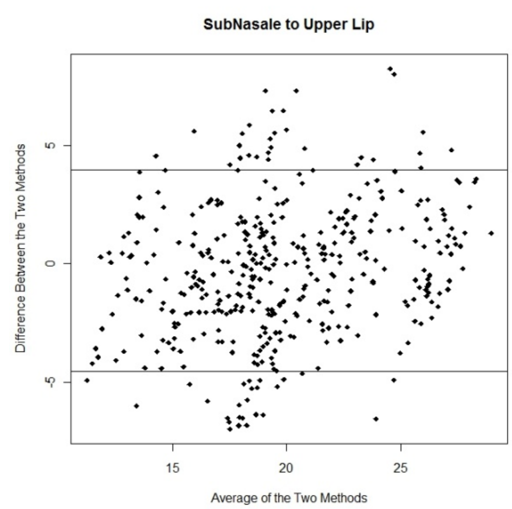

Soft Tissue B Point - Gnathion
Difference Between the Two Methods
-10 0 10 20 30 40
25 30 35 40 45
Average of the Two Methods

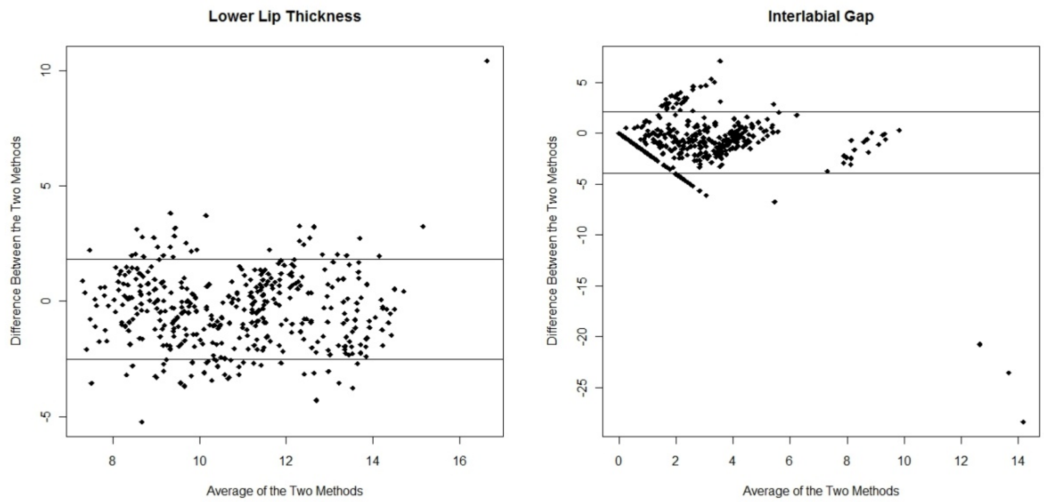

SubNasale - Left Commissure
Difference Between the Two Methods
-80 -60 -40 -20 0 20
30 40 50 60 70 80
Average of the Two Methods


Mouth Height
Difference Between the Two Methods
-80 -60 -40 -20 0
20 30 40 50 60
Average of the Two Methods


SubNasale - Right Commissure
Difference Between the Two Methods
-40 -20 0 20 40
20 30 40 50 60
Average of the Two Methods


Tip of Nose - Upper Lip
Difference Between the Two Methods
-80 -60 -40 -20 0
30 40 50 60 70
Average of the Two Methods


SubNasale to Upper Lip
Difference Between the Two Methods
-5 0 5
15 20 25
Average of the Two Methods


Lower Lip Thickness
Difference Between the Two Methods
Average of the Two Methods
Interlabial Gap
Difference Between the Two Methods
Average of the Two Methods

ADDENDUM F

Original IRB approval for #2083 and addendum/ consent form approval requesting to use dental students to establish the reliability of the measurements performed in the project.

OFFICE OF RESEARCH COMPLIANCE

MARQUETTE UNIVERSITY

November 22, 2010

Dr. Jose Bosio
School of Dentistry

Dear Dr. Bosio:

Your protocol number HR-2083, titled, *"Soft Tissue Changes in Response to Orthodontic Treatment"* was expedited on November 18, 2010, by a member of the Marquette University Institutional Review Board.

Your IRB approved adult consent form, parent permission form, and child assent form are enclosed with this letter. Use the stamped copies of these forms when recruiting research participants. Each research participant should receive a copy of the stamped consent form for their records. Each participant, or legal guardian, must also complete the HIPAA authorization form.

Subjects who go through the consent process are considered enrolled participants and are counted toward the total number of subjects, even if they have no further participation in the study. Please keep this in mind when conducting your research. This study is currently approved for 2000 subjects.

If you need to increase the number of subjects, add research personnel, or make any other changes to your protocol you must submit an IRB Protocol Amendment Form, which can be found on the Office of Research Compliance web site: http://www.marquette.edu/researchcompliance/research/irbforms.shtml. All changes must be reviewed and approved by the IRB before being initiated, except when necessary to eliminate apparent immediate hazards to the human subjects. Any public advertising of this project requires prior IRB approval. If there are any adverse events, please notify the Marquette University IRB immediately.

Your approval is valid until November 17, 2011. Prior to this date, you will be contacted regarding continuing IRB review.

A Protocol Completion/Termination Report must be submitted once this research project is complete. The form should be submitted in a timely fashion, and must be received no later than the protocol expiration date.

If you have any questions or concerns, please do not hesitate to contact me. Thank you for your time and cooperation.

Sincerely,

Amanda J. Ahrndt, RN, MS, MSN, CIM
IRB Manager

cc: Dr. Rebecca Bardwell, IRB Chair

Enclosures (3)

ROOM 102 560 NORTH 16TH STREET P.O. BOX 1881 MILWAUKEE, WISCONSIN 53201-1881 TELEPHONE (414) 288-7570 FAX.(414) 288-6281

MARQUETTE UNIVERSITY

Be The Difference.

