THE UNIVERSITY
of EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

The Heritability
of
Facial Morphology

Helen Langstaff

PhD

University of Edinburgh

2015

Acknowledgments

Firstly, I’d like to thank Dr Elena Kranioti for her support, guidance and trust over the last six years. My appreciation and thanks goes to all the participants of this research for their time and great interest in what I was doing. I’d like to thank the University of Edinburgh for providing me with the equipment needed to carry out my study and my special thanks goes to Karan Howie for making it a lot easier when times got tough. I am so grateful for the friendship and support of all the osteo post graduate girls during my time at Edinburgh but my particular thanks go to Mara Karell for her total willingness to help and for being my sounding board. My most sincere thanks go to my husband for having total faith in me and supporting me in so many ways over the last few years. To my mum I’d like to extend the utmost thanks for encouraging me, believing in me and mostly for all your comments and corrections without which I wouldn’t be where I am. Most of all I’d like to thank God for with him all things are possible.

I hereby declare that this thesis has been
composed by me and save where acknowledgement
is made, is based on my own work.
This work has not been submitted for any other degree or qualification.

Abstract

Facial recognition methodologies, widely used today in everything from automatic passport controls at airports to unlocking devices on mobile phones, has developed greatly in recent years. The methodologies vary from feature based landmark comparisons in 2D and 3D, utilising Principal Component Analysis (PCA) to surface-based Iterative Closest Point Algorithm (ICP) analysis and a wide variety of techniques in between. The aim of all facial recognition software (FCS) is to find or match a target face with a reference face of a known individual from an existing database. FCS, however, faces many challenges including temporal variations due to development/ageing and variations in facial expression. To determine any quantifiable heritability of facial morphology using this resource, one has to look for faces with enough demonstrable similarities to predict a possible genetic link, instead of the ordinary matching of the same individual's face in different instances. With the exception of identical twins, this means the introduction of many more variables into the equation of how to relate faces to each other. Variation due to both developmental and degenerative aging becomes a much greater issue than in previous matching situations, especially when comparing parents with children. Additionally, sexual dimorphism is encountered with cross gender relationships, for example, between mothers and sons. Non-inherited variables are also encountered such as BMI, facial disfigurement and the effects of dental work and tooth loss.

For this study a Trimmed Iterative Closest Point Algorithm (TrICP) was applied to three-dimensional surfaces scans, created using a white light scanner and Flexscan 3D, of the faces of 41 families consisting of 139 individuals. The TrICP algorithm produced 7176 Mesh-to-mesh Values (MMV) for each of seven sections of the face (*Whole face, Eyes, Nose, Mouth, Eyes-Nose, Eyes-Nose-Mouth*, and *Eyes-Nose-Mouth-Chin*). Receiver Operated Characteristic (ROC) analysis was then conducted for each of the seven sections of the face within 11 predetermined categories of relationship, in order to assess the utility of the method for predicting familial relationships (sensitivity/specificity). Additionally, the MMVs of three single features, (*eyes, nose* and *mouth*) were combined to form four combination areas which were analysed within the same 11 relationship categories.

Overall the relationship between sisters showed the most similarity across all areas of the face with the clear exception of the mouth. Where female to female comparison was conducted the mouth consistently negatively affected the results. The father-daughter relationship showed the least similarity overall and was only significant for three of the 11 portions of the face. In general, the combination of three single features achieved greater accuracy as shown by Areas Under the Curve (AUC) than all other portions of the face and single features were less predictive than the face as a whole.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Facial perception, or the ability to recognise the faces of other human beings, is a neurological mechanism which we are born with, and an extension of this inherent capacity is the ability to observe a family resemblance between some or all family members. ‘Isn’t he like his father’ and a multitude of similar comments are a common topic of conversation in most families. However, while we might all agree that a child is the ‘spitting image’ of his father, we have never yet been able to quantify that similarity in a measurable, scientific way. This study of 41 UK families seeks to determine whether it is possible to predict a genetic relationship by comparing and analysing the facial shapes of parents and their offspring.

While there are apparent links between the subject matter of this study and the increasingly widespread practice of the use of facial recognition technology, their aims are very different. Facial recognition technology, as used by Border Force, the Police and social networking sites, aims to match the image of a person presenting at passport control or appearing on Close Circuit Television (CCTV) with an existing record of an individual’s face, held on a database. This study, however, aims to demonstrate a statistical similarity between the facial shapes of two individuals such that a genetic link can be predicted or excluded.

The technology used in conventional Facial Recognition Software today is much more basic than that used in this study, most software compares the relative position, size and shape of a limited number of landmarks on the face using two-dimensional images. Both facial expressions and certain camera angles present problems with two-dimensional technology but despite these limitations such technology is in widespread use searching for ‘persons of interest’. The facial scanning used in this study creates a detailed three-dimensional ‘mesh’ of each individual’s face which replicates each curve and indentation in addition to using a range of recognised landmarks. The search for an inherited facial shape is however far more complicated than the search for an exactly matched facial shape, as this study will explain.

Aims of the study

This study aims to quantify the similarities in facial morphology between actual family members in order to be able to predict possible genetic relationships between groups of people not known to be genetically linked. Clearly, with the advent of DNA testing, proving a genetic link has become much more straightforward, but DNA testing is not always available, affordable or appropriate. There may be scenarios where it is quicker and more cost effective to reduce the pool of people needing to be tested by analysing facial scans. Cases which show little similarity in facial shape and therefore a very low probability of genetic relationship can then be ruled out, and the remainder of people with possible genetic links predicted by the analysis of facial scans could then be tested using DNA.

The specific aims of this study are:

1. To assess the ability of Trimmed Iterative Closet Point algorithm (TrICP), a recognised facial recognition technique, to recognise existing genetic relationships, and not only, as is currently the case, to match a target image to an existing image of a known individual.
2. To assess the impact of sex variation on the recognition process
3. To assess the impact of age variation on the recognition process.
4. To assess the impact of BMI variation on the recognition process.
5. To assess the influence of specific relationship pairs, for example mothers to daughters versus mothers to sons.
6. To assess whether analysis of certain facial features, alone or combined with others, are more effective at identifying a genetic match than each other or, than the face as a whole.
7. To assess the impact of scan quality on all of the above.

Review of the Literature

The extensive literature review in this study extends over three chapters, and will outline the findings of previous studies related to the question being researched in this one. Chapter two covers skeletal growth and the development of the skull and facial bones, going on to review soft tissue growth in the face. The studies reviewed include those relating to both developmental and degenerative growth and the sexual dimorphism seen in both. In order to understand facial heritability, it was necessary to have an in depth understanding of the normal growth process from the human embryo through into adulthood. Also considered are those factors, such as malnutrition, which adversely affect the normal growth of an individual. The literature review covers seminal works such as Goldstein's (1939) study of American Jews and Posen's (1967) work on the growth of the nose before coming right up to date by drawing upon Sforza et al.'s (2009, 2010a, 2013) extensive body of work covering age and sex related changes in facial shape. The author draws attention to the fact that terminology and definitions relating to facial landmarks and measurements differ and that these differences in nomenclature mean that not all studies can be accurately compared with each other. The literature review shows how the bones and soft tissue of the face develop at different rates for males and females across childhood and adolescence until adulthood. Once adulthood has been reached the shape of the face does not remain static as degenerative changes then start to take place. Many of the studies analysed were very large scale ones with in excess of a thousand subjects (Ferrario et al. 2000, Zankl et al. 2002) and all of the work that has been reviewed has also been collated into helpful tables so one can see at a glance the size and subject matter of the studies being reviewed. On a lighter note, and while not particularly relevant to this study many of the studies reviewed confirm the widely held view that older people do indeed have larger noses.

Chapter three is a review of facial recognition which starts with studies describing how the human brain can recognise other human beings without it being a conscious process (Cartoux et al.1989), the chapter shows how human beings have been involved with the business of facial recognition many years before there was any attempt to automate the process. Border control is an obvious example, as is the presentation of any form of photo identification. While Border Force staff are highly skilled at recognising individuals from photographs, the same cannot be said for everyone else and some studies have shown that the ability to recognise strangers from a photograph can be quite poor (Kemp et al.1997). Human ability to recognise others varies, with some showing an extraordinary ability to recognise (Davis & Valentine 2015) while at the other end of the spectrum are those unfortunate people suffering from prosopagnosia, or the impairment of facial perception. People suffering from prosopagnosia cannot even recognise the faces of their own family members.

The origins of two-dimensional facial recognition as a science began in the 1960s with simple comparisons of photographs, this moved on to the comparison of features and landmarks followed by the superimposition of one photograph onto another (Davis & Valentine 2015). Over time landmark placement became more sophisticated as did the analytical tools used in facial recognition. Eigenfaces, developed by Turk & Pentland (1991) were an important breakthrough in the science of facial recognition. Turk & Pentland used Principal Component Analysis (PCA) to produce a basic facial image known as an Eigenface. Following this came Fisherfaces which were another methodology used in facial recognition which some (Belhumeur et al.1997), claimed were even less prone to error than previous analytical tools.

All two-dimensional facial recognition methods suffered from a variety of problems including poor lighting and the shadows caused by this, the angle of the head when the image is taken and differences in facial expression. All of these reduced their efficacy.

The first three-dimensional images were displayed in 1851 at the Great Exhibition when David Brewster introduced his stereoscope (Brewster 1856). It was not however until three-dimensional scanners became widely available almost 150 years later that three-dimensional facial recognition technology became a reality. Most of the problems encountered in two-dimensional technology, illumination, pose and size are removed at a stroke with three-dimensions as the image is totally manoeuvrable and a virtual light source can be used so shadows are no longer a problem.

The advantages of the two-dimensional facial recognition methods are that they do not require the cooperation of the subject, they are quick and relatively inexpensive to conduct. Three-dimensional methods are more accurate but are also more expensive, take more time and, currently, require the cooperation of the subject.

Chapter four is a review of heritability in the craniofacial region. Twenty previous studies of facial heritability have been considered in some depth, many of these are twin studies and the vast majority used lateral cephalographs as their principal data source. In addition to the twin studies there were a few that also examined the relationships between singleton siblings and between parents and children. Of all these studies only a single one Weinberg et al. (2013) used structured light scanning similar to this study. The results of these studies are too variable to draw any overall conclusions from, but very few were able to demonstrate clear facial heritability especially between parents and offspring.

This review of heritability studies clearly demonstrates that this study is unique as there are no published peer reviewed studies using structured light scanning to compare facial morphology between parents and offspring and thus attempt to predict a genetic relationship.

Methodology

As described in chapter five, this is a study of 41 families, comprising 139 individuals, all White British. All subjects volunteered for the study and the children took part with their parents' full permission. A signed declaration from each parent was obtained, in which they stated that the children scanned with them were their biological offspring, mothers and fathers completed these declarations in separate locations and the forms were placed immediately in a 'ballot' style box. There was no further verification of genetic relationship, such as DNA testing, as this would have been too expensive and over intrusive.

Each participant underwent a facial scan and completed and signed a questionnaire. The questionnaire included personal details including: age; Body Mass Index (BMI), ethnicity including that of parents and grandparents; history of disease or trauma affecting the face; dental morphology and diet.

Detailed specifications of the technology and software used in this study can be found in Chapter five but briefly they are as follows: The scans were undertaken using a portable white light surface scanner. The equipment consisted of a white light projector and two cameras mounted on a stable but adjustable tripod. These were all connected to a laptop running FlexScan3D 3.1 and each piece of equipment was calibrated before each scanning session. Other software used in the course of the study included Viewbox and Morpho J. Viewbox 4.1.0.0 Beta allows landmarks to be placed on the surface of three-dimensional meshes and the subsequent alignment of meshes using a Trimmed Iterative Closest Point algorithm. Morpho J software is used to analyse the three-dimensional co-ordinates of the facial landmarks which have been identified.

Multiple scans were taken of each face in order to create a final three-dimensional image or ‘mesh’. The image of each scan was visually reviewed by the researcher immediately after it had been taken so that images of poor quality could be deleted and replaced immediately with a better image. The whole scanning procedure took about seven minutes per subject and the questionnaire took up to ten minutes to complete. Facial hair was a problem because of its light reflecting qualities and this could lead to the scanner not being able to pick up the shape of the underlying skin. Fathers with facial hair were a particular problem but they were scanned and the images used wherever possible.

Once the individual ‘meshes’ had been created they were analysed. In all, 7,176 Mesh-to-Mesh Values (MMVs) were calculated using Trimmed Iterative Closest Point Algorithm (TrICP) for each of the seven identified sections of the face which were named as follows *Whole face, Eyes, Nose, Mouth, Eyes-Nose, Eyes-Nose-Mouth,* and *Eyes-Nose-Mouth-Chin*. Receiver Operated Characteristic (ROC) analysis was conducted for each of the seven sections of the face within 11 predetermined categories of relationship. The MMV of three single features, (*eyes, nose* and *mouth*) were combined to form four combination areas which were analysed within the same 11 relationship categories. The variation caused by age difference and BMI difference between relative also analysed through correlation calculation.

Results

Chapter six details the results for each of the 11 relationship categories. Initially Principal Component Analysis (PCA) is found to be inferior to Trimmed Iterative Closest Point (TrICP), single features, joined features and combined featured are compared both within and against one another. The study shows that whilst it is not possible to determine a genetic relationship, parent-offspring or sibling-sibling, by analysis of facial morphology, using the methodology outlined, it is possible to exclude the possibility of a genetic relationship. Over all relationship categories the combined features are shown to have the greatest predictive power. Sisters have the closest heritable links and are shown to be more similar to each other than other relationship types.

Whilst age difference between parents and children was shown to have an effect on heritable similarities, there was a much stronger effect between siblings. BMI is shown in itself to not be an impediment to the similarity of family members but difference in BMI can impede the calculation of heritability.

Discussion

In chapter seven the results of this study are compared to the other most relevant published studies of facial recognition cited in the literature review. There is an acknowledgment that advances in technology have led to a change from using lateral cephalographs for heritability studies to a range of new methods including photographs, two and three dimensional images and casts. The relative merits of these approaches are discussed and the author reflects on the methods used in this study and compares the results from this study to those from other three-dimensional heritability studies, notably those conducted by Djordjevic et al. (2013) and Naini & Moss (2004). In common with Djordjevic et al. this study also found that Trimmed Iterative Closest Point algorithm (TrICP) was better than Principal Component Analysis (PCA) when assessing the heritability of facial features.

The possible causes for the difference in results between the preliminary study and the final study are discussed, including the possibility that larger samples may show a slight decrease in accuracy.

Very few heritability studies use the same methodology and therefore it is difficult to compare results. Previous studies do not point towards a particular area of the face showing greater heritability than others, although a significant number of studies did show consistently poor results for the mouth and possible reasons for this are discussed.

Recognition rates for this study are poor when compared to others but the author makes the point that the purpose of this study was entirely different, namely to link two different faces and not to match two images of the same face.

This study found that the best results came from comparing combinations of multiple areas of the face and this echoed similar findings from Smeets et al. (2010), the possible reasons for the superiority of this method are discussed in chapter seven.

The effects of varying ages and of varying BMI values are also discussed.

Conclusion

The conclusions set out in chapter eight are self explanatory, the principle ones being that while one cannot predict a genetic link using the methodology in the study one can certainly exclude such a link. One of the other interesting conclusions is that this methodology could be usefully applied to the matching of commingled bones.

Chapter 2: Aging of the Human Face

The chapter will focus first on the embryonic, foetal and childhood growth of the skull before focusing of the soft tissue growth of the main features of the face. This will cover the developmental growth during childhood and adolescence as well as the degenerative growth during adulthood. The sexual dimorphism of each feature will also be discussed. All of this will aid in the understanding of the differences between the faces of children and parents of both sexes.

2.1 Skeletal Growth and Development of the Skull and Facial Bones

This section will focus on the developmental growth of the skull. To understand the effects of age on the face we need to know how the underlying structures form and develop through life.

> “As poles are to tents and walls are to houses, so are bones to all living creatures, for other features naturally take their form from them and change with them”
>
> Galen (Roman physician circa 129-199AD)

As Subtelny (1961) states, you need an understanding of the skeletal structure before you can assess the impact and change in the overlying soft tissue

2.1.1 Growth

Firstly, what is growth? Growth is visualised by changes in size and morphology over time. Growth refers to a combination of increases both in size and in developmental maturity. From birth to adulthood the facial features increase in size and change shape until they reach adult proportions. Although these two aspects of growth are interdependent, at certain times in life one of these aspects may be more dominant than the other. (Scheuer & Black 2004). The disparity can be in either direction with the skeletal maturity more progressed than the size of an individual or the inverse where size has progressed more than skeletal maturity. In order to assess and study both of these aspects they will be related to chronological age.

There are a number of variables that can affect growth as a whole or one aspect of it. Sex is a very obvious one as it is well known that females mature sexually and skeletally at a younger age compared to males, due to hormones. Genetics can affect the rate of growth as well as the final outcome, a good example being the inheritance of height, genetics are also responsible for the many observable variations between population groups (Scheuer & Black 2000). Malnutrition can significantly affect growth rates and have a detrimental effect on the final height of an individual. Disease and illness, often linked with malnutrition in the developing world today, can have similar consequences as in both cases the body uses the nutrients that should be laid down for the growth of bones and other tissues, for other more immediate purposes. Even the seasons have been shown to affect the rate of growth (Scheuer & Black 2000) although this may be influenced by the effect the seasons have on the diet of populations reliant on the land for food. Due to the number of variables and the disparity between size and developmental maturity with chronological age it is hard to use growth as an accurate measure of actual age.

2.1.1.1 Embryonic development

The embryonic development refers to the first eight weeks of growth after fertilisation (Baker et al. 2005; Cohen 2005) after which it is referred to as a foetus and bears more resemblance to the adult human form. Once implanted into the uterine wall the embryo is connected to the mother and can start to grow. The first organs to form are the spinal cord, heart, gastrointestinal tract and the brain. The brain starts to show activity at six weeks and the growth of the cranial portion of the embryo then starts to take on its own distinct development. The cranial portion of the neural tube enlarges and separates into distinct features with the first areas to form being the olfactory, hearing and vision organs, the nose, ears and eyes (Fazekas & Kósa 1978). The middle layer of mesenchymal forms around these organs and develops into cartilage. Typically, the morphology of this cartilage resembles the basic shape of the bone which develops later (Fazekas & Kósa 1978). This is what enables the parts of the skull to be identified and differentiated at such an early stage. The cranial portion is both large and recognisable when compared to the rest of the body at this period of gestation, this is due to the fast growth of the brain and the need to protect it.

2.1.1.2 Bone growth

When compared to adults, children have a much larger proportion of organic component in the bone (Baker et al. 2005). Bone develops in one of two ways. Intramembranous ossification, also defined as direct ossification or desmal ossification (Fazekas & Kósa 1978), is the apposition of bone on tissue within an embryonic connective tissue membrane (White & Folkens 2000). Endochondral ossification is where bones are preceded by a cartilaginous model. The shape of this precursor is normally very similar to that of the bone in its final form (White & Folkens 2000)

Whichever form of bone growth is occurring they all start from small centres of ossification and the growth then spreads out from each centre. The number of centres varies between the bones and is noted in Table 2.1 along with the type of bone ossification. As shown in the table below, intramembranous ossification occurs in most of the cranial vault, this is because it is a simpler process than endochondral ossification and can occur faster and earlier in growth thus enabling the protection of the brain. The neurocranium during the foetal stage can be eight times as large as the viscerocraniam (Brigg & Martakis 1998).

From foetal development all the way through perinatal, neonatal, infanthood, childhood and into adolescence the skull continues to form and grow. Each of the separate bones grow at a different rate but most undergo rapid growth in utero and during the first year of life and then continue to grow at a much reduced rate until total fusion of the sutures occurs and the size of the skull is fixed in adolescence or later. Proportionally to the rest of the body the head is large in the early years of life and as the extremities grow a more adult ratio is achieved.

Bone	**Primary centres**	**Type of ossification**
Frontal	2	Intramembranous
Parietal (Paired)	1	Intramembranous
Temporal (Paired)	8 (3 by birth)	Intramembranous and Endochondral
Occipital	5 (4 by birth)	Intramembranous and Endochondral
Sphenoid	6	Intramembranous and Endochondral
Maxilla	4 (2 by birth)	Intramembranous
Palatine (Paired)	1	Intramembranous
Vomer	2	Intramembranous and Endochondral
Inferior Nasal Concha (Paired)	1	Endochondral
Ethmoid	3	Endochondral
Lacrimal (Paired)	1	Intramembranous
Nasal (Paired)	1	Intramembranous
Zygomatic (Paired)	3 (1 at birth)	Intramembranous
Mandible	2	Endochondral

Table 2.1 Ossification of the skull, created with information from White & Folkens 2000, Fazekas & Kósa 1978 and Baker et al. 2005

2.1.1.3. Specific Bones

The growth of each of the bones of the skull will now briefly be described to give an overview of the development of the skull as a whole and how this might impact upon the growth of the face.

Frontal

The frontal bone, forming the underlying shape of the forehead and brow area, develops as two separate, symmetrically identical parts. These are separated by the frontal or metopic suture until between one and four years when, in the majority of cases, it fuses; for a minority of the population fusion will not occur until adulthood or occasionally there are cases where the metopic suture is retained as a non-metric trait throughout life, which according to Fazekas & Kósa (1978) occurs in approximately 10% of the population.

Ossification of each side of the frontal bone starts in the seventh to eighth week of embryonic development from a single ossification centre at the glabella and radiates outwards (Fazekas & Kósa 1978, Gilbert & Segal 1958). By the third lunar (28 day) month the bone is only recognisable by the orbital margin on its inferior surface, however a dramatic change occurs over the following two months so that by five lunar months the metopic suture, supra-orbital notch, temporal line and characteristic curved shape of the fully developed adult bone are recognisable (Fazekas & Kósa 1978). Over the next five lunar months the bone grows upwards and backwards (Fazekas & Kósa 1978). The most anterior section of the metopic suture is starting to fuse and only separates at the superior portion to allow for the anterior fontanelle.

Parietal

The two parietal bones encase the brain, each formed in the same manner from a single ossification centre. Ossification begins at two lunar months in the region of the parietal eminence (Fazekas & Kósa 1978). Ossification then radiates in all directions with the bone appearing circular for the first few months and gradually changing to a more angular shape bordered by the sagittal, lambdoid and coronal sutures. Being the site of the primary ossification centre and consequently having the longest growth period, the eminence or parietal boss thickens and becomes more prominent over time, this boss is at its most prominent when the foetus approaches term. (Fazekas & Kósa 1978).

Temporal

Each temporal bone forms from four distinct parts, the squamous portion, tympanic portion, petrous portion and mastoid process. These four areas encompass the auditory canal and the three ear ossicles and form part of the lateral portion of the cranial vault. The squamous portion of the temporal bone, the thin vault like part of the bone extending above the auditory meatus, arises from three ossification centres (Fazekas & Kósa 1978). The first of these forms the zygomatic process and appears in the third lunar month and becomes a clearly recognisable feature before birth (Baker et al. 2005). The tympanic portion is a ring that fuses with the squamous portion around birth whilst the petrous portion, containing the inner ear, fuses into the other parts during the first year of life (Baker et al. 2005). After birth the mastoid process starts to grow posterior to the external auditory meatus and continues to grow into puberty.

Occipital

The occipital, the most inferior and posterior portion of the cranial vault, starts life in five parts. Two of these fuse before birth to create the squama (Baker et al. 2005), the four ensuing pieces, the squama, two lateral and a basilar part, then fuse together during childhood. First the lateral parts fuse to the squama at around four years followed by the basilar part at seven years leaving a foramen for the spinal cord.

Sphenoid

The sphenoid is the bone that holds the rest of the skull together. This bone articulates with no less than twelve other bones and is situated in the centre of the skull. Formed of three parts, the body and the greater and lesser wings, the body is fused to the lesser wings by birth but the greater wings do not fuse until the first year of life (Baker et al. 2005).

Lacrimal

Forming part of the orbit the lacrimal bones take until three years of age to develop recognisable adult shape (Baker et al. 2005).

Zygomatic

The zygomatic bones form a large part of the forward projection of the cheeks. Developing early and having adult shape by birth, the zygomatic is hidden in children by the thick soft tissue of infant cheeks.

Maxilla

During week six of foetal development the first ossification centre appears above the canine tooth germ and spreads out. By birth the characteristics of the adult bone are attained although they are smaller in size (Baker at al 2005). From birth to six years the maxilla extends forward at a rate of approximately 3.5mm per year, thereafter it only grows forward by another 4mm over the following ten years (Gilbert & Segal 1958). The maxilla undergoes constant change and remodelling from birth until puberty due to the constant loss and eruption of teeth (Baker et al. 2005). The curve of the maxilla needing to grow to hold the additional permanent dentition.

Mandible

The least connected part of the facial bones, the condyles of the mandible only articulate with the temporal bones at the mandibular fossae. The mandible forms in two parts split symmetrically and is recognisably U shaped by the third foetal month (Baker et al. 2005). Like the maxilla, constant development and change occurs in the mandible whist the dentition is still forming and erupting.

Nasal

The nasal bones form the defining angle of the nose as it protrudes from between the eyes, in the fifth week of foetal development a single ossification centre appears and the surrounding cartilage slowly transformed to bone (Gilbert & Segal 1958). Adult shape is attained early (Baker et al. 2005) but the nasal bones continue to grow until 14-17 years of age (Gilbert & Segal 1958).

Ethmoid

The ethmoid forms both part of the orbit and part of the nasal cavity and is adapted to accommodate the olfactory nerve. Ossifying from three centres it is predominatly cartilaginous at birth. Fusion of the three centres and most of the ossification is complete by three years, with only the perpendicular plate remaining to ossify slowly during childhood (Baker et al. 2005).

Inferior Nasal Conchae

Forming part of the lateral walls of the nasal cavity the inferior nasal conchae are recognisable by birth (Baker et al. 2005).

Palatine

The palatine which forms the posterior of the upper portion of the mouth and lateral parts of the nasal cavity attains adult form before birth but takes until puberty to reach adult size (Baker et al. 2005).

Vomer

Forming the lower part of the septum, the vomer begins ossification during the eighth foetal week. In its earliest form it is part of a thick cartilaginous septum. Ossification moves from the inferior posterior portion and moves forwards and upwards. The vomer completes ossification between the 13th and 15th year (Gilbert & Segal 1958).

Skull

After the appearance and fusion of ossification centres and the attainment of adult shape of each bone, the skull then grows and eventually fuses to form one solid structure. As mentioned above the relative proportion of adult size is much greater for the skull than the postcranial skeleton during childhood and only grows slowly during late childhood and adolescence to complete the last portion of growth. Over time, the increase in the height of the face is greater than that of the width, thus elongating the face. The face is also thrust forwards by the anterior growth of the facial bones (Briggs & Martakis 1998). Both the height and anterior growth is seen most obviously under the orbital region, the orbits as part of the brain casing reach their adult dimensions earlier than the rest of the face. The actual fusion of the sutures varies between individuals but the growth of the skull ceases around puberty.

2.2 Facial Soft Tissue Growth

Whilst it is important to have a knowledge of the growth and changing shape of the skull, it is clear that the observable changes in the shape of the face throughout life are also, and perhaps more significantly, affected by changes in the soft tissue of the face. In order to assess these changes and the reasons for them, for the purposes of discussion, the face will be split into the four main features; Eyes, Nose, Mouth and Chin. The reason for this is that most papers focus on one or two of these areas rather than the face as a whole. The features themselves also change in different ways and are affected by different issues, whether by the comparability of the different means of data capture or by the variations in the naming and placement of landmarks. Where studies provide data on more than one facial feature they will be split and examined in all of the relevant sections. Whilst this might mean that an overall view of the changing face as a whole is difficult, due to the proximity of the main features each section runs into the next only leaving the cheeks and forehead unstudied. A small review of total facial studies will provide some insight into the less studied more featureless areas such as the cheeks. A large number of developmental studies have been conducted which relate to the resultant effect on facial shape following the application of dental work. Subjects in these studies are often classed as I, II, III which refer to the occlusion and alignment of the upper and lower dentition (I = normal, II = overbite, III = under bite). The differences in morphology and growth between these classifications is sometimes mentioned in the reviews to show the normal variation but where orthodontic work has been carried out the subsequent effect is not discussed here, as the focus is on the normal changes in the human face. For each of the four features growth is separated into developmental growth, approximately birth to eighteen years, and degenerative growth, from eighteen years onward. The reason for creating a division between developmental and degenerative growth in this study is that the factors which influence these two types of growth are quite different, and it is important to qualify them before they can be applied to results in this study. Growth in the children and adolescents in this study is developmental, whereas growth in the older adults in this study is also degenerative. Sexual dimorphism is then approached separately for all ages as this will give insight into the variation in this study both between the children of opposite sex and between children and adults of the opposite sex.

2.2.1 Orbital Region

When it comes to the eyes and the orbital region there are very few studies which focus on changes related to age and sex. In a number of studies of the whole face, the orbital region is mentioned only briefly while the nose and mouth are covered in great detail (Albert et al. 2007). Due to the lack of published research in this area, this review of changes to the orbital region will have to be minimal. Studies which refer to measurements involving the eye itself have either been excluded from this review or only those sections referring to the external structure of the orbit region have been included. This is due to the methodology of the present study, explained in full in the methodology chapter, where the position of eye itself will not be captured but rather the external surface of the closed eye and its immediate surrounds will be the focus of investigation.

Table 2.2 provides the basic facts for each of the studies considered in this review of age and sex changes of the orbit region. The other reason for the lack of comparative studies is the smaller number of definable features in this area, thus reducing the number of landmarks and measurements possible.

2.2.1.1 Growth

For all sections an attempt has been made to split the review of ageing into two sections developmental growth, approximately birth to 18 years, and degenerative changes, 18 years and over. Many studies however have populations that span the age ranges mentioned above and not all studies present their results in ways that allow the resultant effects of the two forms of ageing to be recognised and separated. Where this is the case they will be presented in the section that best fits.

Study	**No of Subjects**	**Age (yrs.)**	**Sex**	**Longitudinal /Cross sectional**	**Measurements taken**	**Means of data collection**	**Population**
Fledelius & Stubgaard 1986	454	5-80	Both	Cross Sectional	Interpupillary distance, Orbital width, Hertel value	2D physical, exophthal mometer	Danish
Kunjur et al. 2006	26	18-15	Both	Cross Sectional	Eyebrow thickness, Size of upper eyelid	2D Photos	White, Indian, Chinese
Albert et al. 2007	Review of multiple studies	20-60+	Both	Cross Sectional	Surface patterns	Visual	Multiple
Sforza et al. 2009	888	4-73	Both	Cross Sectional	Biorbital width, Intercanthal width, Orbit height, Eye fissure length, Orbit area.	3D landmarks	Italian Caucasoid
Sforza et al. 2013	654	4-30	Both	Cross Sectional	Biorbital width, Intercanthal width, Orbit height, Eye fissure length	3D landmarks	Northern Sudanese

Table 2.2 List of studies which examine the sex and age related changes to the orbital region. Note that the population and measurements are recorded as described by the authors of the studies.

2.2.1.1.1 Developmental Growth

Fledelius & Stubgaard (1986) studied the interpupillary distance, orbital width and Hertel value of over 400 Danish subjects. Whilst the interpupillary distance, the distance between the pupils when in a relaxed forward focus, cannot be apply to the present study, the overall change in the distance between the eyes can be inferred from the result. The interpupillary distance, orbital width, also known as eye fissure length (Exocanthion to Endocanthion) and Hertel value (the protuberance from the lateral point of the orbit rim to the apex of the cornea) all increased from 2 to 18 years of age at a steady rate.

In the first of Sforza et al.'s (2009) studies a general agreement with Fledelius & Stubgaard (1986) was seen with all measurements increasing throughout childhood before levelling out in later adolescence. Sforza et al.'s (2013) second study, using the same methodology but on a different population, also shows this increase but in one measurement, the intercanthal distance (the distance between the endocanthions) a lifetime peak was reached at 16 years before decreasing.

2.2.1.1.2 Degenerative Changes

Each of Fledelius & Stubgaard's (1986) three measurements gave rise to a different process of change during adulthood. The interpupillary distance increased throughout adulthood while the optical width decreased and the Hertel value levelled out. This implies that the eyes themselves moved apart whilst each eye became narrower while remaining level in terms of its protrusion.

Albert et al.'s (2007) review of a number of studies summarises the visual variations in the area around the orbit from the age of 20 to over 60. Whilst these variations do not indicate any metric changes they do give some insight into the alteration in the surface texture. Alterations in surface texture, such as wrinkles, could be an issue when using surface-based data collection methods such as surface scans. In the third decade of life the early stages of crow's feet (horizontal lines at the outer corner of the eye) and vertical grooves between the eyebrows can be seen. These deepen in the thirties and by the forties the eyebrows start to drop and the upper eye lip starts to droop. By the age of fifty the area under the eye can start to bag and form a pouch while all the previous characteristics are accentuated.

In accord with Fledelius & Stubgaard's (1986) growth in optical width, both of Sforza et al.'s (2009, 2013) studies showed that the eye fissure length decreased throughout adulthood from approximately 18 years onwards. Although employing slightly different measurements, the biorbital width showed a similar pattern to the interpupillary distance described by Fledelius & Stubgaard (1986) that of levelling out or showing a slight increase in adulthood. Finally, the intercanthal distance levelled out with only a slight increase shown in females in their fifties. Overall the eyes become slightly wider apart whilst each eye also narrows.

2.2.1.2 Sexual Dimorphism

Kunjur et al. (2006) not only showed that there was a significant variability between races when it comes to the dimensions of the orbit but also showed that males had thicker eyebrows than females across the board. However, in all cases, with the exception of the lateral point in white people, females were shown to have a greater distance between the upper eyelid and the lower curve of the eyebrow. The distance from the exocanthion to the lateral point of the eyebrow was shown to be larger in males for the white population, which could mean a greater vertical distance but might be due to a greater horizontal distance between the two points due to a longer eyebrow.

In both of Sforza et al.'s studies (2009, 2013) variation in the developmental growth of the sexes was observed. The 2009 study saw growth spurts in females at both seven and eleven years of age in both the biorbital width and the eye fissure length, whereas in males this occurred at nine and thirteen years. The intercanthal distance simply increased at a steady rate in males from four to fifteen years while female growth levelled out at eight years of age. The second study (2013) does not show the same pattern. The biorbital width was seen to increase at a steady rate in both sexes with females reaching a peak at 14 years of age before decreasing. The intercanthal distance showed no sexual dimorphism and the eye fissure length increase steadily in males increasing by up to 8mm, whilst showing no overall change in females. There was however a sharp spike in the recorded length of the intercanthal distance at 14 years in females before returning to the length recorded at 13 years. This is likely to be an outlier caused by the cross sectional style of the study. In both studies male dimensions were larger than female except the previously mentioned intercanthal distance in the 2013 study which showed no significant variation.

2.2.2 Nose

While nasal proportions have been studied for the purposes of evaluation and diagnosis of dysmorphic syndromes such as Down’s syndrome (Farkas et al. 2001) and for the assessment of childhood developmental growth, the studies into the normal dimensional variation of adults are limited (Zankl et al. 2002). Many of the studies of the nasal area suggest to the reader that due to the obvious visual variation of nasal dimensions between races, studies can only give insight into populations comparative to that of each study (Uzun et al. 2006). This variation is shown clearly in Heidari et al.’s (2009) study of two ethic groups from Iran and Ercan et al.’s (2007) study of seven Turkish populations.

Table 2.3 provides the basic facts for each of the studies considered in this review of age and sex changes of the nasal area. As can be clearly seen for most of the last century lateral cephalograms were the main source of data which facilitated the analysis of the forward projection and the height and length of the nose, but not the width, surface area or volume of the nose. These additional metrics were first introduced by Ferrario et al. (1997) and have enabled a holistic view of the changing shape of the nose as growth and sex determine its growth.

2.2.2.1 Growth

2.2.2.1.1 Developmental Growth

It is the generally believed that a child’s nose is broad and flat whilst an adult’s nose is long, narrow and high bridged (Goldstein 1939), but in order to test that statement and understand the process between the two states, developmental studies have needed to be conducted and analysed. Below is a review of the most well-known and renowned studies in this field.

Study	No of Subjects	Age (yrs.)	Sex	Longitudinal /Cross sectional	Measurements taken	Means of data collection	Population
Goldstein 1939	500	3-21+ 74	Male	Cross sectional	Shape of nasal arch	2D Profile gauge	American Hebrew
Meredith 1958	80	5-14	Both	Longitudinal	Nose height	Lateral Cephalogram	American white
Posen 1967	30	3mth-18yrs	Both	Longitudinal	Nose height, nasal bone length, nose length, nose depth, angle of nasal dorsum.	Lateral Cephalogram	American Caucasian
Chaconas 1969	46	10-16	Both	Longitudinal	Multiple	Lateral Cephalogram	American Caucasian
Buck & Brown 1987	48 162	6-18 21-30	Both	Longitudinal	Nose area	Lateral Cephalogram	Northern European Caucasian
Meng et al. 1988	40	7-18	Both	Longitudinal	Nose height, nose depth, angle of nasal profile	Lateral Cephalogram	American Caucasian
Buschang et al. 1993	37	6-14	Female	Longitudinal	Angle of nasal dorsum	Lateral Cephalogram	French Canadian
Ferrario et al. 1997	503	6-12 11-14 19-32	Both	Longitudinal Longitudinal Cross sectional	Nose height, nose length, protrusion, nose width volume, surface and angles.	3D Landmarks	White Italian
Zankl et al. 2002	2500	0-97	Both	Cross sectional	Nose length, Nasal protrusion	2D Physical	Swiss
Ferrario et al. 2007	20	19-29	Both	Cross	Multiple	3D Landmarks vs 3D surface	Italian Caucasoid
Ercan et al. 2007	150	18-39	Both	Cross sectional	Multiple	2D landmarks, photos	Turkish
van der Heijden et al. 2008	Review	12-18	Both	Longitudinal	Nose length, nose height and nasal protrusion	Multiple	White
Sforza et al. 2010a	859	4-73	Both	Cross sectional	Multiple	3D landmarks	Italian Caucasoid

Table 2.3 List of studies which examine the sex and age related changes to the nose. Note that the populations listed here are as described by the author.

Goldstein was one of the early researchers who looked into the changing shape of the face through childhood and into old age. In his 1939 study he studied the shifting contour of the bridge of the nose through childhood and adolescence (3-21 years) and then on into old age (mean of 74 years). On a sample of American born Hebrews, he plotted the profile of the nose at the level of the lower rim of the orbits. Although not the normal position described as the ‘bridge’ of the nose, it was possible to use a profile gauge at this level and so procure more accurate measurements than previous studies had employed using wire bent around the bridge. The problem with using this equipment was that the resultant pinching and thus deforming of the nose meant that the width measurements may have been erroneous. The study concluded that at all depths from the apex, the width of the nasal bridge decreases from age three to 17 years of age with rapid decreases between three to five years and 11 to 13 years. This decrease is not proportional, with the greater decrease being further from the apex, showing that the angle of the nasal bridge becomes more acute with time. From 17 to 21 years this pattern starts to reverse and there is a slight increase in widths at all depths. Interestingly, Goldstein also indicated that there is no correlation between the width of the nasal bridge and the width of the alae, or of the whole face, showing that a narrow bridge does not necessarily indicate to a narrow nose to the alae, or a narrow face. Although not the purpose of Goldstein’s study, the steady increase of 10mm or more shown in nasal protrusion is calculated by the increased number of measurements possible at the lower depths between three and 21 years of age.

Meredith’s (1958) study into the height of the skeletal nose (nasion-anterior nasal spine) saw a steady and consistent increase in this measurement from five to fourteen years. Of the population studied 86% exhibited either a linear or a concave growth graph implying either constant growth or a slight decrease in growth as the top age of 14 was approached. An increase in growth was observed but it only rarely in this study population.

Posen (1967), along with Subtelny (Subtelny 1961), is a landmark author in the field of nasal growth. Posen is referenced by many authors and is constantly used for comparisons by more recent studies. He studied a group of American children from three months to 18 years of age taking regular (6 to 12 months) lateral cephalograms of the children in order to describe the growth of multiple dimensions. This remains one of the most extensive childhood studies of facial dimensions in terms of the age range of the participants. The length of the nasal bones, length of the external dorsum of the nose, the depth of the nose, the height of the nose and the angle or inclination of the nasal dorsum were the main dimensions considered. Posen showed that 90% of the final length of the nasal bones is attained by 13 years of age. This is achieved by steady linear growth from three months old to 13 years. Growth then ceases with the exception of two small growth spurts at 15 and 18 years old. The growth of the external dorsum of the nose (soft tissue nasion–pronasale) was constant throughout the period of the entire study. The depth or protrusion of the nose was measured both from the skeletal and soft tissue profile of the face with almost identical results. The depth of the nose showed a dramatic decrease from three months to six months of age before this trend is reversed and the depth of the nose increasing at a constant rate until 15 years old thus accounting for 95% of the total growth. Growth then halts for two years before a further increase at around 17 years. Both the height of the external nose and the anterior nasal cavity were measured, while they showed slight variations from each other the overall growth rates were similar. A steady and constant increase in nasal height from three months to 14 years accounted for 90% of the total height, the last 10% being achieved by erratic periods of growth and non-growth from 14-18 years. Both the angle of the upper (nasion-rhinion) and the total dorsum (soft tissue nasion-pronasale) were calculated. For the first 14 years the change in angle was the same for the two measurements. From three months to one year the angle decreased sharply which was consistent with the decrease in depth noted earlier. There followed a plateau over the next six month period when the angle did not change. From 18 months there was a period of consistent angle increase until 14 years of age. From this point forward the total dorsum angle sees little change, whilst the upper dorsum angle continues to increase suggesting that the profile of the nose will either straighten or start to hump around the rhinion position. Overall Posen’s findings showed a general anterior and downward growth of the nose and that while a large proportion of the growth of the face has occurred by the age of 13 years, there is still significant nose growth to follow after this age.

Chaconas (1969) took multiple measurements from longitudinal lateral cephalograms of 46 children taken yearly from aged 10 until 16 years old. All the nasal measurements varied significantly with age showing that growth is still occurring between each year. The least significant finding was the length of the nasal bones, suggesting that the nasal bones have stopped or are slowing in their growth in comparison to the soft tissue areas of the nose, before the study was conducted at age 10. A difference was noted in the directional growth of the nasal tip between class I and class II subjects. Whilst the soft tissue tip grows in a forward direction for class I subjects it grows in a more downward direction for class II subjects, leading to a higher probability of an elevated bridge. In correlation studies of each of the measurements taken, it was shown that a long nose was linked to a having a large mandible and that the rate of growth of the nose length matched that of the upper lip height, suggesting that they grow as a unit.

Buck & Brown (1987) studied the area of the nose as seen in profile from the age of six to 30 years in a combination of two longitudinal studies. The area is delineated by the lines between soft tissue nasion and nasion; nasion and anterior nasal spine (approximately); anterior nasal spine and low point on nose (not Subnasale) and then along the anterior profile of the nose to the soft tissue nasion. Whilst the method for calculating the area was not the most ‘high tech’ consisting as it did of cutting the traced area out of acetate and weighing it, it did reveal that from the age of six the nose grows until 18 years of age. In comparison with the second study population the average size at 30 years was not significantly different to the size at 18 years suggestive of a cessation of growth at aged 18.

Meng et al. (1988) focus on three measurements to evaluate the changing morphology of the nose through childhood and adolescence, nose height, depth and inclination. The depth of the nose and nose height increased across the entire age span of seven to 18 years, although a constant ratio of 3:1 was maintained between the upper and lower portions of the nose throughout this growth. The inclination of the upper nose (soft tissue nasion - pronasale) increased with the profile rotating counter clockwise whilst the angle of the lower nose (pronasale - projected anterior nasal spine) remained in the same plane or rotated clockwise dependent upon sex. Overall there was no great variation in the average rate of growth and it was noted that between 70% and 80% of the growth had already occurred before the study commenced at age seven.

In Buschang et al.'s (1993) longitudinal study of females, lateral cephalograms were taken at six, ten and fourteen years of age and the variation between each was assessed. By plotting the position of the nasion, rhinion and pronasale, the angle of both the upper (hard and soft tissue) and lower nasal dorsum could be addressed separately. The upper dorsum was shown to rotate counter clockwise over the age span in all subjects whilst the lower dorsum varied dramatically between subjects. Both clockwise and counter clockwise shift was noted which resulted in cancelling each other out and showed very little variation in an averaged population. The overall angle of the total dorsum increased over time, straightening out the nose. If, as previous cross sectional studies had shown, the variation of the lower dorsum was considered as an overall average, then the change in angle would be accounted for by the change in the upper dorsum. However, due to the longitudinal method of study, the change in angle can be more closely correlated with the lower dorsum. All of the changes occurred evenly over the age span studied.

In one of the first studies of nasal growth to use three dimensional data capture Ferrario et al. (1997) plotted five landmarks and took all possible measurements between them. They used two longitudinal study populations of children (6 to 14 years) and a cross sectional adult study population to assess both age and sex differences in the growth of the nose. While all linear dimensions increased, overall there was a greater increase in vertical dimensions than horizontal, indicative of a change from a wide short nose in childhood to a long narrow one in adulthood.

In their large cross sectional study of 2,500 subjects Zankl et al. (2002) studied the growth of nose length, nasal protrusion and philtrum length. They covered both developmental change and degenerative change with an age spectrum of zero to 97 years. There were 50 subjects included in each of the following groups: each year group for subjects aged zero to 28 years; each 'two year' group for subjects aged 29 to 44 years; each 'three year' group for subjects aged 45 to 86 years; a single group of 50 for the last grouping of 87 to 97 years. The three relevant measurements were taken directly using a calliper. Whilst the measurements from the nasal protrusion (subnasale - pronasale) and the philtrum length (subnasale - labiale superius) were similar to those from other studies, this was not the case with the length of the nose. The nose length is an unusual measurement, rather than the normal sellion or soft tissue nasion to pronasale measurement used by others (Ferrario et al. 1997, Heidari et al. 2009, Posen 1967) Zankl and colleagues used sellion to lower nasal tip. The lower nasal tip is defined as ‘the lowest visible point of the nose when looking at the calliper at a 90-degree angle’ (Zankl et al. 2002, pp389), in most cases this method of measurement will produce a longer length than using the pronasale. Zankl et al. showed that both nose length and nasal protrusion increased at a steady rate during childhood and only start to slow in later adolescence with a dramatic slowing at around 20 years of age. The results of philtrum length are discussed in Section 2.2.3.2 on the lips and mouth.

Van der Heijden et al.*’*s (2008) review of nasal growth studies showed how few there were and highlighted the difficulty in finding comparable dimensions across studies. They attempted to combine the results of a number of studies to calculate the age at which the nose reaches maturity and dimensions stop changing. They suggested 15.8 years for girls and 16.9 years for boys, although they did note that many other studies had shown a continuation of growth long into adulthood but these were seen as degenerative rather than developmental changes.

Sforza et al. (2010a) conducted an extensive study of nearly 900 subjects from the age of four to 73 years. Using the same population as that from their lip study (Sforza et al. 2010b), this gives a good overview of the age and sex related changes in the nose. The expected growth during childhood and adolescence was noted for all dimensions measured with the exception of one. The nasal tip angle, the angle between the pronasale and the two alae, decreased from four years all the way through the study changing from a flat nose in childhood, to a more pointed nose by early adulthood and continued to change in that way throughout adulthood. At the beginning of the study some measurements had achieved a greater percentage of their adult dimensions than that of others, indicating that parts of the nose grow earlier and faster than other parts and also suggesting that the proportions of the nose will change over time. By the age of four the nasal width and protrusion of the nose had achieved over 70% of their adult dimensions (the average in the 18-30 year age bracket), while nasal height only reached 62-67% of adult dimensions dependent upon sex; nose bridge length reached 54-58% and the area and volume of the nose achieved less than 50% of the adult dimensions by the age of four.

Although many of the same words are used to describe landmarks and measurements there is a good deal of discrepancy between authors as to the exact location of points such as the pronasale or the naming of measurements such as nasal height and length. The pronasale as defined by Rynn & Wilkinson (2006) and Stephen et al. (2003) is the most anterior point on the nose when the head is positioned in the Frankfort plane. However, as also encountered by Rynn & Wilkinson (2006) in their review of pronasale projection methods, different authors use different planes of reference to locate pronasale meaning that they are not directly comparable. In this review alone Posen (1967), Meng et al. (1988), Chacanas (1969) and Buschang et al. (1993) all use different planes to orientate their images, while others do not explain how they selected the landmark. A similar issue arises with the positioning of the soft tissue nasion. While some use the sellion when the skeletal nasion is not visible, most using lateral cephalograms project a plane through the skeletal nasion and onto the external surface in order to place it. However, once again the plane used to do this differs and some use the closest point to the skeletal nasion on the external surface whichever plane it is in. The wording of measurements can also change, whilst most authors refer to the depth of the nose as the distance from the anterior nasal spine (or a horizontal plane through it) to the pronasale, others (Ferrario 1997) call this the protrusion of the nose which most take to mean the distance from subnasale to pronasale. The same confusion can occur with height and length of nose while most refer to the height as nasion (skeletal or soft tissue) to subnasale and length as nasion to pronasale, Uzun et al. (2006) for example, swaps these definitions around. All of this can make it impossible or at least very difficult to make valid comparisons between studies.

While direct comparisons between studies are problematic the general trends in growth studies can be compared and reviewed. Most authors report a steady increase in nose length across childhood with Ferrario (1997) suggesting that there is a slight increase in growth rate from 11 to 14 years. The overall result for nose height is very similar. The nasal bone length, unlike the soft tissue nose length, does seem to slow or stop growth during adolescence. Chaconas (1969) suggests a halting or slowing at 10 years while Posen (1967) believes that whilst most of the growth is achieved by 13 years there are further small increases until 18 years. Nasal protrusion increases steadily throughout childhood while nasal depth is shown by Posen (1967) to decrease from three months to six months before then increasing, with Posen and Meng et al. (1988) indicating a plateau in growth at about 14-15 years of age before continuing a slow increase. Although not commonly measured, as it is not possible on lateral cephalograms, the width of the nose is shown by Ferrario et al. (1997) and Sforza et al. (2010a) to increase during childhood, although this may appear to counteract Goldstein's (1939) assertion that the nose gets longer and narrower, and provided that the rate of increase in height is faster than the increase in width, this is still possible. Sforza's decrease in nasal tip angle, and Ferrario's increase in height to width ratio is further evidence of the observed narrowing and lengthening of the nose as age progresses.

2.2.2.1.2 Degenerative Changes

Goldstein's 1939 study notes a very slight increase in the width of the nasal bridge between 21 and 74 years of age. This is most prominent at 10mm and 15mm of depth from the apex rather than at the crest or the fusion point with the cheeks.

Zankl et al. (2002) showed that though the rate of growth of both the nose length and nasal protrusion is dramatically reduced in comparison with the developmental rate, they do both continue to grow throughout life. In accordance with the general belief that older people have larger noses, the nose length is shown to increase at a steady rate from 30 to 90 years with an average 1cm increase over that time. The nasal protrusion does increase overall but shows a slight decrease between 30 and 50 years before increasing again over the rest of the lifetime. The actual increase from 30 years to 90 years is however only 1mm.

Running to the age of 73 years Sforza et al.'s (2010a) study confirmed that almost all nasal dimensions continue to grow throughout adulthood, again confirming the commonly held belief that old people have larger noses. The rate of that continued growth does depend on the dimensions themselves, the nose height and nose length growth rates both slowed dramatically after 30 years of age but did continue to grow gaining approximately 5% more in size than the recorded dimensions at 18-30 years. Nose width, protrusion volume and area all increased at a steady rate in later life reaching between 112% and 129% of the adult dimensions (the average in the 18-30 year age bracket) by the end of the study.

All authors discussed above who used the measurement demonstrated an increase in nose length throughout adulthood and agreed that the rate of growth is markedly slower than that seen in childhood. Although only accounted for in one study (Sforza et al. 2010a) the same pattern was true of nose height. Nasal protrusion in adults is reported slightly differently across studies, Sforza et al. (2010a) suggest it increases at a steady rate while Zankl et al. (2002) show a dip in growth at around 50 years before an increased growth rate until 90 years old. Nasal depth is not measured in any of the adulthood studies but it is expected to follow approximately the course of nasal protrusion. Nasal width also continues to increase all the way through adulthood as shown by Sforza et al. (2010a) but the nasal tip angle decreases showing the nose becoming more pointed and visually appearing narrower. In accord with all the two dimensional measurements, the surface area and volume also increase throughout life. Due to changes in the angles of the nose, it is shown by some studies that in certain people the nasal bridge can become more prominent during late adolescence and adulthood thereby forming a hump. All of these findings lend credence to the commonly held belief that older people have larger noses.

2.2.2.2 Sexual Dimorphism

Meredith (1958) interestingly noted that he saw no significant variation between the boys and girls, aged 5-14 years in his study of nose height, either between the actual measurements or the rates of growth.

In all of Posen’s (1967) dimensions, he saw no significant variation in the rates of growth between girls and boys. As expected he did observe greater dimensions overall in boys compared to girls at all ages with only a few exceptions. The depth of the nose only showed boys to have greater depth until two years, after that age the average depths were equal for both sexes. The angle of the nasal dorsum showed significant variation between the sexes from six to 16 years with girls having a greater angle than boys, both before and after this age band the angles are roughly similar between the sexes. This greater angle is not shown in the depth of the nose, suggesting that there must be some anterior movement in the lower portion of the face to mask this.

In his study of nasal growth Chaconas (1969) showed that from 10-13 years overall female nasal measurements increased at a faster rate than between 13-16 years and also faster than boys of the same age. Boys, however, grow more overall across the whole six year span. Whilst girls' nasal proportions changed very little over this time of growth, males showed a significant downward shift of the nasal tip between the ages of 13 and 16.

Whilst the overall trend of males having larger dimensions was seen in Buck & Brown's (1987) study the actual dimorphism changed dependent upon age. Whilst at 15 years old the depth and area of the nose were significantly different between the sexes, by the age of 18 the height and length of the nose also showed significant differences. The area of the nose did increase and grow from the age of six to 18 for both sexes but the rates of growth differed. The female growth rate was at its peak between nine and 12 years, while for males is was 12 to15 years. The study also revealed that while male noses grow proportionally, females do not, indicating that the proportions of the female nose change with age rather than simply becoming larger.

Meng et al. (1988) showed the nose height and depth to be larger in males than females, female growth was mostly completed and had started to plateau aged 14-15 years whilst male growth continued until the end of the study at 18 years old. Whilst the angle of inclination of the upper nasal profile was very similar across the sexes in the younger years, at approximately 14 the females' angle of inclination cease to move, while males continue in the counter clockwise rotation. The angle of the lower nose remains constant in males while increasing and rotating clockwise in females. In combination with the upper angle this means that the angle of the nose at the pronasale decreases, becoming more pointed, this is more exaggerated in females than males.

In Ferrario et al.'s (1997) study the most obvious sexual dimorphism was seen in nasal volume, females showed a large growth spurt at 11 years whereas males showed a less extreme but more consistent growth increase between 11 and 14 years of age. By the end of the childhood study at 14 years female nasal volume was almost that of the adult, whereas males had a large amount of growth still to occur.

For all proportions females were closer to adult proportions in comparison to males by the age of 14 demonstrating the earlier age of maturity.

Zankl et al. (2002) show males' dimensions (nose length and nasal protrusion) to be generally larger overall than females but the rate of growth to be almost identical between the sexes. Females do reach the point of developmental maturation at a slightly younger age.

Ercan et al. (2007) in direct contrast to almost all other studies found that females in their Turkish population showed a greater nose height than males. The other measurements were more in line with other studies having larger male dimensions in comparison to females, with approximately 50% of the dimensions taken showing significant sexual variation. It is likely that this is a race specific deviation from the norm, but is a reminder that population specific data are always needed when applying metric information to an individual.

Whilst Sforza et al. (2010a) principally focused on the age variation, they did report the major points of sexual dimorphism seen within their study. For all linear measurements, the surface area and volume of the nose, males displayed larger values than females within the same age category with very slight exceptions for a few measurements such as nose height and length, where in early adolescence (12-13 years) females briefly overtook males. The only other significant difference is that of the width to height ratio which indicates that males have a wider, shorter nose in comparison to females' longer thinner noses. The angles measured showed no significant sexual dimorphic variation.

In summary, the majority of the reviewed studies found that male dimensions were larger than female dimensions at almost every stage of life (Buck & Brown, 1987; Ferrario et al., 1997; Meng et al., 1988; Posen, 1967; Sforza et al., 2010a; Zankl et al., 2002). The only study which did not accord with these findings was that of Ercan et al. (2007) although they suggest that this is likely to be a population specific characteristic. There is however a variety of views on the rate of growth, while Meredith (1958), Posen (1967) and Zankl et al. (2002) report little or no variation in growth rate between the sexes, Meng et al. (1988), Buck & Brown (1987), Ferrario et al. (1997), Zankl et al. (2002) describe females as completing growth at a younger age than males. In situations where growth continues throughout life, the attainment of adult proportions, as defined by the average size between 18 years and 30 years, is reached at an earlier age and females have a higher percentage of completion from approximately 10 years onwards. This earlier completion of growth by females is caused by their earlier growth spurt, noted by Chaconas (1969), Meng et al. (1988), Buck & Brown (1987), Ferrario et al. (1997) and Sforza et al. (2010a) who generally agree upon an increased growth rate in females between 10 and 13 years of age whereas the male growth spurt is observed at between 13 and 16 years.

It terms of shape changes and variation between the sexes Chaconas (1969) indicates that female growth is proportional, meaning that the overall shape does not change significantly over time, while males show a more downward movement of the nasal tip which is supported by Meng et al. (1988) as they report a change in angle of the lower border of the nose in males. Sforza et al. (2010a), although the only study to suggest it, described a shape variation between the sexes with males having a shorter and wider nose relative to the longer and narrower nose of females. Due to the nature of Sforza et al.'s (2010a) study which uses three-dimensional landmarks, and covers most of the human life span (four to 73 years) and including, as it does, nearly 900 subjects, it would be prudent to rate this result quite highly as there are few studies which give similar insight into the shape variations of the nose.

2.2.2.3 Health Issues and Deformity:

The height of the nasal cavity in relatively unaffected by fractures of the nasal bone during youth, though it is thought that when nasal trauma or surgery does occur, the younger the individual is at this time the more likely it is that permanent damage will occur (Gilbert & Segal 1958).

2.2.3 Mouth and Lips

The changes in lip shape due to age and sex have been studied for a variety of reasons including: knowledge required for plastic surgery on trauma or genetically damaged lips (Zhu et al. 2008); minimisation of the effect of orthodontic procedures on the lip's appearance (Burstone 1967, Mamandras 1984); creation of more lifelike results from cosmetic implant procedures (Iblher et al. 2008); ageing of children in pornographic videos (Cattaneo et al. 2009); improved results of facial reconstruction (Stephan & Henneberg 2003, Stephan & Murphy 2008); building a database of population specific changes (Sforza et al. 2010b).

Studies can be difficult to compare due both to the nature of the data collection (three-dimensional, two-dimensional, lateral view, frontal view, photos, radiographs and direct measurements) but also the variation in the naming of landmarks and measurements especially when comparing modern studies to older ones. This variation in naming and positioning is much more widespread than the issues previously identified with the nasal area, for this reason comparisons are made as each study is reviewed as there a very few situations where direct comparisons between measurements can be made.

Many studies include, or are focused on, the skeletal changes of the maxilla and mandible; while this is of great importance in understanding the changes in shape of the labial area this review will mainly focus on studies which demonstrate some information relating to the soft tissue changes as this is more relevant to this study.

Table 2.4 provides the basic facts for each of the studies considered in this review of age and sex changes of the mouth and lips. As can be seen, until the turn of the millennium the only useful data came from lateral cephalograms providing two dimensional data of the horizontal proportions, and the anterior/posterior measurement at the midline, but these contain no information on the transverse dimensions of the mouth. From the year 2000 onwards, multiple studies have given insight into the third dimension and the overall shape changes that result from the combination of linear variation in multiple directions.

Study	No of Subjects	Age (yrs.)	Sex	Longitudinal /Cross sectional	Measurements taken	Means of data collection	Population
Vig & Cohen 1979	50	4-20	Both	Longitudinal	Lip height	Lateral Cephalogram	British
Oliver 1982	40	12-15	Both	Longitudinal	Lip thickness	Lateral Cephalogram	American
Mamandras 1984	28	8-18	Both	Longitudinal	Area of lips	Lateral Cephalogram	White Canadian
Mamandras 1988	32	8-18	Both	Longitudinal	Lip length and thickness	Lateral Cephalogram	White Canadian
Park et al. 1989	15	9-20	Both	Longitudinal	Multiple	Lateral Cephalogram	American
Ferrario et al. 2000	1347	6-32	Both	Longitudinal / Cross sectional	Mouth width, vermillion heights, lip heights, areas, volumes	3D Landmarks	White Italian
Zankl et al. 2002	2500	0-97	Both	Cross sectional	Philtrum length	2D Physical	Swiss
Lévêque & Goubanova 2004	100	20-80	Female	Cross sectional	Mouth width, lip height	2D physical	Russian
Iblher et al. 2008	182	5-83	Both	Cross sectional	Mouth width, Lip height, vermillion height	2D landmarks Photos	Caucasian
	60	20-35 65-80	Female		Lip height, Lip thickness, area	MRI	
Zhu et al. 2008	1500	2-12	Both	Cross sectional	Philtrum length, area and widths of upper lip.	2D physical	Chinese Han people
Sforza et al. 2010b	918	4-73	Both	Cross sectional	Mouth width, philtrum width, vermillion heights, lip height, areas, volumes	3D landmarks	Italian
De Menezes et al. 2011	40	21-34 45-65	Both	Cross sectional	Lip thickness, vermillion area, lip volume	3D Casts and Landmarks	Italian Caucasian

Table 2.4 List of studies which examine the sex and age related changes to the mouth. Note that the population and measurements are recorded as described by the authors of the studies.

2.2.3.1 Growth

2.2.3.1.1 Developmental Growth

Vig & Cohen (1979) use serial lateral cephalograms from the age of four to 20 years of age to assess the growth of the upper and lower lip in relation to anterior lower facial height. They discovered that the lower lip grows more vertically, both actually and proportionally, than the upper lip especially from the age of nine to 11 years. They also showed that the growth of the total lip in height, a combination of upper and lower, exceeds that of the anterior lower facial height indicating that, visually, the lips appear larger on the face with age rather than growing in proportion to the rest of the lower face.

Much like Vig & Cohen (1979), Mamandras (1984) used serial lateral cephalograms from the age of 8 to 18 years of age to measure facial growth. Although restricted to a single type of measurement due to the two dimensional aspect of the study and having only a lateral cephalometric view, Mamandras' (1984) study is longitudinal and therefore provides an insight into the ageing of specific individuals rather than an averaged population. The area of the cross section of the upper and lower lip were measured and results indicated that both increase with age from 8 to 16 years when both start to level out. As these cephalograms were only viewed in two-dimensions it is difficult to infer what effect there would be if viewed in three-dimensions, however there is a wide ranging accord with other studies that there is general growth until adolescence.

In a parallel study Mamandras (1988) used the same population to evaluate upper and lower lip length and thickness. These measurements are not directly comparable with some other studies as they are measured using different landmarks than the thickness and length measurements in Oliver (1982), Park et al. (1989) and De Menezes et al. (2011) among others. The thickness measurement, but only in the upper lip, is most similar to that used by Iblher and colleagues (2008) which is also taken at what they call the 'mid lip' but the similarity cannot be confirmed as Iblher et al. (2008) do not describe how they locate the mid lip. The 'length' or 'height' as many studies describe it uses the palatal plane to form the uppermost point and thus can only be directly compared to other studies that incorporate combined skeletal and soft tissue data such as Vig & Cohen (1979). It was found that both the upper and lower lip length increased between 8 and 18 years at slightly different rates while the thickness of both increase between 8 to 16 years and then started to decrease in the case of the upper lip and level out in the case of the lower lip.

Ferrario et al. (2000) use the Ferrario method which uses optoelectronics to digitise landmarks in a three-dimensional space in order to assess the lengths, areas and volumes calculable from six landmarks. Children from the age of 6 to 18 years were assessed in a mixed longitudinal and cross-sectional study. The growth of all facial dimensions was assessed and compared and it was found that, although all dimensions increased with age, the rates of those increases were very varied. Total lip volume, lower lip volume and lower lip vermilion area all doubled in size over the age range considered, whereas the upper lip volume only increased by one third. The upper vermilion area, mouth width and all heights; vermilion, upper lip and total lip, increase by a fifth. The lower lip height was not measured in this study so it not possible to confirm Vig & Cohen’s (1979) findings that the lower lip height grows more than the upper. It is however possible to confirm that, due to the variability in growth rates the morphology of the face must, of necessity, change across this age range and does not simply grow in a proportionate fashion. More insight into the variability of growth within the age range is also given in this study. The attainment of adult dimensions of the upper lip are reached faster and earlier than those of the lower lip, while this does not directly agree with Vig & Cohen (1979) it could be deduced that as the lower lip continues to grow for a longer time period than the upper, then it must reach a greater size overall as shown with the area and volume measurements. For all the measurements assessed 95% of the adult dimensions were reached by 13 to 14 years in females and 15 to 18 years in males. In general agreement with Zhu et al. (2008) who write about the comparison between the mouth width and height and their proportional growth, the Ferrario method suggests that the vermilion height reaches its adult proportions earlier than the mouth width. Note that the adult proportions described by Ferrario et al. (2000) are those of young adults and do not take into account any further changes in dimension caused by degenerative aging.

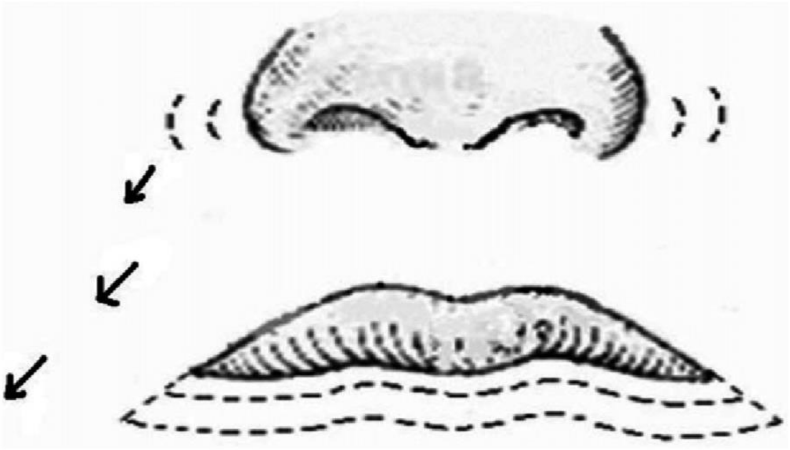


Figure 2.1 Diagram of the development of the upper lip taken from Zhu et al. 2008

In Zhu et al. (2008) a study on healthy Chinese children from two to 12 years based on direct facial measurements, they focused on the area between the top lip and the nose, in this study referred to as the upper lip. All measurements taken showed an increase with respect to age, however the rate of that increase changed dramatically and only some measurements showed any statistical significance. All facial measurements showed their greatest increase between two to three years. The length of the cheilion to the related crista philtri increases dramatically after five years of age and the area of the upper lip, excluding the philtrum, shows a faster rate of growth at six and then again at ten years of age. This shows that even within the small area of the upper lip, different parts grow at specific ages, changing the overall shape of the lips dramatically over this age span. The one area that seems to change very little is the height of the philtrum, visually this means that although the lips themselves may grow, the distance between the base of the nose and the top of the upper lip stays the same. This is shown in Figure 2.1. In accordance with other studies they conclude that the width of the mouth grows at a significantly increased rate compared to the height giving the impression of an increasingly thinner mouth over the course of childhood.

Overall the authors in this field agree that the upper lip is smaller (Ferrario et al. 2000; Mamandras 1984; Mamandras 1988) grows at a slower rate (Ferrario et al. 2000; Vig & Cohen 1979) and reaches adult dimensions at an earlier age than the lower lip (Ferrario et al. 2000). The growth of the upper lip tends to reach its climax at 16 years of age (Mamandras 1984) and can even decrease (Mamandras 1988) while for the lower lip it is 18 years. In summary the lips grow more horizontally than vertically (Ferrario et al. 2000; Zhu et al. 2008) creating an impression of thinning lips with increasing age and compared to the length of the face the lips grow more, thus making them look larger in comparison (Zhu et al. 2008)

2.2.3.1.2 Degenerative Changes

Lévêque & Goubanova’s (2004) study into the more visual aspects of ageing studied women from 20 to 80 years of age. Their conclusions are in accord with the general view that the visible height of the lips decreases with age after the third decade. In this particular case the ‘lips’ refers only to the vermilion height. The width of the mouth is shown to increase with age at a steady rate throughout the range of this study. In addition to most other studies the dimensions of: contraction; a pursed pose; a grin; and an extension of the mouth were also recorded. The contracted width demonstrates a similar statistical increase to the normal mouth width whereas the extended width shows no significant variation across the ages. In order to understand fully what occurs with ageing further calculations were also undertaken. The extensibility of the lip, (calculated as (extended width-normal width)/normal width), was found to be negatively correlated with age while contractibility calculated from the pursed width was insignificant. The combination of these results showed that the increase in the purse width results from the increase in normal width and that the actual extension of the mouth does not change but the variation between the lengthening normal width and the grin simply becomes less.

Although not as clearly observable as size and shape are to the changing morphology of the lips, changes to the skin surface are also symptomatic of possible similarities or variations between the ages and can be examined in a number of ways. This review will only focus on those that would be identified by a surface scan therefore hydration, colouration and surface patterning will not be discussed. The number and visibility of the radial wrinkles on the vermilion surface were analysed. Both the number and visibility of wrinkles, and their multiplication product are correlated positively with age. The product shows an exponential increase beginning at about 30 years of age and the visibility of wrinkles increases dramatically in the forties.

Iblher et al.’s. (2008) MRI study of two groups of women, 20 to 35 years old and 65 to 80 years, indicates that the thickness at both the vermilion border and the midpoint of the upper lip decrease between these age groups and in accord with Mamandras’ (1984) study of the cross section of the upper lip, shows no statistical variation. This possibly shows what the volumetric studies (De Menezes et al. 2011; Sforza et al. 2010b) have shown, that the thickness and height of the vermilion decreases over time, flattening the lateral view, the volume is not lost but redistributed. The study also shows an increase in upper lip length, defined as subnasale – stomion, between the two groups. This result is further evidenced by an adjacent study consisting of photographs of both sexes from five-83 years. This increase in the upper lip length is caused by an increase in the prolabium (Subnasale – Labial superius) as the height of the upper vermilion (labiale superius – stomion) in accordance with other studies (Sforza et al. 2010b) actually decreases in the ageing face. This is contrary to the proposal of Zhu et al. (2008) that the distance between the subnasale and the top of the upper vermilion does not change. This discrepancy may be due to the variation in the age ranges studied but may also be due to the disparity of data collection methodology, two-dimensional photos verses three-dimensional direct landmark measurements. Iblher et al. (2008) in accord with Penna et al. (2009) suggest that the prolabium height does not actually change but that the visible height, when viewed from the front in Frankfort plane, increases due to the flattening and drooping of the upper lip obscuring the vermilion.

Penna et al.’s (2009) study of autopsy specimens examining the histological changes of the upper lip, concluded that the generally accepted loss of lip volume with increasing age is incorrect and that what has been observed is a drooping of the upper lip.

Sforza et al. (2010b) conducted a cross sectional ageing study of Italians using landmark data collected by an electromagnetic digitizer from four to 73 years of age. From the nine landmarks collected, the volumes, areas, heights, widths and ratios were calculated and regression analysis was performed on the average measurements for specific age ranges. In contrast to some studies the results showed that all measurements, with the exception of the vermilion height to mouth width ratio and vermilion height of upper lip, both of which decrease from birth, increased during childhood as would be expected due to the general growth of the face. Around adolescence the growth rate of mouth width and lip volume both slow and the vermilion area and height start to decrease so that by young adulthood the lip height growth reduces and stabilises.

Using a similar methodology to Iblher et al. (2008), De Menezes and colleagues (2011) conducted a study based on two groups of subjects belonging to two age ranges. The first group of 21 to 34 year olds is almost identical to the group used by Iblher et al. (2008) whereas the second group were younger at 45 to 65 years. Both groups, in complete contrast to Iblher et al., were of mixed gender. De Menezes and colleagues employed an overly complicated method, which consisted of using stone labial models cast from both the dental layer and external labial area. These casts were then digitised using a microscribe to calculate the volume, thickness and area of both upper and lower labial areas. Although the volumes showed no significant changes between the two age groups, the area and thickness of both the upper and lower lips decreased with age across both sexes. This agrees with Iblher et al.’s (2008) results for thickness and Sforza et al.’s (2010b) results for area.

Possible causes for the changes in lip morphology after initial growth have been suggested by a number of authors, however most are merely conjecture and have yet to be studied or conclusively proved. The suggestions include: gravity (Sforza et al. 2010b); loss of elasticity due to degeneration of collagen fibres (Penna et al. 2009); hormonal changes; repetitive use and loss of hydration (Lévêque & Goubanova 2004).

Caisey and colleagues (2008) were able to endorse the loss of elasticity in the lips using ballistometry but added that it was only affected by normal ageing and not by onset of the menopause or hormonal replacement therapy. Lévêque & Goubanova (2004) do however suggest that hormones can play a role due to the nonlinear decrease of total lip height that they observed beginning in the third to forth decade, as this is the age when female hormones start to change. The study however provides no evidence for this speculation. The lengthening width of the mouth could simply be caused by the repetition of lip extension and the increase of wrinkles, owing to the increasing inability of the lips to recover from the deformation caused by speaking and eating (Lévêque & Goubanova 2004).

The above studies all demonstrate that the lips grow during childhood but become thinner, wider and lose surface area, or the area is hidden, as the adult face ages. This conflicts with the belief of those women who opt for lip fillers to counteract the effect of ageing as the volume of the lips actually increases throughout life.

2.2.3.2 Sexual Dimorphism

A number of the studies referred to in the section below have already been discussed above in the review of age related changes as there are few studies solely focused on sexual dimorphism. It is accepted that where the sexes concur and show the same changes in shape or in size, this will have been discussed as a consequence of growth or degenerative ageing, therefore only data which show variation between the sexes will be discussed in this section.

Burstone (1967) whilst looking at the positional changes to the lips caused by malocclusion of the teeth noted that the upper (subnasale – stomion) and lower (stomion – gnathion) lip height is greater in males than females in an adolescent population.

Oliver (1982) provides yet more evidence to show that the variation between the teenage sexes occurs or is significantly increased somewhere between 12 and 15 years. Although not the purpose of the study, Oliver (1982) notes that he saw no significant variation between the sexes in his pre-treatment measurements of lip thickness at age 12 but by 15 years old and after treatment, there was a significant difference between the males and females involved.

Although mainly focused on the skeletal elements seen on the lateral cephalograms, Bishara et al.'s (1984) longitudinal study of 5 to 25 year olds showed that the greatest variation of facial dimensions in females occurs between 5 to 15 years whereas in males the variation is equally spread from 5 to 25 years. The extra range in age up to 25 years helps to show the influence of puberty on the growth rates and ages, demonstrating that females complete growth both quicker and earlier.

Mamandras' (1984) longitudinal study of lateral cephalometric radiographs of 8 to 18 year olds showed an increase in the area of the maxillary and mandibular lip area in both sexes, the rate of this increase however differs between the sexes. The maxillary lip area was always found to be greater in males and the percentage increase from 8 to 18 years was twice that of females. The mandibular lip area was larger in females until the age of 12 when the position was reversed and from then on was larger in males. It has been suggested that this may be due to the younger age at which females reach adolescence and to the later growth spurt of males.

In his follow on study of the same population Mamandras (1988) noted that males' length and thickness measurements are greater than the females' at all ages with one exception. As would be expected from the results of the mandibular lip area in the preceding study, the mandibular length also appears to be larger in younger females before being overtaken by the males. The age at which this occurs is slightly different from that of the area results with females showing more length at eight years, the sexes then become equal at 10 and 12 years before the males then overtake at 14 years. The only statistically significant results between the sexes shown in this study are those for the maxillary lip length and thickness and for the mandibular length from age of 14 to 18, showing that the variation between males and females increases as they grow. The mandibular thickness showed no significant sexual dimorphism across the entire study. It is interesting that neither of Mamandras' studies (1984, 1988) shows any sign of female growth retardation due to the earlier adolescent growth spurt completing. This may be due to the study ending at 18 years of age, perhaps if the study had carried on for a greater span of years this this would have been observed.

In correlation with most growth studies, males were larger than equivalent aged females in all the measurements taken in Ferrario et al.'s (2000) study. These measurements were not significant in the six to eight year age range, thus showing similarities, although to a lesser degree, with Mamandras' (1984, 1988) findings, even though the measurements taken in the two studies are not directly comparable. Ferrario et al's (2000) study did however show a significant difference between 8 to 10 years displaying disparity with Mamandras (1984,1988) but the findings once again show similarities for the 11 to 14 year range which present no significant difference between the results of the two studies. At this stage it is suggested that the female growth spurt and attainment of adult proportions is occurring. The only specific exception to the sex based variation is observed for the vermilion height to mouth width ratio which shows no significant sexual dimorphism across the entire age ranges studied. This indicates that although the mouth area might be growing differently in male and females the proportions of the mouth remain constant for both sexes.

Although mainly focused on the nose, and discussed in more detail in section 2.2.2.1.1, Zankl et al. (2002) included philtrum length (Subnasale-Labiale superius) in their study. Whilst both sexes showed an increase in philtrum length during childhood, followed by a decrease and further increase after the age of 30, the age at which the decrease in length commences, varies. In females the decrease starts at approximately 12 years and reaches the shortest length at 23 whilst in males the decrease starts much later at 18 and plateaus at 24 years before increasing again.

Sforza et al.'s (2010b) study into the variation of the lips showed that at all ages total, upper and lower lip volumes, mouth width and lip height were all larger in men than in women. The one slight anomaly was vermilion height (Labiale superius-Labiale inferius) to mouth width (Cheilion left – Cheilion right) ratio, which is greater in girls than boys but is reversed around age ten and is then larger in adolescent males than females before switching again at about 17 years with females then having a greater ratio than males into old age. Visually this means men have thinner lips than women throughout life, with the exception of the adolescent years.

In De Menezes (2011) study all the dimensions, volume, thickness and area were greater in males than females in both the adult age groups studied. This is in line with all the studies cited above and all have shown that, overall, adult males are always larger than adult females although this is not always the case in childhood and adolescence.

Overall almost all studies confirm that the male dimensions are larger than females', the only exception to this is seen in the lower lip around 10 to 12 years where females briefly outsize males (Mamandras 1984, 1988). This is thought to be due to the earlier growth spurt and attainment of adult proportions seen in females (Bishara et al. 1984, Ferrario et al. 2000, Oliver 1982, Sforza et al 2010b). However, although females reach adult proportions earlier their rate of growth is slower than males (Oliver 1982, Mamandras 1988).

2.2.3.3 Health Issues and Deformity

When it comes to the effect that dental work, in particular extraction of the teeth, has on the external shape of the soft tissue there is much debate. A number of studies have been undertaken but most contain factors that render their results open to question. In Park et al.'s. (1989) study of North American Black patients undergoing extraction of all four premolars, the pre and post treatment lateral cephalograms were compared for significant changes. The only change in the soft tissue which was of significance was the increase in nasolabial angle (the angle formed by two lines intersecting at subnasale, one tangent to the lower border of the nose and one passing through labrale superius). Assuming that the nose is unaffected this simply means that the protrusion of the upper lip decreases, which is what would be expected as the anterior teeth will recede to fill the gaps created by the extraction. However, this study was conducted on patients who were between the ages of 9 and 16 years at the beginning of treatment and no standard was used to take into account the normal changes due to growth. In addition, no explanation of the reasons for the extraction of teeth was given. Burstone (1967) shows that deformity and misalignment of orthodontics can greatly affect the positioning of the lips, however the lips themselves are generally unaffected. Oliver (1982) states that subjects with thicker lips, the top 25% of the study population, are unaffected by incisor retraction (the use of braces) while those with thinner lips, the bottom 25%, are affected when it comes to the resultant effect on the soft tissue positioning. The study was unable to conclude whether it was simply the positioning or the size of the lips themselves or both of these factors that caused this variation. It could be inferred from this result that boys, having thicker lips at the age that the majority of orthodontic work is performed, are less affected when it comes to the final positioning of the labial soft tissue than girls.

2.2.4 Chin

Due to the severe lack of recognisable landmarks on the surface of the chin there is only a limited pool of studies focused on this area and those that do rely almost entirely on soft tissue depths at the pogonion. Table 2.5 provides the basic facts for each of the studies considered in this review of age and sex changes of the chin.

Study	**No of Subjects**	**Age (yrs.)**	**Sex**	**Longitudinal /Cross sectional**	**Measurements taken**	**Means of data collection**	**Population**
Subtelny 1961	NA	3-18	Both	Long	Multiple	Lat Ceph	NA
Singh 1990	60	10-21	Both	Long	Soft tissue thickness	Lat Ceph	White American
Nanda et al. 1990	40	7-18	Both	Long	Pogonion thickness, Mandible protrusion, Angle of chin	Lat Ceph	Caucasian
Formby et al. 1994	47	18-42	Both	Long	Pogonion thickness	Lat Ceph	White American
Prahl-Andersen et al. 1995	82	9-22	Both	Long	Pogonion thickness	Lat Ceph	Dutch
Franklin & Cardini 2007	79	1-17	Both	Cross	Ramus height	3D Landmarks	South African, African American
Franklin et al. 2008	79	1-17	Both	Cross	Ramus height, Gonial angle, Mandible shape	3D Landmarks	South African, African American
Thayer & Dobson 2010	180	Adult	Both	Cross	Curvature, height	Mandible 3D physical	Multiple
Tilotta et al. 2010	47	20-40	Female	Cross	Chin surface	3D landmarks, surfaces	Caucasian

Table 2.5 List of studies which examine the sex and age related changes to the chin. Note that the population and measurements are recorded as described by the authors of the studies.

2.2.4.1 Growth

For all ages the growth of the external chin is shown in numerous studies (Subtelny 1961, Koudelová et al. 2015, Ksiezycki-Ostoya et al. 2009 and Nanda et al. 1990) to relate closely to the growth of the mandible. This is due to the relatively thin layer of soft tissue in this area. The depth of the soft tissue is shown to decrease the more inferior to the pogonion it is. Due to this close relationship of skeleton and external skin surface, and lack of soft tissue studies, inference can be taken from the skeletal growth itself and applied to the external chin.

2.2.4.1.1 Developmental Growth

Subtelny (1961) referred to forward growth of the chin from the flat chin of babies to the pronounced chin of adults. The chin's soft tissue increase during developmental growth was not nearly as large as that seen in the nose and mouth (Nanda et al. 1990). However, in Nanda et al.'s 1990 study there was a slight increase observed during childhood and adolescence. The mandible was seen to grow attaining a more anterior position and subsequently increasing the forward angle of the chin. Franklin & Cardini's (2007) and Franklin et al.'s (2008) study saw an increase in ramus height, lengthening of the lower face, a proportional narrowing of the posterior mandible and a reduction of the gonial angle across the 1 to 17 year age span.

2.2.4.1.2 Degenerative Changes

Tilotta et al. 2010 noted that the stability of the chin and the minimal change in size and shape over adulthood makes it good for reconstruction estimates. The one thing they refer to as more likely to cause variation in older age is the attainment of additional fat. The chin is additionally mentioned by Albert et al (2007) as an area prone to excess fat deposit in older age with the production of a double chin.

2.2.4.2 Sexual Dimorphism

Subtelny (1961) commented that females did not reach the level of protrusion of males but achieved most of their growth before puberty, whereas males had a dramatic growth spurt at puberty and continued to grow subsequently. In accord with Subtelny (1961), Nanda et al.'s (1990) study of children and adolescents showed that for all dimensions females attained adult size (that reached by 18 years) at an earlier age than males. For the soft tissue thickness, the protrusion of the mandible and the angle of inclination of the chin male and female values were similar until the age of 15 when female growth levelled out and males experienced a growth spurt. Singh (1990) also saw that the growth of the female chin levelled out at 15 years of age and had less soft tissue thickness values than males. They commented that the females showed a more even growth across the full height of the chin. Both Formby et al. (1994) and Prahl-Andersen et al. (1995) saw a reduction in the pogonion thickness in females while in males this increase. This is contrary to what Nanda et al. (1990) observed although that might be due to the variation in the age ranges studied. Franklin & Cardini (2007), and Franklin et al. (2008) also agrees with the similarities between the sexes until 15 years, where they diverge and become dimorphic. Thayer and Dobson's (2010) study of adult mandibles showed that males showed higher heights of the anterior portion of the mandible and that the anterior inferior portion was more prominent overall then in females. The tubercular laterale were also more prominent, increasing the anterior position of the whole lower jaw but also giving the front surface more definition and squaring.

Chapter 3: Facial Recognition

3.1 Facial Recognition

Facial recognition, the ability to recognise a specific individual amongst a group of people, is something that humans and animals are able to perform to a high level of accuracy without being aware of the process taking place (Cartoux et al. in 1989, Duvdevani-Bar et al. 1998, Papatheodorou & Rueckert 2007). However, the field of facial recognition is one that spans a wide number of disciplines, from psychophysics, physiology and psychology to pattern analysis, computer vision and forensics (Zhao et al. 2003, Curio et al. 2011). Although the disciplines overlap and develop from each other there are two broad areas which they cover: firstly, the ability of the human brain to recognise individuals, facial characteristics and expressions (Tanaka et al. 1998, Simonian et al. 2001, Pelc et al. 2006), and secondly the ability of computers to recognise and identify multiple images of the same individual from a large data pool.

This recognition ability of the human brain is able to overcome numerous adversarial situations such as lighting variation, change in expression, difference in pose and position, covered or concealed parts of the face, adaptations caused by facial accessories (glasses, piercings and headbands) and surface coverings (make up and tattoos). Since the first semi-automatic techniques emerged in the 1960s (Papatheodorou & Rueckert 2007, Curio et al. 2011, Federal Bureau of Investigation 2015), computerised methods have been attempting to catch up with the ability of the human eye and brain. As humans are already accomplished at the task of facial recognition and are much less affected by variances and errors within images, then why not use them to compare images with each other or an image with the actual individual? In a number of situations that is exactly what is done; common occurrences of people being asked to verify the identity of an individual in an official capacity are (Davis & Valentine 2015):

- Border control situations, which use comparison of photo to an individual.
- Situations involving proof of age or identification (i.e. licensed premises, vehicle leasing, domestic travel, secure premises), which use comparison of photo to an individual.

- CCTV operation, which uses comparison of a still photo with video images or another captured still image.
- Police/security surveillance, which uses comparison of still image or CCTV capture to an individual, often in challenging circumstances.
- Eyewitness identification, which uses comparison of a remembered image or vision with a still photo, video image or individual.
- Court situations involving jurors, which use comparison of CCTV captured images to an individual.
- Court situations involving facial image experts, which use comparison of a still photo with video images or another captured still image to present to a court.

So why not leave this task to humans? The main reason is that while humans are generally good at this, as even from a very young age as babies have been shown to recognised their mothers' faces at only five weeks old (Wilkinson 2004), they are by no means perfect. Although the human ability to recognise faces is highly accurate for familiar faces, such as family and friends, when it comes to unfamiliar faces which make up almost all of the scenarios above, the error rate increases (Bruce 2012, Burton et al. 1999, Burton & Jenkins 2011 and Hancock 2012). The error rate further increases if significant time occurs between seeing the first image and the second image, as in the case of an eyewitness. These circumstances can all lead to costly mistakes, as in the case of Jean Charles de Menezes who was wrongly identified as suspected terrorist Hussain Osman by the Metropolitan Police and subsequently shot dead (Dodd 2008). Even in cases where there is familiarity, mistakes can be made; this was true in the case of the missing Dr Richards Stevens. His wife and family all positively identified a man from good quality CCTV images at John Lennon Airport as him, but they were eventually proved wrong (British Broadcasting Corporation News 2003). In Kemp et al.*'s* (1997) study, cashiers working at a supermarket were asked to confirm or reject the identity of an individual with a photo on a card. They were correct in only 67% of instances, with 7% of photos being incorrectly rejected and 64% of incorrect but similar photos being accepted. This should have been one of the easiest of such tasks to perform as they had both the photo and individual in front of them at the same time, much like a border control official.

The ability to recognise faces is also variable across the population, with at one end of the spectrum those diagnosed with prosopagnosia, face blindness, and at the other super-recognisers, who show an extraordinary ability to recognise faces both familiar and unfamiliar (Davis & Valentine 2015). Wilkinson & Evans (2009) proved that training does play a part: expert facial image analysists were consistently better at facial recognition than the general public, especially when the face was obscured in some way. Recent studies have also now been conducted to test the ability of humans directly with that of computerised facial recognition algorithms. In the situations of unfamiliar faces, algorithms are level with or exceed human ability to recognise, identify and eliminate faces (O'Toole & Phillips 2015). Rice et al. (2013) however show that that in very difficult cases, where there is a great deal of similarity between mismatching faces and little similarity between matching faces, humans rely, almost entirely, on the body shape rather than the features of the face to match images. When limited to only viewing the face, with the rest of the image blanked out, humans performed only slightly better than chance, while the algorithm performed well, though not perfectly. In addition to being prone to error, involving humans in this process is also costly in terms of both time and money. If automated computerised methodologies can surpass the positive recognition ability of humans, in addition to the speed at which they are able to perform, then all of these tasks could be accelerated, and have statistical probabilities attached to them making them more acceptable in courts under the Daubert criteria and as recommended by the Crown Prosecution Service (Crown Prosecution Service 2014). While computerised methods enable statistical error rates to be produced, in the case of the courts, it is still the thought of some that the final decisions on identity are likely to be made by an expert witness for the foreseeable future (Edmond et al.2015).

Advances in technology, mostly the improving the quality of images and the processing power of computers, plus the input of money from the commercial sector over the last couple of decades has enabled the area of automatic facial recognition to expand, leading to the boom in the widely used and accepted power of identification that exists today. Facial recognition systems are now used around the world in a variety of forms, from airport security, automatic passport control and CCTV suspect ID (Sparkes 2014) to Facebook automatic photo tagging and mobile phone unlocking systems.