Office of Research Compliance

Schroeder Complex, 102
P.O. Box 1881
Milwaukee, Wisconsin 53201-1881

P 414.288.7570
F 414.288.6281
W www.marquette.edu/researchcompliance

June 9, 2011

Dr. José A. Bósio
Dental School

Dear Dr. Bósio:

The amendment you submitted on May 23, 2011, for your protocol number HR-2083, titled, "*Soft Tissue Changes in Response to Orthodontic Treatment,*" received expedited approval on June 9, 2011, from a member of the Marquette University Institutional Review Board.

This amendment adds study procedures to establish which soft tissue measurements are repeatable and accurate. This portion of the study adds 40 subjects (dental students) for a total of 2040 subjects, adds email recruitment, and adds a new consent form specific to this portion of the study.

Your IRB approved informed consent form is enclosed with this letter. Use the stamped copies of this form when recruiting research participants. Each research participant should receive a copy of the stamped consent form for their records.

Your protocol is valid until November 17, 2011. Prior to this date, you will be contacted regarding continuing IRB review. Any public advertising of this project requires prior IRB approval. If there are any changes in your protocol or adverse events, please notify the IRB immediately.

If you have any questions or concerns, please do not hesitate to contact me. Thank you for your time and cooperation.

Sincerely,

Amanda J. Ahrndt, RN, MS, MSN, CIM
IRB Manager

cc: Dr. Rebecca Bardwell, IRB Chair

Enclosure

Protocol Number: HR-2083

MARQUETTE UNIVERSITY
STUDENT PARTICIPANT CONSENT FORM

Soft Tissue Changes in Responses to Orthodontic Treatment.

Jose A. Bosio, BDS, MS, Nikolay Molov, DDS, T. Gerard Bradley, BDS, MS
Marquette University School of Dentistry, Department of Developmental Sciences/Orthodontics

You have been invited to participate in this research study. Before you agree to participate, it is important that you read and understand the following information. Participation is completely voluntary. Please ask questions about anything you do not understand before deciding whether or not to give us permission to participate. Whether or not you choose to participate in this research will not affect your grades or standing with the Marquette University School of Dentistry or the Department of Developmental Sciences/Orthodontics.

PURPOSE: The purpose of this portion of the study is to establish the reliability of measurements taken either via a caliper (clinically) or done on a photograph. The overall study is designed to examine soft tissue changes in response to orthodontic treatment in patients at the Marquette University School of Dentistry Department of Developmental Sciences/Orthodontics.

PROCEDURES: You will be assigned to one of two groups if you agree to participate in this study. One group will measure other student participants and the other group will be measured.

There will be 2 sessions for this study:
<u>Session 1</u>

1. Measurements of facial landmarks will be made in a clinical environment using a digital caliper
2. Participants being measured will be positioned in a cephalostat (a radiographic machine) and a photograph will be taken. No radiograph (x-ray) will be taken. The machine will be used simply for positioning purposes.

<u>Session 2</u> Will be exactly the same as Session 1 but will be done two to three weeks later. This will be done in order to determine the consistency of the measurements over a period of time.

DURATION: Your participation will consist of the above described two sessions. The sessions will take place in the Department of Developmental Sciences/Orthodontics at Marquette University School of Dentistry. The amount of time that you will be actively participating will be about one hour per session (approximately two hours total).

RISKS: The risks associated with participation in this study are minimal. For those being measured, measurements will be taken with an 8" digital caliper. The caliper has pointy arms which may poke your cheek if used without care. However, measurements will be always taken with the caliper parallel to the face to prevent such problem. The photographs taken will be matched to your name for data entry purposes but all final results will be represented in terms of averages and no individual names will be identifiable. The photographs will be stored on Dr. Bosio's and Dr. Mollov's computers at the school and no other personal will have access to

Protocol Number: HR-2083

them. There are no known risks for participating in the group that will be conducting the measurements.

BENEFITS: There are no direct benefits to you for participating in this study. Your participation in this portion of the study will help the researchers better understand which soft tissue clinical measurements are consistently repeatable.

CONFIDENTIALITY: All information you reveal in this study will be kept confidential. When the results of the study are published, you will not be identified by name. Data collected from this portion of the study will be maintained indefinitely and may be used for future research purposes. Your research records may be inspected by the Marquette University Institutional Review Board or its designees, and (as allowable by law) state and federal agencies.

COMPENSATION: You will be compensated for your participation with research points. If you participate in both sessions you will receive 30 research points. Research points may also be obtained through activities other than this study.

VOLUNTARY NATURE OF PARTICIPATION: Your participation in this study is completely voluntary and you may withdraw from the study and stop participating at any time without penalty or loss of benefits to which you are otherwise entitled. You will only receive research points for the time you participate in the study.

CONTACT INFORMATION: If you have any questions about this research project, you can contact Dr's Jose A. Bosio, Nikolay Molov or Gerard Bradley at Marquette University School of Dentistry, Department of Developmental Sciences/Orthodontics 1801 W. Wisconsin Ave. Milwaukee, WI, 53233. The phone number is (414) 288-7271. If you have questions or concerns about your rights as a research participant, you can contact Marquette University's Office of Research Compliance at (414) 288-7570.

I HAVE HAD THE OPPORTUNITY TO READ THIS CONSENT FORM, ASK QUESTIONS ABOUT THE RESEARCH PROJECT AND AM PREPARED TO PARTICIPATE IN THIS PROJECT.

______________________________ __________________

Participant's Signature Date

______________________________ __________________

Researcher's Signature Date