One may question the investment into facial recognition, when there are other, more established biometric identification techniques such as fingerprint and hand print analysis, and iris recognition and retinal scans, which perform as well or better. However, the principal advantage of facial recognition over other methods is that it can be used without the subjects' knowledge or permission, such as in a surveillance situation (Mian et al. 2007). Difficulties can arise when facial recognition is conducted without the cooperation of the subject, however, including parts of the face being obstructed from view, size variability, pose variations or poor lighting. Despite the possible complications, analysis can be carried out on image data already regularly recorded like CCTV surveillance videos and stills, satellite images, drone recording, passport and ID card photos, all of which combat the issue of non-compliance and allow for facial recognition to link with the mass surveillance infrastructure which is already in place in many countries.

3.2 History of Two-dimensional Facial Recognition

Facial recognition initially emerged as a form of identification employing manual, human-based methods**.** The manual two-dimensional facial recognition methods can be split into three categories as suggested by Edmond et al. (2015): photo-anthropometry, morphological comparison and photographic superimposition. These automated or semi-automated techniques started to emerge in the 1960s. The basic principle of each of these is described below with particular emphasis given to the progress of the automated algorithmic model, along with its advantages and disadvantages, as it is the quantifiable, objective methods that are of interested for this study.

3.2.1 Human-based Methods

3.2.1.1 Photo-anthropometry (Photogrammetry)

This method uses simple facial dimensions to rule a photo for comparison in or out. Photographs to be compared will be normalised before comparison in order to reduce the effect of size and scale, see below for an overview of the issues, grids, distance and angles which are then used to align and compare two photographs. Problems may be encountered if attempting to compare two photographs in which the pose is substantially different. Although photo-anthropometry has the benefit of simplicity, its accuracy, even with the best quality images taken under known conditions, is less than optimal. The ability of this technique to accurately identify or eliminate is limited and the Facial Identification Scientific Working Group (2012), one of the many Forensic Scientific Working Groups which advise on the suitability of scientific evidential analysis for courts worldwide, recommends that it should not be used in the court room.

3.2.1.2 Morphological Comparison

This method was first introduced to the scientific and criminal world by Alphonse Bertillon in the late nineteenth century and has been used in courts since, however, in more recent times courts are less inclined to accept it as there are no known frequency rates of the features described (Davis et al. 2015). In this method morphological features or landmarks are identified and classified in distinct types, the photos can be of differing size as proportions can still be used for comparison. Again this can rule in or rule out a possible match, however the likelihood of a match can be based on the number of features that agree and specifically which features agree, as some have a higher rating than others (Facial Identification Scientific Working Group 2014). Pose can obviously cause problems as the measurements will alter dramatically if the pose varies. Ethnicity is also an issue, the dataset should reflect the ethnicity of the individual, as features individuating people of one ethnicity may vary for another ethnicity (Davis et al. 2012).

3.2.1.3 Photographic Superimposition

In a very similar way to video superimposition where a skull of unknown identity is superimposed with a photo of a possible individual to assess whether they are compatible, photographic superimposition compares two photos. Normally using a video scaler and displayed on a screen, these can be faded from one to the other, wiped in all directions or quickly ‘flicked’ from onto the other. It is thought that from all these, any incompatible areas will be noticed by the observer. Whilst it is sometimes possible, with some confidence, to eliminate people who vary greatly, a positive identification is still only an opinion. Some suggest that individual features such as scars and moles provide a higher level of accuracy, but others note that unless the probability of the general population having those features is known, then no conclusions can be drawn. Worldwide there is a great deal of variation in the practice of this method with no protocols and standards in place. This technique was first used in a UK court in 1990, but some judges and scientists claim it is subjective and not scientific (Davis et al. 2012). In 2015 Ibáñez et al. conducted a review of the protocols used around the world by craniofacial superimposition specialists in a multi lab study. They were able to use the results of this to produce the first ever standard for this field which suggests the best practises (Damas et al. 2015).

All of the above techniques have the issue that there are no facial databases that compare to the DNA or fingerprint databases held across the world. In court situations this dramatically limits the ability for them to calculate likelihood ratios, it is for this reason that in most Western courts experts limit themselves to either describing the similarities between two images to the court, letting them interpret them or to placing their interpretation on a descriptive scale of support. An expert is very unlikely to ever positively identify an individual as this would leave them in a difficult position if stronger contradictory evidence were to come to light subsequently.

3.2.2 Automated Methods

3.2.2.1 Geometric Feature Based Algorithms

Photo-anthropometric techniques have proved poor when analysed by humans and simple mathematic comparison, however the data it collects lends itself to automated facial comparison. Landmarks on the face can be identified and mapped and the distances and angles between them automatically calculated by specialist software.

There are three main steps for a fully automated system; face detection, feature extraction/face normalization and identification/verification (Abate et al. 2007). In most systems the first two often combine and occur in unison as they both require the locating of specific points of the face from the source image. In order to locate specific points, lines or curves of a face in a two-dimensional image a number of methods have been created.

Firstly, a template method (Hallinan 1991, Yuille et al. 1992), applies a manually developed template depicting the eyes and/or mouth placed over the general area of the face and through iterative adjustments the template is moulded by alignment of the points of high curvature (peaks and valleys), shown by change in intensity in the image, to minimise the energy function. Once moulded to the best fit position, the position of the eyes and mouth can be transposed from the template.

Secondly, Active Shape Models can be used. This method was first devised by Cootes et al. (1995) and expanded by Milborrow & Nicolls (2008). Active Shape Models use the intensity (changes in the colour of the image, normally caused by shadowing and so indicative of a change in direction of the facial surface) of the image to find points. The software is actively primed with sample images which have manually placed landmarks. The samples images are all warped, stretched and twisted so each landmark is in the same position and the same distance apart, enabling the landmarks to align with each other across all the images. The information relating to intensity is collected from all the sample images, thus removing both size and shape as variants and only considers those changes in texture caused by shape. One of the benefits is that parameters can be placed on the model only allowing it to be deformed in ways typical of the face so no unnatural deformation can occur. This information can then be applied to a new image and the landmarks subsequently placed with relative accuracy.

Lastly manual placement, although not feasible in a large sample or in an automated program, manual landmark placement is often conducted when research is being carried out into the other variables of facial recognition (Amor et al. 2006, Cook et al. 2004, Lu et al. 2004). It is both accurate and relatively simple to do and is also required when verifying automated procedures.

Once the features have been detected in some way, the face needs to be normalised or registered to allow for comparison and consequently, recognition. Recognition techniques can either use the entire face (holistic), specific features (structural) or a mixture of both (hybrid) to compare faces. The dominant analytical tool is Principal Component Analysis (PCA) and Zhao et al. (2003) list the many variations developed for facial recognition including the best known: Eigenfaces and Fisherfaces. Eigenfaces developed by Turk and Pentland (1991) were the first well established and effective automatic facial recognition system. Features of the face, more than just the major features, described as Eigenvectors were given weightings and it was the vector weightings that were summed and compared using Euclidean distances to identify a match. Fisherface method (Jing et al. 2006) based on Robert Fisher's 1936 equation developed for taxonomic classification (Belhumeur et al. 1997) was developed by Belhumeur et al. (1997) for use in facial recognition. The method which uses subspace projection prior to Linear Discriminant Analysis projection is class specific. Belhumeur et al. (1997) and Yang (2002) both concluded that the Fisherface method out performed others including that of Eigenface.

3.2.2.2 Problems

Whilst the early techniques can perform well, this is generally only the case when tested on well lit, frontal or profile images like those found in the accessible databases such as the Face Recognition Technology (FERET) database and the Face Recognition Vendor Test 2002 (FRVT 2002) database. However, when it comes to angled poses, poor lighting conditions and variable sizes and scales of image the problem becomes much more challenging. Some of the solutions to illumination, pose and scaling are discussed below, whilst temporal change between source and reference images and variations in facial expression do cause difficulty, these issues will be discussed later in the section on three-dimensional facial recognition, as the approaches applied are similar and the three-dimensional resolution has more relevance to this study.

3.2.2.2.1 The Illumination Problem

As many of the facial recognition programs feature detection elements rely on the variation in colour or grey scale, any dramatic change to the illumination of a subject will cause complications. Whether this is due to the amount of light or the direction of light, both humans and computers can find it hard to pinpoint features of the face in an over, or under, exposed image. This can also occur when the light source is not in the expected, superior position. The solutions to the illumination problem are wide ranging they include: a) Removing the first few principal components presuming they are solely caused by variation in lighting (Belhumeur et al. 1997) b)‘Training’ the algorithm by imputing images of the same individual under multiple lighting conditions, although there is a limit to the number of different lighting conditions that can be inputted; c)Using three-dimensional generic models to establish the position of a light source for the two-dimensional image and then using that information to locate the features assuming that “differences in the 3D shapes of different face objects are not dramatic” (Zhao et al. 2003)

3.2.2.2.2 The Pose Problem

Whatever the image capture process, a two dimensional image of a three dimensional object such as the face, is going to appear differently for every variation of the angle of capture. This change in angle can obscure parts of the face and distort others making them seem larger or smaller in comparison to other facial features. Obviously this variation in dimension between facial features is going to cause problems when attempting to match images of the same individual taken at different angles and the comparison itself is only possible if the feature detection aspect of the process has been able to locate accurately the identifying features. Where the face is facing forward but is at an angle, a simple ‘in plane’ rotation is all that’s required to align the photo to the normal. To calculate the variation from normal or position landmarks in a pictures which are not front facing one has to remember that the two-dimensional image depicts a three-dimensional object and by knowing the angle the picture was captured at the variation in feature position can be inferred. In most cases, where the change of angle is less than about 60° (Zhao et al. 2003), inputting multiple images at known angles into the dataset is all that is required (Duvdevani-Bar et al. 1998). As most of the features, both eyes and nose for example, will still be visible to the camera at these angles and so are detectable, the angles between these features can be used and compared samples in the dataset, and the angle of view calculated. Once this is known, then the comparison of dimensions can be conducted against other images taken at that particular angle of view, or the dimensions can be adjusted to approximate to what they would be in a normalised situation and then compared.

Most commercial algorithms at present rely on frontal images and generally require the cooperation of the individual at the point of image capture in order to obtain a usable image.

3.2.2.2.3 The Size, Distance and Scaling Problem

Photographs are often normalised using proportional indices measured from easily identifiable landmarks (Jayaprakash & Pritam 2014). This in itself only goes some way towards providing comparative data, as the focal length of the lens and the distance of the subject from the camera still poses a problem due to the distortion of the face. If the face is close to the camera at the point of capture the nose seems large in comparison to other features and the ears themselves can be hidden from view whereas when at a distance from the camera the nose appears smaller and the ears larger. In some situations, background objects in the probe image (the image from a scene or CCTV capture for which the true dimensions are not known) which can be recognised, and measured such as furniture or items of a known size, such as coins, may be used to decipher the focal length and angle.

3.3 Three-dimensional Facial Recognition

The idea of three dimensional imaging was first brought to the world by David Brewster at the Great Exhibition in 1851 with the stereoscope (Brewster 1856). In the stereoscope two photographs of the same scene were taken from points that were approximately 7cm apart, this being the average distance between the midpoint of the right and left eye. These two photographs were then placed side by side on a card and viewed through an apparatus which incorporates both mirrors and lenses. People viewing the image through the stereoscope were able to see a scene which appeared to be three dimensional, rather that the flat two dimensional form of the traditional photograph. Since then three dimensional imaging has moved on to incorporate video imaging in addition to still images but the principle of capturing and viewing from multiple angles simultaneously, still remains. The invention and subsequent commercialisation of three-dimensional scanners, either structured or white light and laser based, has enabled the study of facial recognition to enter into the three-dimensional world. The price of three-dimensional scanners has reduced since they first came onto the market and continues to do so, that said, they still outprice two-dimensional forms of image capture and therefore most real world situations still require a two-dimensional solution at present. There are numerous studies covering the comparison of different types of three-dimensional image which are described later in this chapter (Al-Osamini et al. 2009, Cartoux et al. 1989, Cook et al. 2004, Guo et al. 2003, Lu et al. 2004, Maurer et al. 2005, McCool et al. 2010, Mian et al. 2007, Mohammadzade & Hatzinakos 2013, Tan & Zhang 2006, Vivek & Sudha 2007). Authors have also offered a solution for real world situations by attempting to bridge the technology gap by comparing two-dimensional source images to three-dimensional reference images (Bowyer et al. 2006, Chang et al. 2006).

Considering the issues affecting two-dimensional facial recognition systems such as the illumination of subjects, pose and size (Cook et al. 2004) an increasing amount of research is being conducted into the area of three-dimensional facial recognition, which provides a solution to these problems. Scientists, engineers and software programmers alike have been attempting to solve the numerous issues faced in designing an accurate, fast and robust facial recognition system. Even if a method of facial recognition can be designed to identify accurately the same individual in a variety of poses, lighting conditions and data collection methods, two significant factors still remain to be addressed. Both variability in age between the comparative images and the changeability of facial expression, change the intention of the method from recognising two identical representations of the same face to recognising two representations of the same face that may differ greatly from each other. These are addressed differently for two-dimensional to three-dimensional comparison and three-dimensional to three-dimensional comparison. An overview of the methods of facial recognition along with the advantages and disadvantages of the three-dimensional realm will now be discussed, followed by the approaches taken to solve the complications of facial expression variation and the ageing process.

3.3.1 Advantages of Three-dimensions

Although the passage into three-dimensions creates a number of new issues, three of the most problematic faced in two-dimensional facial recognition, illumination, pose and size, are solved simply by the nature of the three dimensional form. Variation in illumination of the subject in a two-dimensional image can affect the visualisation and placement of landmarks, as it is the change in tone and shade which is used to identify specific landmarks. Although illumination can cause problems when capturing a three-dimensional image due to the two or more cameras being unable to adequately capture their respective two-dimensional images (Bowyer et al. 2006), once the three-dimensional image has been captured and created, illumination is no longer a factor. In most software which enables the viewing and manipulation of a three-dimensional model, there is also a facility for virtual illumination. This feature allows the user to place and direct a virtual light source onto the image from any angle within the three-dimensional space. The manoeuvrability of the light source removes the issue of poor and obtrusive lighting encountered with two-dimensional images and also enables the optimum illumination for visualisation and placement of each landmark in turn.

The second issue of pose is solved very simply by the ability to move the three-dimensional model within space into which ever pose is required. When dealing only with three-dimensional models the pose of the face does not need to be known. With two-dimensional models the reason for computing the pose is to be able to extrapolate the facial dimensions through normalisation of the face, as the facial dimensions can be measured independently of the pose. The only limitations to this are those linked to the initial creation of the three-dimensional model. When a three-dimensional model is created it is done either by multiple cameras capturing multiple views of a subject, or by a moving laser capture method. In both cases the resulting model can only include surfaces captured successfully by the scanner from the views available. In the case where a specific pose is required in order for a comparison of some sort to be conducted, when applying a three-dimensional reference to a two-dimensional source image for example, then the initial scan must have included the correct views to create that pose.

Size and distortion, as seen earlier in this chapter, can cause problems when comparing two or more two-dimensional images taken at different distances from the camera or at different angles of the lens. As with the issue of pose variation, the images need to be normalised in order for the facial dimensions to be compared or measured. The distance from the camera and angle of the lens also need to be known or calculated. As the creation of three-dimensional images requires the distance from the camera and the angle of lens to be set and the scanner calibrated to those known settings, then the issue of size is resolved. Three-dimensional images are therefore always of known dimensions and can be measured and compared directly within three-dimensional space without the need to scale the images first.

3.3.2 Disadvantages of Three-dimensions

Whilst the main advantage of facial recognition in comparison with fingerprints and iris scans is the ability to collect data without subject cooperation, this is not yet possible for three-dimensional data collection. The acquisition time for a three-dimensional image is at present slower than that of two-dimensional meaning that the subject would be required to stand still, albeit for only a matter of a couple of seconds, in order to capture the three-dimensional image. There are three main categories of scanner set up; two cameras in stereo, one camera and a structured light or laser source or two cameras and a structured light or laser source. All three use triangulation either between the two cameras or the camera and light source. The stereo approach attempts to pair the same point in both images to assess the position in three-dimensional space while the structured light is projected in known sizes and so the pictured distance is used to calculate depth. Whilst all three methods are fast enough for accurate capture of faces, i.e. people are able to stay still for the duration of the image capture, the fastest approach is two or more cameras in stereo (Boehnen and Flynn 2005). As there is no need to wait for the time it takes to project a series of light patterns onto the subject, the cameras without structured light are faster, however the accuracy of depth perception is improved by the use of structured light so a compromise has to be made (Bowyer et al. 2006). With both structured light and laser scanning, the subject is fully aware of the process as bright light has to be shone directly onto the face and so again, is not suitable for a non-cooperative application. Some scanners however use an infrared light source which is invisible to the human eye although the accuracy of this method cannot yet be compared to that of white light (Boehnen & Flynn 2005). It is likely that the speed of acquisition will be improved over time and that the issues of cooperation or non-cooperation may be resolved.

Artefacts can occur on scans and are described as either ‘holes’ or ‘spikes’ (Bowyer et al. 2006). ‘Holes’ are essentially areas of missing data where, for some reason, the scanner was unable to read the surface being scanned, the most common reason for these ‘holes’ on the face is facial hair, such as the eyebrows. In contrast ‘Spikes’ are created from erroneous data that tend to spike forwards or backwards from the plane of the face, these can only be seen clearly when viewing the image perpendicular to the original angled view of the scanner. The normal cause of ‘spikes’ are reflective areas where the light bounces back to the camera at unusual angles, these can be due to greasy or sweaty skin, metallic objects such as earrings and the moist glassy surface of the eyes. Whilst ‘holes’ can be left, although it is possible to fill them in by estimating the missing surface ‘spikes’ need to be removed from the data before analysis can be conducted. At present ‘hole’ filling can be automated but needs to be visually checked and ‘spike’ removal is best done manually, though this will undoubtedly improve with more research.

While the ability of three-dimensional recognition may be an improvement on two-dimensional recognition, the computational time of three-dimensional is, at present, a hindrance. Two-dimensional based algorithms are extremely fast allowing for real time comparisons and verification thus enabling them to be used in commercial situations. Three-dimensional comparison on the other hand is very time consuming and processor heavy, meaning that not only does it take longer but the infrastructure required needs to be more high powered and thus more expensive.

The most serious disadvantages of three-dimensional methods, when compared to two-dimensional methods, are at the point of data collection rather than the analysis itself. All the disadvantages mentioned above, however, are likely to be resolved in the near future through the improvement of scanning technology and the adaptation of the facial recognition algorithms to speed up the alignment process.

3.3.3 Overview of Three-dimensional Recognition

The subject of three-dimensional face recognition was first addressed by Cartoux et al. in 1989. They used range finder images and translated the three-dimensional data into two-dimensional data depicting the profile of the face. It was the two-dimensional profiles that were subsequently used for comparison by means of correlation coefficient and mean quadratic distance. For the following decade the area of three-dimensional face recognition was very quiet with no real progress until the end of the 1990s. When it comes to the methodologies used and researched in the field of facial recognition they can be split into three broad categories, as described by Papatheodorou & Rueckert (2007) in their review of the field; the three categories are surface, statistical and model based approaches. Three-dimensional data can be viewed and used in a number of ways, the most common depictions of data are Ranges Images, Point Sets, Surface Meshes and Feature Vectors (Bowyer et al. 2006). Whilst the data collection method does not limit the approach or specific algorithm used, some data forms do pair better with certain approaches for instance PCA and range images.

3.3.3.1 Surface-based Approach

The surface-based approach uses the geometry of the facial surface itself, whether that be the whole surface or curves, lines or landmarks extracted from the surface, and uses these for direct comparison between two surfaces. The algorithms used to match the two surfaces are numerous but include minimum distance, Hausdorff (Guo et al. 2003, Tan & Zhang 2006, Vivek & Sudha 2007), Hidden Markov Models (McCool et al. 2010), Gaussian curves (Cartoux et al. 1989) and Iterative Closest point (ICP) (Al-Osamini et al. 2009, Chang et al. 2006, Cook et al. 2004, Lu et al. 2004, Maurer et al. 2005, Mian et al. 2007, Mohammadzade & Hatzinakos 2013). While some authors take on the matching of the entire face as a single model (Al-Osaimi et al. 2009, Cook et al. 2004, Maurer et al. 2005) others section the face into specific features (Chang et al. 2006, Lu et al. 2004, Mian et al. 2007, Mohammadzade & Hatzinakos 2013). The main purpose for sectioning the face is to discard data from the non-rigid portions of the face such as the mouth and chin and so that only the rigid structures of the eyes, nose and forehead remain (Lu et al.

2006, Papatheodorou & Rueckert 2007). It has been suggested (Chang et al. 2006, Mian et al. 2007, Smeets et al. 2010) that by discarding the non-rigid areas of the face the robusticity of the technique can be improved in the light of variation in facial expression to be discussed later.

3.3.3.2 Statistical Approach

The most common statistical methods used in facial recognition are Geometric Morphometrics and Principal Component Analysis (PCA). These have both been widely used in the types of two-dimensional facial recognition described above and they are also used as analytical tools in three-dimensional facial recognition. As with two-dimensional methods, additional information such as colour, depth, texture, age or weight can be incorporated into the equation to give insight into the variation caused by those as well as the shape of the face (Hesher et al. 2003, Tsalakanidou et al. 2003). Procrustes superimposition first removes the variation due to size and position through scaling and rotation/registration and leaves only that due to shape variation. PCA then simplifies the shape variation and attributes it to different variables through component scores. From these scores the effect of variables such as age, height or sex can be calculated.

3.3.3.3 Model-based Approach

The idea of the model approach is to use a generic three-dimensional model and morph specific faces and their features to the generic model, calculating the deformation (the amount and direction of change) that was required to transform from the specific to the generic. It is the deformation parameters for each individual that are then compared rather than the actual surfaces as with the surface-based approach. It is possible for the generic model to be two-dimensional allowing for three-dimensional images to be transform to two-dimensions and thus used for comparison with real world images such as those from CCTV (Blanz & Vetter 2003).

3.3.4 Three-dimensional Facial Recognition Databases

The different approaches to face recognition are all valid, and improvements are constantly being made as the field grows. Comparing the different approaches however is a difficult task as there are few available three-dimensional face datasets, making direct comparison of techniques hard, and the metrics and statistics used to evaluate the current methods also vary. There is some progress with the introduction of a few standardised databases which are made available to select researchers.

- The Facial Recognition Grand Challenge database (FRGC v2.0) contains 4007 laser scanned images of 466 subjects exhibiting a variety of facial expressions and a time lapse of six months between a selection of the subjects. (Phillips et al. 2005)
- The Grupo de Algorítmica para la Visión Artifical y la Biometría database (GAVAB) contains 549 laser scanned images of 61 Caucasian subjects displaying accentuated facial expressions taken in different poses.
- The Binghamton University 3D Facial expression database (BU-3DFE) contains 2,500 scans of 100 subjects of varying ethnicities collected using a stereo-camera system. There are 25 scans of each subject demonstrating six facial expressions with differing degrees of intensity.
- The Bosphorus database contains 4,666 scans of 105 subjects showing up to 34 different expressions, 13 poses and four variations of occlusion. They are collected using a structured light scanner. (Berretti et al. 2013)

Across the board these databases all have limitations, in all cases the number of scans does not necessarily indicate that the entire face containing all features is captured. In many instances one scan may refer to a single viewpoint (sometimes referred to as a 2.5 dimensional image) capturing perhaps only the left side of the face or a scan with large portions of data missing due to hair, glasses or facial hair. All of the databases have been developed in order to focus on facial expression variation meaning that the large proportion of the images contained within them show some degree of non-neutral facial expression. If one only wants to use scans from these databases that display neutral expressions, then the number of scans available for use in creating or testing new techniques is dramatically reduced. As there is no standardisation of the neutral expression used across the databases then it is also difficult to combine them to create a larger dataset.

The most studied of these databases is the FRGC v2.0 mainly due to the competitive challenge which was the reason for its production. A number of the resultant studies will be discussed later in this chapter.

3.3.5 Variation Problems

As the issues of illumination, pose and scaling are resolved by three-dimensional application, the key matters which need to be addressed for three-dimensional facial recognition to be applied to real world situations are the variations due to the ageing of an individual between the capture of the first and subsequent images and the variation in facial expression between images. If the recognition process is unable to cope with any variability caused by either of these factors, then it is unlikely to be successful in a real situation as the aim is to realise accurate facial recognition that does not require the cooperation of the subject. Both these issues are addressed in some way by almost all authors on this subject but priority is given to facial expression variation. Studies of variation due to ageing tend only to look at relatively short time periods of six months to a year between original and subsequent scanned images (Chang et al. 2006, Paone et al. 2014). Such scans are more likely to capture variations in hairstyle, make-up, jewellery and neck line, due to time lapse rather than any variation caused by ageing itself.

Facial expression, by contrast, has been widely studied, especially within the last few years, with only a few of the studies not incorporating it into their remit and all of the large databases including non-neutral images. There are a few different approaches when it comes to addressing the problem of facial expression variation with different levels of complexity. Bowyer et al. (2006) describe the three approaches most frequently used and the issues faced by each of them. When looking at the face as a rigid structure and applying algorithms that do not deform, such as the surface-based approaches, the simplest technique is to remove the portions of the face which naturally deform most due to expression. These are generally accepted to be the mouth and cheeks, leaving the forehead, eyes and nose as more rigid comparable structures. The chin is seldom mentioned in studies, possibly due to the fact that it has very limited features and if the mouth is removed it is simpler to continue by also removing the chin rather than being left with two separate portions of the image. This approach allows the use of databases of reference images with non-neutral expressions for comparison. The main limitation of this technique is the extra manipulation of the image needed to remove the non-rigid portions. This option is employed by Chang et al. (2006) Lu et al. (2006) and Mian et al. (2007) all described later in the chapter.

Another option for surface-based and statistical-based approaches is to flood the database with multiple images showing a wide variety of expressions for each individual. The idea is that when a probe image is submitted for comparison, there will already be an image of the same individual showing the same facial expression, in the database. There are two issues with this method, firstly, that it is impossible to capture and include an image of every possible variation of facial expression for each individual and secondly, that due to the large number of reference images needed this could only be used in an authentication/verification situation where the subject is fully cooperative.

The third option as discussed by Bowyer et al. (2006) is relevant to those taking a model-based approach. This technique could work in two ways, either the generic model shows a neutral expression and both the expressive reference and probe images are adapted to fit, or there are multiple generic models showing different expressions and the images showing those expressions are moulded to the relevant model before comparison. It may prove difficult when using the second of these methods to account for the extreme variation across the population in the forming of the ‘same’ expression. Put simply, not only do we smile in a different way to the people around us but we also smile differently at different times, so having one generic model for ‘smiling’ would not accurately illustrate the individual factors of the face. For the first of the model-based approaches to work, information detailing how the features and surfaces of the face move between the neutral and expressive would need to be known so that the reverse can be applied to any image in order to mould it back to a neutral expression. This information can be collected through multiple landmarked three-dimensional images (Al-Osamini et al. 2009), video motion capture with (Curio et al. 2011) or without markers (Walder et al. 2011) although both of these methods are in the early stages of development. This technique is use by Al-Osamini et al. (2009), Smeets et al. (2010) and Mohammadzade & Hatzinakos (2013) described later in this chapter.

3.3.6 Preregistration

Whichever approach is taken whether statistical, model or surface-based, the issue of preregistration (alignment, correspondence, matching) applies. For each approach the two faces to be compared, either two individuals or one individual and one generic face, need to be placed in relatively close proximity to each other and be aligned in roughly the same direction for the algorithms to work. In two-dimensional this is a relatively simple task, excluding the issue of pose and size, requiring a few corresponding landmarks, placed either manually or automatically, to be aligned to their counterparts. Three dimensional images bring an additional dimension of complexity when applying some kind of automatic or semi-automatic procedure to preregister images. More landmarks or points of correspondence are needed for successful preregistration of three-dimensional shapes in comparison to two-dimensional (Papatheodorou & Rueckert 2007) and the identification and placement of the landmarks is more time consuming.

Whilst not practical for large scale datasets, for research purposes many authors do use manual landmark based preregistration along with least squared fitting, a form of regression, or PCA to perform the initial rotation and translation, before applying their methodology for the subsequent comprehensive alignment (Amor et al. 2006, Chang et al. 2006, Lu et al. 2004). This is normally done to save time and provide a level of accuracy where the preregistration element is not the main focus of the study.

The profile line, line of quasi-symmetry of the face, can be used to preregister a face. This is located either through placement of midline landmarks or by the identification of a line for which the sum of curvature on either side of the said line is roughly equal (Cartoux et al. 1989).

As an alternative to the use of manually placed landmarks, which is a time consuming task, it is possible to apply surface preregistration which involves the measurement and alignment of areas of curvature. The curvature of the surface is measured and areas of high curvature in one image, such as the eyes and nose, can be aligned to areas with a similar level of curvature in the other image. The principles of the technique are similar to the registration of two-dimensional images using the intensity of images. Curvature can be measured in a number of ways, if the general position of the face is known and the face is in the normal position, then simply taking the maximal *z* coordinate in a range image should identify the tip of the nose or pronasale (Cartoux et al. 1989, Mohammadzade & Hatzinakos 2013). Issues arise with this method if something is obscuring the nose, the image is not in a normal position or another artefact (e.g. hair) is anterior to the pronasale. Mohammadzade & Hatzinakos (2013) use PCA to confirm that the nose has correctly been identified. Once the position of the pronasale is identified then the Hotelling transform (a form of image processing transformation) or PCA can be used to then rotate the rest of the face into alignment as employed by Mian et al. (2007). The curvature of a number of areas can be measured and then analysed to identify shape indices such as the troughs and ridges depicting the eyes and nose as in Zhao et al. (2011).

Some authors use Iterative Closest Point algorithms (ICP), to perform preregistration and then go on to use another method of quantitative comparative analysis. These are discussed within the Facial Recognition implementing ICP section below (Cook et al. 2004, Maurer et al. 2005)

The simplest method, though very time and personnel heavy so only suitable for studies where preregistration is but a side issue, is total manual alignment. This is simply done by rotating and translating each of the faces so that they are all in general visual alignment with each other. This is obviously a subjective technique, however as it is not the final stage of alignment and its purpose is to place the faces in approximate alignment, total accuracy is not required at this stage in the process. This method is used by both Gökberk et al. (2006) and Papatheodorou & Rueckert (2004).

3.3.7 Iterative Closest Point (ICP) Algorithms

Overall there are so many possible variants in the approaches taken in addressing the facial recognition problem that it is impossible to review all of them in detail. In light of the purpose of this study, to assess the similarity between relatives rather than between multiple images of the same person, the review will be limited to include only Iterative Closest Point (ICP) algorithm based approaches.

It was Besl & McKay (1992) who first compiled and collated the small body of work that had been completed in the field of shape registration before the 1990s. Until then shape registration often referred to as alignment, matching or motion estimation, had been studied by a small number of people across a variety of disciplines, each with their own particular problem to solve. Shape registration studies included both two-dimensional and three-dimensional data; curves, points and surfaces and each boasted an assortment of equations and algorithms used in attempting to answer their questions. Essentially they were addressing the same task, that of taking a pair of shapes and by rotating and translating them, aligning them in the way that minimises the distance between them. This then allows a calculation of the similarity between the shapes using a mean square distance.

Using all the previous work as a baseline, taking its best points and heeding its problems Besl & McKay created the ICP algorithm. An improvement on many of the previous algorithms, ICP is capable of aligning point sets, line segment sets, implicit curves, parametric curves, triangle sets, implicit surfaces and parametric surfaces with the same algorithm. The analysis of a variety of data types is possible as non-point based data sources can first be transformed into point sets by either changing the triangle vertices or end points of lines or curves. ICP was also an improvement on previous methods as no pre-processing, such as smoothing of the three-dimensional point data, was required. In simple terms ICP takes a model shape ‘M’ and a data shape ‘P’ both composed of points and aligns them. The quality of alignment is calculated by pairing each point on M with the closest point on P, measuring the distance between each pair and then calculating the mean square error (MSE) for the entire shape. In order to improve and lower the MSE to get the least MSE, ideally down to 0 if the objects are identical, M is locked in space while P is moved through translation and rotation. For each iteration of movement, the mean square distance is calculated. This continues until the change in the mean square value falls below a set value or until a certain predetermined number of iterations have been completed, depending on whether accuracy or time is the most important constraint. At this point a final value can be computed for the least mean square error and it is this figure that can be used as a quantification of the difference between the two shapes. Whilst solving some of the issues faced by earlier scholars in this area Besl & McKay's ICP algorithm still lacked speed due to computational complexity (Kapoutsis et al. 1999), the ability to cope with noise and outliers in the data. There were also problems with shapes that had only a partial overlap.

Kapoutsis et al. (1999) attempted to improve on ICP with the use of the Voronoi diagram to create the Morphological ICT algorithm. Both the P and M shapes are encompassed within a volume tessellated to the P shape. This means that the amount of space that the rotations and translations can occur in is limited, reducing the possibilities and thus reducing the time and Random Access Memory (RAM) needed to compute the optimal registration.

3.3.7.1 Trimmed Iterative Closest Point (TrICP) Algorithm

In 2002 Chetverikov et al. addressed the issue of speed, partial overlap and shape defects with the invention of the Trimmed Iterative Closest Point Algorithm (TrICP). By making use of Least Trimmed Squares (LTS) instead of Least Mean Square Error (LMSE) or Least Median of Squares (LMedS) they were able to speed up the process of registration, enable registration in cases with overlap of less than 50% and make the algorithm more robust, enabling data with defects and outliers to be registered. LTS as described by Rousseeuw & Zomeren (1990) was previously used for initial estimation of alignment parameters and in contrast to LMedS based ICP it does not require randomisation of points, so all datum points are considered rather than a proportion of them, increasing the accuracy (Chetverikov et al. 2002). The algorithm has roughly the same structure as ICP only using LTS and with an additional feature of selecting a stopping point, for when the trimmed MSE falls below a certain value.

TrICP has improved the original ICP in many ways but still has some limitations. The model and data shape must be roughly preregistered beforehand as TrICP has only successfully been applied where rotation of up to 30% has been required from the initial registration (Chetverikov et al.2002). The overlapping part of the two shapes must have features and not be too smooth or symmetrical in order to be clearly aligned (Chetverikov et al. 2002). Although the convergence of the two shapes is faster and more accurate, it is still recommended that, in addition to preregistering the shape, one uses multiple initial positions. In order to select the best trimmed MSE it is advisable to run the algorithm several times as this will yield the most accurate results.

3.3.7.2 M-Estimation Iterative Closest Point (M-ICP) Algorithm

Kaneko et al. (2003) from the field of sensing robotics, attempted an improvement of ICP specifically in cases of outlying data. Their method of adding another iterative loop to the ICP algorithm did improve two-dimensional and three-dimensional cases where outlying data had caused the original ICP to fail. This was mainly due to improving the accuracy of convergence of the two shapes using M-estimation. One of the consequences of using this method was to increase the processing time to an unfeasible length. At the same time the issues relating to stopping points, initial registration selection and the degree of overlap still remain unresolved.

3.3.7.3 Partial Iterative Closest Point (Partial ICP) Algorithm

Partial ICP is essentially the same as plain ICP but is applied to fewer points. For every 10 points in a point cloud the central one is selected and that is used in the application of ICP. Very simply the speed of alignment is improved as it become computationally easier but the degree of accuracy is decreased as the surface has essentially been smoothed (Zhao et al. 2011)

3.3.7.4 Mixed Iterative Closest Point (Mixed ICP) Algorithms

Different versions of ICP can be combined and either conducted one after another or in series with one iteration of each type being conducted alternately. This technique is referred to as Mixed ICP (Lu et al. 2004, 2006).

3.3.8 Facial Recognition Implementing Iterative Closest Point (ICP) Algorithms

Medioni & Waupotitsch (2003) were some of the first researchers to use an ICP like algorithm in the area of facial recognition. They applied Chen & Medioni’s (1992) algorithm to a sample of 100 individuals with 700 scans in total, these showed a variety of poses but all displayed at least one neutral expression. They are able to achieve 97.5% Verification at a 1.0% False Acceptance rate but do state there is very little difference between the scans of each individual and they were really only comparing their methodology to commercial two-dimensional methods.

Whilst some used very detailed laser scans Cook et al. (2004) opted for the cheaper, more attainable, safer and user friendly structured light scanning. Where the comparison of two-dimensional faces was relatively simple in technique and processing power, the comparison of three-dimensional structures became a much more complicated and time consuming procedure. The initial issue faced was pre-registration of the two scans to be compared. Cook et al. (2004) opted for feature based common axis alignment using the profile of the nose to get all the faces facing the same direction before the application of ICP. In this case ICP was only used to align the faces and the values from this calculation were not used for the study. Instead, feature extraction and Gaussian Mixture Models, a type of landmarking using depressions and curves common to both scans, was used and analysed using geometric morphometrics and PCA on the distances between these common points rather than the entire scan. An error rate of 2.67% was achieved from a 30 person trial.

In their 2004 study Lu et al. attempted to use ICP in matching 18 three-dimensional face models with 113 2.5D scans of the same 18 people. 2.5D in this and other studies refers to the x, y, z coordinate data from a single three-dimensional scan captured from one angle, once merged with additional scans from other viewpoints it creates a three-dimensional model. In a method similar to that used by Cook et al. (2004), Lu et al. initially used facial features to roughly preregister the faces, employing least square fitting based on manually selected landmarks. Automatic landmarks detection was attempted with range images but not perceived to be accurate enough so manual placement was used. Lu and colleagues (2004) then use a hybrid of Besl & McKay’s (1992) Point to Point ICP algorithm and Chen & Medioni’s (1992) Point to Plane ICP algorithm, running them on alternate iterations. Rather than using the whole face, only selected portions of the face were used in order to reduce the variation in the mouth and cheeks caused through facial expression. A matching error rate of 4.4% was achieved. In a later study Lu et al. (2006) increased the sample size to 200 three-dimensional models and slightly less than 600 2.5D images of the same 200 people. They used the point of three features to preregister and created a short list of the top 30 possible matches based on linear measurements between the features. Only the short listed scans went on to be compared with the hybrid ICP used in the smaller study. The study both analysed the ability to match 2.5D scans with three-dimensional scans from different angles of view and with and without a smiling expression. Ninety-eight percent accuracy was achieved from a frontal view with neutral expression. This dropped to 96% for a profile view taken at a 60° angle to the frontal plane. The introduction of a smiling expression in the 2.5D image drastically decreased the accuracy of matching to 68% and 76% for frontal and profile respectively. However, the addition of texture comparison through linear discriminant subspace analysis did improve all of the results by up to 9%.

Papatheodorou & Rueckert’s 2004 study is slightly over promoted with the claim of 4D facial recognition. What they refer to as 4D data is actually a combination of three-dimensional mesh data and two-dimensional texture data. Although almost all other authors in this field use, or at least have access to, this combination of data, they describe it as, three-dimensional not 4D data. Once their data from 62 individuals were collected the three-dimensional mesh and two-dimensional texture were aligned automatically by their scanning system, the mesh was then manually cleaned by drawing an ellipse on the two-dimensional image and deleting all data outside that ellipse in both the two-dimensional and three-dimensional images. This removed outlying hair, ear and neck data whilst leaving the majority of the rest of the face. The faces were then moved within the three-dimensional space to be roughly in the same position. They used a combination of ICP on the three-dimensional mesh and intensity weighting of the two-dimensional texture simultaneously to create a “4D” registration technique. The matching score was also formed from a combination of Euclidean distances of the three-dimensional meshes and intensity weighting similarity as described for two-dimensional recognition. Each subject was scanned in frontal view, at a 45° angle and a 20° backwards tilt with a neutral expression, then in frontal view with two expressions, smiling and frowning. The addition of texture weighting improves the accuracy of correct matching in frontal view, while it decreases the performance of the algorithm when looking at angled views. It is suggested that this is due to the illumination variation in the two-dimensional image. Although not mentioned, this may be because the target scan is a frontal view rather than a full face so there would be less overlap between that and the angled scans. A match can only be assessed through the overlapping data of scans, when the overlapping data are limited due to angle differences accurate matches can prove difficult. Where expression variation occurred the addition of texture information did also increase the accuracy, dramatically in respect to smiling from 57% to 88% whilst only minimally for frowning images with an increase of 3%.

Like Cook et al. (2004), Maurer et al. (2005) use ICP to preregister the two faces being compared. However, instead of using Mean Squared Error to quantify the variation between the faces, they employ a distance mapping technique to create a score of similarity. This approach resulted in an accuracy of 87% at a false acceptance rate of 0.1% when applied to the FRGC v2.0 database.

Gökberk et al. (2006) conducted a comparative study looking at both deformation model techniques using Thin Plate Spline warping (TPS) and ICP. The Thin Plate Spline requires landmarks to be placed, this is initially done with a 10 landmark template (focused in the eyes, nose and mouth area) and then manually adjusted. The faces are then warped to an average face model. The ICP registration is conducted after cropping the face using an ellipsoid centred at the pronasale, the faces are first preregistered manually through basic translation, placing the centre of the nose of each face in the same position. From a total of 106 subjects and 579 images, mostly of neutral expression and with only slight variation from the frontal view, four datasets were produced with one, two, three or four images of an individual in the respective sets. Each of these was tested not only for the two techniques but also across a number of feature input techniques, point clouds, surface normal, profile set, shape index and depth images. For both TPS and ICP the surface normals perform best, closely followed by point clouds all with accuracies over 92%, PCA and Linear Discriminant Analysis of the depth images perform worst in both cases with accuracy as low as 45% in the case of TPS PCA. Profile sets and shape indices lie between the others with shape index performing slightly better for TPS and profile sets ranking higher for ICP. The comparison of the four datasets show that for all scenarios the use of multiple images of individuals in the training sets (the original database of images the method is based on) increases the accuracy of matching.

When comparing the techniques themselves ICP outperforms TPS in all input types, with the exception of shape index, indicating that the warping process loses discriminative information from the face. The best result of 99.17% accuracy was achieved with ICP of surface normal. In addition to the individual test Gökberk et al. (2006) also investigated the fusion of results and achieved mixed results. In the datasets with only one or two images of each individual the fusion of result improved the accuracy of match but where more images were included the fusion of methods did not outperform the highest ranked individual technique.

During their three-dimensional study into the recognition of faces under varying facial expressions Chang et al. (2006) concluded that ICP outperformed PCA in its ability to identify. For two frontal scans with neutral expression and up to a year interval between acquisitions, the accuracy using ICP is 91%. The data are derived from 449 individuals with 2349 neutral scans and 1590 non-neutral scans. Rather than simply focusing on the whole face, the authors went on to select smaller areas of the face, mainly focused around the nasal area, which when registered with the entire face improved the accuracy to 95-96%. This accuracy rate is still based on scans termed as neutral, yet, expressional variation is still possible although obviously far less than between opposing expressions. The observed increase in the recognition rate when smaller sections of the face are analysed could be attributed to the minimisation of the expressional variation shown on the partial facial features when compared to the whole face. The proposed reason for focusing on the nasal area and its immediate surrounds is the minimal malleability and that ‘noise’ and errors in scans, which commonly occur around the hairline, eyes and ears, are minimised by deleting these areas. The outperformance by the nasal area over the whole face is repeated when variation of facial expression is introduced, from 61.5% to 84% respectively. In addition to the improved accuracy of recognition, the reduction of the facial area consequently reduces the data size and so speeds up the computational processing. Processing time is one of the major limitations of the transition to three-dimensions as the data involved area much larger.

Amor et al. (2006) attempted to authenticate the basic ICP approach of facial recognition. In the normal manner they apply a two-step alignment with the course alignment performed through manual landmark selection on the gallery and probe images, before the fine ICP alignment. Their database consists of 50 individuals each with a full three-dimensional face image of neutral expression and 400 2.5D images of the same individuals showing a variety of expressions. A holistic, in as much as the 2.5D images can be holistic, approach was taken with all the comparisons and the lowest ranked image had an accuracy rating of 92.68% produced by a 'dramatic change' in expression. The accuracy ranged from that to 100% for the probe images with neutral expression.

Mian et al. (2007) used a combination of 2D and three-dimensions and applied a "short list methodology" to reduce the list of possible matches from a database using most of the face and then use feature based comparison to identify the individual. Mian et al. (2007) contributed to, and used data from, the FRGC v2.0 database, mentioned earlier, which contained 4007 scans. They use Spherical Face Representation and Scale-Invariant Feature Transform for the initial elimination. The spherical face was prepared by finding the pronasale and then selecting a sphere with a radius of 80mm centred on the pronasale. This generally incorporated the eyes, nose, mouth and chin with a small portion of the forehead and cheeks although this did vary between individuals as the section was metric rather than proportional. The authors of this study refer to this as a 'holistic' or 'whole face', although the ears and portions of the outer face region are not included which some might argue is not the whole face. Once the spherical face was produced and aligned Scale-Invariant Feature Transform was used on the two-dimensional texture layer along with Spherical Face Representation to speedily conduct the 1,608,166 possible comparisons and reduce this figure before performing the more complicated and time consuming three-dimensional comparisons. Simply, the mean of the Euclidean distance between approximately 80 automatically identified descriptors on the two-dimensional texture layer and the Euclidean distance between the Spherical Face Representations was combined and compared enabling the elimination of 97% of the initial dataset. Once the dataset is reduced, similarly to Chang et al. (2006), smaller areas of the face are then used for comparison, in this case they use the eye and forehead region plus the nasal area, again chosen for their invariability during expression. These are automatically segmented based on the inflection points near the nose, the nasal area is essentially the trapezoid with the endocanthions and lateral alaes as the corners. The forehead and eyes section is a semi-circle formed by the curvature of the spherical face with the base line running horizontally around the base of the orbits and the nasal area removed. The mean square area for each region is calculated using ICP and then the two are combined to achieve a final matching score. This method achieved an accuracy (verification rate) of 98.5% with a false acceptance rate of 0.1% for, neutral vs non-neutral and 99.4% when only comparing neutral expression. They also note that the nasal area was less affected by variation in expression than both the eyes and forehead. Although they performed well on a large dataset they do note that a number of studies have seen a large decrease in the accuracy rate when transferring their methodologies from a small to a large dataset. Mian et al. (2007) also resample the faces to equal the resolutions which helps as they use a point to point version of ICP which is more accurate with models of similar resolution.

Using an expression deformation model, where an observed change in expression is known, quantified and repeated Al-Osamini et al. (2009) employed ICP to achieve 94.1% accuracy for facial recognition between a neutral expression image and the deformed prediction of neutral expression from the expressive image. This was conducted using the FRGC v2.0 database.

Smeets et al. (2010) used a combination of isometric deformation models and ICP on sources from part of the BU-3DFE database. They analysed data from 100 individuals shown in 900 three-dimensional images, depicting both neutral and expressive faces. To allow for isometric deformation, the faces are cropped to show the same area and be the same size, enabling direct point to point comparison. This is done using a spherical area with the nose at the centre but still depicting the whole face. For the region based ICP closer cropping is conducted, still based around the nose at both 30mm and 90mm. Each approach is ranked for every probe face analysed and the ranks are combined through summing, multiplication and taking the minimum rank. The best approach was concluded to be the sum of the isometric deformation and both ICPs at 30mm and 90mm, this gave a rank one recognition rate of 94.48% and equal error of 5.85%. This method performed better than other non ICP based approaches conducted on the same database.

Zhao et al. (2011) initially used shape index comparison to find both endocanthions and the pronasale from a selected area and apply course alignment (pre-registration). Partial ICP is then conducted to reduce the 100,000 points of each face to approximately 10,000 and increase the speed of comparison. The main focus of Zhao et al.'s research is to remove the variance of scale. This is an unusual problem as almost all data collection methods provide accurate scale at the point of capture, so unless this information has been lost during exportation or comparison with two-dimensional images of unknown scale are needed, it appears redundant.

Mohammadzade & Hatzinakos (2013) initially used nose tip detection and PCA to preregister the images from the FRGC v2.0 database. In a similar way to Mian et al. (2007) and Smeets et al. (2010) they crop the facial image in order to remove poor or erroneous data and non-facial artefacts such as clothing. They initially use a sphere with a radius of 100mm which was larger than the one used by Mian et al. An 80mm by 80mm square grid in the *xy* plane, with the nose tip at the centre, is then selected and any points forward of the nose tip *z* coordinate are removed. In this method holes are filled automatically in the scans which have them, either due to scan error or because they are natural orifices. Rather than calculate the distances between every face in the database and every other, this study is proposing the use of a single generic face and the comparison of the distances between that and each other face. The idea of this is to dramatically reduce the number of comparisons needed thus reducing the processing time. Rather than using a simple ICP algorithm, they have adapted it to recognise specific features such as the eyes and the lips. This is therefore capable of finding, for example, the closest distance between the outer edge of the eye on the generic face to the outer edge of the eye on the input face. This is an improvement to the previous technique which does not measure the distance between specific features but merely measures the distance between the nearest points wherever they happen to be on the face. To enable this method, denoted as Iterative Closest Normal Point Algorithm, they employ the texture images in correspondence with the three-dimensional meshes to identify features. Many adaptations of the methodologies were evaluated in this study and a number of conclusions put forward. Firstly, they concluded that PCA is not as useful as ICP type methodologies when it comes to facial recognition. This suggests that size might be a useful feature of recognition, which is particularly pertinent where there is variation in expression. Secondly, the more scans of each individual contained in the dataset the better the verification rate, again especially when dealing with expressive images. Thirdly they concluded that their Iterative Closest Normal Point algorithm outperforms the simpler ICP by introducing feature correspondence, again this was more pronounced with expression variation. Fourthly they showed that it is possible to conduct three-dimensional facial recognition through the use of a generic face thus reducing the processing time.

Author	**Number of Subjects**	**Number of Images (2D, 2.5D or 3D)**	**Size of images**
Medioni & Waupotitsch 2003	100	700	NA
Cook et al. 2004	30	30	NA
Lu et al. 2004	18	113	NA
Lu et al. 2006	200	598	NA
Papatheodorou & Rueckert 2004	62	806	NA
Maurer et al. 2005	466	4007	480x640 pixels
Chang et al. 2005	466	4007	480x640 pixels
Gökberk et al. 2006	106	579	4000 points
Amor et al. 2006	50	400	7000 vertices, 13000 triangles
Mian et al. 2007	466	4007	480x640 pixels
Al-Osamini et al. 2009	466	4007	480x640 pixels
Smeets et al. 2010	100	900	NA
Zhao et al. 2011	30	90	10,000-100,000 points
Mohammadzade & Hatzinakos 2013	466	4007	480x640 pixels

Table 3.1 List of ICP based facial recognition studies with the number of subjects and images included in each study.

3.4 The Performance of Facial Recognition Systems When Encountering Twins

However effective facial recognition systems are reported to be, almost all the databases and datasets used are of unrelated individuals or at least the presence or absence of relatives is not reported. Although rarely used for criminal purposes, there are many reported cases of twins or siblings using each other’s identities for travel, work and proof of age where their own may be lost or have some kind of restriction or endorsement attached. As facial recognition systems are used more widely they need to be able to accurately identify differences between individuals however similar they may appear to be, in order to prove that they are robust (Amor et al. 2006).

The only case where twins are noted in an otherwise normal dataset is in Bronstein et al. (2005). They suggest that the robusticity of technique is high as they are able to differentiate between the twins using rigid surface matching and canonical form matching but not Eigenface.

No other datasets found, report that there are any other form of genetic relationships between the subjects and it is normal for the age of the participants not to be mentioned by authors. It can be assumed that they are all adults unless otherwise mentioned but the age and sex profile of datasets seems to remain unspecified.

Sun et al. (2010) conducted a two-dimensional study of 134 individuals (67 pairs of twins) with approximately 20 images per person. They compare iris recognition and fingerprint analysis to facial recognition and their ability to identify when a large number of twins (both identical and non-identical) are imported to the database. While the iris and fingerprint recognition programs showed little variation in performance when comparing twins, the facial recognition system performed poorly with an Equal Error Rate (EER) of 13%.

Seven existing facial recognition systems were put to the ultimate test by Paone et al. (2014) in a study using 126 pairs of identical twins. There are 17486 2D images collected in one year and a further 6864 images from 120 pairs of twins (including 48 of the original pairs) collected one year later at the Twins Day festival in Ohio in the years 2009 and 2010. All twins were over 18 years of age. Before the addition of twins, the algorithms performed with an average (EER) of 1.1%, this increases to 15%-20% after the twins are added with a 12% EER variance across the seven systems. When the variance of ageing is added to the equation all the systems decreased in performance with EERs of between 25%-40%. The results of this study show that although possible in ideal conditions (i.e. no illumination, age, pose or expression changes) the identification of twins using facial recognition systems is no simple task. The slightest variation between images challenges the best two-dimensional algorithms available on the market, meaning that twins are incorrectly identified as each other on a number of occasions.

Vijayan et al. (2011) applied the challenge of twin identification to the three-dimensional facial recognition algorithms. One hundred and seven pairs of identical twins were each scanned twice, one scan showing a neutral expression and the other with the twin smiling. These scans were collected at the Twins Day festival in 2010. Four ‘state of the art’ algorithms were then tested and compared. They found that all the algorithms performed less well than their reported levels on the FRGC v2.0 database and that where the state of expression varied between probe and gallery images, the accuracy was further reduced. The worst case was where a smiling image of Twin A and neutral image of Twin B are attempting to match a gallery containing a neutral image of Twin A and smiling image of Twin B. The algorithms are confused by the matching expression and often match a smiling Twin A to a smiling Twin B rather than to its own neutral image.

Chapter 4: Heritability Studies of the Craniofacial Region

Heritability and genetics have been widely studied since Darwin, in a manner of different ways and for a variety of different reasons. Early studies were limited to the physical manifestations of genetic inheritance whereas since Crick and Watson's breakthrough in the 1950s, the specific genomes responsible for every aspect of all living organisms are slowly being identified.

The face has been studied with respect to both the physical dimensions, (see table 4.1) and genomes (Boehringer et al. 2011, Hammond et al. 2012 and Liu et al. 2012). The main aspect of the face where the genomes have been studied and understood are those linked to eye colour (Zhu et al 2004) and more are being discovered every day. For this study the area of interest is the physical manifestation of genes as seen in parent and sibling relationships. Although this topic has not been widely studied the main aspect of research has been through the analysis of siblings specifically twins (Djordjevic et al. 2012, Weinberg et al. 2013). Relatively few studies have focused on or considered the relationship between parents and their offspring. This may partly be due to the difficulties and limiting factors such as of tooth loss orthodontic treatment (AlKhudhairi & AlKifide 2010) and the practicalities of data collection. Even fewer studies enable the total view of family relationship by studying both the parent-offspring and sibling relationships together (Saunders et al. 1980, Jelenkovic et al. 2010). In this chapter the two types of relationship will be addressed separately, siblings first and then parent-offspring, and subsequently compared with each other. The basic details of the studies being reviewed can be found in table 4.1.

Study	**Twin/Sibling/ Parent**	**Number of subjects (M-Monozygotic, D-Dizygotic)**	**Age of children (yrs.)**	**Means of data collection**	**Type of measure**	**Skeletal/soft tissue**	**Population**
Hunter et al. 1970	Parents	38 families	17-21+	Lateral cephalogram	2D	Skeletal	American
Nakata et al. 1973	Parents & Twins	64 families	8-17	Lateral cephalogram	2D	Skeletal	American
Saunders et al. 1980	Parents & Siblings	147 families	14-21	Lateral cephalogram	2D	Skeletal	Canadian
Lunstörm & McWilliam 1987	Twins	56 pairs 28M 28D	13-20	Lateral cephalogram	2D	Skeletal	Swedish
Lunstörm & McWilliam 1988	Twins	55 pairs 28M 27D	12-20	Lateral cephalogram	2D	Skeletal	Swedish
King et al. 1993	Siblings	104 pairs	9-22	Lateral cephalogram	2D	Skeletal	American
Kitahara et al. 1996	Parents	985 families	6-17	Lateral cephalogram	2D	Skeletal	Japanese
Manfredi et al. 1997	Twins & Siblings	30 pairs 10M 10D 10 Sibling	10-13	Lateral cephalogram	2D	Skeletal	Italian
Savoye et al. 1998	Twins	79 pairs 33M 46D	9-16	Lateral cephalogram	2D	Skeletal	Belgian
Carels et al. 2001	Twins	79 pairs 33M 46D	9-16	Lateral cephalogram	2D	Skeletal	Belgian
Eguchi et al. 2004	Twins	78 pairs 44M 34D	12-22	Dental casts	3D	Dental	Australian
Naini & Moss 2004	Twins	26 pairs 10M 16D	12 (mean)	Laser scan	3D	Soft tissue	British-mixed races
Johannsdottir et al. 2005	Parents	363 children	6-16	Lateral cephalogram	2D	Skeletal	Icelandic
Gelgör et al. 2005	Parents	120 families	9-17+	Lateral cephalogram	2D	Soft tissue depths	Turkish
Baydaş et al. 2007	Siblings	138 pairs	18-28	Lateral and posteio-anterior cephalogram	2D	Skeletal and Soft tissue	Turkish
Amini & Borzabadi-Farahani 2009	Twins	50 pairs 25M 25D	13-20	Lateral cephalogram	2D	Skeletal	Iranian
AlKhudhairi & AlKifide 2010	Parents	24 families	17+	Lateral cephalogram	2D	Skeletal	Saudi
Jelenkovic et al. 2010	Parents & Siblings	122 families	13 +	Physical	2D	Skeletal and Soft tissue	Belgian
Djordjevic et al. 2012	Twins	37 pairs 19M 18D	15	Laser scan	3D	Soft tissue	British
Weinberg et al. 2013	Twins	21 pairs 10M 11D	5-12	Structured light scan	3D	Soft tissue	American Caucasian

Table 4.1 Review of facial heritability studies.

4.1 Heritability Assessed through Twin Studies

Twins studies are used across a wide variety of genetic studies due to the unique insight they can give. Twin studies are able to use the variation between the two types of twins, monozygotic and dizygotic, to assess genetic versus environmental influences. Identical or monozygotic twins arise from the same egg, which splits to form two embryos with the same genotypes. Dizygotic twins are each formed from a different egg which happen to be fertilized at the same time, making them genetically the same as any other siblings with the exception of being born at the same time so there is no age variation. For craniofacial dimensions and morphology, heritability can be assessed for each measurement individually and across wider areas. If the same similarities are seen in both monozygotic twins and dizygotic twins then the trait is thought to be highly heritable. It is thought that whilst all siblings are exposed to similar environmental factors, their impact on the body is associated with genetics. In the light of this, the response of monozygotic twins to external affects will be the same or very similar while dizygotic twins may respond differently to each other. Therefore, if a trait is only seen in monozygotic twins then it does still show some heritability but is also subject to environmental factors which affect the dizygotic twins differently. If traits vary between the monozygotic twins then there is not heritable effect and the cause must be environmental, the most likely cause is a functional response to a specific action one of the twins performs.

The testing methods for detecting whether twins are monozygotic or dizygotic are varied in both manner and accuracy. The studies reviewed below range in their methods of assessment from the analysis of genomes (Eguchi et al. 2004) to a simple questionnaire (Weinberg 2013). Some start with the correct assumption that all monochorial twins, sharing a placenta, are monozygotic (Savoye et al. 1998) and that twins of opposite sex are dizygotic (Carels et al. 2001). It is the subsequent determination of the remaining same sex twins that varies in its accuracy. Eguchi et al. (2004) and Djordjevic et al. (2012) both use genetic testing, while Savoye et al. (1998) and Amini & Borzabadi-Farahani (2009) use both blood groups and five simple DNA polymorphisms and Naini & Moss (2004) just use blood groups. The least precise test was conducted by Weinberg et al (2013) who asked parents to complete a 15 question questionnaire about similarities and differences of the same sex twins. Whilst the specific questions are not published other studies have been known to use tongue roll, phenylthiocarbamide taste, hair and eye colour to determine zygosity (Lauweryns et al. 1993, Nakata et al. 1973).

Nakata et al. (1973) were attempting to use the facial size of parents to predict the final dimensions of their offspring. They looked into both the relationship between parents and children, and between the twin offspring. A total of eight measurements were collected and of them five expressed significant variation between twin pairs suggestive of environmental influences. They did however state that the angular measurements exhibited more heritability than the linear suggesting that heritability may be a more complex issue than can be understood by just two dimensional linear measurements.

Lundström & McWilliams produced two publications based on the same population of twins, one considered at the linear distances of the lateral skeletal form (1987) while the other assessed the proportions (1988). Overall the conclusion they formed was that the vertical dimensions had higher heritability values than the horizontal dimensions. In the later paper the dimension with the least accurate value of heritability was the anterio-posterior incisal edge relationship. The suggested causal factor for this dissimilarity was the environmental influence of the dentition.

Savoye et al. (1998) conducted a very similar study to Lundström & McWilliams (1987, 1988) and gained very similar results, even though they mistakenly report the opposite result at one point in their publication. Carels et al. (2001) used the same study population as Savoye et al. (1998) but were able to expand the study to give insight into specific dimensions. They again found that the vertical dimensions expressed more heritability than the horizontal, they found that all the variables were inherited through additive genes (multiple genes that each code for the same trait and rather than one exhibiting dominance they work together in addition) with the exception of the mandibular body length. The mandible body length was judged to be controlled by dominant alleles and so showed very high heritability which might explain the obvious similarity of chin protrusion seen in some families such as the Habsburg Spanish imperial family. In concordance with this the lower dentition displayed higher heritability than the upper, although they both showed significant genetic control. The anterior facial height was the only dimension to show a significant sexual variation, the genetic control of this dimension was shown to be stronger in males than females.

Eguchi et al.’s (2004) study of the dental arch confirms that the morphology of the dental area does have a genetic influence. The height, length and breadth of both the upper and lower dental arches all had high levels of heritability whilst tooth crowding, rotation and overbite were environmentally influenced. Variation in the level of heritability was seen between the posterior and anterior teeth with the posterior showing more similarity between twin pairs. The suggested theory that functional and habitual activities cause displacement of the shorter and single rooted anterior teeth is further evidenced by the low heritability of the horizontal position of the anterior teeth as described by Lundström & McWilliams 1988 and Savoye et al. (1998).

Naini & Moss (2004) were among the first to take the study of twins into the three-dimensional world with surface laser scans. They used inter landmark measurements and surface shape analysis to look at the variation between 26 pairs of twins. The inter landmark distances confirmed the results of previous authors (Carels et al. 2001, Lundström & McWilliams 1987, 1988, Savoye et al. 1998). All of these results found that vertical proportions of the face showed greater heritable influence than the horizontal, this was seen particularly in the lower face. The highest levels of heritability were seen for the intercanthal, orbit and nasal width and the nasal height. The surface analysis consisted of the automatic calculation of the type of surface curvature for all areas of the face, using Gaussian and mean curvatures. Examples of the type of curvatures exhibited were domes, ridges, saddles and troughs and each of these was colour coded so that the overall distribution could be visualised. A visual comparison of the colour coded scans of each twin pair was conducted and the overall areas of similarity and disparity noted. Monozygotic twins showed similar surface types in the triangular area with the apex just below the nose and the base across the forehead just above the eyes. The dizygotic similarities differed between the sexes. Female same sex twin pairs exhibited the most similarity under the eyes while male same sex twin pairs resembled each other around the eyes, the nasal root and the chin. No similarities were seen between the dizygotic twins of opposite sex. The area around the mouth showed little or no similarities across any of the twins, female twins especially exhibited pronounced disparity in the mouth area. It must be noted that the surface element of this study is entirely visual and observer based.

Amini & Borzabadi-Farahani (2009) agreed with previous authors (Carels et al. 2001, Lundström & McWilliams 1987, 1988, Savoye et al. 1998) who concluded that the vertical dimensions were more controlled by genetics than the horizontal. This was particularly true in the lower portion of the face as described by Naini and Moss (2004). The anterior portion of the face had higher heritability values than the posterior which counteracted Eguchi et al.'s (2004) findings although they were only based on the dental arch. In contrast to Naini and Moss (2004) the overall dimensions of the lower face, and not simply the vertical dimensions, were seen to be more inherited than those of the upper face. Low levels of inheritance were calculated for the dento-alveolar area compared to the rest of the face which appeared to differ from the results of Eguchi et al. (2004) and Carels et al. (2001) who both observed high levels of heritability for the dental region. In particular, the length of the mandible was shown by Carels et al. (2001) to be controlled by dominant genes and was seen to have a low level of heritability.

Using the Avon Longitudinal Study of Parents and Children to recruit participants, Djordjevic et al. (2012) conducted laser scans of 15 year old twins and applied landmark based Procrustes analysis and a surface-based Trimmed Iterative Closest Point algorithm (TrICP), to explore the similarities between same sex twins. Procrustes analysis was unable to detect any significant variation between the two types of twins and so was not helpful in assessing the heritability of specific areas of the face. The surface-based TrICP was however able to quantify the heritability of the face as a whole and create colour maps to visualise the specific parts of the face with the highest and lowest heritability. The average monozygotic and dizygotic face for each sex were compared, the areas showing the least variability are thought to be more inherited than the areas showing greater variability. The males' faces coincide most in the lower forehead, supra and infra orbital ridges, the bridge of the nose and the lower lip, while the females' faces are most similar around the eyes, supra and infra orbital ridges, the philtrum and the lower cheeks. Overall these results concur with those published by Naini and Moss (2004) with regard to the focus around the eyes but add additional areas of heritability such as the lower lip for males and lower cheeks and philtrum for females. Although it is difficult to compare the skeletal dimensions shown by most of the previous studies to the soft tissue facial form shown in this study, the heritability of the lower lip in males concurs with Carels et al.'s (2001) result showing the lower dentition to be under more genetic control than the upper dentition.

Most recently Weinberg et al. (2013) used landmarked three-dimensional structured light scans, geometric morphometrics and principal component analysis to calculate heritability. Moderate to high heritability was displayed by principal component scores and were seen in the interorbital region, the breadth of the nose and the height and protrusion of the upper lip. The similarity and heritability of the interorbital region and nasal breadth are in total concurrence with the results of Naini and Moss (2004) but the upper lip results differ from all the other three-dimensional and two-dimensional studies. It should be noted that the lower mouth and chin area suffered from very poor scan quality in this study and therefore landmark placement was very limited in this area. This being understood the absence of heritable features in this area should not be seen as an absence but as an unknown, so the similarities of the lower lip and lower cheek shown by Djordjevic et al. (2012) and of the chin shown by Naini and Moss (2004) can neither be confirmed or denied.

Taken as a whole the two-dimensional lateral cephalographic studies all observed higher levels of heritability in the vertical dimensions compared to the horizontal (Amini & Borzabadi-Farahani 2009, Carels et al. 2001, Lundström & McWilliams 1987, 1988, Naini & Moss 2004, Savoye et al. 1998). The prominence of the anterior face as opposed to the posterior, may suggest that the frontal facial area depicted by the limited 3D studies might show higher levels of heritability as they are fairly closely related. The dentition, especially the anterior dentition (Amini & Borzabadi-Farahani 2009, Lundström & McWilliams 1987, 1988, Savoye et al. 1998) was shown to have lower heritability than the rest of the face as was it subject to environmental and functional variation. The upper dental area when compared to the lower, was also less controlled by genetics (Amini & Borzabadi-Farahani 2009, Carels et al. 2001) a result that was replicated in three-dimensions by Djordjevic et al. (2012). There remains some dispute as to the specific areas of the facial surface which have the highest levels of heritability but the single area that all authors agree upon is the eyes and their immediate surrounds. For all same sex comparisons, the eyes are a common feature of heritability for both males and females, the exception to this is the similarity between opposite sex twins. Opposite sex twins show no significant similarities in Naini and Moss (2004) and were not considered by Djordjevic et al. (2012) or Weinberg et al. (2013). It may be that the variation caused by sexual dimorphism is great enough to mask any similarities due to heritability or it might be a phenomenon of the specific age of the participants of that study, the mean age of which was 12 years.

4.2 The Use of Singleton Sibling Pairs

The use of siblings rather than dizygotic twins is unusual when attempting to assess the heritability of features and morphology. The most common reason for using siblings is availability. Whilst there are numerous twin studies and meetings set up around the world, the most famous being the Twins Day Festival in Twinsburg Ohio, it is still difficult to establish a large dataset of twins. The number of twins has increased over the last few decades, however this is mostly due to the increase in In vitro fertilisation (IVF). There is a one in five chance of twins from IVF but the majority of these are due to multiple embryos being implanted and so result in dizygotic twins whereas there is only a one in 80 chance of twins from natural conception, of which only one in five are monozygotic (National Health Service 2015). In general, this means that the increase in monozygotic twins is only very slight, so it continues to be a difficult task to find and secure enough twins for a study to be conducted. So while the use of siblings is not ideal, for some research it is the only way to get a large enough sample for comparison where there is not readily available access to large twin datasets.

While Saunders et al.’s (1980) study mainly focuses on the relationship between parents and their children they do also investigate the relationship between sibling pairs. In general accord with the twin studies (Amini & Borzabadi-Farahani 2009, Lundström & McWilliams 1987, 1988, Savoye et al. 1998) the significance of the measurements involving the dentition was low while the other portions of the skull had greater significance and varied little from each other. The most interesting findings were seen in the overall number of significant results for each of the sibling categories. The brother-sister pairs, that is siblings of the opposite sex, had a greater number of significant results than siblings of the same sex. One would expect that the sex difference would cause greater disparity between opposite sex siblings and would show fewer dimensions of significance than those of same sex siblings. There are a number of possibilities as to why this might have occurred, but as neither BMI, age gap or environmental factors were not considered in this study it would be impossible to confirm if any of those factors played a role.

King et al. (1993) attempted to rectify gaps in previous research caused by selection bias. Malocclusions were thought to be a factor of environmental influence, but selection bias in previous studies caused by excluding those with the most severe malocclusions due to treatment having been received, meant that the true picture had not been assessed. Their study of siblings concluded that whilst malocclusions are acquired, the predisposition to have the facial shape and behavioural factors which increase the likelihood of malocclusion, are inherited.

In a combined skeletal and soft tissue study Baydaş et al. (2007) assessed the heritability of the facial proportions of a Turkish population. All sibling pairs were combined so no variation between same sex and opposite sex pairs, or male and female pairs, is described. The proportions which exhibited the highest levels of heritability were the soft tissue chin thickness and the soft tissue facial angle, the angle of the nasion-pogonion line. The anterior facial height had moderate heritability as was previously seen by Carels et al. (2001) and the vertical measurements displayed more genetic control than the facial depths, which findings concur with almost all the twin studies (Amini & Borzabadi-Farahani 2009, Carels et al. 2001, Lundström & McWilliams 1987, 1988, Naini & Moss 2004, Savoye et al. 1998). Baydaş et al. also described a greater genetic control of the lower lip in comparison to the upper lip, which agrees with the findings of Carels et al. (2001) and Djordjevic et al. (2012) but disagrees with those of Weinberg et al. (2013) although this may be due to the data error in that study, discussed above. Overall a greater genetic control was seen in the soft tissue measurements than the skeletal ones.

One of the few studies to compare twins and singleton siblings in the same study is that of Manfredi et al. (1997) although only same sex siblings were considered. The results mirrored those seen in both the twin and siblings studies: the anterior measurements showing greater genetic control than the posterior (Amini &

Borzabadi-Farahani 2009); the vertical more than the horizontal (Amini & Borzabadi-Farahani 2009, Carels et al. 2001, Lundström & McWilliams 1987, 1988, Naini & Moss 2004, Savoye et al. 1998); the dento-aveolar area displaying poor heritability (Lundström & McWilliams 1987, Savoye et al. 1998, Amini & Borzabadi-Farahani 2009); the anterior facial height standing out as a good measure of heritability (Baydaş et al. 2007). Nakata et al. (1973) had given some indication that the shape showed more heritability than the size and this is seen in the mandible in this study. The main point of disagreement between this and other studies is the suggestion that the lower third of the face is under stronger genetic control than the upper two thirds, whilst Djordjevic et al. (2012) had seen this in their Procrustes analysis the of three-dimensional studies which saw the most similarity in the upper portion of the face around the eyes. The main point to note from this study is the finding that the results for singleton siblings are comparable with those of dizygotic twins and so can be used in their place. Monozygotic twins are still required to examine thoroughly the influence of genes upon the structure of the face.

4.3 Heritability Assessed through Parental Studies

Hunter et al. (1970) was one of the first studies to look at dimensions and their heritability rather than the simple the presence or absent of traits. Using lateral cephalographs they confirmed that dimensions did show heritability and were able to compare fathers vs mothers and sons vs daughters. They found that whilst all dimensions showed significance for a father's offspring only the correlation coefficient for the anterior face height was significant for a mother's offspring. This was true for both mothers-daughters and mothers-sons and meant that the other measurements which involved the mandible showed little correlation between mothers and their children. The results were also able to demonstrate that there is little difference in the correlation between daughters and sons. Hunter et al. (1970) did make an interesting point about the combination of the measurements of both parents. They said that while the combined 'midparent' can provide more accurate results overall, the most significant variable can be missed as this often derives from a specific parent. It is not known to which parent each child might most closely correspond.

Nakata et al. (1973) expand upon an existing twin study by collecting parental data in addition through the acquisition of lateral cephalolgraphs. They combined the dimensions of monozygotic and same sex dizygotic twin pairs and compared them with the ‘midparent’ dimensions formed from the combination of both parents’ data. In agreement with Hunter et al. (1970) father-offspring were seen to have higher correlation confidences and more significant values than mother-offspring. There was no consistent pattern found for daughters and sons with father-son averaging greater correlation than father-daughter while mother-daughter outperformed mother-son. The authors noted that age was also a strong predictor of variation in the offspring. Since craniofacial growth is also controlled by age and sex the genetic influence may change at different ages.

Saunders et al. (1980) looked into all possible family relationships. When it came to the parent-offspring dimensions there was little variation between the sexes of either generation. The only slight variance was seen in the mother-son comparison which had a lower statistical significance than the other relationships. It was suggested that this may be due to an evolutionary trend in growth meaning that children of both sexes would be more similar in size to their fathers rather than their mothers. They also confirmed what Hunter et al. (1970) had suggested, i.e. that ‘midparents’ were more significant overall. When it comes to separating the different dimensions measured, the dental measurements were much less significant than the remainder of the face. The mandible was the least significant area for mothers-sons which corresponds to Hunter et al.’s (1970) findings for mother-offspring as they did not separate their results into son and daughters.

Kitahara et al. (1996) were able to look at the variations in similarity between parents and offspring across three age ranges; six to nine years, ten to 13 years and 14 to 17 years. They advised the use of average growth rates to predict the growth of children, as offspring from the same family do not always show the same growth patterns. To assess the heritability of facial shape, disparity from a mean face is calculated for each category in turn. Due to a previous study where no effect on disparity was seen for sex variation in offspring, they were combined into a single category rather than being split into daughters and sons. No significant difference was found between the mother-offspring or father-offspring disparity in any of the age ranges, although the father-offspring category did have slightly stronger significance. For both parents the combination of all the measurements increased in significance as age increased, with no significance in the six to nine year age group. The maxilla-facial dimensions were shown to display more heritability than the dento-alveolar and of the dento-alveolar the upper showed less heritability than the lower dentition. This compares with the results of both Hunter et al. (1970) and Saunders et al. (1980) who suggest that this is because dental shape is more affected by diet and mastication than by genetics. It must be noted that all of the child subjects of this study suffered from malocclusion which is how they became known to the authors.

Johannsdottir et al. (2005) were able to study children longitudinally with a ten year age gap between the data collection dates. Lateral cephalographs were taken at six and 16 years old and compared to parental ones. It is not clear whether the parental data were only collected at the first point of contact when the children were six or whether a second radiograph was taken at the later date. This might positively influence the strong relationship seen between age and the genetic similarities. If radiograph was only taken of the parents when the child was first pictured, then the age gap would be reduced from the comparison of the six year old to the comparison of the 16 year old child. This in itself would account for the increase in similarity without taking any account of genetics. Although the result might be influenced it does sit in agreement with other studies (Kitahara et al.1996, Saunders et al. 1980). In harmony with Saunders et al.'s (1980) results daughters were seen to be more similar than sons to both parents. The father-daughter and mother-daughter correlation were very similar, with the father-daughter displaying a slightly greater significance as seen in Kitahara et al.'s (1996) study. At the same time the mother-son correlation was greater than that of mother-daughter which is contrary to Nakata et al. (1973), Saunders et al. (1980) and Hunter et al. (1970). Dentition was again held responsible for the areas of low heritability with the function of breathing and mastication, and the influence of nutrition being thought of as greater influences on tooth position than genetics.

Gelgör et al. (2005) was one of the very few studies to measure soft tissue depth rather than skeletal dimensions. Even though they were looking at different measurements the results of the study overall are very similar to those seen in Johannsdottir et al. (2005). Generally, mothers were more similar to their offspring than fathers, daughters were more similar than sons and similarity increased with age. The greatest correlation was seen in upper lip thickness, chin thickness and the soft tissue facial angle although the parental influence differed between them. Upper lip thickness and soft tissue facial angle had a greater correlation for father-offspring while the chin thickness was greater for mother-offspring.

AlKhudhairi & AlKifide (2010) compiled a study of ‘perfect’ families all consisting of father, mother, son and daughter. Father pairings to both sons and daughters showed stronger values than mother pairings with the daughters displaying higher values than sons for both parents. Whilst the superiority of the daughters’ relationship to their parents accords with that seen in most studies when the variation was considered (Gelgör et al. 2005, Johannsdottir et al. 2005, Saunders et al. 1980), the father to mother comparison continues to change from study to study. For these results, this study confirms the findings of Hunter et al. (1970) and Nakata et al. (1973) and is similar to those shown by Kitahara et al. (1996), but differs from those of Johannsdottir et al. (2005) and Gelgör et al. (2005). Focusing more on individual dimensions the mandible offers greater significance than other measurements for all relationships which conflicts with Hunter et al.’s (1970) results for mothers. The lower facial height is shown to exhibit a greater genetic influence than the upper face.

Using data originally collected in the 1960s Jelenkovic et al. (2010) compared the dimensions of numerous measurements taken directly from the subjects. Whilst they approached the question of genetic influence differently to the other studies, by having age and sex as variables rather than splitting the relationships into categories, they were able to see genetic similarities. The most interesting result was that the skeletal traits were seen to be influenced to a greater extent by genetics than by the soft tissue depth and the soft tissue traits. Although this is in direct opposition to the results of Baydaş et al. (2007) it is possible that it is due to the early growth of the craniofacial skeleton compared to the soft tissues, allowing for environmental factors to have had more of an influence on the soft tissue growth. Across all the measurements, age and sex variation were able to explain between 7% and 46% of the variation, suggesting that if they can be controlled in some way then the genetic influence would be easier to see.

The main areas of agreement across the parental studies are: the increase in heritability with an increase in the age of the children; the superior heritability of daughters compared to sons irrespective of the parental relationship; the low levels of heritability seen in the dento-alveolar area contrasting with high levels associated with the mandible.

The major area of disagreement across parental studies is the proportional levels of heritability shown by mothers and fathers; some studies suggesting that father-offspring relationship displays greater heritability than the mother-offspring relationships (AlKhudhairi & AlKifide 2010, Hunter et al. 1970, Nakata et al. 1973) while others concluding that the opposite is the case (Gelgör et al. 2005, Johannsdottir et al. 2005) while yet others suggest that the levels of heritability between mothers and fathers are even (Kitahara et al. 1996, Saunders et al. 1980).

4.4 Overview of Sibling and Parental Heritability

Whilst there are some features which display heritability both between siblings and between parents and offspring, due to the varying degrees of expression between the ages, sexes and parental relationships it is better to consider each relationship in their own right. The two studies which covered both types of relationship together (Nakata et al. 1973, Saunders et al. 1980) are however able to give a small insight into the comparative levels of heritability. Nakata et al. (1973) was able to show that while there was heritability in the parental-offspring dimensions it was less evident than that seen between the twins themselves. Saunders et al. (1980) agreed with Nakata in rating the correlation seen between siblings as greater, and of more significance, than that of parents-offspring. The results of both studies agree with the expected outcomes as siblings, especially twins, are likely to have a more similar genetic makeup than that seen between a parent and a child. The variation caused by age is also likely to be greater in a parent-offspring comparison unless a generational longitudinal study could be conducted, collecting data from both the parent and the child at the same chronological age.

Chapter 5: Materials and Methodology

5.1 Study Population

The study population consists of living people, adults and their offspring. The subjects were scanned at some point between May 2012 and March 2014. All relevant data relating to the subjects, for example their weight and height, were collected at the time of each scan so no subsequent variation could occur between each piece of datum being collected for an individual. Eligibility for the study was based on a number of factors; ethnicity, age, familial relationship and the presence or absence of any facial disfigurement.

Location of study: The scanning sessions were conducted in Kent (England), Norfolk (England) and Edinburgh (Scotland). Recruitment to the study was through a variety of social networks including school and church groups.

Ethnicity: The population studied is white British families. To classify as white British the two generations above the parents needed to class themselves as white British, meaning that four generations were white British. Although recruited from both countries, families were not classified as Scottish or English as there was much crossover even within the small number of generations in question.

Family: The minimum classification of a family used in the study is one biological parent and one child, although many of the families consisted of both parents and several children.

Age: The age of the parents was not limited in anyway, and ranged from 39 to 62 years. The ages of the children were however restricted as at least one of the children in each family scanned in the study was required to be between the ages of 14 and 25 years. This parameter was determined due to the cessation of growth in facial bones (see Chapter 2). Other siblings outside this age range were scanned if available to increase the study population and their age was then taken into consideration when analysing the results.

Facial disfigurement: Subjects with any form of facial disfigurement were also scanned but careful consideration was taken when analysing the data to see if this could be an influencing factor and avoid biases.

Facial hair: Another factor which would limit the use of a family member for the study was facial hair. The scanner does not have the capacity to scan hair successfully. This does not present a problem with head hair as the area beyond the hairline is not needed for facial recognition, but facial hair of any type presents a challenge. Any subjects with facial hair longer than 1mm (one or two days of growth for an average male) would create a problem in capturing data during the scanning process. Both the hair and the underlying skin reflect the light source from the scanner, but the scanner cannot differentiate between the two and so it is not possible to produce an image of the shape of the face under a beard. In some cases, it was possible to obtain a reliable image from a subject with ‘designer stubble’. All fathers with facial hair were scanned and images were used wherever possible.

Make-up: In most cases cosmetic foundation cream worn by female subjects did not affect the scan, however in one or two cases where the brand of make-up contained reflective particles, this adversely affected the quality of the image, and in one case the data could not be used for the study. Where participants were comfortable to do so they were asked to remove make-up that affected the scan (make-up remover was available if required).

After the removal of any unsuitable participants, see 5.2.2.1, a total 41 families were used for this study, a total of 139 participants categorised in Table 5.1. Within the male children there are two sets of twins, one dizygotic and one unknown type though thought to be monozygotic.

No of Children	Families	Parents			<14 yrs.		14-25 yrs.		Total	
		Both	Mother	Father	M	F	M	F	M	F
1	18	10	8	0	1	0	9	8	10	8
2	19	15	4	0	7	4	17	10	24	14
3	3	3	0	0	1	1	2	5	3	6
4	1	1	0	0	1	0	2	1	3	1
Total	41	29	12	0	10	5	30	24	40	29
TOTAL		70			15		54		69	

Table 5.1 Summary table of study population M=Male, F=Female

5.1.1 Ethical Approval and Accountability

No ethical approval was needed as all participants volunteered themselves for the study and any children under the age of 18 were included with their parents' full permission.

The study did not seek to verify the biological relationships in the same family using Deoxyribonucleic Acid (DNA) or blood group testing as this would have been both prohibitively expensive and over intrusive. Instead, parents were asked to sign a declaration stating that they were the biological parents of the children scanned with them. The declaration was at the end of their information questionnaire (see Appendix 1).

The father was asked to testify that to the best of his knowledge he was the biological father of the child/children scanned. The mother was asked to certify that she is the biological mother of the child/children scanned and also to testify that the father scanned is the true biological father of the child/children scanned.

In conducting the study care was taken to ensure that mothers and fathers were physically separated from each other while completing the questionnaires. This was usually achieved by having the 'mothers' questionnaire in one room and the 'fathers' in another. Once completed questionnaires were put immediately into a 'ballot box' type container which was only opened later by the researcher. These precautions were taken to ensure that in the event of a mother or father not being the biological parent for whatever reason, (adoption, infidelity, sperm /egg donation etc.) they could record this discreetly. A research assistant was available in the room to oversee the process and help participants with any questions that they were unsure of.

These parental declarations were completed separately, see above, and a confirmation from the author was included on the form stating that no information given would be released to any other participant.

All participants included in the study completed both parental declarations where both parents were scanned.

5.2 Methodology

5.2.1 Equipment

A wide range of both physical and digital tools was required to create the three-dimensional images, in addition to that a variety of computer software was used for data acquisition, manipulation and analysis. A white light scanner and Flexscan3D were used to produce three-dimensional surface representations of the participants. Once obtained these three-dimensional surface scans had pre-determined landmarks placed upon them using Viewbox software (dHal company) and the coordinates of these landmarks were then exported to Excel and inputted into Morpho J software to enable analysis.

Scanner: The equipment consisted of a white light projector mounted on a solid mount with two cameras mounted below at a distance of 20 cm each side of the projector lens. The mount was positioned on an extendable tripod allowing for height and angle adjustment. With the cameras at distance of 40cm from each other, the optimum position of the subject is 120cm in front of the system. The cameras and projector were connected to a laptop running FlexScan3D3.1.

<u>Technical specifications of the scanning equipment:</u>

- Optomo HD66 white light projector: 2500 lumens, 1280 x 720 resolution, 4000:1 contrast ratio.
- Fujinon CCTV Ueye lens - HF12.5HA-1B: 12.5mm focal length, 1:1.4 maximum relative aperture, manual aperture and focus.
- Toshiba Satellite L670-1DN: Windows® 7 Home Premium 64-bit, Intel® Pentium® processor P6100, DDR3 RAM (1,066 MHz).
- 3D3 solutions camera and projector tilting mount: five camera positions, 30˚ angle variation.
- Manfrotto 804RC2 three-way tilt head and tripod: -30˚ to +90˚ front tilt and lateral tilt, 35cm min height, 146cm max height.

FlexScan3D: Version 3.1.9.109 – Enables capture of surface data in conjunction with cameras and projector which are linked up. Once captured, alignment of multiple scans is possible through mesh-to-mesh alignment and best fit calculation. These aligned scan can be merged into a single scan with adjustable degrees of hole filling and accuracy. Surface data can be exported in .obj or .stl files. Flexscan can also calculate the deviation between two forms of surface data and outputs the relevant statistics.

Viewbox: Version 4.1.0.0 Beta – Allows for three-dimensional images to be input and landmarks to be placed upon the surface of these images. The images can be rotated in all three dimensions, a computerised light source can be altered to give a range of light intensities and directions and specific views and settings can be saved. A template can be formed to aid in the speed of landmark placement and functions are available to select the most prominent or most depressed point of the surface in a specific area. Once landmarks have been selected the software allows for exportation in the form of three-dimensional coordinates of the landmarks in a .txt or .xsl format. Variables can be input to each patients file at this stage if required and exported along with the three-dimensional coordinates. A Trimmed Iterative Closest Point (TrICP) algorithm can be utilised to compare the similarity of any two three-dimensional surfaces.

Morpho J: Version 1.05f with Java 1.7.0_25. Used to analyse the three-dimensional coordinates of the landmarks. Variables can be added to the original data and Procrustes fit, PCA and regression can be performed.

MedCalc: Version 15.8. Used to conduct Receiver Operating Characteristic Analysis (ROC)

Set of scales: John Lewis Digital Scale with LCD screen, 150kg capacity, one decimal place.

Height scale: Black and Decker, solid metal 3m rule.

5.2.2 Data acquisition

Each participant is given a unique ID to keep track of the family relationships. The unique ID is in the form of a family number, a family position and a child number where necessary.

For example, if the first family consists of a mother, father and two children they would be coded as F1M, F1F, F1C1 and F1C2 respectively. Where there is only one child in a family the code for that child is still C1 rather than just C. These codes are use throughout the data acquisition and analysis enabling results to be easily assessed.

5.2.2.1 Questionnaire

The initial data collected was in the form of a questionnaire. This was completed by each participant separately, the same questions asked of all subjects, mothers, fathers and children, with the final declarations of paternity and maternity being the only exception (see 5.1.1 for more details). The observer recorded the sex of the children at the time of scanning and added this to the information collected on the questionnaires.

In addition to personal and contact data which has been excluded from the study to keep the subjects anonymous, the questions covered six main areas; Age, Body Mass Index (BMI), Ethnicity, Disease or Trauma, Dental morphology and Diet. These were all possible variables which could mean that the subject needed to be removed from the study or that the variables needed to be taken into consideration when analysing the data.

The questionnaires were completed in hard copy at the scanning sessions.

5.2.2.1.1 Age at time of scan

The date of birth of the participant and the date the scan is taken are both recorded allowing for the age of each participant to be calculated accurately to the day.

5.2.2.1.2 Body Mass Index (BMI)

The participant’s height and weight are collected. This was conducted at the time of scanning on the same set of scales which was checked and calibrated between each scanning session. The height scale is made of a resilient material but was also checked randomly against two other length scales. The weight was taken with the participants clothed but footwear; bulky outerwear and other personal items are removed beforehand. Footwear is also removed when measuring the height of participants. BMI is calculated using the formula BMI = Weight (kg)/Height^2 (m).

5.2.2.1.3 Ethnic origin

The ethnic origin was asked of each family member as well as that of their parents and grandparents, this means that the ethnic origin of four generations of participant families has been accounted for. An example of White British was given in the questionnaire but apart from that, the specific manner of defining ethnicity was self-declared, no list of options was provided.

5.2.2.1.4 Disease and Trauma

Participants were asked whether they had ever sustained any injuries to the face, head or neck area, prompts were also provided which included examples such as a broken nose. A broken nose would not necessarily mean exclusion from the study as it is often such a minor defect that it can be discounted, especially if it occurred a substantial length of time before the scan was undertaken. They were asked if they had had any surgical procedures on those same areas, be they medical or cosmetic. They were also requested to disclose any diseases or illnesses that they had experienced and the treatment received. As in the case above prompts were included such as arthritis, as this commonplace condition is often overlooked when asking about disease. It was hoped that the responses to these questions would provide information about anything that may have affected or altered the morphology of the face of a participant in any way other than genetic variation.

5.2.2.1.5 Teeth

The participants were asked to specify which third molars had erupted in the jaw. A brief instruction of how to identify the number of molars a subject has was provided so they were able to work this out for themselves.

> ‘These are the teeth at the very back of your mouth. If you have them, you would have eight teeth in each quarter of the mouth. If you have seven or less, then you may be missing them. The teeth at the back of your mouth are the molars, they look square and flat. You would have three in each quarter of the mouth if your wisdom teeth are there; if you don’t and there are no gaps in the row of teeth you are missing them’

To verify the ability of the participants to carry out this procedure, the instructions were initially given to a number of volunteers with no dental, medical, forensic or osteo-archaeological knowledge. They were asked to assess their own molars and this was then checked by the observer, trained in forensic anthropology. In all cases, ten in total, the volunteers had successfully identified which third molars they had in accord with the trained observer. This was taken as verification that accurate results could be obtained with the participants conducting the count themselves and no change was made to the instructions.

As well as the number of third molars, as these are the most likely teeth to be missing, subjects were asked if they had at any time had any other teeth removed, lost or been told of any other abnormality by their dental practitioner. Major dental work including orthodontic braces should have been recorded in this section, as would the presence of any supernumerary teeth, as missing teeth, extra teeth and braces applied at a young age can all dramatically alter the morphology of the mandible and maxilla. All subjects with major dental work were excluded from the study.

In addition to the instructions given above the researcher or a research assistant was also available to help participants if they had problems counting their teeth.

5.2.2.1.6 Diet

Lastly participants were asked to disclose any major dietary changes or variation from an average omnivorous western diet. These changes may have been the result of allergy, intolerance or personal choice. Vegetarianism was given as a prompt and details of the length of time that any of these changes had affected the individual were asked for.

5.2.2.2 Surface Scans

The method of collecting surface data chosen for this study is white light scanning. This was chosen for its speed and quality and for the transportability of the equipment. The equipment is set up in the manner stated in 5.2.1. The equipment is calibrated each time it is moved so the different lighting and background area can be taken into account. Calibration is conducted by placing a calibration board on a stable surface in front of the scanner in the area where the subject's head will be when scanning. The calibration board consists of a printed two-dimensional grid pattern made up of $15mm^2$ alternating light and dark grey squares, nine squares by 12 squares, backed with a solid metal board providing rigidity, see Figure 5.1. The projector is then turned on and set to project a pattern of small black and white squares which facilitate the setting of the projector's focus to the distance between the projector and the calibration board. This is performed manually and the visual accuracy of the focus is judged by the observer. For consistency this was undertaken by the same observer throughout the study. In the same manner each of the two cameras is focused at the same distance whilst their output is viewed on the monitor. Once all instruments are correctly focused, the aperture of each camera needs to be set so that the subject is neither too dark to be seen nor too bright to define features. The lighting in the room must be fixed at this point and not changed dramatically throughout the scanning session conducted with that calibration. The apertures are set while the projector is emitting solid white light, as this is the highest light level used during scanning and the apertures must be set to allow just enough light through at the level where it is possible to see all the squares on the calibration board with no area overexposed, as this would result in loss of detail. The focus and aperture settings are then locked so they cannot be altered accidently.

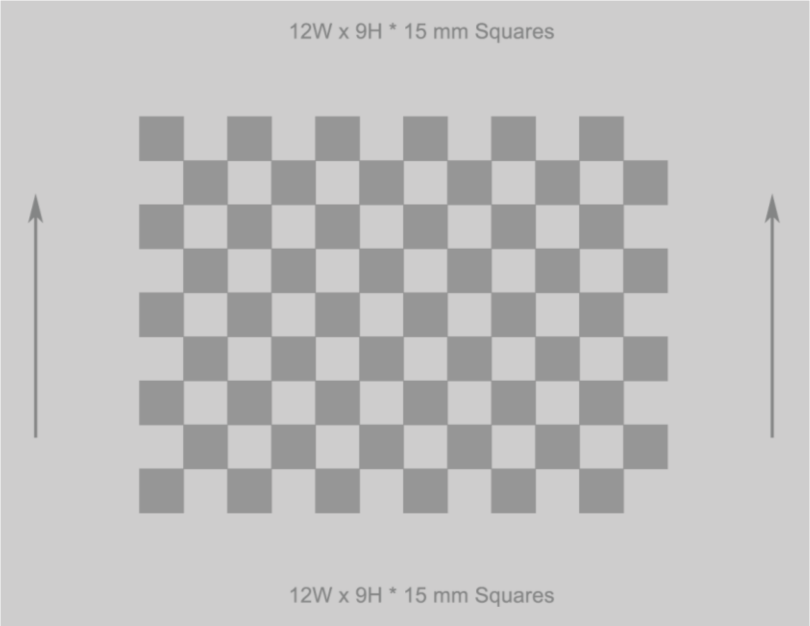


Figure 5.1 Calibration board

The entire grid pattern of the calibration board must appear within the viewfinders of both cameras and must be fully illuminated by the projector for a good image to be acquired. A series of images is then taken with the calibration board in differing positions. For the best result the board should be tilted around the x, y and z axis, moved to the limits of the horizontal and vertical space allowed by the cameras and brought to the fore and back of the focused area, an example of a series of images can be seen in Figure 5.2. A minimum of 30 good quality images were taken for each calibration. The software is designed to detect all the corner points of the grid pattern and will reject any image which does not detect all of the corner points. An image may be rejected due to the grid being out view of either camera, the angle of the board being too great to allow definition between the squares, being over or under exposed or being too out of focus to clearly pinpoint the corners. Once the images have been taken the software can calibrate all the images and scanning can commence.

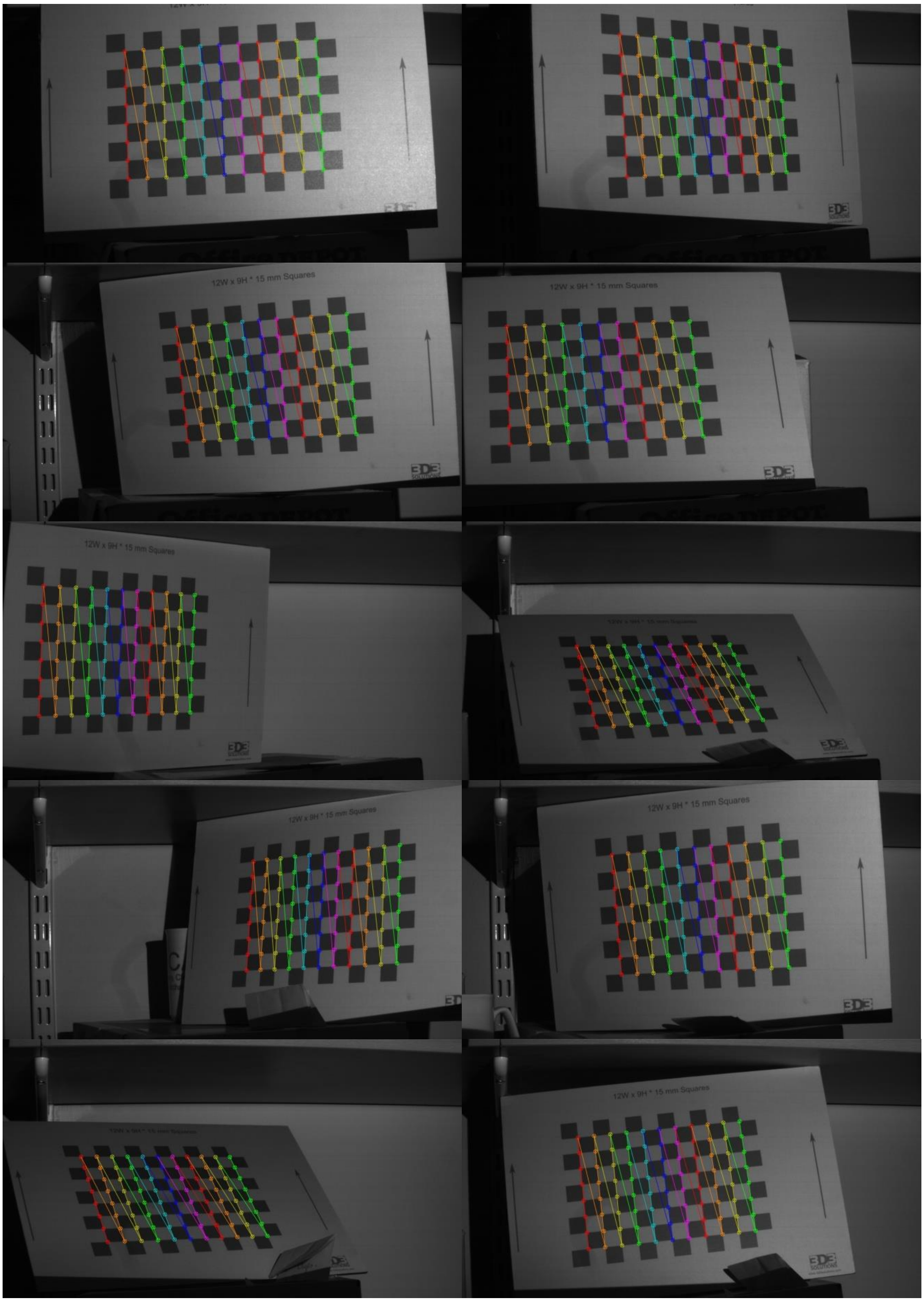


Figure 5.2 Series of calibration images, the coloured graphic shows the software finding the corners of the squares

Before being scanned, participants were asked to remove any jewellery worn on the face such as nose rings or piercings. Small stud earrings were allowed to be retained as they would not deform the ear in any way that could affect subsequent landmark placement. Glasses were removed but contact lenses were left in as the eyes were closed during scanning. Where necessary head hair was kept from the face with a wide fabric hair band, loose enough to not stretch and tighten the facial area. Excessive eye make-up was removed so no false shadows could be created and any false eye lashes had been removed. Glossy make-up had been removed, see above, as it could create over exposed areas due to intense light reflection, however a small amount of powdered make-up was allowed as it can improve the capture of the surface. High collared clothing was removed or turned down so the neck was visible and clear. Occasionally when scanning was conducted in a warm room, perspiration on the face had to be dried because, as with glossy make-up, it could cause overexposed areas due to high levels of light reflection.

The subject was then seated in front of the scanning equipment and the equipment was positioned with the cameras level with the eyes of the participant. The spirit levels in the tripod were used to confirm the equipment was level, so one camera was no higher than the other. To ensure uniformity of the facial morphology across participants and between single scans of individuals, the participants were requested to place the lower teeth in occlusion with the upper whilst not gritting them, so the muscles were not too contracted. Ideally the eyes would be left open allowing for easy identification of the exocanthion and endocanthion and enabling the position of palpebrale superius and inferius, the highest and lowest point at the midpoint of the margins of the eyelid.

The light of the projector was however too bright for people to look towards without squinting, and thus affecting the morphology of the face especially the position of palpebrale superius and inferius. It was therefore decided to have the participants close their eyes, removing the ability to place palpebrale superius and inferius, but meaning none of the other landmarks would be affected. As with the teeth, the eyes are closed but not clenched so the surrounding area is as relaxed as possible. The participant is then asked to sit up straight and face the scanner, relax and stay as still as possible whilst each scan is being taken.

The software can only produce surface data from areas within the visual field of both the cameras; as the face is curved multiple scans from differing angles are needed to get an accurate representation of the whole face. With the first few participants, the participant was left seated in place and the scanner was moved around them to gain the additional angles needed, as it was believed that this method would reduce the amount of variation due to participant movement between scans.

This method proved to be time consuming with additional problems resulting from the cameras being knocked out of place, or out of focus or the level of the tripod mounting being affected. The process was modified and instead of the equipment being moved the subjects rotated on the same spot. To enable the alignment of the individual scans to form a single surface scan of the whole face, there needs to be an adequate overlap between each scan. After testing, the optimum number of scans to collect data from the anterior portion of the ear on the left through to the anterior portion of the ear on the right, was ascertained to be seven. These seven scans were taken at 30˚ intervals (Lu et al.2006) as seen in Figure 5.3 where the arrows show the direction the participant is facing towards.

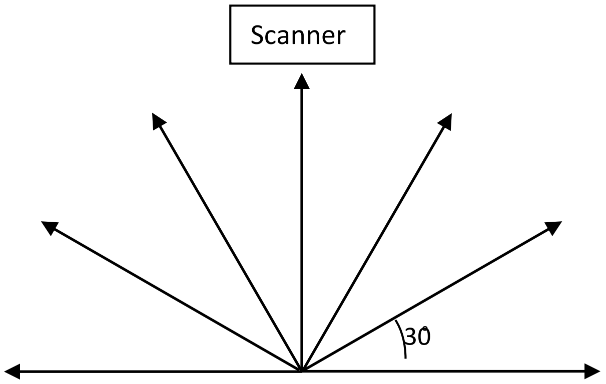


Figure 5.3 Scanning Angles, the arrows indicate the direction the subject is facing.

The angles were measured before initiating scanning and markers were placed on the floor indicating which direction participants were to face in for each consecutive scan.

The product of each surface scan is a three-dimensional point cloud or ‘mesh’. This mesh represents a collection of exact coordinates in three-dimensional space (x, y, z) appearing on a digital screen as a contoured surface.

Each scan takes under two seconds, so the possibility of the subject moving during those two seconds exists, but in most cases movement was so minimal that the scans were unaffected. Where movement was noted, either visually whilst observing the scan or when reviewing the scan when the subject was still in situ, another scan was taken at the appropriate angle. Each of the seven scans is quickly reviewed to check for missing data and repetitions were completed where necessary.

5.2.2.3 Aligning

After the data capture has been completed, each of the individual scans needs to be aligned with the others to create a single surface image of the individual in question. This is conducted in Flexscan 3.1 by locking one scan in place then manually manoeuvring a second scan in alignment with the first, as close as possible to the correct position. There must be some overlap of each scan with the next so it is best to work from one lateral angle to the other. Once moved manually the software can perform the alignment procedure. For each datum point in the locked surface mesh, the software calculates the distance to the closest point on the surface of the second mesh. This is done for all points and the mean total of the distances between these paired points then calculated. The second mesh is then automatically rotated by the software, in order to minimise the mean total which indicates the least possible variation between the two surfaces. The deviation is calculated and any variation below 0.5mm is classified as a good result. This variation was chosen as this is the range one would achieve when merging scans of a solid uniform object. A visual check should also be performed at the same time to ensure the software has performed the alignment correctly. The first two scans will then both become locked in the virtual space and the same procedure repeated with all the remaining scans of an individual. Once all the scans are initially aligned, fine alignment, which iterates the alignment procedure several hundred times for all the loaded scans as a group, was employed to correct any minor errors.

Outlying and background data can often be picked up by the scanner if the wall behind a person is in range or light bounces off an object in the room. Areas of clothing, often around the neck, can also be detected by the scanner. If this occurs data generated by clothing should be deleted at this stage in the process. If this data is retained it may cause confusion when trying to align scans as the software may be attempting to align the collar of a shirt for instance, in addition to the face itself, and movement between the face and collar may have occurred between subsequent scans creating a false alignment. Such data can easily be selected and deleted allowing for improved alignment.

While every attempt was made to maximise the quality of each scan, repeating scans where necessary, there remain some imperfections in the scans of some individuals. For example, where facial hair has occluded the view of the underlying skin.

5.2.2.3.1 Removal of neck data

The surface data forming the area of the neck of individuals is collected as part of the scanning process. However, as no clear landmarks can be defined on the neck these data are not required. As the scanning proceeds and the subject shifts between the directional angles for each scan to be taken, the position of the head in relation to the neck can vary to a much greater degree than movement within the face itself. A study was done to see if the removal of the neck area from the surface data before alignment took place would improve the alignment of the remainder of the face data, leading to a better representation of the actual face of the individual. The initial mesh-to-mesh distance (the average distance between the two meshes calculated using TrICP) was calculated with the neck data still in place for 32 individuals. The variation in meshes ranged from 0.112mm to 0.655mm with an average of 0.3271mm. This can be seen in Table 5.2 with the green highlighted distance showing the greatest variation and the blue showing the least. The neck data up to the level of the curve between the gnathion and ear lobe about 2cm below the lower jawline were then removed and the alignment and calculation repeated with the revised surface data. The results then ranged from 0.100mm to 0.289mm with an average of 0.1587 over the 32 specimens. The maximum improvement was 0.493mm which was a 75% improvement from the initial alignment with the neck data. Overall the alignment improvement ranged from 10.7% to 75.3% with an average improvement of 46.9%. As the alignment was only ever improved by conducting this procedure it was applied to the manipulation of all data from this area of the neck.

5.2.2.4 Merging of Partial Scans

Flexscan 3.1 is used to merge the, now aligned, scans together to form a single piece of surface data, or mesh, which can be exported to other analytical programmes. The software allows for parameters to be set when merging. The ‘precise’ merge is selected as opposed to the ‘smoothing’ merge as this retains the original data and does not alter it. The hole filling scale is set to ‘none’ which also assures that no additional data are added by the software. This will produce a single piece of surface data which can then be exported in .obj or .stl files for analysis performed by other programmes. The meshes produced are composed of an average of 600,000 vertices.

Scan number	Initial distance between meshes (mesh-to-mesh) (mm)	Mesh-to-mesh after neck removal (mm)	Improvement (mm)	Improvement percentage
1	0.234	0.116	0.118	50.4274
2	0.222	0.146	0.076	34.2342
3	0.276	0.207	0.069	25.0000
4	0.266	0.151	0.115	43.2331
5	0.21	0.184	0.026	12.3810
6	0.392	0.17	0.222	56.6327
7	0.655	0.162	0.493	75.2672
8	0.338	0.183	0.155	45.8580
9	0.196	0.17	0.026	13.2653
10	0.501	0.289	0.212	42.3154
11	0.388	0.193	0.195	50.2577
12	0.342	0.174	0.168	49.1228
13	0.313	0.144	0.169	53.9936
14	0.112	0.1	0.012	10.7143
15	0.225	0.11	0.115	51.1111
16	0.206	0.151	0.055	26.6990
17	0.309	0.161	0.148	47.8964
18	0.224	0.151	0.073	32.5893
19	0.36	0.147	0.213	59.1667
20	0.45	0.124	0.326	72.4444
21	0.472	0.164	0.308	65.2542
22	0.309	0.146	0.163	52.7508
23	0.181	0.113	0.068	37.5691
24	0.226	0.119	0.107	47.3451
25	0.304	0.154	0.15	49.3421
26	0.473	0.147	0.326	68.9218
27	0.235	0.141	0.094	40.0000
28	0.363	0.12	0.243	66.9421
29	0.412	0.205	0.207	50.2427
30	0.242	0.121	0.121	50.0000
31	0.516	0.151	0.365	70.7364
32	0.516	0.263	0.253	49.0310
Average	0.3271	0.1587	0.1685	46.8983

Table 5.2 Improvement due to removal of neck data.

5.2.3 Landmarks

Thirty-two landmarks are placed on the surface of each facial scan, these fall into two categories, midline points which are a series of single points down the centre of the face and symmetric points, where there are matching points on the left and right side of the face. There are 10 midline points, from top to bottom: Glabella; Sellion; Pronasale; Subnasale; Labial superius; Stomion; Labial inferius; Sublabiale; Pogonion; Gnathion. There are 22 symmetrical points, 11 on the left and 11 on the right, from top to bottom: Superciliare; Endocanthion; Exocanthion; Otobasion superius; Tragion; Alare; Subalare; Otobasion inferius; Crista philtre; Chelion; Protrusion of mental tubercle. The landmarks can be seen in Figure 5.4 and Figure 5.5 and are defined in Table 5.3. These definitions are taken from Demayo et al. (2010), Abrahams et al. (1998) and Bass (1971) and are sometimes adapted to fit a three-dimensional structure.

A template showing the names and order of the landmarks is created. This names each of the landmarks to be placed, specifies the order in which they are placed and creates a guide to the general placement of the landmarks. The template enables the resulting coordinate data to be easily combined and exported for multiple subjects, as the data will all be in the same format. The template can be used to manually place each landmark within it, the template itself facilitating the naming of the next point to be placed. Conversely the placing of landmarks can be automated with the software projecting the template of landmarks onto the loaded surface data. These points will then need to be manually adjusted to place them accurately on the loaded individual scan. Both methods give the same end result and it is merely a matter of the observer’s preference as to which one to use.

Both methods were used but preference was given to the manual placement of landmarks in the majority of cases.

5.2.3.1 Positions

The landmarks are placed using Viewbox. Having loaded a single facial scan, the scan is manually manoeuvred into the Frankfort plane, where the right and left portion, the top of the external auditory meatus, and the bottom of the left orbit are all aligned in the Frankfort horizontal (White & Folkens 2005). Once in this plane, both left and right norma lateralis views, viewed from the side perpendicular to the Frankfort horizontal, and the norma frontalis view, viewed from the front parallel to the Frankfort horizontal, are saved as a ‘View’. A View is a saved view of the facial scan where the position in the three-dimensional space, the angle of sight, the zoom and the light setting are all saved. This enables the observer to place landmarks both faster and more accurately.

5.2.3.1.1 Norma Lateralis

The landmarks placed whilst in the norma lateralis view are the landmarks that are the most anterior or posterior points on the midsaggital plane. This view gives a profile of the face allowing for easy placement of these landmarks: Glabella; Sellion; Pronasale; Sublabiale; Pogonion. The Tragion is first placed in this view (left and right specific) but the accuracy of this placement is checked when in the norma frontalis as in this plane it is the most lateral as well as the most posterior point of the tragus.

5.2.3.1.2 Norma Frontalis

The landmarks placed whilst in this view are superior, inferior, lateral or medial to features of the face or the midpoints of a feature so a frontal view is helpful for placement. These landmarks are relatively easy to place as there are strong features around them indicating their position.

Superior:	Superciliare, Crista philtri
Inferior:	Gnathion
Lateral:	Exocanthion, Alare, Chelion
Medial:	Endocanthion
Midpoints:	Labial superius, Labial inferius

The Stomion is placed whilst in this view but requires verification through manipulation of the scan to ensure that in is in the correct position in the sagittal plane as this is difficult to observe from norma frontalis.

No	Landmark	Code	Description
1	Glabella	Gl	The most prominent point in the midsagittal plane between the superciliary arches.
2	Sellion	Se	Most posterior point of the nasofrontal angle on the midsagittal plane, when aligned in the Frankfort plane.
3	Pronasale	Pn	Most protruded anterior point of the nasal tip, when aligned in the Frankfort plane.
4	Subnasal	Sn	Point in the midsagittal plane where the lower border of the nasal septum merges into the upper lip.
5	Labial superius	Ls	The midpoint of the vermilion border of the upper lip.
6	Stomion	St	Point in which the midsagittal plane crosses the horizontal labial fissure between closed lips.
7	Labial inferius	Li	The midpoint of the vermilion border of the lower lip.
8	Sublabiale	Sl	Most posterior point between the Pogonion and the Labial inferius on the midsagital plane, when aligned in the Frankfort plane.
9	Pogonion	Po	Most anterior point of the chin, located on the skin surface in the front of the identical bony landmark of the mandible, when aligned in the Frankfort plane.
10	Gnathion	Gn	Lowest, most inferior point in the midline on the lower border of the chin, when aligned in the Frankfort plane.
11,12	Superciliare	Sc	The most superior point of the upper margin of the midline portion of the eyebrow.
13,14	Exocanthion	Ex	Most lateral point at the outer commissure of the eye fissure.
15,16	Endocanthion	En	Most medial point at the inner commissure of the eye fissure, should be placed at the medial end of the lacrimal caruncle.
17,18	Alare	Al	Most lateral point of the nasal alar.
19,20	Subalare	La	The point at the lower limit of each alar base, where the alar base disappears into the skin of the upper lip.
21,22	Crista philtri	Cp	Most superior point of the upper lip.
23,24	Chelion	Ch	Most lateral point at the outer commissure of the labial fissure.
25,26	Protrusion of mental tubercle	Mt	Protrusion of mental tubercle.
27,28	Otobasion superius	Pa	The point of attachment of the ear helix to the facial surface.
29,30	Tragion	Tr	Most lateral and posterior point of the tragus of the ear.
31,32	Otobasion inferius	Sa	The point of attachment of the ear lobe to the cheek.

Table 5.3 Landmarks

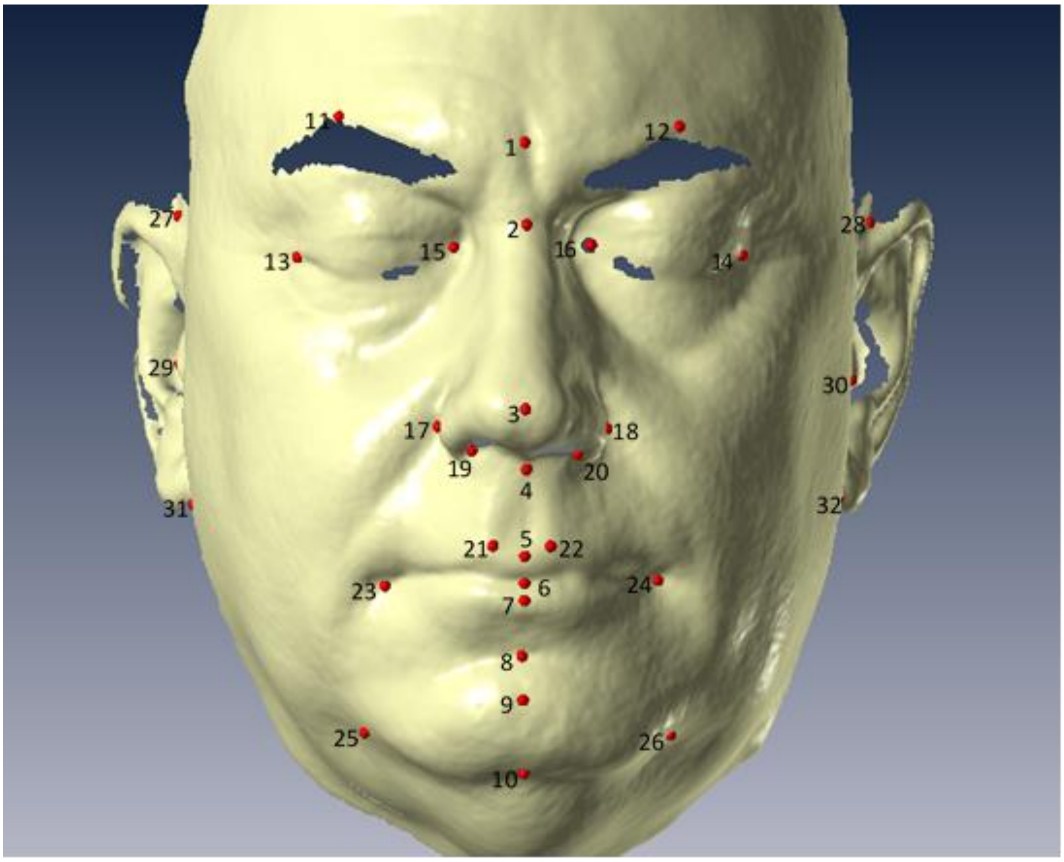


Figure 5.4 Landmarks from a frontal view

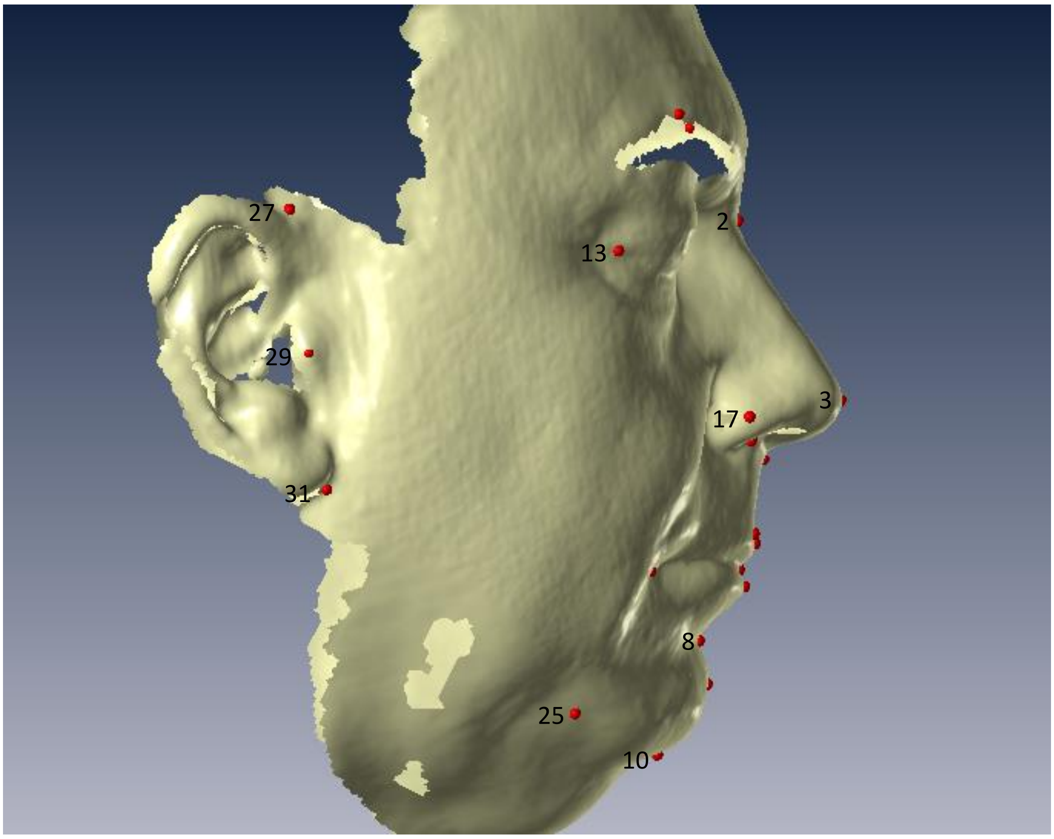


Figure 5.5 Landmarks from a lateral view

The Protrusion of the mental tubercle is perhaps the most difficult of all the landmarks to place. It is something that is much more easily observed on the mandible when in its skeletal form and the addition of soft tissue obscures the feature. Its correct position is best visualised in the norma frontalis view as in other views this landmark may appear to be more posterior.

5.2.3.1.3 Other Views

The Subalare, Otobasion superius and Otobasion inferius are all defined by a change in the feature or the meeting of two features; this allows them to be placed whilst in any view that give a good sight of the feature in question, normally a high level of magnification aids their placement. As with the Tragion and Stomion it is often helpful with these landmarks to place them and then rotate the scan to check the accuracy of their placement.

5.2.3.2 Lighting

Lighting can be used to improve the placement of landmarks. The computerised light source can be moved around within the virtual space to represent differing directions of a light source. The intensity and proximity of the virtual light source can also be altered. All of these allow the observer to gain the best shadow and light variation on the feature they are looking at and enabling enhancement of the landmark they are attempting to locate.

5.2.3.3 Prominent and Depressed Points

Viewbox allows for multiple positions on surface data to be plotted for any one landmark and can then calculate the most prominent point from the average of these points or most depressed point when between two or more promontories. This aims to reduce the error in landmark placement and is more accurate than simple manual placement. It is in placing landmarks where the definitions refer to the ‘most anterior’ or ‘most posterior’ that this technique is most useful. This technique is used for six landmarks in total.

Sellion and Sublabial are both ‘most posterior’ points where the depression of the curvature is only wanted in the midsagittal plane as the surface at this point is saddle shaped and so falls away posteriorly as you move laterally away from the midsagittal plane. To allow for this, two points are placed as close to the presumed position of the landmark as possible, both in the midline, and the most depressed setting is used.

Subnasale is a typical depression point where the surface rises on all sides of it, so a simple, most depressed point is used. This is achieved by placing at least four points equally spaced around the suspected position of the landmark and using the most depressed setting.

Pogonion and Pronasale are both the opposite of Subnasal, they are both typical most prominent points and the placing of the landmark is conducted in a similar manner to Subnasal. At least four points are placed equally around the suspected position of the landmark and the most prominent setting used.

Tragion differs slightly from the above but only in the number of points placed. Three points are placed as it is difficult to place any more than this equally on the pyramidal shape of the tragus.

5.2.3.4 Missing Landmarks

Sometimes due to the subject or the scanning process sections of the face are missing from the surface data, as discussed in 5.2.2.3. This can make it difficult or impossible to place landmarks as the relevant surface data the landmark is based on is missing. Where the missing data are limited and the void is small it is sometimes possible to deduce the position of the landmark with relative accuracy based on the surrounding data available. For example, the underlying surface data at the Stomion seems to be lost occasionally as it is quite recessed and the cameras find it hard to capture, however often the remainder of the surface of the lips is available and the fissure line up to the point of the Stomion is visible. In cases like this it is possible to extrapolate the fissure line to get the position in the transverse plane, taking the midpoint of the labial area based on the position of the Cheilion and Crista philtri to get the position in the medial-lateral plane and then following the curve of the upper and lower lips to deduce the point at which they meet to find the position in the sagittal. This positioning of the landmark is conducted by initially placing the landmark on the closest point of surface datum and then manually editing the coordinates whilst viewing the landmark's position in a number of different planes.

Where there are a substantial amount of missing data the observer cannot deduce the position of a landmark by inferring from local data, in such instances the landmark cannot be placed as the accuracy will be too low to aid the analysis and will in fact deplete the value of the final result. When a landmark is not placed the exported coordinate will be exported as a blank cell in .xls file and just not entered in the .txt file. Care must be taken to make sure that the relevant point is missed and that the following points in the template are correctly placed. Once exported, these blank cells are replaced with the figure 9999. This figure is widely recognised by Morpho J and other analytical programmes as representing a missing point and are thus removed from any further analysis.

5.2.4 Morpho J Regression Analysis:

Once the landmarks had been placed they were input into Morpho J, where Procrustes superimposition was first applied to scale, translate and rotate all of the landmarks of each individual to allow for comparison. Once Procrustes superimposition had been applied the data were checked for outliers. Each of the outliers reported by the software was manually checked to ascertain if a landmark had been misplaced. Any mistakes were rectified and the process repeated until the outliers were reduced and the result of the individual's facial shape rather than landmark misplacement.

The data were split in two ways to allow for more detailed analysis of the results. First the population was split into eight smaller groups as well as the population as a whole. The groups allowed for variations between males and females and between the generations to be analysed separately. The groups are:

- All Individuals
- Adults
- Mothers
- Fathers

- Children
- Sons
- Daughters
- Females
- Males

Secondly the landmarks themselves were split into groups based around specific features of the face to analyse whether the different portions have responded differently to different genetic and environmental factors. The five areas of the face are:

- Whole faces (all 32 landmarks)
- Eyes (Exocanthion-left and right, Endocanthion-left and right, Sellion)
- Nose (Alare-left and right, Subalare-left and right, Sellion, Subnasale, Pronasale)
- Mouth (Labial superius, Labial inferius, Chelion-left and right, Crista philtri-left and right, Stomion)
- Profile (Glabella, Sellion, Pronasale, Subnasale, Labial superius, Stomion, Labial inferius, Sublabiale, Pogonion, Gnathion)

Each combination of people group and landmark group was separately analysed using Principal Component Analysis and Regression after the covariates data of BMI, Age and Sex were input for every individual.

5.2.5 Areas of the Face

As each of the main features of the face, eyes, nose, mouth and chin grow and change in different ways, depending on age and sex as well as other factors, it is expected that some may out-perform others when looking for parallels deriving from genetic similarity. In order to assess different areas and combinations of areas of the face for their heritability the face was divided into 14 different areas.

These are: The Whole face; Removed forehead; Removed ears; Removed inner eye; Eyes; Nose; Mouth; Chin; Eyes-Nose; Eyes-Nose-Mouth; Eyes-Nose-Mouth-Chin;

Nose-Mouth; Nose-Mouth-Chin; Mouth-Chin. Each discrete area is described below, these areas were devised for this study by the author and checked for intra observer error by repeating the procedure on the first 15 individuals for every area and comparing them using the fine alignment tool. The variation between sections never exceeded 0.001mm.

Whole face: the raw scan of the entire face unedited with the exception of removal of the neck, background and outlying data as mentioned previously to allow for more accurate merging of the scans.

Removed forehead: this is the Whole face as noted above with the forehead area above the bulge of the eyebrows removed. The forehead is removed above a line which runs between the hair line on each side of the face parallel to the ridge of the Glabella but approximately 2.5cm above it. The thinking behind removing this section of the forehead is to remove the variability of the hair line, whilst retaining the Glabella area.

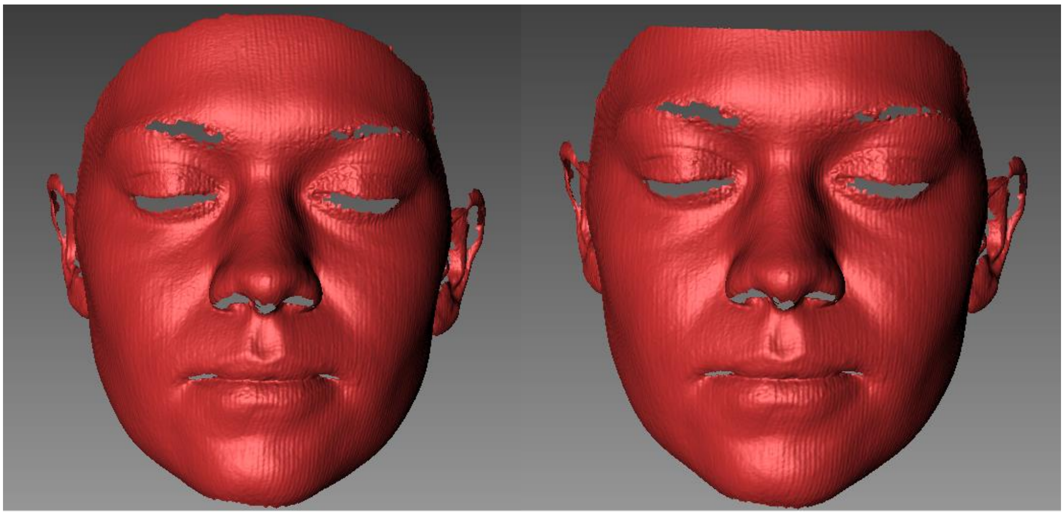


Figure 5.6 Removed forehead. The left shows the natural superior edge of the scan showing the hair line of the individual, the right shows the cropped mesh with the majority of the forehead removed.

Removed Ears: This section begins with the Removed forehead mesh as mentioned above, the ears are then removed along the line formed by the change in angle between the cheek and the lobe and helix of the ear. The Tragus is retained in the mesh. The idea behind removing the ear is to remove the possible confusion and variation between scans caused by poor scanning as, unlike the rest of the face, accurate scanning of the entire ear would require numerous scans from a huge number of angles due to its complex shape and overlapping structure.

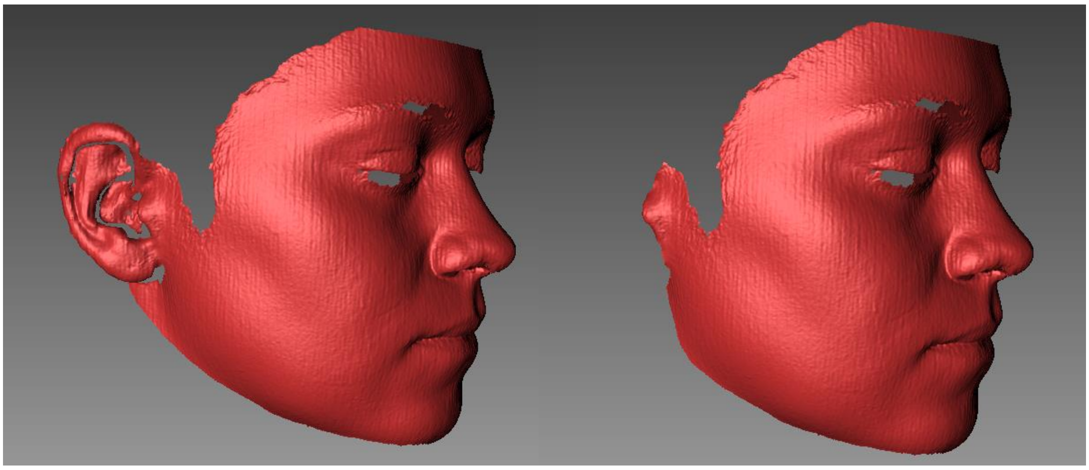


Figure 5.7 Removed ears. The left picture shows the ear in position whereas the right picture shows the lateral edge of the face after the removal of the ear.

Removed inner eyes: The Removed ears mesh is taken and the inner portion of the eye area removed. The eye lids are retained as much as possible but the area where the eye lashes have caused the scan to blur and fragment is removed. As with the ears the thinking behind this is that this imperfection in the scanning process may cause variations between the scans of individuals which does not accurately represent the three-dimensional surface scans of those subjects' faces.

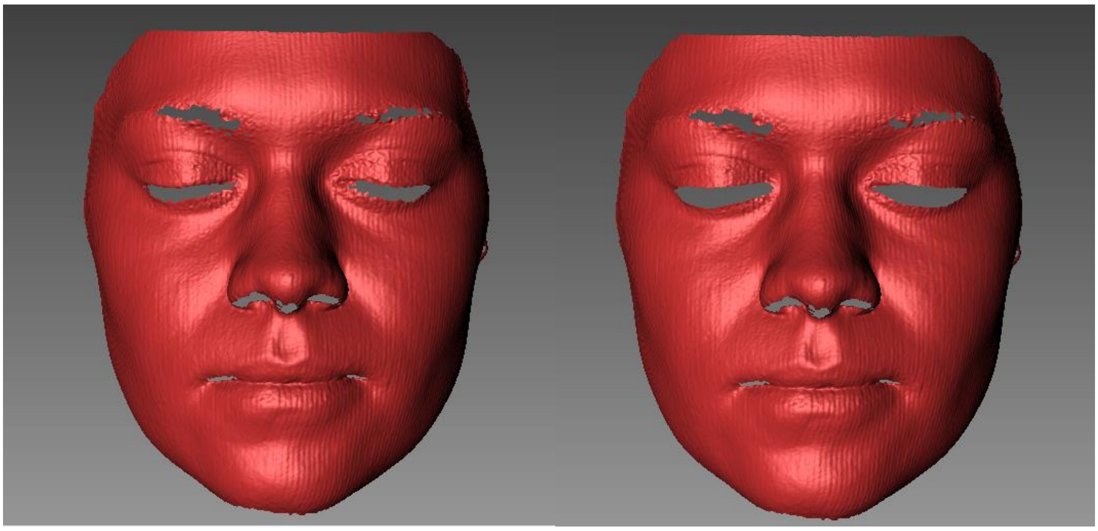


Figure 5.8 Removed inner eyes. The left picture shows the eyes as scanned originally where the right picture shows the eyes after the trimming of the eye lash area.

Eyes: The eyes are defined by the area encompassing both eyes and the connection between them formed by the bridge of the nose. They cover a very similar area to that used by Hammond (2012). The border of the area starts at the Nasion, follows a straight line to the left Endocanthion then curves, tracking the crease of the lower left eye lid until level (horozontally) with the exocanthion. The upper section is delineated by a straight line from this point to the lateral edge of the eyebrow, along the arc of the superior edge of the eyebrow until its most medial point. The same border is followed for the right eye and the two medial borders of the eyebrows are joined by a straight line to complete the edge of the eye area. (The deficient area caused by the eyelashes as mentoned above is removed in this mesh as well).

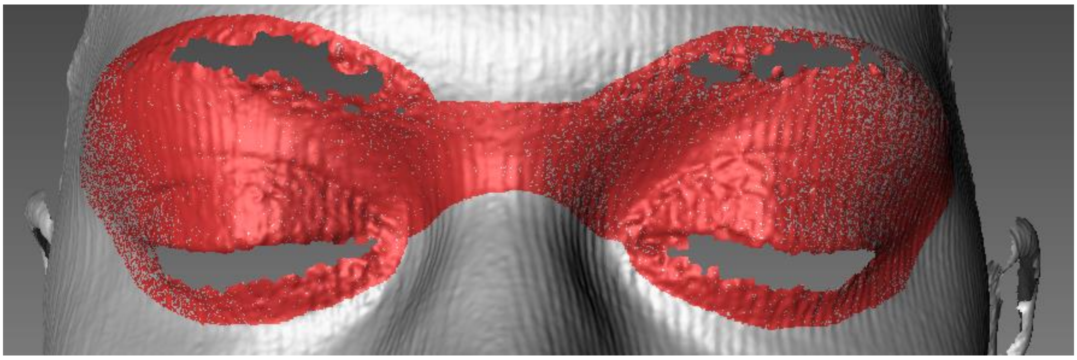


Figure 5.9 The Eyes, the red area denotes the section referred to as the eyes.

Nose: The area of the nose is similar to that used in Chang et al. (2006) and Hammond (2012) and almost identical to that used by Ferrario et al. (2007) in their surface-based analysis, the only variation is due to the difference in data collection methodology. Where Ferrario et al (2007) are able to define the lateral contour of the nose by taking stone casts directly from the face, this study defines the lateral contour of the nose by using the line between the Endocanthion and the Alar crest. As can be seen clearly in Figure 5.10 the Nose area is defined by a border from: the Nasion to the Endocanthion, the Endocanthion to the Alar crest, the Alar crest to the Subalare following the curve of the Alar, the Subalare to the Subnasale. This is then mirrored on the other side.

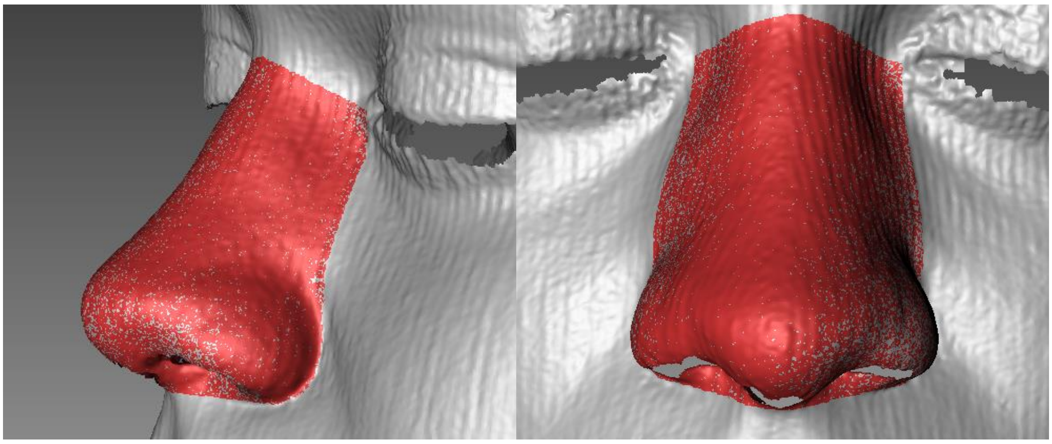


Figure 5.10 The Nose, the red area denotes the section referred to as the nose.

Mouth: The Mouth area in this study differs from that used by most other authors in studies of this facial feature. In part this is due to the lack of, and variability between, similar 3D studies but also due to the results obtained from preliminary studies conducted on a few families in the early stages of this study. The best results occurred from the larger area encompassing the entire upper and lower lips as described below, rather than the smaller areas used by Sforza et al. (2010b), Ferrario et al. (2000) and DeMenezes et al. (2011). The mouth section is defined by a line from the Subnasale to the Subalar, the Subalar to the Alar crest following the curve of the Alare, the Alar crest via the nasolabial groove until it fades (normally at a point lateral to the Chelion) and then in a straight line to the lateral appearance of the labiomental groove, finally from this lateral appearance to Sublabiale, following the labiomental groove. This is then mirrored on the other side as shown in Figure 5.11.

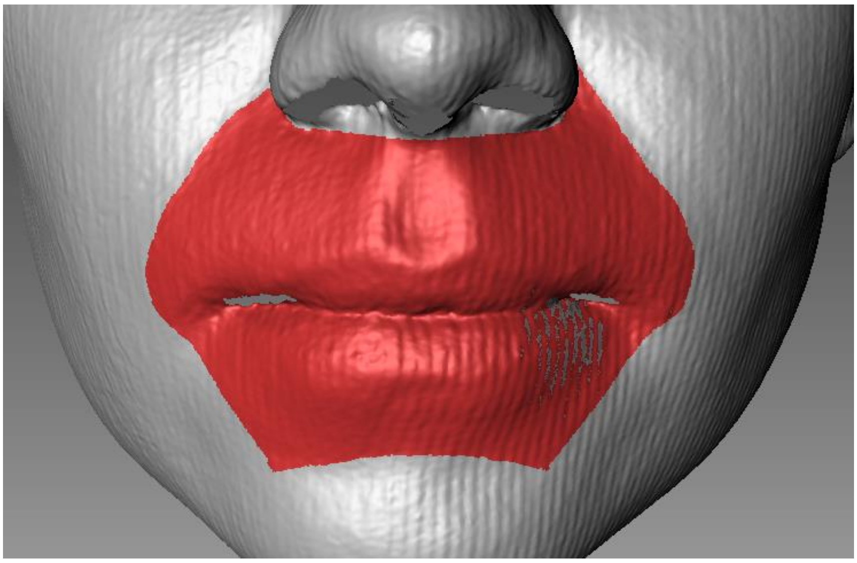


Figure 5.11 The Mouth, the red area denotes the section referred to as the mouth.

Chin: The chin appears so little in literature of this kind and has so few clear landmarks that a simple method was followed to create the best repeatability of delineating this area. The uppermost border is formed by the labiomental groove and where this fades the line is continued down on both sides following the coronal plane as seen in Figure 5.12.

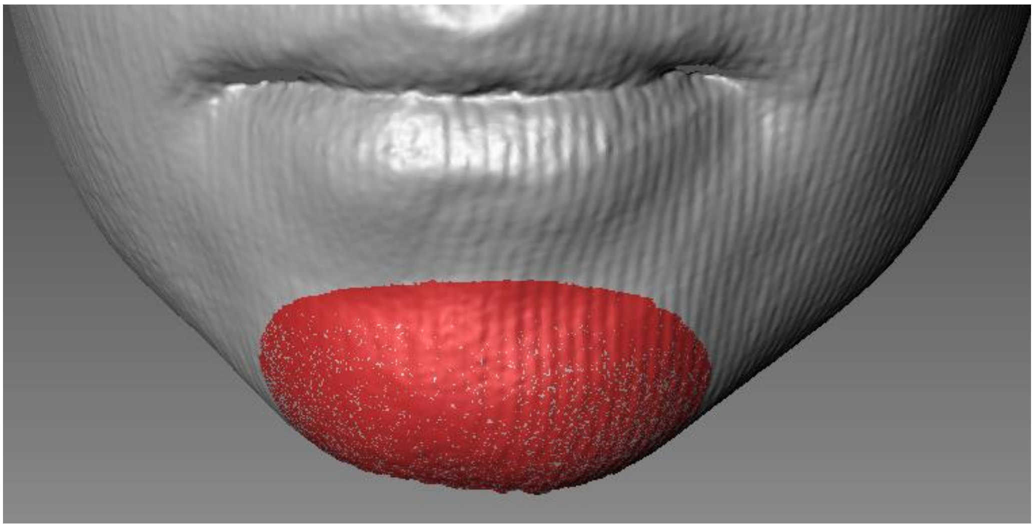


Figure 5.12 The Chin, the red area denotes the section referred to as the chin.

Mouth-Chin

Nose-Mouth

Nose-Mouth-Chin

Eyes-Nose

Eyes-Nose-Mouth

Eyes-Nose-Mouth-Chin

The areas on the left are formed by a simple combination of the individually described areas. They can be seen in Figure 5.13.

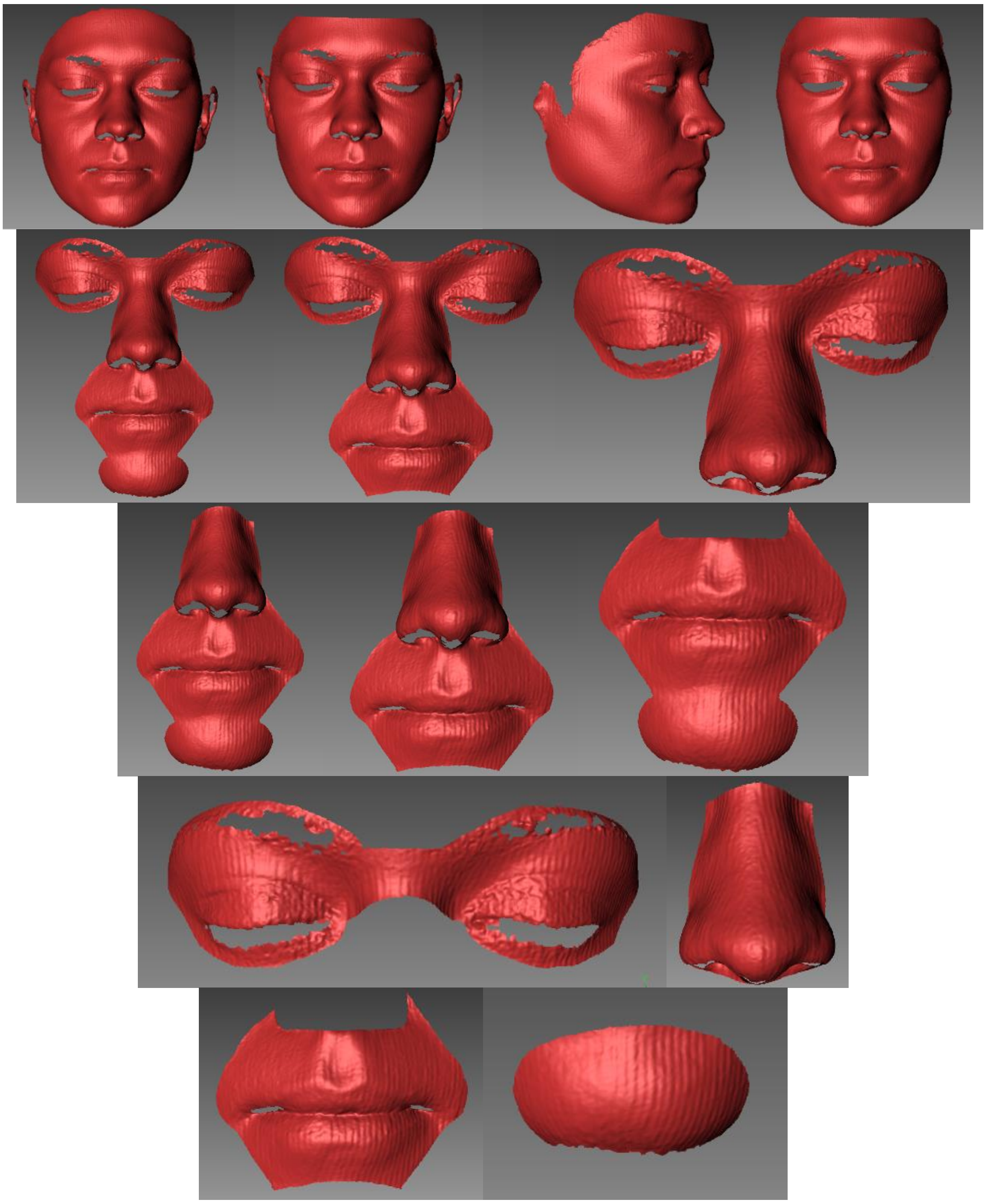


Figure 5.13 All areas of the face. All 14 variants of the face are shown from top to bottom left to right they are Whole face, Removed forehead, Removed ears, Removed inner eyes, Eyes-Nose-Mouth-Chin, Eyes-Nose-Mouth, Eyes-Nose, Nose-Mouth-chin, Nose-Mouth, Mouth-Chin, Eyes, Nose, Mouth, Chin.

5.2.6 Mesh-to-mesh Value Analysis

Mesh-to-Mesh analysis is carried out in Viewbox 4.1.0.0 using the ‘Mesh similarity’ function. This applies a Trimmed Iterative Closest Point algorithm and outputs a single distance value. The distance value indicates the similarity between the two meshes or surfaces with a smaller value specifying greater similarity. Due to the size of the whole face meshes a decimation function is used to reduce the number of vertices in each mesh and decrease the amount of computational time required. The whole faces are reduced from around 600,000 vertices to 300,000, the sizes of the other areas are shown below in Table 5.4

Facial area	**Approximate Size of mesh (vertices)**
Whole face	300,000
Removed forehead	550,000
Removed ears	505,000
Removed inner eyes	502,000
Eyes	50,000
Nose	40,000
Mouth	40,000
Chin	20,000
Eyes-nose-mouth-chin	150,000
Eyes-nose-mouth	130,000
Eyes-nose	90,000
Nose-mouth-chin	100,000
Nose-mouth	80,000
Mouth-chin	60,000

Table 5.4 Approximate size of facial meshes.

Viewbox has a number of settings which can be set and altered. The following settings are selected for the analysis of this study. The first rough alignment is conducted using point to point matching with a sample of 1% of the points available. Twenty different initial positions are assayed when conducting rough alignment of the two meshes. Once roughly aligned, more detailed final alignment is conducted, this time using point to plane matching for 100% of the points with exact accuracy. For both the rough and final alignment up to 100 iterations can be performed. Either single meshes can be selected for alignment or all the meshes within a folder can be compared to each other. The Mesh-to-mesh values (MMVs) of each comparison are output into a txt. format.

5.2.7 Receiver Operating Characteristics (ROC) Analysis

ROC analysis is a discriminative technique which uses ROC curves, described below, to calculate Sensitivity and Specificity at multiple classification thresholds. The efficiency of the technique can be assessed using both the shape and the Area Under the Curve (AUC). ROC analysis is mainly used as a diagnostic tool to select optimal models of classification.

ROC analysis was first developed in the 1950s for the analysis of RADAR signal detection (Van Erkel & Pattynama 1998, Greiner et al. 2000). Since then it has been developed and the areas of use have vastly increased. ROC is now used in the analysis of Radiology & Medical imaging (Swets 1979), sequence matching of proteins in chemistry (Gribskov & Robinson 1996), BMI calculations (Rankinen et al. 1999), psychiatric screening tests (De Jesus Mari & Williams 1985), psychology (Swets 2014) and even in linking burglaries to an offender's modus operandi (Bennell & Canter 2002) as well as many more. The development of the technique to include cut off points or thresholds and the comparison of multiple ROC curves expanded the range of subject matter to include immunology (Greiner et al. 1995) and veterinary medicine (Greiner et al. 2000).

ROC analysis is especially useful for datasets with skewed class distribution (Fawcett 2006) where one may have very few positive results, for example when studying a disease where the prevalence is normally low. This study has a typically skewed class distribution with over 7,000 possible mesh-to-mesh comparisons and just over 150 of those being positive matches. Due to this ROC analysis is the most suitable analysis to carry out when looking at the single mesh-to-mesh values.

When performing ROC analysis care needs to be taken to avoid selection bias (Van Erkel & Pattynama 1998). Selection bias is a particular problem when being used for medical decision making as the sample pool might be affected by previous test or screening processes. In this study bias is avoided as the families are chosen at random and not because they do or do not have visual facial similarity.

ROC analysis will be employed in the evaluation of a single variable (mesh-to-mesh value) to evaluate whether facial morphology as depicted in a three-dimensional mesh, rather than landmark data, can reveal genetic links between individuals. The hypothesis being tested is whether two individuals known to be genetically related, show a true (positive) match or show no match at all (negative).

5.2.7.1 Confusion Matrix:

Two classifications are used within each of two classes, as shown in Figure 5.14 below, in order to understand the analysis being carried out. The two classifications used here are Positive and Negative, in this case ‘positive’ refers to a true genetic match which could be Parent-Child or Sibling-Sibling and ‘negative’ accounts for all other possibilities where there is no genetic link. Parent-Parent is never compared as there should never be any positive results. Both these classifications can be assigned in the Actual Class, the known data collected in the study, often referred to as the ‘diagnosis’ in medical applications of this method, Predicted Class, i.e. the results predicted by this analysis.

		Actual Class	
		Positive	Negative
Predicted Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Figure 5.14 Confusion matrix adapted from Fawcett 2006

As depicted in the confusion matrix in Figure 5.14 these classifications give rise to four possible outcomes. If both the actual and the predicted are positive, the test is categorised as true positive (TP) while when the actual is positive and the predicted is negative it is considered a false positive (FP). Similarly, an actual negative with a negative prediction is a true negative (TN) and an actual negative with a positive prediction is a false negative (FN) (Fawcett 2006).

The TP and TN show results that are correct while FP and FN show result where the ROC analysis was unable to predict correctly.

Below in Figure 5.15 are all the equations that are used for ROC analysis and that will be employed in the analysis of this study, each is discussed in the following sections.

Equations:

True positive rate = TP / Total Positives = TP/(TP + FN) = Sensitivity

True negative rate = TN / Total Negatives = TN/(TN + FP) = Specificity

False positive rate = 1- Specificity = FP/(FP + TN)

Positive Predictive Value = TP/(TP+FP) = Precision

Negative Predictive Value = TN/(TN+FN)

Figure 5.15 ROC equations.

5.2.7.2 Sensitivity and Specificity:

There are a number of ways in which the results of ROC analysis can be presented, the simplest of these being the True Positive Rate (TP rate) and the True Negative Rate (TN rate).

Not all analysis programs use TP rate and TN rate as terms. TP rate can be referred to as ‘true positive probability’; ‘true positive fraction’; ‘recall’; ‘hit rate’ or ‘sensitivity’, whilst TN rate is often called ‘specificity’ but is sometime referred to as ‘true negative fraction’. From this point onwards the terms sensitivity and specificity will be used. Sensitivity will express the proportion of actual cases for which the positives are correctly predicted by the test, this can be expressed as a decimal between zero and one or as a percentage from 0% -100%. In an ideal scenario the sensitivity would be 1.0 or 100% meaning that all of the actual positive cases have been identified by the methodology and classified as predicted positive. However, for a methodology to be robust it also needs to predict the actual negatives as negatives and for this the specificity needs to be known.

The specificity will express the proportion of cases for which the actual negatives are correctly predicted as negative by the test. It is expressed in the same way as for sensitivity and the ideal scenario remains the same with 1.0 or 100% demonstrating that all of the actual negative cases were correctly predicted to be negative. In respect to this study the best possible outcome would be 100% sensitivity and 100 % specificity. This would imply that all the actual positives, the cases where there is a genetic link between the two meshes being compared, are correctly identified by the method and all the other cases, where there is no genetic link between the two meshes, are identified as negatives. Most importantly this situation would mean that there are no false positives, no non-genetic linked mesh comparisons being wrongly identified as positive. Additionally, there would be no false negatives, where a mesh comparison of two family members has been missed and incorrectly classified as a negative.

5.2.7.3 ROC Graphs:

Sensitivity and specificity are used to produce ROC graphs, a simple way of visualising the effectiveness of one's hypothesis. Sensitivity is plotted against 1- specificity (100-specificity when using percentages) with sensitivity taking the y axis and 1- specificity the x axis. 1- specificity can be referred to as the False Positive rate and so in some experiments the True Positive rate is plotted against the False Positive rate giving the same result as sensitivity versus 1- specificity.

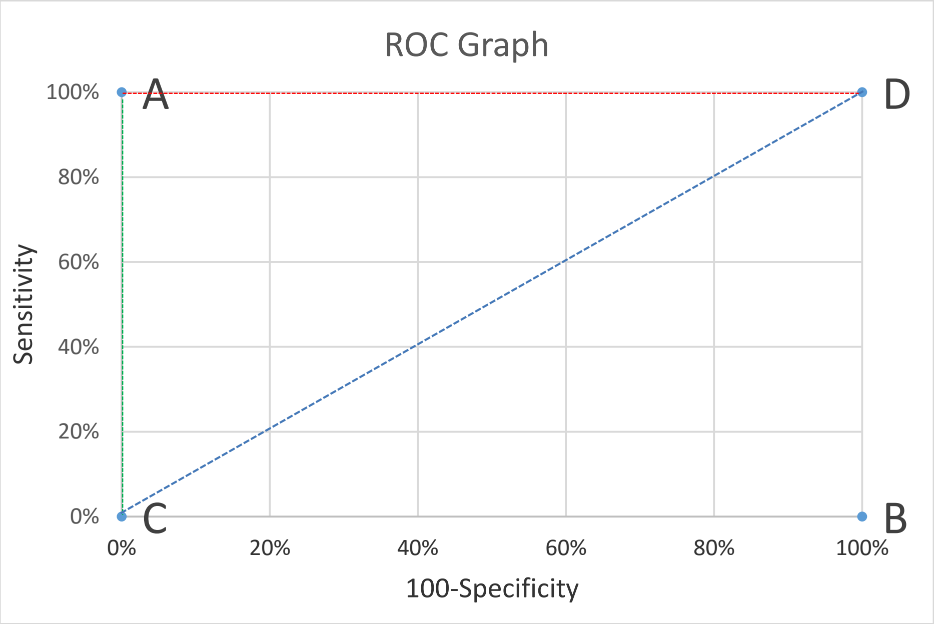


Figure 5.16 Typical ROC graph

A typical ROC Graph is shown in Figure 5.16, in some circumstances the axis may be numeric and if so will range from zero to one, rather than from 0% -100% as shown here. Four points shown on the graph as A, B, C and D are plotted here as examples of the most extreme results possible. Example A shows Sensitivity of

100% and Specificity of 100%, and illustrates the best case scenario, with all the positives identified correctly, all negatives identified correctly and with no False Positives or False Negatives identified. B shows Sensitivity of 0% and Specificity of 0%, in this scenario none of the actual results both positive and negative are predicted correctly, all of them being predicted incorrectly, B however does have a high predictive power as discussed below. C demonstrates Sensitivity of 0% and Specificity of 100%, this illustrates a situation where none of the positive cases have been predicted to be positive but all of the negative cases have been correctly identified as negative. What this actually shows is that all cases have been classified as negative whether right or wrong and there are no positive predictions. In direct opposition to C is D where Sensitivity is 100% and Specificity is 0%. Here all the positive cases have been classified correctly as positive, however all of the negative cases have also been identified, wrongly, as positive, in other words all cases are classified as positive and there are no negative predictions. Both A and D, and the red line between them, can be branded as ‘rule in’ situations where the main purpose is to make sure all the positive cases have been ruled in (Sensitivity 100%) whatever the effect on the specificity. The closer to A, the better the ability to correctly identify the negative cases. Points C and A, and anywhere on the green line, are examples of ‘rule out’ circumstances where the main objective is to make sure no negative cases are falsely classified as positive (Specificity 100%) however this may affect the sensitivity, again the closer to A the better the sensitivity. Points C and D are both on the x = y line, this line shows the position of any random guess, and it is not material where on this line a result falls. For example, if one managed to guess correctly 80% of the positives it would also follow that the False Positive rate would be 80% (100 – Specificity of 20%). Likewise, if one correctly guessed any given percentage of positives the same percentage of False Positives would occur. The further away from the line the results fall, towards A or B, the better the predictive value of classifier (Fawcett 2006)

5.2.7.4 Predictive Values

The positive and negative predictive values can be calculated in addition to the specificity and sensitivity. Unlike sensitivity and specificity which have the ability to rule in, or rule out, the actual cases, positive and negative predictive values are the percentages of the predicted positive or predicted negative assessments that are correct. A low predictive value indicates that a large proportion of the positives or negatives are false and a high predictive value indicates that the majority of the predictions are correct. It is common to get a high number for the negative predictive value whilst having a low positive predictive value or vice versa whichever way round, even with one low value, the methodology may still have some utility in some studies.

5.2.7.5 Curves in ROC space:

When discussing the ROC graphs above only discrete classifiers were considered, however there are cases, for example this study, where continuous data are being assessed, in these cases curves can be employed to help. The ROC curve is obtained in much the same way as plotting a single variable by calculating sensitivity and specificity, and then plotting them in ROC space, except that in this case the sensitivity and specificity of all possible thresholds are plotted to produce a curve where each point in the curve represents the calculation of sensitivity and specificity at a given threshold (Van Erkel & Pattynama 1998). The curve is not necessarily curved in the common sense but rather stepped. The more thresholds that are plotted the more the appearance becomes curved.

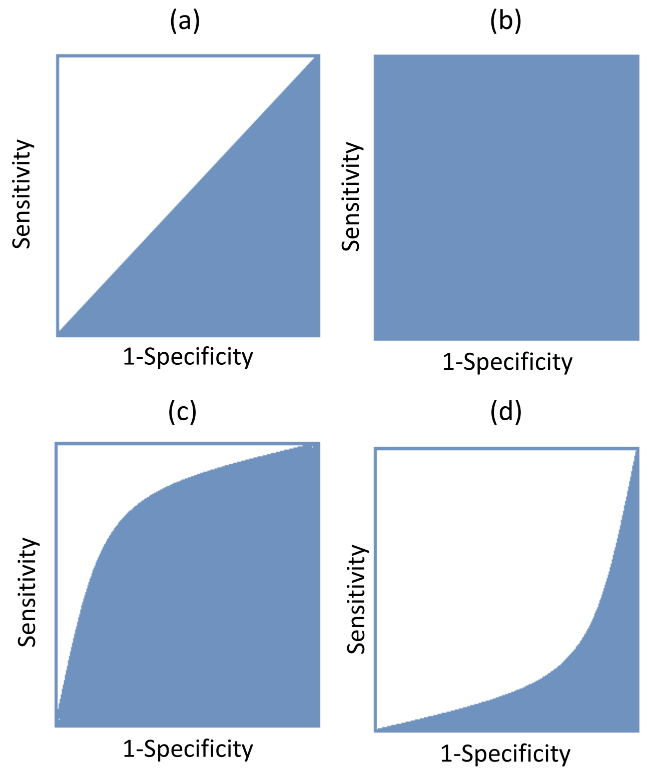


Figure 5.17: Variations of AUC. The blue shaded area represents the AUC.

5.2.7.6 Area Under the Curve (AUC)

The discriminative ability of a method is shown by the position of the applicable ROC curve in ROC space and the AUC is used to calculate that predictive performance. The larger the area under the curve is the better discriminating performance the test has. As with the simple ROC graph the straight line from the bottom left corner to the top right corner indicates that the test has equal true positive and false positive values for all cut-off points which automatically makes it useless for discrimination. The AUC for this scenario is 0.5 or 50% as seen in Figure 5.17(a). So if the curve/plot is away from the x=y line then there is some predictive value in the methodology. In the case of this study it will show that there is more similarity between family members compared to the remaining pool of unrelated people. In a perfect situation the ROC curve would contain the optimal point of 100% sensitivity and 100% specificity and thus make the AUC one or 100% as seen in Figure 5.17(b). The more likely scenario is somewhere in between a and b as is shown in by Figure 5.17(c) where the AUC is between 0.5 and 1. It is possible for a method to produce a curve with an area of less than 0.5, as shown in Figure 5.17(d), this indicates that there is predictive value in the method and by simply reversing the classifications, a curve can be flipped to resemble Figure 5.17(c). Swets (1988) suggested categories of discriminative power based on the AUC in order to compare multiple ROC analyses with each other and against other discriminative tests. He suggested that the following values of AUC could be categorised as:

$AUC = 0.5$	Non Informative
$0.5<AUC\leq0.7$	Less accurate
$0.7<AUC\leq0.9$	Moderately accurate
$0.9<AUC<1$	Highly accurate
$AUC = 1$	Perfect test

5.2.7.7 Optimising the Threshold

In order to apply the information yielded by the ROC curve to the situation in hand, a specific threshold needs to be chosen and applied to the rest of the population with the specificity and sensitivity that come with it. In the case of this study it would be a specific value of mesh-to-mesh value that if a comparison is less than that value it predicts related and if greater it predicts unrelated. The sensitivity and specificity calculated at that threshold would then apply for calculating the likelihood of the prediction being correct. As this study is using continuous data, the mesh-to-mesh values, the optimisation of the threshold is possible (Van Erkel & Pattynama 1998).

In basic terms the optimal threshold is the value that produces the point on the curve closest to the ideal, the top left hand corner of ROC space. This is easily calculated by various software programs. Nevertheless, the optimal threshold as calculated by the software might not be the best threshold for the specific question being asked of the analysis, it might be further along the curve in either direction. In order to determine the optimal threshold for the question being asked, it is necessary to see what effect changing the threshold has upon the relationship between sensitivity and specificity. Once the effect is known a decision can be made. As Metz (1978) says the ‘ROC curve describes the compromise that can be made between sensitivity and specificity’.

This process can be demonstrated graphically in ROC space or expressed as a probability distribution graph. Figure 5.18, adapted from Van Erkel & Pattynama (1998), shows the effect of five different threshold values (1,2,3,4.5 in Figure 5.18). The same five threshold points can be plotted in ROC space (Figure 5.18(a)) and on the probability distribution chart (Figure 5.18(b)). If this were to be replicated in this study then threshold number one, the highest threshold (this may be the lowest numerically, dependent upon whether high or low numbers equate to a positive classification) indicates all subjects being classed as genetically related whereas at threshold five, the lowest threshold, all subjects have been classed as genetically unrelated. These threshold positions are seldom useful as they show no discriminative power and would therefore not yield any helpful results. The calculated ideal threshold is depicted by threshold three, closest to the top left hand corner in Figure 5.18(a) and at the crossover point in Figure 5.18(b) minimising the number of False Positives and False Negatives. Threshold two, also shown in Figure 5.18(c), shows the effect of increasing the threshold from the graphical optimal. This increases the specificity, reducing the number of False Positives but consequently decreasing the sensitivity so the number of False Negatives rises. Threshold four, the effect of which is shown in Figure 5.18(d), is lower than the calculated optimal which increases the sensitivity but subsequently lowers the specificity. This in turn increases the number of False Positives, reduces True Negatives, and decreases the number of False Negatives, in this case to zero, meaning that the True Positives are 100%. This means that by selecting threshold four it is possible to have a ‘rule in’ policy accounting for all the True Positives or actual relatives, but in the process increasing the number of false positives.

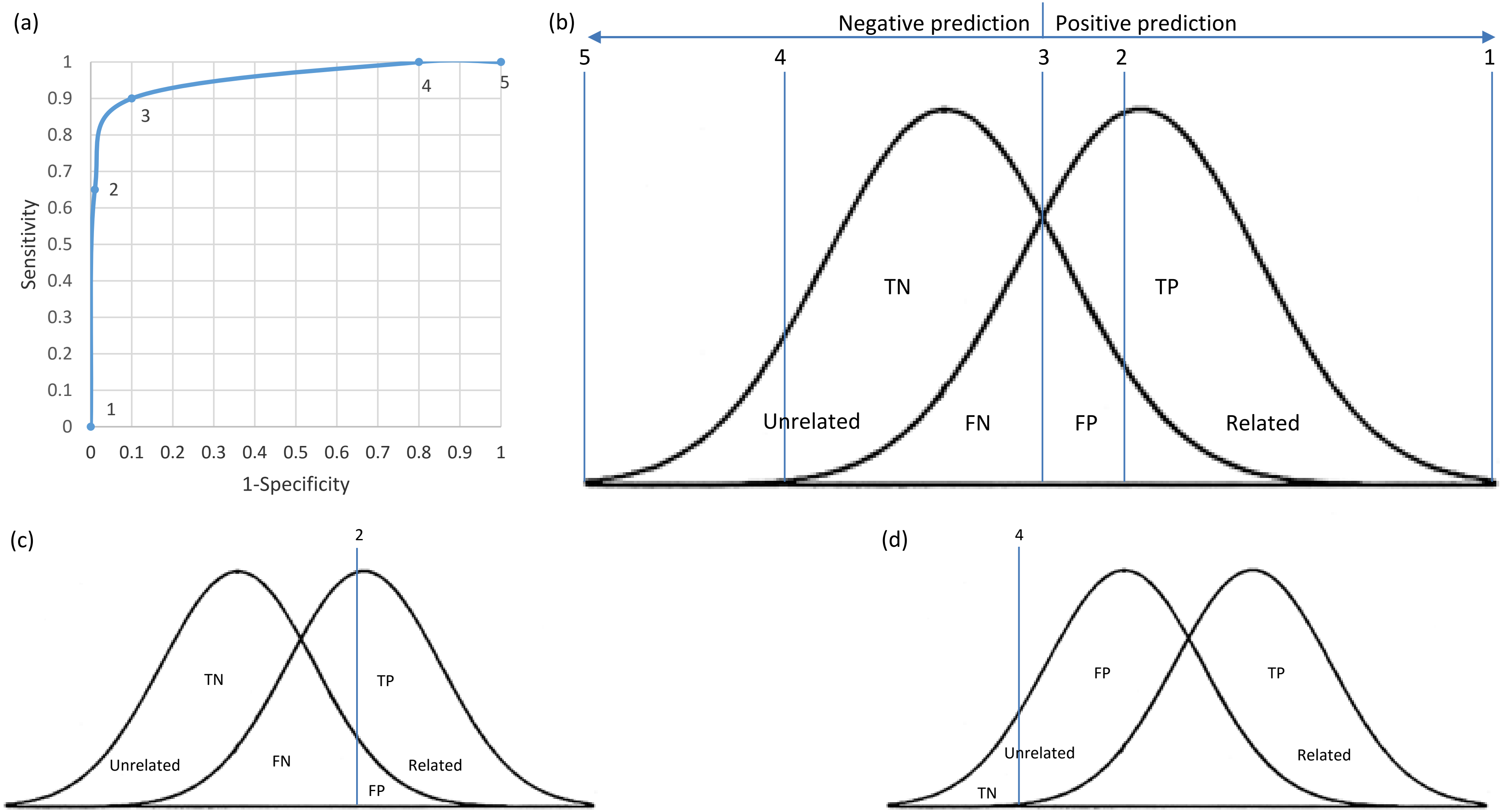


Figure 5.18: Adapted from Van Erkel 1998 and Metz 1978. (a) ROC graph with multiple thresholds. (b) Probability distribution showing multiple thresholds and the effect of threshold 3 on the classifications numbers. (c) Effect of threshold 2. (d) Effect of threshold 4.

5.2.6.8 Comparing ROC Curves

As the author will be applying ROC analysis to a number of different areas of the face and to a variety of population groupings it is important to be able to compare the ability of each of these to assess for familial similarity in order to determine which is most efficacious. To do this the ROC curves and results will need to be compared with each other both numerically and visually within ROC space.

As with threshold optimisation this is not always a simple question. In some cases, one ROC curve outperforms another at all points as in Figure 5.19(a). Curve A outperforms curve B at all thresholds, at any given Specificity the Sensitivity of A is greater than B and vice versa. Curve A also has a greater area under the curve (AUC) than curve B. The decision becomes more difficult if the curves cross each other at one or more points in ROC space as in Figure 5.19(b). In this circumstance it depends on the situation and question being asked of the analysis as to which curve provides the best performance. The Area Under the Curve might be a useful indicator, but as shown in Figure 5.19(b) this can be equal. If the situation demands high sensitivity, then curve B is the best option as it outperforms A in the high sensitivity area at the top of the graph. Whereas if specificity is more important than curve A is the better option. So the question to be answered in this study is whether it is more important to correctly identify related family members and keep them in the possible pool of people or to rule out a greater number of people as unrelated with the risk of not identifying a genetically related family member.

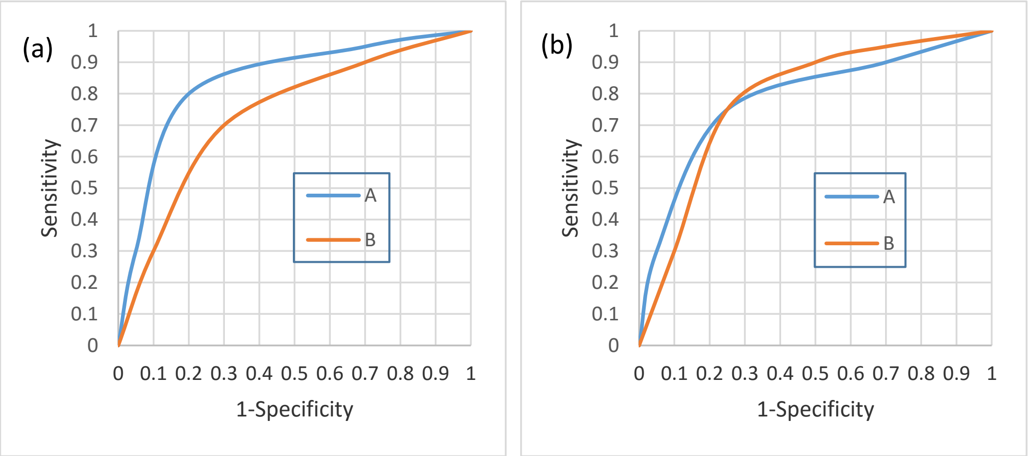


Figure 5.19 Area Under the Curve examples of sensitivity vs specificity

In performing ROC analysis, the mesh-to-mesh values exported from Viewbox and entered for analysis. Before doing so they are each given a simple classification of ‘1’ or ‘0’, ‘1’ meaning a true related match (Actual Positive as seen on the confusion matrix) and ‘0’ indicating unrelated (Actual Negative in the confusion matrix). The thresholds and the diagnostic characteristics of the mesh-to-mesh values (Sensitivity, Specificity, Positive and Negative predictive values, AUC) were calculated with MedCalc Version 14.12.0.

5.2.8 Preliminary Study of Areas of the Face:

As seen in both facial recognition and the heritability studies of the face, specific areas and features of the face are affected differently by ageing, sex variation, weight gain, disease, health issues and other variables. In order to assess whether heritability has differing effects on features, each area of the face and combinations of them, will be assessed separately for similarity.

The process of comparing the faces or areas of the face is very time consuming and has a long processing time, therefore, in order to minimise this, a preliminary study of ten families was conducted initially and only the areas of the face which yielded the best results were used for the full study.

Ten families consisting of one or both parents and one or more children were scanned using a portable and user friendly white light surface scanner (3D Flex Scan). The ten families consisted of ten mothers (M), four fathers (F), eight daughters (D) and eight sons (S). Each scan was manipulated in order to produce three-dimensional meshes of the 14 areas listed in Areas of the Face, 5.2.5.

For each of the 14 areas of the face ICP was performed for all possible combinations of the ten families, with the exception of the adults who were not compared with each other as no genetic relationship existed between any of them. For each area this meant there were 344 possible combinations and so 344 MMVs, 120 of these are just between children, 160 of them are between mothers and children and 64 of which are between fathers and children. Of these 344 combinations, 29 are actual positives, these being: seven between siblings, 16 between mothers and children and six between fathers and children.

Once the ICP had been conducted using Viewbox and all 4816 MMVs had been exported, the results for each area of the face were split into 11 categories before ROC analysis was carried out. The splitting of the tests is to ascertain whether heritability of facial morphology is equivalent across the sexes and generations, or whether certain relationships display greater inherited similarity. The 11 categories are:

- Mothers to all children (M-D/S)
- Mothers to all female children (M-D)
- Mothers to all male children (M-S)
- Fathers to all children (F-D/S)
- Fathers to all female children (F-D)
- Fathers to all male children (F-S)
- All children to all children (D/S-S/D)
- Children of opposite sexes – female to male (D-S)
- Children of the same sex – female-female and male-male (D-D, S-S)
- Children of the same sex – female-female (D-D)
- Children of the same sex – male-male (S-S)

ROC analysis was then carried out for each area of the face and for each grouping of subjects using MedCalc. The only group not to be studied was the children of opposite sex (D-S) as, of the ten families selected at random, only one contained an actual positive for this category and so ROC analysis was unable to be conducted. Optimal sensitivity and specificity, as calculated by MedCalc and area under the curve (AUC) were all considered in this preliminary study in order to compare the results for the different areas of the face. The 14 areas were considered in two groups. First the Whole face, Removed forehead, Removed ears and Removed inner eyes were compared as these are broadly similar in that they all show the majority of the face but with specific features incrementally removed.

Figure 5.20 shows an example of Mesh-to-Mesh comparison, the colour chart describes the variation with green showing very similar morphology and blue and yellow describing variation in both directions. The MMVs increase from right to left and top to bottom with the first two images depicting the sibling matches for F11C1.

Table 5.5 shows the results (sensitivity, specificity, AUC) for each of the first four areas of the face divided into the ten remaining categories and includes the total number of MMVs and actual positives for each category. Where the optimal sensitivity differed between areas, the most common sensitivity and its respective specificity is shown to allow for easier comparison. Where this is the case the percentages are shown in italics.

Seven out of the ten remaining categories show the Whole face to have the highest AUC. As seen in Table 5.5, (with two minor exceptions of just 0.001, which are to be found in the categories of All children to all children (D-S/S-D) and Mothers to all children (M-D/S), the numeric order of the AUC and the specificity in these seven categories follows the incremental sequence of the removal of parts of the face, with Whole face at the top and Removed inner eye at the bottom. These results indicate that manipulating the scans, by removing parts thought to interfere and confuse the comparison, is not actually beneficial and that using the Whole face gives superior results when looking for genetic similarity. This may be for a number of reasons; the position or shape of the ears, which even if not scanned, may affect the alignment of the face; the shape of the forehead or position of the hairline may aid comparison, and strange though it might seem the position of the eye lashes when scanning may also be a factor linked to genetics.

The remaining three categories do not follow the same pattern and so raise some doubts about the robustness of this possibility. The male same sex siblings (S-S), mother to son (M-S) and father to daughter (F-D) all display a different order for the areas when comparing AUC. All three place the Whole face as lowest whilst the top ranked area varies across all three of them with no single area with a higher ranking than others.

As seven of the categories show the Whole face to have the highest AUC and results for the remaining three categories were non concordant, it became clear that using the Whole face for the full study would be the best approach. Another advantage of using the Whole face was that this would not require any manual manipulation of the facial scans, thus making the while process simpler and more automated.

Table 5.6 shows the results (sensitivity, specificity, AUC) for each of the first 10 areas of the face split into the 10 remaining categories and includes the total number of MMV and actual positives for each category.

For each category the highest AUC calculated was selected, these are shown in bold. These show the optimum areas of the face for finding a genetic similarity for that particular category of relationship. The first issue to become apparent was that varying categories of relationship show very different results for specific areas of the face. No two categories have the same values or ranking. For instance, when looking at the highest AUC the Nose ranks highest with three categories rating it as top, the next three areas all incorporating the nose Eyes-Nose; Eyes-Nose-Mouth and Eyes-Nose-Mouth-Chin, ranked highest within two categories each and both Eyes and Mouth ranked highest in one category each.

These results illustrate that the remaining four areas of the face, Nose-Mouth; Nose-Mouth-Chin; Mouth-Chin and Chin will invariably yield results that are inferior to those of at least one other area from each category of relationship, and are therefore less effective tools in the search for genetic similarities. In order to save time in both mesh adaptation and the mesh-to-mesh processing time, data from these four discrete areas of the face will not be taken forwards to be processed and analysed as part of the full study.

The full study will process and analyse data from seven areas of the face, Whole face, Nose, Eyes-Nose; Eyes-Nose-Mouth, Eyes-Nose-Mouth-Chin, Eyes and Mouth, which will be addressed for each of the eleven relationship categories for ROC analysis.

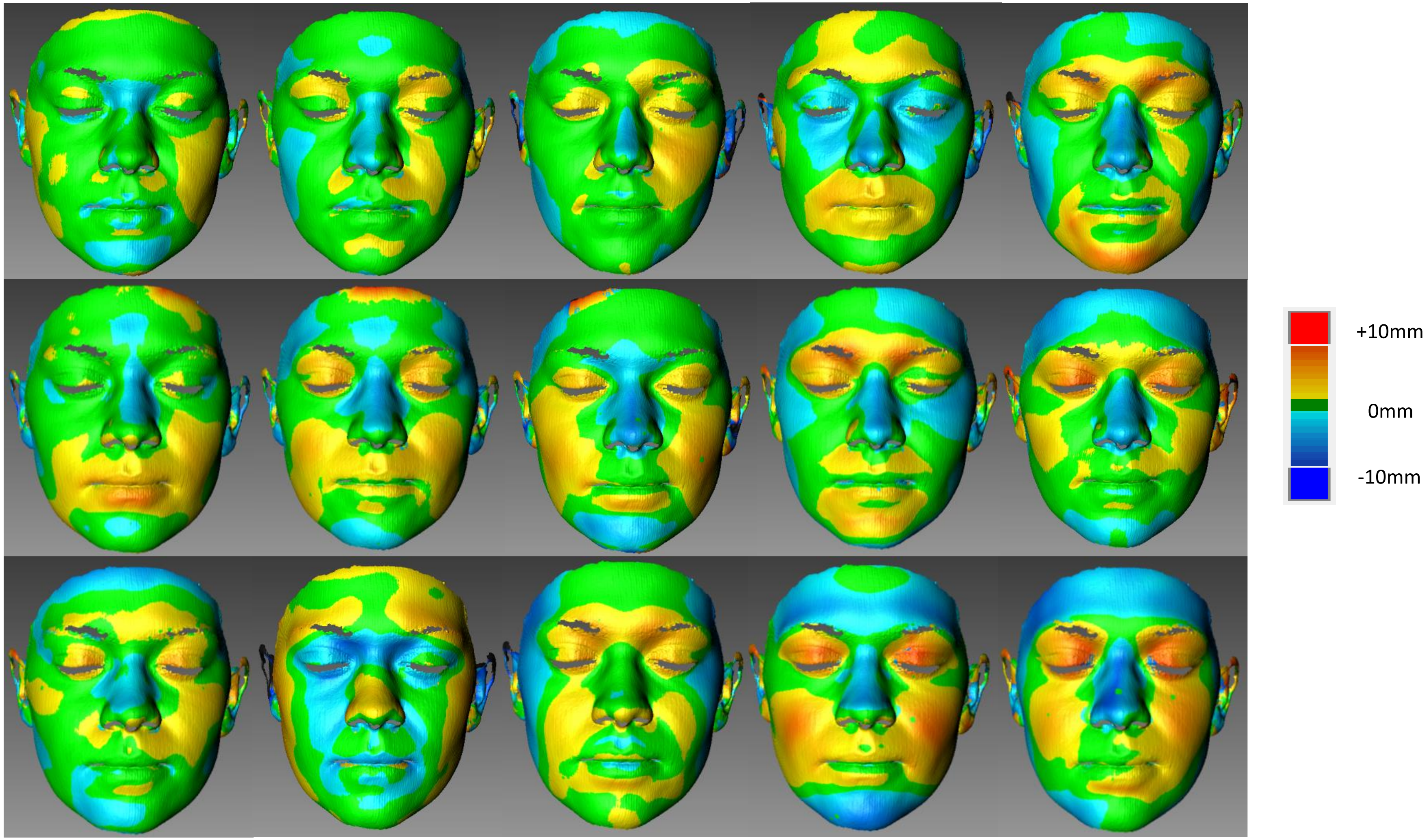


Figure 5.20 Example of Mesh-to-Mesh comparison, Family 11 Child one compared to all children, MMV increase left to right, top to bottom.

Category	All Children to all children			Same Sex Siblings			Same Sex Daughters			Same Sex Sons			Mothers to all Children		
	Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches	
	120	7		56	6		28	4		28	2		160	16	
Results	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC
Whole face	71.43	84.96	0.771	*66.67*	*94.00*	0.797	100.00	79.17	0.927	50.00	100.00	0.558	68.75	61.11	0.669
Removed Forehead	71.43	81.42	0.760	66.67	94.00	0.787	100.00	75.00	0.885	50.00	100.00	0.596	*68.75*	*64.58*	0.665
Removed Ears	71.43	78.76	0.742	66.67	92.00	0.773	100.00	70.83	0.875	50.00	96.15	0.577	*68.75*	*65.28*	0.666
Removed Inner Eyes	71.43	78.76	0.743	66.67	92.00	0.773	100.00	70.83	0.875	50.00	96.15	0.577	*68.75*	*64.58*	0.663

Category	Mothers-Daughters			Mothers-Sons			Fathers to All Children			Fathers-Daughters			Fathers-Sons		
	Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches	
	80	8		80	8		64	6		32	4		32	2	
Results	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC
Whole face	100.00	59.72	0.800	62.50	63.89	0.583	*66.67*	*43.10*	0.557	100.00	25.00	0.598	100.00	86.67	0.900
Removed Forehead	100.00	58.33	0.763	*62.50*	*66.67*	0.592	*66.67*	*37.93*	0.514	100.00	42.86	0.670	100.00	76.67	0.850
Removed Ears	100.00	56.94	0.756	62.50	72.22	0.595	66.67	17.24	0.511	100.00	42.86	0.688	100.00	76.67	0.850
Removed Inner Eyes	100.00	55.56	0.748	*62.50*	*69.44*	0.593	66.67	17.24	0.509	100.00	42.86	0.688	100.00	73.33	0.833

Table 5.5 ROC results for the Whole face, Removed forehead, Removed ears, Removed inner eye.

Category	All Children to all children			Same Sex Siblings			Same Sex Daughters			Same Sex Sons			Mothers to all Children		
	Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches	
	120	7		56	6		28	4		28	2		160	16	
Area	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC
Eyes	57.14	76.11	0.649	66.67	72.00	0.660	100.00	62.50	0.823	100.00	65.54	0.692	81.25	45.14	0.648
Eyes-Nose	57.14	93.81	0.757	66.67	96.00	0.837	100.00	95.83	**0.990**	100.00	61.54	0.692	75.00	60.42	0.678
Eyes-Nose-Mouth	71.43	73.45	0.757	66.67	88.00	0.827	100.00	87.50	0.958	100.00	46.15	0.577	75.00	64.58	0.701
Eyes-Nose-Mouth-Chin	57.14	83.19	0.753	66.67	78.00	0.770	75.00	83.33	0.813	100.00	46.15	0.654	50.00	93.06	**0.723**
Nose	100.00	69.91	**0.869**	100.00	78.00	**0.923**	100.00	79.17	0.906	100.00	80.77	**0.904**	81.25	53.47	0.637
Nose-Mouth	85.71	68.14	0.776	66.67	88.00	0.777	100.00	83.33	0.896	100.00	30.77	0.538	68.75	59.72	0.653
Nose-Mouth-Chin	57.14	92.92	0.747	66.67	90.00	0.693	75.00	87.50	0.729	50.00	100.00	0.615	56.25	82.64	0.648
Mouth	71.43	62.83	0.612	83.33	60.00	0.673	75.00	54.17	0.573	100.00	80.77	**0.904**	43.75	76.39	0.537
Mouth-Chin	57.14	81.42	0.612	66.67	84.00	0.700	75.00	79.17	0.615	100.00	65.38	0.788	68.75	45.83	0.544
Chin	85.71	48.67	0.621	83.33	52.00	0.617	75.00	58.33	0.604	100.00	57.69	0.596	6.25	70.14	0.501

Table 5.6 Page 1 of 2 ROC results for preliminary study.

Category	Mothers-Daughters			Mothers-Sons			Fathers to All Children			Fathers-Daughters			Fathers-Sons		
	Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches		Sample	Matches	
	80	8		80	8		64	6		32	4		32	2	
Area	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC
Eyes	100.00	37.50	0.684	100.00	26.39	0.618	50.00	18.97	0.532	75.00	78.57	0.625	100.00	83.33	**0.867**
Eyes-Nose	87.50	47.22	0.649	62.50	76.39	0.708	100.00	37.93	0.641	100.00	46.43	**0.661**	100.00	76.67	0.850
Eyes-Nose-Mouth	75.00	61.11	0.682	75.00	68.06	**0.719**	83.33	56.90	**0.658**	75.00	67.86	0.652	100.00	83.33	0.833
Eyes-Nose-Mouth-Chin	87.50	62.50	**0.785**	62.50	93.05	0.682	83.33	50.00	0.626	100.00	35.71	0.607	100.00	53.33	0.733
Nose	100.00	47.22	0.674	50.00	81.94	0.583	16.67	62.07	0.506	100.00	35.71	0.554	100.00	56.67	0.717
Nose-Mouth	100.00	26.39	0.623	50.00	87.50	0.688	83.33	51.72	0.621	75.00	75.00	0.625	100.00	43.33	0.650
Nose-Mouth-Chin	50.00	91.67	0.696	62.50	77.78	0.628	83.33	43.10	0.534	100.00	39.29	0.554	100.00	36.67	0.583
Mouth	62.50	68.06	0.609	62.50	77.78	0.684	66.67	72.41	0.655	75.00	60.71	0.607	100.00	83.33	0.833
Mouth-Chin	37.50	91.67	0.523	37.50	88.89	0.625	83.33	53.45	0.572	75.00	57.14	0.527	100.00	60.00	0.633
Chin	75.00	61.11	0.573	87.50	44.44	0.575	83.33	43.10	0.572	100.00	42.86	0.545	50.00	100.00	0.633

Table 5.6 Page 2 of 2 ROC results for preliminary study.

5.2.9 The Combination of Single Features of the Face.

In addition to the seven areas of the face selected from the preliminary study, four combinations of the single features will also be subject to ROC analysis in a similar method to that conducted by Smeets et al. (2010). The MMV of the eyes, nose and mouth will be added together in the four combinations below:

Eyes and nose = MMV of eyes + MMV of nose

Eyes and mouth = MMV of eyes + MMV of mouth

Nose and mouth = MMV of nose + MMV of mouth

Eyes and nose and mouth = MMV of eyes + MMV of nose + MMV of mouth

Each of the four combinations will be then be analysed in the same way as the seven other areas of the face.

Chapter 6: Results

When using the terms ‘children’, ‘sons’ or ‘daughters’ in this study the descriptions will always refer to those offspring included in this study, whose ages range from 7 to 24 years. The terms used do not refer in any way to the legal status or developmental classification of these offspring. In the same way the terms ‘parents’, ‘adults’, ‘mothers’ and ‘fathers’ refer to the parental categories with an age range of 39 to 62 years.

When referring to the portions of the face three classifications are used to describe the type of portion, or segment, in addition to the whole face. A ‘Single feature’ describes each of the three features in their own right, these are the Eyes, the Nose and the Mouth. ‘Joined features’ are each a single surface which comprise more than one feature, a hyphen is used between the features to describe a joined feature. ‘Combined features’ are formed from the combination of the MMVs of three single features and are always referred to with an ‘and’ between the names of the features.

6.1 Landmark Based Regression

Principal Component Analysis was conducted for the different groups of individuals and for each of the landmark groups described in the methodology. Whilst analysis of the principal components could give some insight into the variation caused by BMI, age and sex, no clear component could be identified in which family members were close to each other spatially.

The results of regression analysis of BMI, Age and Sex are shown in tables 6.1-6.3. Each regression was carried out for the different groups of individuals and for the varying landmark groups. Regression was conducted after Procrustes superimposition and the generation of a covariance matrix. Overall the percentage predicted by the covariates are low and many are not significant (P value >0.05), shown in bold. There are some results worth noting, the percentage predicted by BMI for the eyes in the sons group is 17.18%. This suggests that a large amount of the variance in the shape of the eyes, between the male children is caused by a variation in BMI. The same is true of the mouth but this time across all the male participants, with 15.59% predicted. The results for the same group and landmarks, the male mouth, but regressed with respect to age is also high with 29.57% predicted. It is likely therefore that the variation in BMI is a consequence of the generation gap and that the percentage predicted by age is higher as a result of the difference in the overall BMI between the younger male sons and older male fathers. The mouth is also the area where age is a high predictor for females, 22.61% and for all individuals, 23.78%. These are the three groups which include both generations in their makeup and so again the variation of the mouth seems to be explained by the change between childhood and adulthood. The regression results for the mouth where only one generation is considered are much lower. Sex, Table 6.3, does not have any high predictive percentages and most of the results are not significant.

BMI	**All landmarks**		**Nose**		**Eyes**		**Mouth**		**Profile**	
	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value
All individuals	2.88%	0.0007	1.82%	0.0154	1.31%	**0.1096**	5.66%	0.0002	3.10%	0.0004
Adults	3.35%	**0.0501**	0.86%	**0.7926**	0.87%	**0.6939**	1.86%	**0.2776**	3.65%	0.0250
Mothers	5.79%	0.0205	2.66%	**0.3690**	2.81%	**0.3389**	3.31%	**0.2298**	5.60%	0.0266
Fathers	8.66%	**0.0804**	5.78%	**0.1261**	2.46%	**0.6308**	6.81%	**0.1308**	5.20%	**0.3202**
Children	3.81%	0.0221	3.86%	0.0138	6.72%	0.0002	0.88%	**0.6654**	1.89%	**0.2699**
Daughters	7.08%	**0.0681**	7.84%	0.0382	8.87%	0.0303	2.62%	**0.5742**	8.28%	0.0336
Sons	6.91%	0.0207	4.97%	0.0498	17.18%	<0.0001	1.75%	**0.5928**	3.77%	**0.1902**
Females	4.52%	0.0013	1.55%	**0.3685**	1.21%	**0.5146**	1.36%	**0.4079**	4.26%	0.0045
Males	6.16%	0.0010	4.16%	0.0081	5.03%	0.0056	15.59%	0.0001	4.44%	0.0124

Table 6.1 Regression analysis of three-dimensional landmarks with respect to BMI.

Age	**All landmarks**		**Nose**		**Eyes**		**Mouth**		**Profile**	
	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value
All individuals	5.97%	<0.0001	3.81%	0.0001	1.79%	0.0369	23.78%	<0.0001	7.69%	<0.0001
Adults	2.98%	**0.0974**	1.27%	**0.5360**	0.46%	**0.9140**	1.53%	**0.3954**	1.28%	**0.6034**
Mothers	6.33%	0.0091	1.43%	**0.8118**	0.86%	**0.8996**	3.65%	**0.1757**	1.63%	**0.7562**
Fathers	5.07%	**0.5083**	3.07%	**0.5567**	0.96%	**0.9437**	2.65%	**0.7131**	2.84%	**0.7982**
Children	4.60%	0.0056	5.62%	0.0008	4.04%	0.0202	1.95%	**0.2281**	5.60%	0.0009
Daughters	6.22%	**0.1508**	2.26%	**0.7254**	2.65%	**0.5823**	3.63%	**0.3790**	4.62%	**0.2608**
Sons	8.20%	0.0076	11.55%	<0.0001	9.34%	0.0053	4.28%	**0.1522**	10.50%	0.0008
Females	6.67%	<0.0001	3.29%	0.0308	1.15%	**0.5453**	22.61%	<0.0001	8.33%	<0.0001
Males	8.22%	<0.0001	5.80%	0.0001	3.85%	0.0269	29.57%	<0.0001	9.55%	<0.0001

Table 6.2 Regression analysis of three-dimensional landmarks with respect to age.

Sex	**All landmarks**		**Nose**		**Eyes**		**Mouth**		**Profile**	
	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value	% predicted	P-value
All individuals	4.37%	<0.0001	1.74%	0.0213	1.47%	**0.0726**	0.80%	**0.3251**	1.45%	**0.0661**
Adults	8.45%	<0.0001	2.75%	**0.0595**	1.99%	**0.2163**	6.01%	0.0036	4.65%	0.0041
Children	3.74%	0.0238	2.31%	**0.1322**	2.04%	**0.2197**	2.90%	**0.0965**	1.32%	**0.5366**

Table 6.3 Regression analysis of three-dimensional landmarks with respect to sex.

	All Children	Children Opposite Sex	Children Same Sex	Children Same Sex Daughters	Children Same Sex Sons	Mothers-Children	Mothers-Daughters	Mothers-Sons	Fathers-Children	Fathers-Daughters	Fathers-Sons
Whole Face	3.4448	3.4830	3.4075	2.7118	3.7697	3.5526	3.2020	3.8067	4.3147	4.4921	4.1860
Eyes	1.5190	1.5284	1.5099	1.3276	1.6047	1.5406	1.4540	1.6034	1.6681	1.6562	1.6768
Nose	1.0321	1.0419	1.0225	0.9230	1.0742	1.1213	1.0431	1.1781	1.2234	1.2058	1.2361
Mouth	1.0633	1.0618	1.0648	0.9279	1.1361	1.1578	1.0735	1.2189	1.2893	1.2412	1.3242
Eyes-Nose	1.7430	1.7692	1.7173	1.5014	1.8298	1.8151	1.6688	1.9211	1.9264	1.9442	1.9136
Eyes-Nose-Mouth	1.8740	1.8912	1.8572	1.6128	1.9843	1.9491	1.8270	2.0376	2.1527	2.1798	2.1331
Eyes-Nose-Mouth-Chin	2.0488	2.0552	2.0426	1.7306	2.2049	2.1565	2.0125	2.2609	2.3297	2.3245	2.3335
Eyes and Nose	2.5511	2.5703	2.5323	2.2506	2.6790	2.6620	2.4971	2.7815	2.8915	2.8619	2.9129
Eyes and Mouth	2.5823	2.5902	2.5747	2.2555	2.7408	2.6984	2.5275	2.8223	2.9574	2.8974	3.0009
Nose and Mouth	2.0954	2.1038	2.0873	1.8509	2.2103	2.2791	2.1165	2.3970	2.5127	2.4470	2.5603
Eyes and Nose and Mouth	3.6144	3.6321	3.5971	3.1785	3.8151	3.8197	3.5705	4.0004	4.1808	4.1032	4.2371

Table 6.4 The Average Mesh-to-Mesh Values (mm) for each type to comparison, Relationship categories vs Areas of the Face. Conditional formatting shows the lowest values in red, through orange to green for the highest values.

6.2 Comparison of Relationship Categories.

6.2.1 Mesh-to-Mesh Value Averages

The average Mesh-to-Mesh Value (MMV) for each relationship category against the areas of the face are shown in Table 6.4. Using the ‘All Children’ column as an example it can clearly be seen that the values for each area of the face vary greatly from 1.0321 for the Nose to 3.6144 for the Eyes and Nose and Mouth. This variation is simply relative to the size of area being compared, the larger the area the greater the MMV as there is cumulatively more variation possible. The same order is generally followed for each of the relationship categories. For this reason, no comparison of MMV should be made across areas of the face but rather each section of the face should be considered separately.

A colour code has been applied by row showing the lowest values in red, increasing through orange and into green for the highest values. The lower the average the more similar the comparative groups are. The comparison of sisters exhibits the lowest values, as denoted by the red shading, for all areas of the face. All comparisons involving Fathers are at the top end of the values, show in green. Figure 6.1 shows these comparisons graphically, the Daughter to Daughter category (yellow) sits well below all the others exhibiting a clear gap. By using the mean MMV values we are simply stating that females under 18 are basically more similar to each other than any other combination in the sample independently of whether they are genetically related or not. This indicates the strong influence of age (ranging from 11 years two months to 21 years three months) and sex to facial morphology.

The other categories tend to remain in the same order for all areas of the face. For the majority of the areas of the face, from lowest to highest the relationship categories are ordered:

1. Children of the same sex – female-female (D-D)
2. Mothers to all female children (M-D)
3. Children of the same sex – female-female and male-male (D-D, S-S)
4. All children to all children (D/S-S/D)
5. Children of opposite sexes – female to male (D-S)
6. Mothers to all children (M-D/S)

7. Children of the same sex – male-male (S-S)
8. Mothers to all male children (M-S)
9. Fathers to all female children (F-D)
10. Fathers to all children (F-D/S)
11. Fathers to all male children (F-S)

The main variation to this order is seen in ‘Mouth’, ‘Nose’ and ‘Nose and Mouth’. Although most of the order remains unchanged, ‘Mothers to all female children’, shown in dark blue in Figure 6.1, moves down to position five with the others moving up to fill the gap.

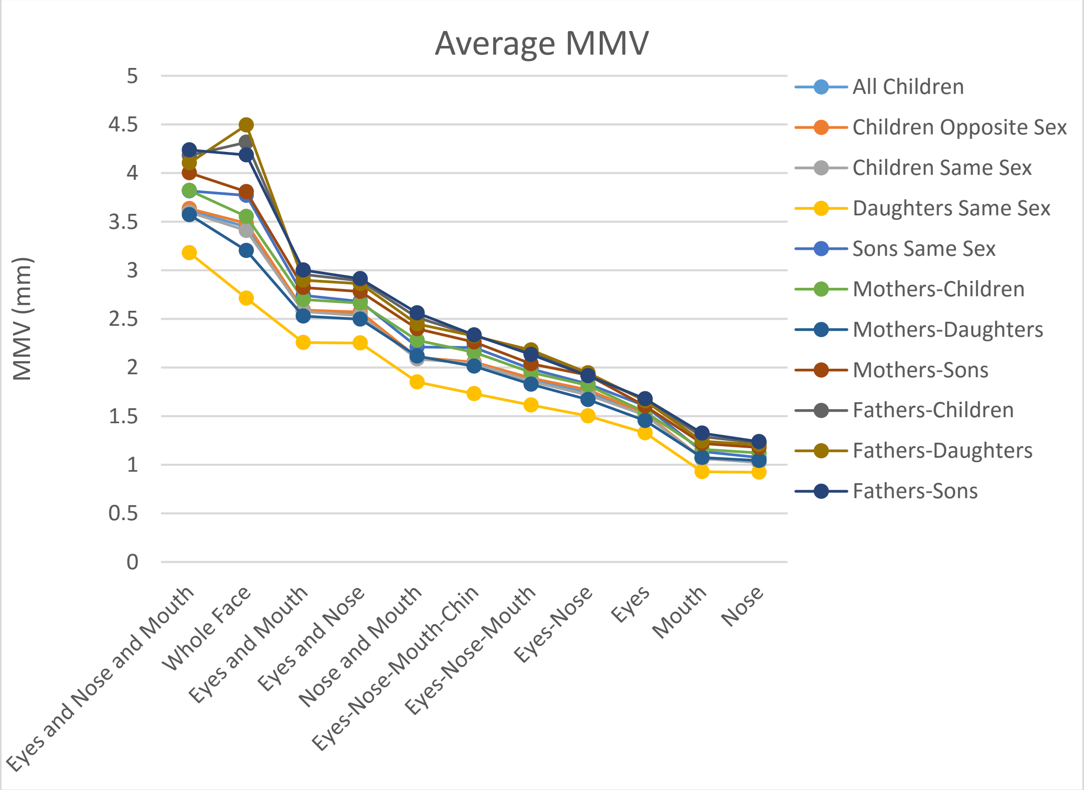


Figure 6.1. The Average Mesh-to-Mesh Values vs Areas of the Face for each Relationship Category. Note the lines between markers are only there to aid the observation of each relationship, they do not indicate any trends.

6.2.2 Area Under the Curve

The Area Under the Curve (AUC) in Receiver Operating Characteristics (ROC) analysis enables comparison across both Relationship categories and Areas of the face. Table 6.5 shows the AUC for all combinations, conditional formatting has been applied across the entire table with green indicating the higher, more predictive, values and red the lower, less predictive, values. Whilst most of the table is varied in colour one Relationship category jumps out as superior in predictive value (dark green), Children of the same sex (Daughter to Daughter comparison) shows much higher AUC values across the board with a maximum of 0.857. There is one sharp exception, the Mouth gained one of the lowest values of 0.519. Figure 6.2 also demonstrates this superiority of the sisterly relationship (Purple) but clearly shows that the remainder of the categories do not exhibit any clear order. The plot lines cross and transect each other multiple times with no pattern discernible across the different Areas of the face. This shows that with the exception of the Daughter to Daughter category, the predictive value of the different areas of the face differs between Relationship categories. In the same manner as Relationship categories, no clear pattern is seen with the Portions of the face. The Mouth does show some extremely low values but other areas change position from category to category indicating that for each Relationship category a different area of the face will give the best predictive power in finding a genetic relationship.

	All Children	Children Opposite Sex	Children Same Sex	Children Same Sex Daughters	Children Same Sex Sons	Mothers-Children	Mothers-Daughters	Mothers-Sons	Fathers-Children	Fathers-Daughters	Fathers-Sons
Whole Face	0.663	0.633	0.689	0.851	0.663	0.704	0.734	0.699	0.608	0.579	0.631
Eyes	0.623	0.562	0.672	0.784	0.628	0.686	0.661	0.707	0.633	0.609	0.651
Nose	0.652	0.614	0.681	0.720	0.687	0.638	0.671	0.623	0.572	0.587	0.563
Mouth	0.643	0.648	0.640	0.519	0.674	0.582	0.503	0.639	0.636	0.565	0.687
Eyes-Nose	0.648	0.591	0.695	0.857	0.654	0.689	0.684	0.701	0.669	0.650	0.684
Eyes-Nose-Mouth	0.648	0.639	0.654	0.816	0.607	0.694	0.674	0.708	0.656	0.667	0.653
Eyes-Nose-Mouth-Chin	0.685	0.668	0.697	0.747	0.713	0.665	0.698	0.643	0.675	0.663	0.685
Eyes and Nose	0.664	0.601	0.713	0.822	0.693	0.714	0.722	0.724	0.644	0.620	0.663
Eyes and Mouth	0.675	0.628	0.709	0.760	0.697	0.709	0.652	0.746	0.671	0.622	0.707
Nose and Mouth	0.673	0.645	0.694	0.658	0.707	0.643	0.602	0.673	0.644	0.586	0.686
Eyes and Nose and Mouth	0.696	0.635	0.74	0.801	0.742	0.726	0.703	0.753	0.677	0.626	0.717

Table 6.5 Area Under the Curve of ROC curve for all Areas of the Face by Relationship Categories.

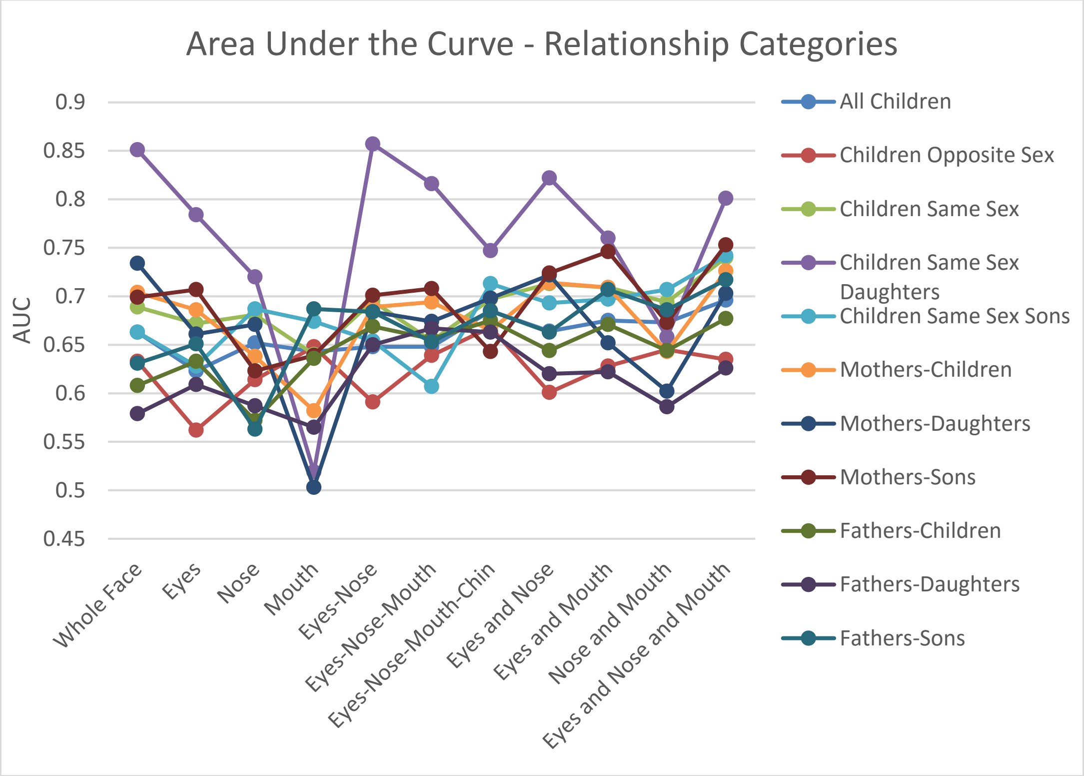


Figure 6.2 Area Under the Curve vs Areas of the Face for each Relationship category. Note the lines between markers are only there to aid the observation of each relationship, they do not indicate any trends.

6.3 Comparative Results of Each Relationship Category.

As seen in the overall AUC comparison, each Area of the face performs differently in each Relationship category. It is likely that if one was searching for relatives within a pool of people, the type of relationship, e.g. Mother-Daughter, would be known. Due to both of these factors, the results will be presented in more detail for each Relationship category.

Appendix three available on the attached CD contains the full ROC results for every relationship category and every portion of the face. The ROC graphs, optimal dot diagrams and Rule In dot diagrams are also available. The full contents of this appendix can be found within the printed appendices at the end of this thesis. Some of the most relevant figures are reproduced within this chapter.

Each section will cover;

- The Area Under the Curve and its significance, allowing for direct comparison between the different areas.

- The optimal specificity (%) and sensitivity (%) combination, calculated as the closest distance to the top left hand corner of the ROC graph.
- The Rule In option, the specificity at 100% sensitivity. This rules in all the true matches and gives the percentage of the total number of comparisons that can be ruled out, defined as true negatives, with confidence.
- All results will be repeated after having removed the ten poorest quality scans from the total population, for each Area of the face in turn. This will indicate whether the quality of scan has an impact on the comparison. The number of comparisons and true matches varies for each Portion of the face so is denoted in the tabled results. The identity of the ten poorest quality scans for each Portion of the face can be seen in Appendix 2
- The mean and range of both age and BMI will also be presented in addition to the age variation between matches.

6.3.1 Children

The category of children includes all possible comparisons between all children of both sexes. In total there are 2,346 comparisons and 34 true matches indicating a sibling relationship. The mean age of the category is 16 years with a total age span of seven years one month to 24 years. The mean BMI is 21.39, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between siblings is three years ten months, ranging from 0 years (twins) to 12 years, nine months.

As shown in Table 6.6, all these results are significant. Accuracy of correct identification of siblings and non-siblings, as calculated by AUC is greatest for Eyes and Nose and Mouth the additional combination of the three separate features. This gives optimal sensitivity of 66.65% and specificity of 66.87% with a threshold value of ≤3.221mm. In Figure 6.3 a dot diagram portrays this result visually, with all the actual negatives on the left classified as zero and all actual positives on the right classified as one. Dots above the threshold line have been predicted as negative whilst those under the line have been predicted as positive. In actual numbers this means 23 of the 34 sibling matches and 1,547 of the 2,312 non-matches are correctly predicted while 11 matches and 765 non-matches are incorrectly predicted.

When a Rule In application is needed so that there are no False Negatives, sensitivity is set to 100% and the relative threshold applied to give the specificity. In this situation Eyes and Nose produces the best specificity of 13.41%. In order to correctly identify all 34 of the sibling matches the threshold is set ≤3.373mm, meaning that 2,002 of the 2,312 non-siblings are wrongly identified as siblings. This enables a reduction of the 2,346 total comparisons considered as possible siblings by

310 to 2,036 a reduction of 13.2%.

The Nose performs best of all the Single features but none perform as well as the Whole Face.

The largest of the joined features Eyes-Nose-Mouth-Chin performed better than the other joined features and overall they are slightly better than any of the single features. Only Eyes-Nose-Mouth-Chin outranks the Whole Face.

The triple combination Eyes and Nose and Mouth is the best performing of the combinations, this outperforms the Whole Face and all of the joined features with the exception of Eyes-Nose-Mouth-Chin.

Table 6.7 shows the results following the removal of the ten poorest scans. The AUC increases across the board and the Rule In percentages dramatically increase for the largest of the three portions; Whole Face, Eyes-Nose-Mouth-Chin, Eyes and Nose and Mouth. The top results for AUC and Rule In changed from the previous ranking to Eyes-Nose-Mouth-Chin and Eyes and Nose and Mouth respectively.

.

Number of comparisons:2346
Matches: 34

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.663	0.0013	50.00	77.16	3.98
Eyes	0.623	0.0042	79.41	44.12	10.47
Nose	0.652	0.0033	76.47	50.17	5.67
Mouth	0.643	0.0038	79.41	46.93	3.07
Eyes-Nose	0.648	0.0014	79.41	47.40	10.47
Eyes-Nose-Mouth	0.648	0.0025	85.29	37.50	5.49
Eyes-Nose-Mouth-Chin	0.685	<0.0001	88.24	44.20	13.06
Eyes and Nose	0.664	0.0002	79.41	47.84	13.41
Eyes and Mouth	0.675	<0.0001	79.41	53.98	4.28
Nose and Mouth	0.673	0.0003	82.35	46.45	12.72
Eyes and Nose and Mouth	0.696	<0.0001	67.65	66.87	7.92

Table 6.6 ROC results for all children compared with all children

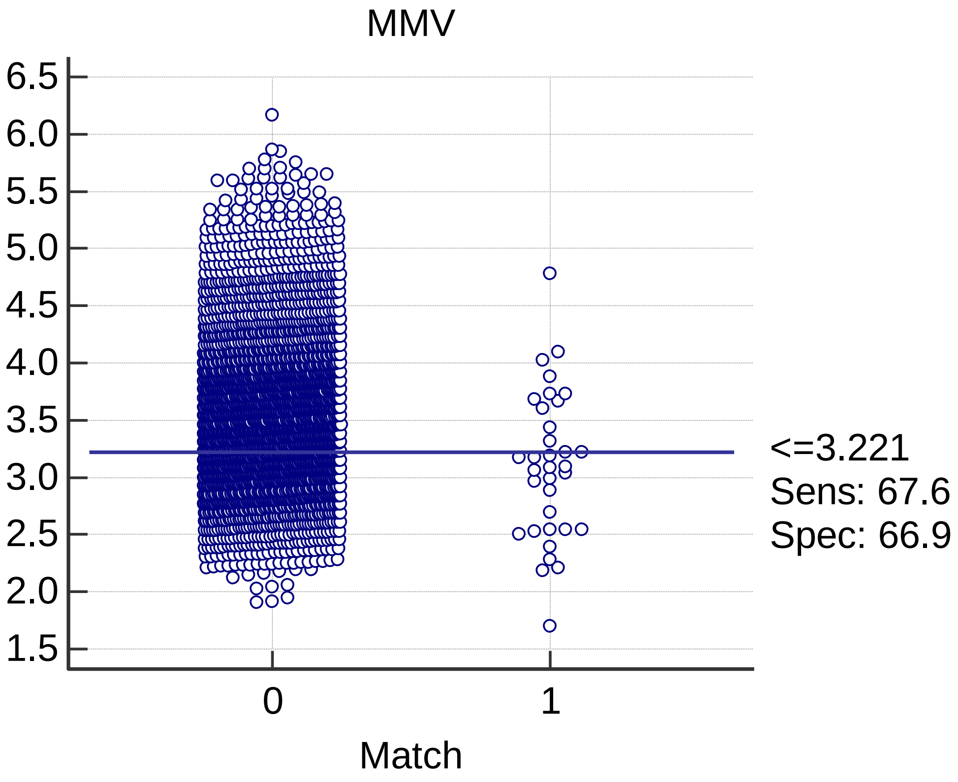


Figure 6.3 ROC Dot diagram for all children compared with all children, Eyes and Nose and Mouth combination with optimal threshold

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	2080	31	0.696	0.0001	54.84	77.01	18.45
Eyes	2016	31	0.628	0.0045	87.10	37.38	9.42
Nose	2016	30	0.682	0.0005	53.33	78.45	5.69
Mouth	2145	29	0.643	0.0083	62.07	63.09	2.22
Eyes-Nose	2016	29	0.689	0.0001	86.21	46.55	9.81
Eyes-Nose-Mouth	2145	29	0.700	<0.0001	75.86	53.92	8.65
Eyes-Nose-Mouth-Chin	2145	29	0.728	<0.0001	93.10	43.43	18.76
Eyes and Nose	1830	29	0.674	0.0005	37.93	88.73	13.71
Eyes and Mouth	1830	26	0.690	<0.0001	92.31	46.78	27.66
Nose and Mouth	1830	26	0.701	0.0001	92.31	42.35	12.47
Eyes and Nose and Mouth	1653	25	0.725	<0.0001	100.00	34.77	34.77

Table 6.7 ROC results for all children compared with all children after removal of the ten poorest quality scans.

6.3.1.1 Children of Opposite Sex

The category of children of opposite sex includes all the comparisons between children of opposite sexes, daughters to sons. In total there are 1,160 comparisons and 15 true matches indicating a sibling relationship of sister to brother. The mean age of the category is the same as with all children, 16 years with a total age span of seven years one month to 24 years as is the mean BMI, 21.39 which ranges from 14.67 to 30.52. The average age difference between siblings is three years six months, ranging from one year to ten years, nine months.

Table 6.8 shows the results for this category. Six of the 11 results are not significant (shown in bold italics) although two of those, Whole Face and Eyes-Nose-Mouth are close to the threshold. The accuracy of sibling identification is shown to be best using Eyes-Nose-Mouth-Chin with AUC of 0.668 giving an optimal sensitivity of 86.67% and specificity of 48.12%. As shown in Figure 6.4 this optimal result is produced with a threshold set at ≤2.0607mm. In real numbers this means 13 of the 15 sibling matches and 551 of the 1,145 non-matches are correctly predicted while two matches (above threshold line, right side) and 594 non-matches are mistakenly predicted.

If wishing to be certain of correctly identifying all 15 sibling matches, then the best area is Eyes and Mouth which gives a specificity of 30.39% when a Rule In status is applied. A threshold of ≤2.892 reduces the False Negatives to zero but only correctly identifies 348 of the 1,145 non-siblings, therefore misclassifying 797 non-siblings as siblings. Overall this could reduce the total number of comparisons being considered from 1,160 to 812 a reduction of 30%.

The Mouth is the most accurate of the single features in this category and the only significant result. It performs better than the Whole Face for both AUC and Rule In values.

The AUC and significance of the joined features improved as the number of included features increased and only became significant when all four features where included as in Eyes-Nose-Mouth-Chin. The Whole Face again performed less well than this selected area, with the Mouth showing results between the two.

The combined areas achieved the better significance as only one of the four was not significant. The Nose and Mouth had the greatest AUC bettering the larger Eyes and Nose and Mouth and the other combinations involving the eyes. This may mean that the eyes themselves are a hindrance in this category as they also show the lowest AUC overall.

Table 6.9 displays the results once the ten poorest quality scans have been removed, the AUC increases for seven of the eleven face portions. The AUC of the Mouth, Eyes and Mouth, Nose and Mouth and Eyes and Nose and Mouth all decrease and become non-significant. The Whole Face improves dramatically becoming significant and gaining the best Rule In specificity of 38.30% while Eyes-Nose-Mouth-Chin remains the highest AUC overall.

Number of Comparisons:1160
Matches:15

Area	**AUC**	**Significance P-value**	**Optimal Sensitivity (%)**	**Optimal Specificity (%)**	**Rule In (%)**
Whole Face	0.633	***0.0634***	93.33	39.30	3.84
Eyes	0.562	***0.2646***	100.00	25.15	25.15
Nose	0.614	***0.1598***	80.00	43.93	6.64
Mouth	0.648	0.0338	80.00	47.16	13.45
Eyes-Nose	0.591	***0.2031***	73.33	49.00	11.62
Eyes-Nose-Mouth	0.639	***0.0696***	73.33	57.38	9.34
Eyes-Nose-Mouth-Chin	0.668	0.0112	86.67	48.12	17.55
Eyes and Nose	0.601	***0.1269***	93.33	31.44	14.85
Eyes and Mouth	0.628	0.0279	100.00	30.39	30.39
Nose and Mouth	0.645	0.0358	80.00	48.38	14.15
Eyes and Nose and Mouth	0.635	0.0344	86.67	41.05	25.33

Table 6.8 ROC results for opposite sex children.

Area	**No of comparisons**	**Matches**	**AUC**	**Significance P-value**	**Optimal Sensitivity (%)**	**Optimal Specificity (%)**	**Rule In (%)**
Whole Face	1026	13	0.668	0.0132	100.00	38.30	38.30
Eyes	1008	13	0.562	***0.2756***	100.00	24.52	24.52
Nose	999	14	0.640	***0.0736***	85.71	41.32	6.70
Mouth	1073	13	0.624	***0.1226***	30.77	92.17	10.75
Eyes-Nose	999	13	0.630	***0.0954***	84.62	48.99	11.26
Eyes-Nose-Mouth	1064	13	0.670	0.0312	76.92	56.14	9.42
Eyes-Nose-Mouth-Chin	1064	13	0.694	0.0047	92.31	47.38	17.79
Eyes and Nose	918	13	0.614	***0.1026***	100.00	28.73	28.73
Eyes and Mouth	924	11	0.607	***0.1152***	81.82	47.21	27.27
Nose and Mouth	918	12	0.626	***0.1222***	83.33	44.70	12.80
Eyes and Nose and Mouth	837	11	0.629	***0.0792***	100.00	35.59	35.59

Table 6.9 ROC results for opposite sex children after removal of the ten poorest quality scans.

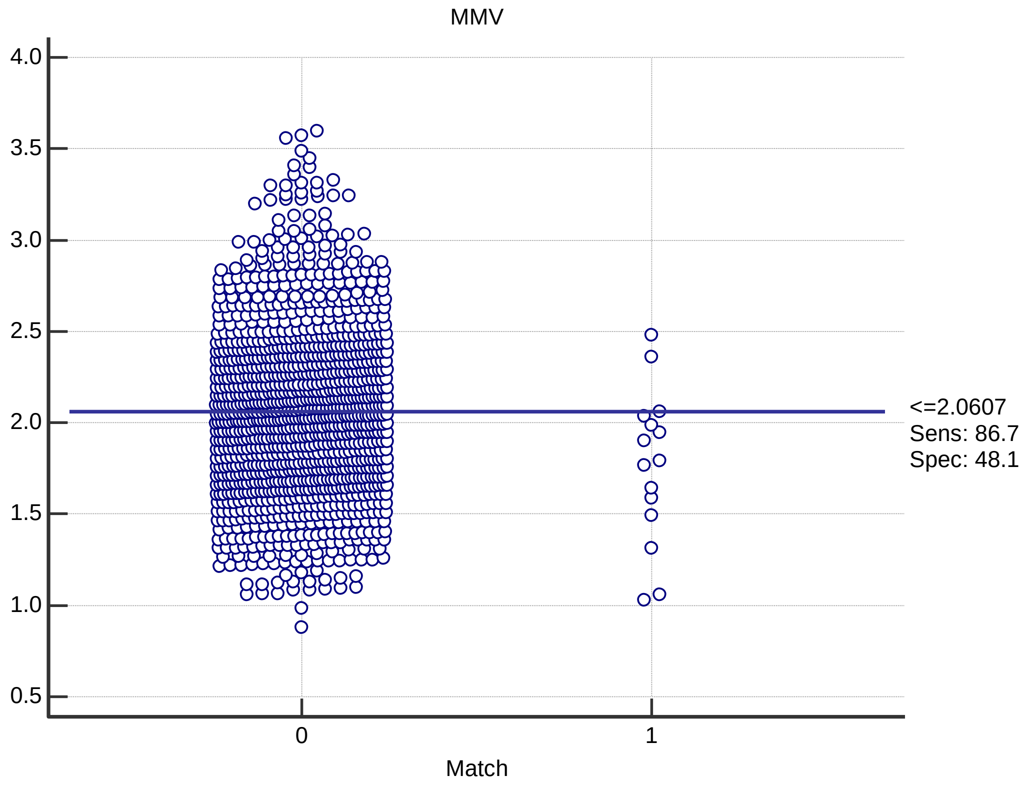


Figure 6.4 ROC Dot diagram for all opposite sex children, Eyes-Nose-Mouth-Chin with optimal threshold

6.3.1.2 Children of Same Sex

The category of children of same sex includes all the comparisons between children of the same sex, daughters to daughters and sons to sons, removing sex as a variant. In total there are 1,186 comparisons and 19 true matches indicating a sibling relationship of sister to sister or brother to brother. The mean age of the category is the same as with all children, 16 years with a total age span of seven years one month to 24 years as is the mean BMI, 21.39 which ranges from 14.67 to 30.52. The average age difference between siblings is four years two months, ranging from 0 years to 12 years nine months.

All the results for this category, shown in Table 6.10, are significant. The accuracy of sibling identification is shown to be best using the combination of Eyes and Nose and Mouth with AUC of 0.740. The optimal threshold, shown in Figure 6.5, is ≤3.3145mm and gives a sensitivity of 84.21% and specificity of 60.41%. In real terms this means 16 of the 19 sibling matches and 705 of the 1,167 non-matches are correctly predicted while three matches (above threshold line, right side) and 462 non-matches are incorrectly predicted.

The top result for a Rule In scenario when not wanting to misidentify any actual siblings is given by Eyes-Nose. At a threshold of ≤2.1781mm all 19 sibling matches are correctly predicted with a consequential specificity of 17.31 %. This correctly identifies 202 of the 1,167 non-siblings, therefore misclassifying 965 non-siblings as siblings. Overall this could reduce the total number of comparisons being considered as siblings from 1,186 to 984, a reduction of 17%.

Of the single features the Nose has the greatest accuracy shown by AUC, however the Whole Face outranks all the single features.

The joined features do not follow a clear pattern, although the largest portion does exhibit the highest AUC, the next highest is Eyes-Nose followed by Eyes-Nose-Mouth. The top two outperform the Whole Face.

All the combined features perform better than the Whole Face and with the exception of the Nose and Mouth they also outperform the joined features. The triple combination Eyes and Nose and Mouth performs best although does not have a good Rule In specificity.

Table 6.11 displays the results after removing the poorest ten scans, the AUC improved for all portions of the face with Eyes and Nose and Mouth staying as the highest portion and increasing to 0.797. The Rule In results increase for all the joined features, the combined features (with the exception of Eyes and Nose) and the Whole Face. There is a vast improvement for Eyes and Mouth from 4.03% to 48.60% specificity when sensitivity is set to 100%. The population, and thus number of comparisons, is reduced by the removal of the poorer quality scans. In the case of the Eyes and Mouth all 15 sibling matches are correctly predicted as positive and 433 of the 891 non-siblings are correctly categorised as negative. This could reduce the 906 possible comparisons by 48% to 473.

Number of comparisons: 1,186
Matches: 19

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.689	0.0089	63.16	71.21	15.85
Eyes	0.672	0.0057	89.47	39.50	9.34
Nose	0.681	0.0071	84.21	51.07	12.68
Mouth	0.640	0.0486	63.16	67.44	3.77
Eyes-Nose	0.695	0.0012	84.21	50.04	17.31
Eyes-Nose-Mouth	0.654	0.0189	94.74	32.13	5.48
Eyes-Nose-Mouth-Chin	0.697	0.0007	89.47	43.10	13.71
Eyes and Nose	0.713	0.0003	89.47	47.56	16.28
Eyes and Mouth	0.709	0.0004	89.47	53.64	4.03
Nose and Mouth	0.694	0.0039	63.16	73.01	13.02
Eyes and Nose and Mouth	0.740	<0.0001	84.21	60.41	7.46

Table 6.10 ROC results for same sex children.

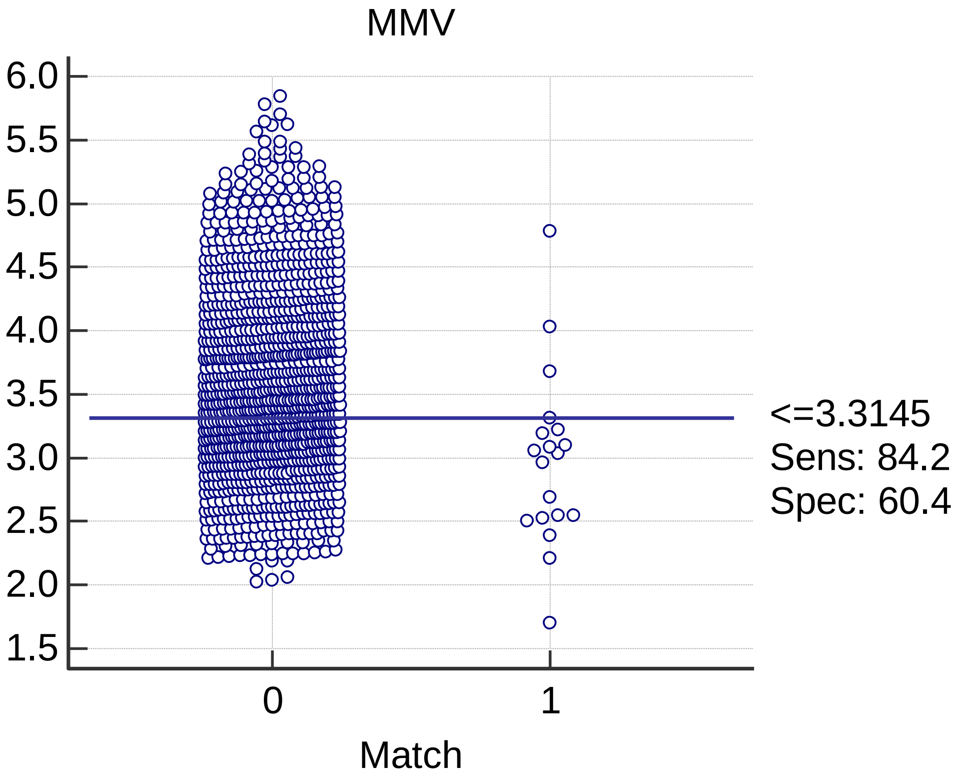


Figure 6.5 ROC Dot diagram for all same sex children, Eyes and Nose and Mouth combination with optimal threshold.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	1054	18	0.717	0.0022	66.67	71.14	17.86
Eyes	1008	18	0.677	0.0068	88.89	38.99	8.18
Nose	1017	16	0.717	0.0021	62.50	77.52	12.59
Mouth	1072	16	0.657	0.0375	68.75	64.68	2.84
Eyes-Nose	1017	16	0.739	0.0001	87.50	49.55	36.36
Eyes-Nose-Mouth	1081	16	0.723	0.0003	100.00	34.55	34.55
Eyes-Nose-Mouth-Chin	1081	16	0.754	<0.0001	62.50	78.40	33.52
Eyes and Nose	912	16	0.721	0.0016	43.75	91.29	13.28
Eyes and Mouth	906	15	0.747	<0.0001	100.00	48.60	48.60
Nose and Mouth	912	14	0.761	<0.0001	71.43	69.38	40.20
Eyes and Nose and Mouth	816	14	0.797	<0.0001	85.71	66.33	35.16

Table 6.11 ROC results for same sex children after removal of the ten poorest quality scans.

6.3.1.2.1 Children of Same Sex, Females Only

The category of children of same sex, females only, comprises all possible comparisons between female children. In total there are 406 comparisons and seven true matches indicating sisters. The mean age of the category is 16 years four months with a total age span of 11 years two months to 21 years three months. The mean BMI is 21.71, a healthy ratio, which ranges from 16.03 (underweight) to 26.95 (overweight). The average age difference between siblings is three years nine months, ranging from one year five months to seven years nine months.

The ROC analysis of the Nose and the Mouth and the combination of these two Nose and Mouth do not show significance but all other portions do, as depicted in Table 6.12. The greatest AUC of 0.857 is given by Eyes-Nose with an optimal threshold of ≤1.1196mm. In Figure 6.6 an optimal sensitivity of 85.71% and specificity of 88.22% give a threshold line which clearly lies below one outlying positive match. Six of the seven sister matches are correctly identified and 352 of the 399 non-sibling comparisons are correctly classified. Only one positive match and 47 negative matches are misclassified.

The best Rule In result for sisters is achieved by Eyes and Mouth. A threshold of ≤2.3842mm correctly identifies all seven sister matches with a resultant specificity of 35.84%; 143 of the 399 non-sister comparisons are correctly predicted as negative meaning that 256 are incorrectly predicted as positive. Overall this enables the reduction of the 406 total comparisons being considered as possible sisters to 263, a reduction of 35%.

The best single feature is the Eyes. The Mouth has a particularly low AUC, just above the guess value of 0.5 showing that is has very little predictive value in this category. The Whole Face is better than all of the single features.

The joined features seem to have more predictive value, the fewer the features which are included. The smallest, which does not incorporate the mouth area has the highest AUC and it is only this portion which outranks the Whole Face.

As with the joined features, the one combination which does not include the mouth area ranks highest for AUC. The Nose and Mouth is the lowest of the three which do include the mouth area and none of the combined features perform better than the Whole Face when AUC is considered.

Table 6.13 has the secondary results after the lower quality scans have been removed. The same portions of the face as before are not significant and overall there is very little change in the AUC values for each facial portion, however the minor changes create a reversal in the top spot with the Whole Face moving up a place. The Rule In specificities mostly improve but by a very small margin and the Eyes and Mouth still show the best result.

Number of comparisons: 406
Matches: 7

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.851	0.0048	85.71	87.97	11.03
Eyes	0.784	0.0060	71.43	79.95	24.06
Nose	0.720	***0.0951***	57.14	92.98	11.03
Mouth	0.519	***0.8634***	28.57	50.38	9.77
Eyes-Nose	0.857	0.0018	85.71	88.22	17.79
Eyes-Nose-Mouth	0.816	0.0014	71.43	83.46	28.32
Eyes-Nose-Mouth-Chin	0.747	0.0375	71.43	83.96	14.79
Eyes and Nose	0.822	0.0065	85.71	88.22	12.28
Eyes and Mouth	0.760	0.0014	85.71	68.17	35.84
Nose and Mouth	0.658	***0.1486***	57.14	72.18	25.56
Eyes and Nose and Mouth	0.801	0.0039	85.71	86.47	18.80

Table 6.12 ROC results for opposite sex children, females only.

Area	**No of comparisons**	**Matches**	**AUC**	**Significance P-value**	**Optimal Sensitivity (%)**	**Optimal Specificity (%)**	**Rule In (%)**
Whole Face	351	7	0.858	0.0036	85.71	89.83	12.50
Eyes	378	7	0.786	0.0052	71.43	80.32	24.80
Nose	351	7	0.716	***0.0994***	57.14	92.73	11.34
Mouth	406	7	0.519	***0.8634***	28.57	50.38	9.77
Eyes-Nose	351	7	0.856	0.0020	85.71	88.08	17.15
Eyes-Nose-Mouth	378	7	0.823	0.0009	71.43	84.64	29.65
Eyes-Nose-Mouth-Chin	378	7	0.753	0.0310	71.43	84.64	15.63
Eyes and Nose	351	7	0.817	0.0065	85.71	86.92	13.08
Eyes and Mouth	378	7	0.758	0.0015	85.71	67.65	36.12
Nose and Mouth	351	7	0.648	***0.1777***	57.14	70.93	24.71
Eyes and Nose and Mouth	351	7	0.797	0.0041	85.71	85.76	19.19

Table 6.13 ROC results for opposite sex children, females only, after removal of the ten poorest quality scans.

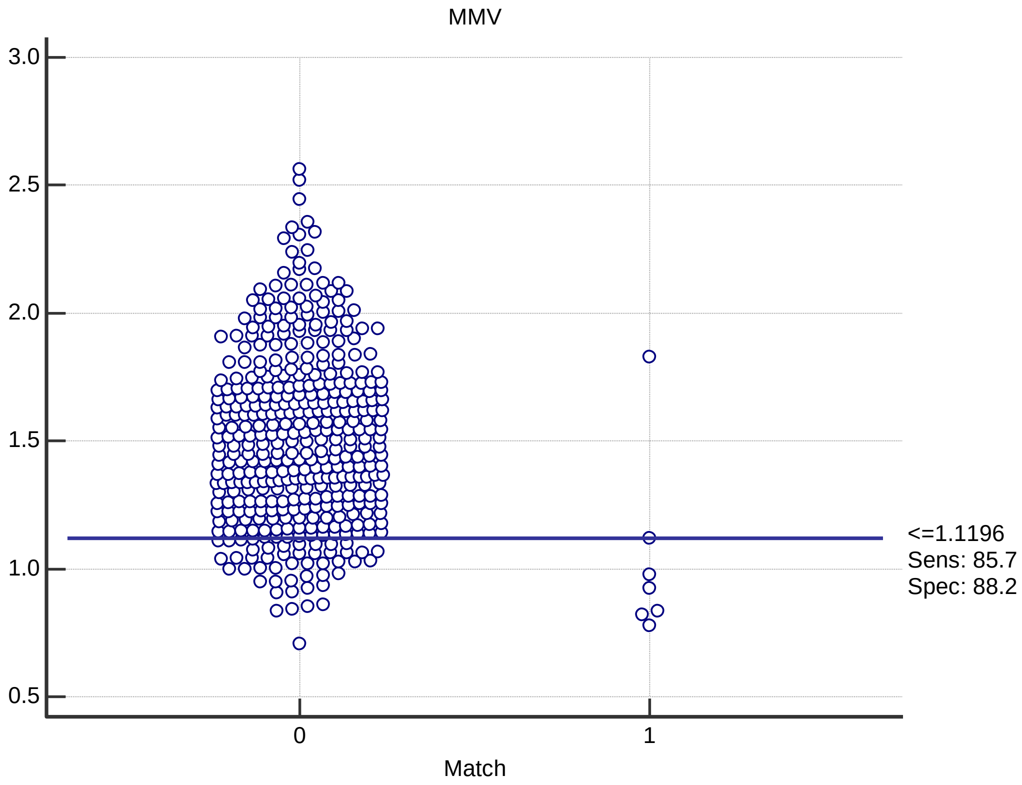


Figure 6.6 ROC Dot diagram for all female children, Eyes-Nose with optimal threshold.

6.3.1.2.2 Children of Same Sex, Males Only

The category of children of same sex, males only comprises all possible comparisons between male children. In total there are 780 comparisons and 12 true matches indicating brothers. The mean age of the category is 15 years ten months with a total age span of seven years one month to 24 years. The mean BMI is 21.16, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between siblings is four years five months, ranging from 0 years to 12 years, nine months.

Three of the 11 portions shown in Table 6.14 are not significant, the Eyes, Mouth and Eyes-Nose-Mouth. Across the whole relationship category, the portion that gives the best predictive values for identifying brothers is Eyes and Nose and Mouth. An AUC of 0.742 has optimal sensitivity of 83.33% and specificity of 71.87%. Figure 6.7 shows the threshold associated with the optimal values, ≤3.3145mm allows for ten of the 12 brotherly relationships and 552 of the 768 non-brother comparisons to be correctly predicted. This leaves the two outlying brother pairs, shown in the top right of Figure 6.8 wrongly classed as negative and 216 unrelated pairs incorrectly predicted as brothers.

If applying a Rule In command the Eyes-Nose portion gives the best result. A specificity of 24.87% at a threshold of ≤2.1781mm correctly predicts 191 of the 768 unrelated pairs and all 12 of the brother pairs whilst incorrectly predicting 577 unrelated pairs as related. This could reduce the total number of possible pairs from 780 to 589 reduction of 24.5%.

Of the single features the Nose was the only significant result and on its own was a better predictor than the Whole Face

The joined features with significant results increased in accuracy as they increased in size, the four feature portion showing the highest value and outranking the best single feature and the Whole Face.

The AUC values for the two feature combinations are all significant and very similar to each other. They all perform better than the Whole Face and the single features. The triple combination Eyes and Nose and Mouth tops the list.

The effect of scan quality is shown in table 6.15. The AUC values all increase after the removal of the worst scans with one exception, Eyes and Nose decreases but only by 0.011. The highest AUC value changes from Eyes and Nose and Mouth to Nose and Mouth although there is only 0.002 between them. Eyes-Nose-Mouth becomes significant and six of the eleven facial portions show a dramatic increase in Rule In specificity whilst the other five only show slight variation. The Eyes and Nose and Mouth combination, although resulting from a greatly reduced number of comparisons, had a Rule In specificity of 66.38%. In real terms this allows for the correct prediction of the seven brother pairs and 304 of the 458 unrelated pairs whilst incorrectly predicting 154 unrelated pairs as brothers enabling a reduction of the total number of comparisons by 65%.

Number of comparisons: 780
Matches: 12

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.663	0.0461	50.00	82.68	22.79
Eyes	0.628	***0.0692***	83.33	51.95	13.41
Nose	0.687	0.0216	83.33	59.64	14.97
Mouth	0.674	***0.0840***	58.33	82.29	5.34
Eyes-Nose	0.654	0.0055	83.33	59.38	24.87
Eyes-Nose-Mouth	0.607	***0.1251***	91.67	43.36	8.33
Eyes-Nose-Mouth-Chin	0.713	0.0009	83.33	61.59	20.57
Eyes and Nose	0.693	0.0010	91.67	57.03	22.14
Eyes and Mouth	0.697	0.0122	83.33	64.97	6.12
Nose and Mouth	0.707	0.0245	58.33	85.16	18.36
Eyes and Nose and Mouth	0.742	0.0009	83.33	71.87	11.07

Table 6.14 ROC results for opposite sex children, males only.

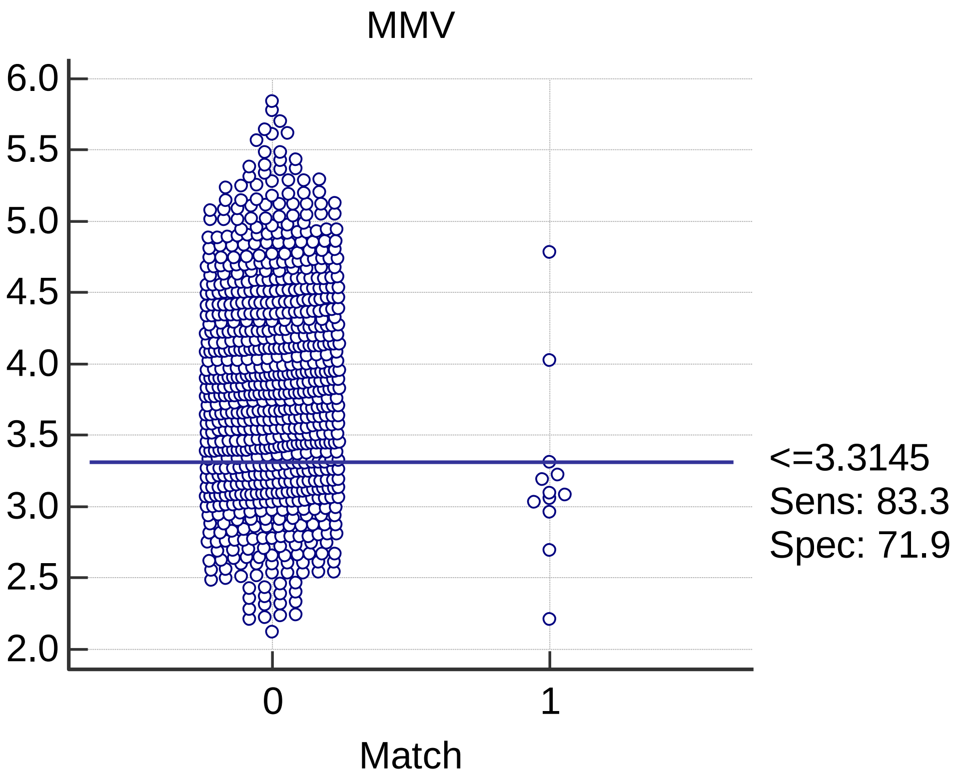


Figure 6.7 ROC Dot diagram for all male children, Eyes and Nose and Mouth combination with optimal threshold.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	703	11	0.687	0.0217	54.55	81.21	23.70
Eyes	630	11	0.630	***0.0905***	81.82	52.02	12.12
Nose	666	9	0.735	0.0072	88.89	57.23	14.61
Mouth	666	9	0.711	***0.0678***	66.67	80.06	4.11
Eyes-Nose	666	9	0.692	0.0005	100.00	47.79	47.79
Eyes-Nose-Mouth	703	9	0.681	0.0039	100.00	44.81	44.81
Eyes-Nose-Mouth-Chin	703	9	0.778	<0.0001	88.89	66.28	47.12
Eyes and Nose	561	9	0.682	0.0255	88.89	53.99	18.12
Eyes and Mouth	528	8	0.751	<0.0001	100.00	60.38	60.38
Nose and Mouth	561	7	0.830	<0.0001	71.43	82.13	51.99
Eyes and Nose and Mouth	465	7	0.828	<0.0001	100.00	66.38	66.38

Table 6.15 ROC results for opposite sex children, males only, after removal of the ten poorest quality scans.

6.3.2 Mothers to Children

The category of mothers to children contains all possible comparisons between mothers and all children of both sexes. In total there are 2,829 comparisons and 69 true matches indicating a maternal relationship. The mean age of the mothers is 48 years five months ranging from 39 years nine months to 59 years three months. The mean age of the children is 16 years with a total age span of seven years one month to 24 years. The mean BMI of the mothers is 26.18 (overweight), ranging from 16.95 (underweight) to 45.18 (obese). The mean BMI of the children is 21.39, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between mothers and their children is 32 years one month, ranging from 22 years five months to 42 years ten months.

Table 6.16 describes the ROC results for all the mother to child pairs. The significance of all the results is very high with a slight variation for the Mouth. An AUC of 0.726 achieved by the combination of Eyes and Nose and Mouth gives rise to an optimal sensitivity of 72.46% and specificity of 64.86%. As seen in Figure 6.8 a threshold of ≤3.5055mm enables the accurate prediction of 50 of the 69 mother-child pairs and 1,790 of the 2,760 unrelated pairs. This means that 19 of the mother-child pairs are misclassified as unrelated and 970 of the unrelated pairs are predicted as related.

If the correct classification of related pairs is required and a Rule In option is applied, then the Eyes give the best result. All 69 of the mother-child pairs are correctly classified as positive when a threshold of ≤1.8977mm is applied. This results in a specificity of 25.36% meaning that 700 of the 2,760 unrelated pairs are correctly classed as unrelated and 2,060 are wrongly predicted as related. This could reduce the total number of pair comparisons from 2829 to 2129 a reduction of 24.7%

The Mouth is the poorest single feature at predicting a maternal relationship while the Eyes are the best. The Whole Face however is superior to all the single features.

The joined features are also all inferior to the Whole Face, the addition of the chin seems to adversely affect the AUC as Eyes-Nose-Mouth-Chin has the lowest value while Eyes-Nose-Mouth has the greatest.

The combined features, with the exception of Nose and Mouth, outrank all the joined and single features as well as the Whole Face with the triple combination exhibiting the best results.

The results following the removal of the poorer scans are shown in Table 6.17 and it can be seen that the result for the Mouth becomes non-significant. Overall there are slight improvements in AUC across the various portions of the face with a few decreasing but only very slightly. The Rule In specificities also show little change, the only portion which shows notable change is the increase in specificity by Eyes and Nose and Mouth. The highest ranked facial portions for both AUC and Rule In specificity remain the same as they were before the removal of the poorest scans.

Number of comparisons: 2,829
Matches: 69

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.704	<0.0001	69.57	61.52	14.38
Eyes	0.686	<0.0001	84.06	47.68	25.36
Nose	0.638	0.0001	73.91	51.05	2.03
Mouth	0.582	0.0183	73.91	43.04	2.97
Eyes-Nose	0.689	<0.0001	76.81	52.79	9.96
Eyes-Nose-Mouth	0.694	<0.0001	72.46	59.82	18.59
Eyes-Nose-Mouth-Chin	0.665	<0.0001	52.17	77.03	5.36
Eyes and Nose	0.714	<0.0001	69.57	64.38	12.43
Eyes and Mouth	0.709	<0.0001	76.81	60.83	16.23
Nose and Mouth	0.643	<0.0001	60.87	62.07	1.12
Eyes and Nose and Mouth	0.726	<0.0001	72.46	64.86	6.38

Table 6.16 ROC results for mothers compared to children.

Area	**No of comparisons**	**Matches**	**AUC**	**Significance P-value**	**Optimal Sensitivity (%)**	**Optimal Specificity (%)**	**Rule In (%)**
Whole Face	2665	65	0.718	<0.0001	70.77	61.58	13.58
Eyes	2560	63	0.683	<0.0001	88.89	42.37	24.23
Nose	2432	61	0.650	0.0001	60.66	65.96	3.25
Mouth	2640	64	0.563	***0.0981***	46.88	67.51	2.17
Eyes-Nose	2496	61	0.715	<0.0001	81.97	53.22	9.82
Eyes-Nose-Mouth	2706	66	0.701	<0.0001	74.24	59.28	17.73
Eyes-Nose-Mouth-Chin	2706	66	0.672	<0.0001	54.55	76.97	5.57
Eyes and Nose	2257	58	0.735	<0.0001	74.14	64.17	16.05
Eyes and Mouth	2379	58	0.711	<0.0001	79.31	57.48	14.13
Nose and Mouth	2257	56	0.631	0.0004	94.64	26.12	3.36
Eyes and Nose and Mouth	2088	53	0.739	<0.0001	79.25	60.54	20.93

Table 6.17 ROC results for mothers compared to children after removal of the ten poorest quality scans.

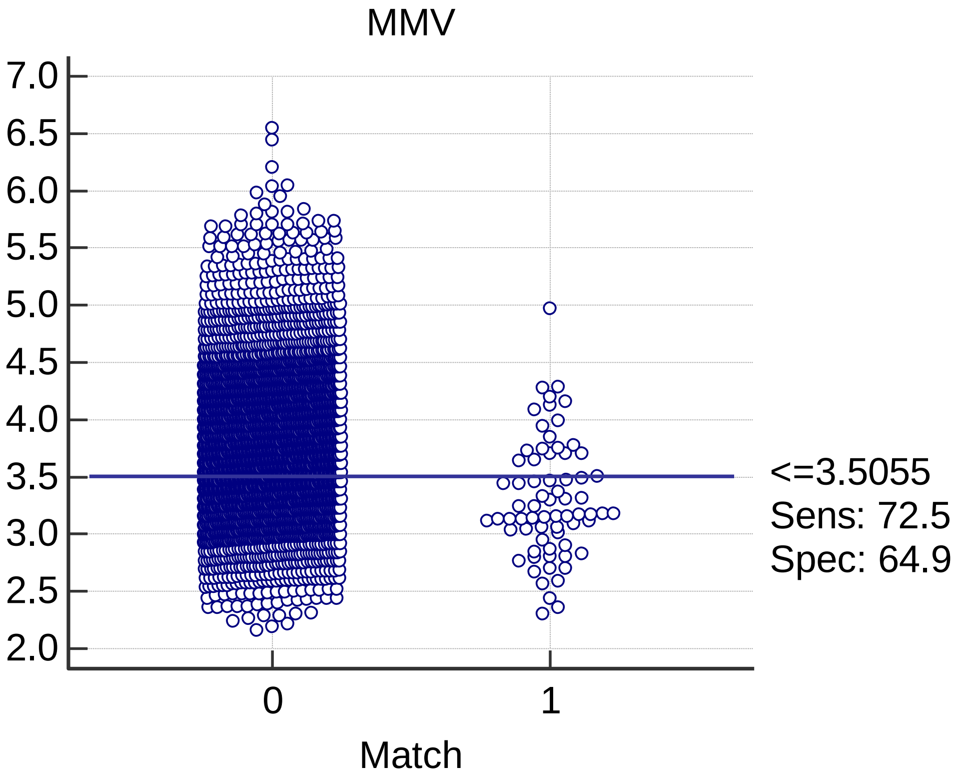


Figure 6.8 ROC Dot diagram for mothers to all children, Eyes and Nose and Mouth combination with optimal threshold.

6.3.2.1 Mothers to Female Children

The category of mothers to female children contains all possible comparisons between mothers and female children only. In total there are 1,189 comparisons and 29 true matches indicating a mother to daughter relationship. The mean age of the mothers is 48 years five months ranging from 39 years nine months to 59 years three months. The mean age of the female children is 16 years four months with a total age span of 11 years two months to 21 years three months. The mean BMI of the mothers is 26.18 (overweight), ranging from 16.95 (underweight) to 45.18 (obese). The mean BMI of the female children is 21.71, a healthy ratio, which ranges from 16.03 (underweight) to 26.95 (obese). The average age difference between mothers and their daughters is 31 years nine months, ranging from 23 years two months to 40 years six months.

The ROC results of the mother-daughter comparisons are shown is Table 6.18. The result of the Mouth is particularly interesting, highlighted in yellow. Although not significant, the threshold is actually flipped so that the lower Mesh-to-Mesh Values are actually more indicative of an unrelated mother and daughter pair rather than a related one as might have been expected. This unusual result is only just over the 0.5 value which is comparable to guesswork. Other than the Mouth the only other non-significant result is that of the combined Nose and Mouth although it is only just over the significance threshold. The Whole Face has the highest AUC of 0.734. A threshold of ≤2.7242mm gives optimal sensitivity of 82.76% and specificity of 60.0%. In real numbers 24 of the 29 actual mother-daughter pairs are correctly predicted while five are misclassified (shown above the threshold line on the right side of Figure 6.9). The unrelated comparison pairs are correctly predicted in 696 of the 1,160 cases, misclassifying 464 pairs.

It is the combination of the Eyes and Mouth that produces the best Rule In specificity of 28.79% although Eyes-Nose-Mouth-Chin comes in a close second. The threshold ≤2.8022mm allows the correct prediction of all 29 actual mother-daughter pairs whilst also accurately classifying 334 of the 1,160 unrelated pairs, however this means that 826 unrelated pairs are wrongly classified as related. The total number of comparisons considered as possible matches can be reduced by 334 from 1,189 to 855 a reduction of 28%.

The better single feature is the Nose according to AUC however it is clearly outperformed by the Whole Face.

The largest of the joined feature portions does have the greatest AUC but the order below does not follow the size of the portion. As it is the top result the Whole Face obviously shows a greater surface area than all the joined features.

With respect to the combined features the mouth area seems to adversely affect the results. The one combination which does not include the mouth, the Eyes and Nose has the greatest value of AUC followed by the triple combination where the influence of the mouth will be more dispersed. Again the Whole Face is the best predictor for mother to daughter relationships.

Table 6.19 describes the secondary results after the removal of the ten poorest scans for each facial portion. The classification of significance is unchanged and the Mouth continues to predict more poorly than guesswork. The Whole Face still ranks highest for the value of AUC and overall there is very little change from the previous results. Most of the Rule In specificities are increased with the largest increases seen for the ‘Nose and Eyes and Nose and Mouth. The top two swap position and the Eyes-Nose-Mouth-Chin takes the top positon.

Number of comparisons: 1,189
Matches: 29

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.734	<0.0001	82.76	60.00	8.36
Eyes	0.661	0.0002	89.66	40.17	18.71
Nose	0.671	0.0011	65.52	65.60	1.03
Mouth	0.503	***0.9512***	48.28	59.66	0.43
Eyes-Nose	0.684	0.0002	62.07	71.21	16.81
Eyes-Nose-Mouth	0.674	0.0002	58.62	70.00	19.83
Eyes-Nose-Mouth-Chin	0.698	<0.0001	75.86	58.28	28.36
Eyes and Nose	0.722	<0.0001	89.66	56.03	6.90
Eyes and Mouth	0.652	<0.0001	86.21	48.45	28.79
Nose and Mouth	0.602	***0.0517***	51.72	69.83	6.90
Eyes and Nose and Mouth	0.703	<0.0001	75.86	62.76	12.93

Table 6.18 ROC results for mothers compared to female children.

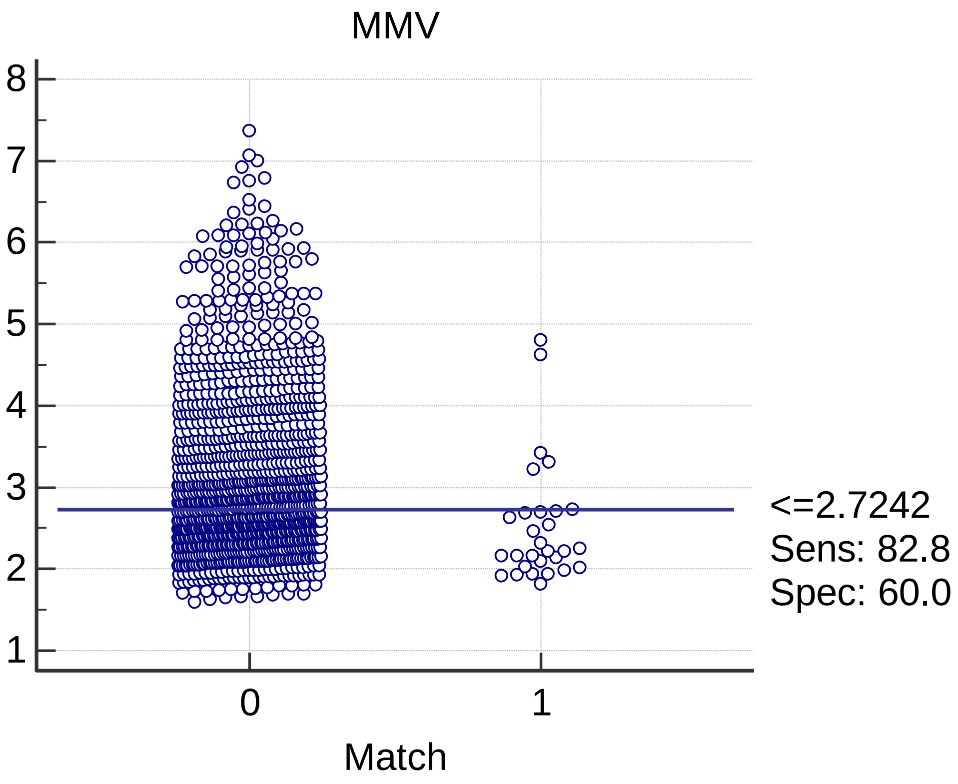


Figure 6.9 ROC Dot diagram for mothers to female children, Whole Face with optimal threshold.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	1107	27	0.760	<0.0001	85.19	61.11	8.61
Eyes	1120	28	0.659	0.0005	89.29	39.74	18.13
Nose	1026	26	0.675	0.0008	65.38	64.60	12.60
Mouth	1160	29	0.511	***0.8406***	48.28	60.74	0.44
Eyes-Nose	1053	25	0.707	<0.0001	64.00	71.69	24.22
Eyes-Nose-Mouth	1148	28	0.693	<0.0001	60.71	70.54	26.34
Eyes-Nose-Mouth-Chin	1148	28	0.716	<0.0001	78.57	59.20	28.57
Eyes and Nose	999	26	0.737	<0.0001	92.31	54.68	10.59
Eyes and Mouth	1092	28	0.646	0.0002	85.71	46.62	27.63
Nose and Mouth	999	26	0.596	***0.0857***	53.85	68.24	14.70
Eyes and Nose and Mouth	972	26	0.709	<0.0001	76.92	60.04	22.62

Table 6.19 ROC results for mothers compared to female children after removal of the ten poorest quality scans.

6.3.2.2 Mothers to Male Children

The category of mothers to male children contains all possible comparisons between mothers and male children only. In total there are 1,640 comparisons and 40 true matches indicating a mother to son relationship. The mean age of the mothers is 48 years five months ranging from 39 years nine months to 59 years three months. The mean age of the male children is 15 years ten months with a total age span of seven years one month to 24 years. The mean BMI of the mothers is 26.18 (overweight), ranging from 16.95 (underweight) to 45.18 (obese). The mean BMI of the male children is 21.16, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between mothers and their sons is 32 years four months, ranging from 22 years five months to 42 years ten months.

The results for the mother-son relationships are shown in Table 6.20, and all facial portions have significant results. The accuracy shown by AUC is best for the triple combination of Eyes and Nose and Mouth with an AUC of 0.753 and optimal sensitivity of 80.0% and specificity of 60.44%. The optimal threshold as shown in Figure 6.10, is ≤3.7776mm which enables 32 of the 40 mother-son pairs and 967 of the 1,160 unrelated pairs to be correctly predicted. This leaves eight actual related pairs and 193 unrelated pairs mistakenly predicted.

The Eyes allow for the best prediction of possible pairs by having a Rule In specificity of 31.94%. The given threshold of ≤1.8627 leads to 511 of the 1,600 unrelated pairs being correctly predicted alongside all 40 actual related pairs. 1,089 pairs are wrongly predicted as related when they are not, however this reduces the overall number of comparisons from 1,640 to 1,129 a reduction of 31%.

The Eyes show better results than any other single area of the face for both AUC and Rule In specificity. The Eyes are the only areas of the face that can produce better results than those obtained from the Whole Face.

As with the mothers to children category the chin seems to impact negatively upon the joined features with Eyes-Nose-Mouth topping the list followed by Eyes-Nose, both of which outperform the Whole Face.

The triple combinations yet again perform better than the doubles and the influence of the eyes is shown by the lower value for Nose and Mouth. It is this lower value alone that falls below than of the Whole Face and the joined features.

Table 6.21 contains the ‘improved’ results having reduced the number of scans by the removal of the poorest in quality. The results from the AUC are little changed with some facial portions increasing and other decreasing in value, but all by only marginal amounts. The triple combination remains the best result with a small increase to 0.769 (AUC). It is the same story for the Rule In specificity, the only vast change seen is for the Eyes and Nose and Mouth combination with an increase of just under 20%. The best performing area in this set of results remain unchanged, being the Eyes.

Number of comparisons: 1,640
Matches: 40

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.699	<0.0001	80.00	52.19	24.63
Eyes	0.707	<0.0001	77.50	56.56	31.94
Nose	0.623	0.0095	65.00	59.94	5.94
Mouth	0.639	0.0033	72.50	52.50	4.56
Eyes-Nose	0.701	<0.0001	65.00	71.19	13.19
Eyes-Nose-Mouth	0.708	<0.0001	70.00	68.87	24.06
Eyes-Nose-Mouth-Chin	0.643	0.0039	45.00	83.56	8.38
Eyes and Nose	0.724	<0.0001	82.50	56.69	22.87
Eyes and Mouth	0.746	<0.0001	87.50	52.94	21.62
Nose and Mouth	0.673	0.0001	60.00	71.37	1.81
Eyes and Nose and Mouth	0.753	<0.0001	80.00	60.44	9.63

Table 6.20 ROC results for mothers compared to male children.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	1558	38	0.702	<0.0001	81.58	51.12	22.89
Eyes	1440	35	0.703	<0.0001	85.71	48.40	30.75
Nose	1406	35	0.635	0.0094	68.57	58.13	4.38
Mouth	1480	35	0.611	0.0391	45.71	79.31	3.25
Eyes-Nose	1443	36	0.733	<0.0001	72.22	71.93	12.94
Eyes-Nose-Mouth	1558	38	0.707	<0.0001	71.05	67.76	22.70
Eyes-Nose-Mouth-Chin	1558	38	0.642	0.0061	47.37	83.03	8.75
Eyes and Nose	1258	32	0.747	<0.0001	84.37	55.87	27.49
Eyes and Mouth	1287	30	0.759	<0.0001	86.67	56.09	18.70
Nose and Mouth	1258	30	0.654	0.003	63.33	66.45	5.37
Eyes and Nose and Mouth	1116	27	0.769	<0.0001	85.19	60.15	29.02

Table 6.21 ROC results for mothers compared to male children after removal of the ten poorest quality scans.

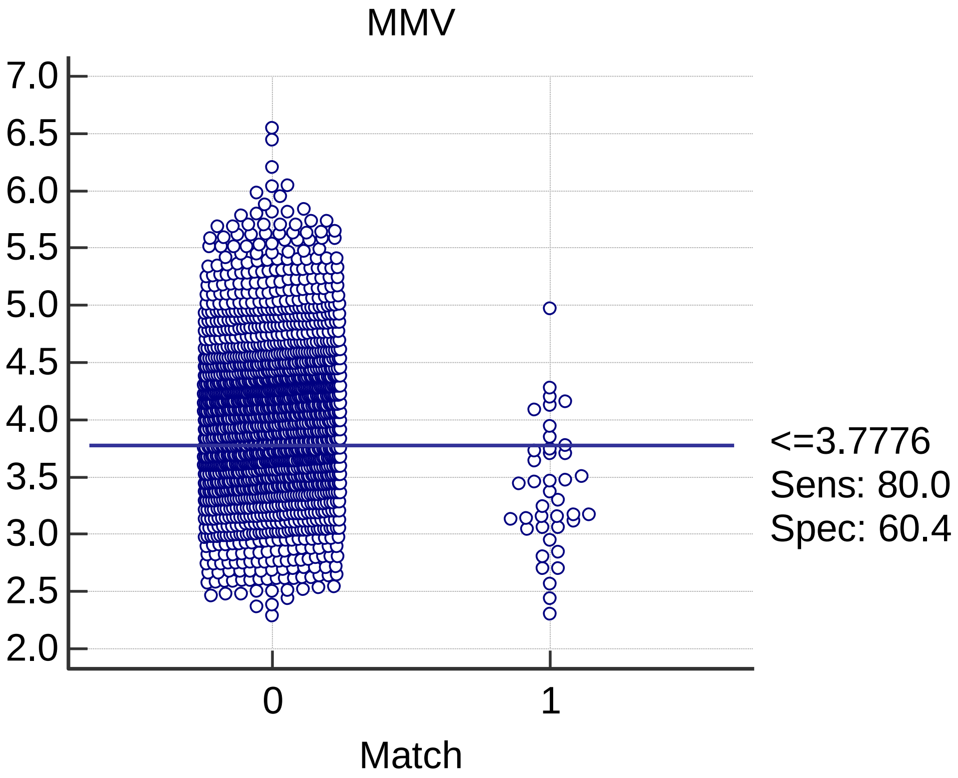


Figure 6.10 ROC Dot diagram for mothers to male children, Eyes and Nose and Mouth combination with optimal threshold.

6.3.3 Fathers to Children

The category of fathers to children contains all possible comparisons between fathers and all children of both sexes. In total there are 2,001 comparisons and 53 true matches indicating a paternal relationship. The mean age of the fathers is 49 years ten months ranging from 41 years one month to 62 years five months. The mean age of the children is 16 years with a total age span of seven years one month to 24 years. The mean BMI of the fathers is 28.73 (overweight), ranging from 24.15 (healthy) to 37.32 (obese). The mean BMI of the children is 21.39, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between fathers and their children is 33 years 11 months, ranging from 25 years four months to 48 years seven months.

The ROC results for the comparison of fathers to all children are shown in Table 6.22. All portions of the face, with the exception of the Nose have significant results and the Nose only just falls over the threshold. Overall the AUC values and Rule In specificity percentages are low compared to other relationship categories. Within this category the triple combination Eyes and Nose and Mouth tops the AUC values at 0.677 giving optimal sensitivity and specificity of 67.92% and 61.91% respectively. Figure 6.11 shows the threshold given the optimal performance as ≤3.9189mm, which in real terms means that 36 of the 53 actual father to child relationships are correctly predicted and 1,206 of the 1,948 of the unrelated pairs are classified correctly as well. This just leaves 17 related pairs and 742 unrelated pairs wrongly predicted.

The Rule In percentages are very low, the best is shown by the Whole Face with a specificity of 7.96%. With a threshold of ≤6.3456mm all 53 father-child pairs are identified but only 150 of the 1,948 unrelated pairs are predicted correctly, misclassifying 1,798 unrelated pairs as related. Overall the reduction of possible comparative pairs is only by 7.5% to 1,851 comparisons.

Ignoring the non-significant Nose, the accuracy of the Mouth is slightly better than the Eyes and both have higher values than the Whole Face.

The largest of the joined feature portions does rank the highest among them, although the other two do not follow the same pattern with the Eyes-Nose coming second. They all have greater AUC than the Whole Face and the single features.

The values of the combined features are little different to the joined features with the largest again taking the top spot for both AUC and Rule In specificity. As with the joined features, they all have greater AUC than the single features and the Whole Face.

Table 6.23 has the results after the poorer scan have been removed and one thing to note is that the Eyes also show a non-significant result alongside the Nose, though both are still just under the 95%. There is little overall change in AUC, approximately half the portions increase while half decrease, and the top positon is acquired by Eyes-Nose-Mouth-Chin’. The Rule In specificities do show a little more variation, whilst they do not all increase, those that do tend to increase by a reasonable proportion; the Eyes and Mouth which improves by the largest proportion but the Nose and Mouth has the best result overall.

Number of comparisons: 2,001
Matches: 53

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.608	0.0056	32.08	84.75	7.96
Eyes	0.633	0.0008	66.04	62.27	6.47
Nose	0.572	***0.0512***	66.04	48.31	1.33
Mouth	0.636	0.0003	69.81	53.95	2.98
Eyes-Nose	0.669	<0.0001	71.70	57.75	7.19
Eyes-Nose-Mouth	0.656	<0.0001	81.13	48.97	5.90
Eyes-Nose-Mouth-Chin	0.675	<0.0001	79.25	55.44	4.26
Eyes and Nose	0.644	0.0001	62.26	60.47	4.16
Eyes and Mouth	0.671	<0.0001	56.60	77.05	2.52
Nose and Mouth	0.644	<0.0001	73.58	52.87	4.67
Eyes and Nose and Mouth	0.677	<0.0001	67.92	61.91	8.11

Table 6.22 ROC results for fathers compared to children.

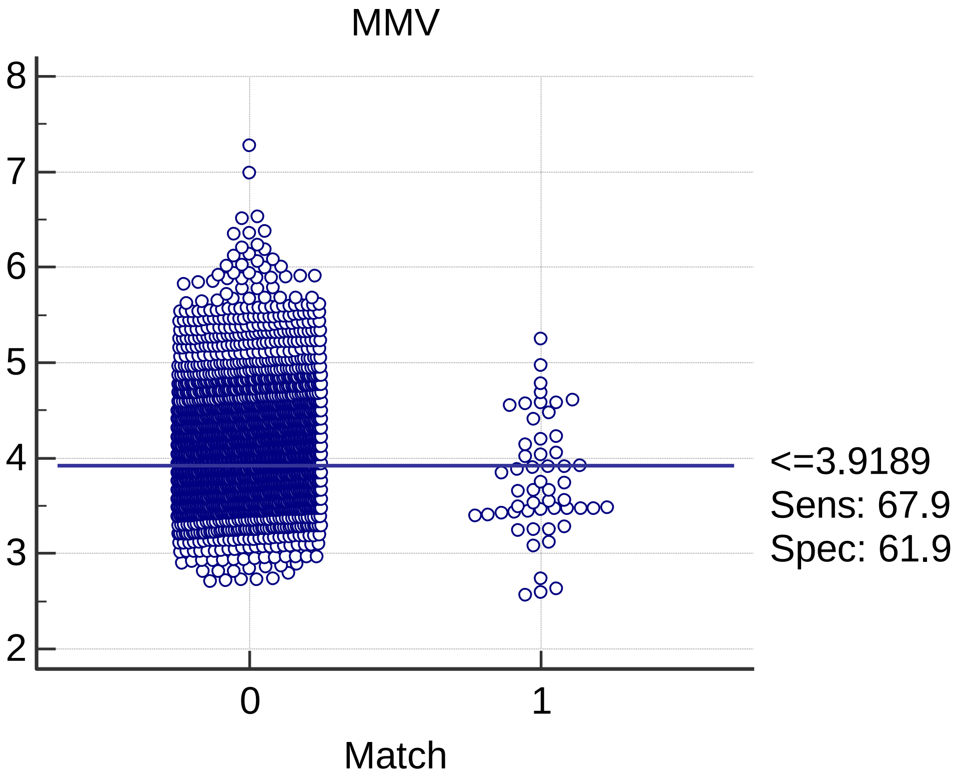


Figure 6.11 ROC Dot diagram for fathers to all children, Eyes and Nose and Mouth combination with optimal threshold.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	1495	40	0.594	0.031	67.50	48.66	15.05
Eyes	1600	42	0.594	***0.052***	52.38	69.45	5.91
Nose	1728	45	0.578	***0.054***	68.89	46.64	10.04
Mouth	1518	41	0.644	0.000	56.10	66.62	6.91
Eyes-Nose	1664	43	0.649	0.000	69.77	58.85	5.86
Eyes-Nose-Mouth	1452	39	0.664	<0.0001	84.62	47.91	6.65
Eyes-Nose-Mouth-Chin	1452	39	0.693	<0.0001	79.49	59.66	5.45
Eyes and Nose	1464	39	0.610	0.019	48.72	72.77	2.32
Eyes and Mouth	1220	33	0.641	0.010	57.58	73.55	15.25
Nose and Mouth	1281	36	0.679	<0.0001	72.22	60.80	15.82
Eyes and Nose and Mouth	1102	32	0.650	0.004	50.00	79.72	8.22

Table 6.23 ROC results for fathers compared to children after removal of the ten poorest quality scans.

6.3.3.1 Fathers to Female Children

The category of fathers to female children contains all possible comparisons between fathers and female children. In total there are 841 comparisons and 23 true matches indicating a father to daughter relationship. The mean age of the fathers is 49 years ten months ranging from 41 years one month to 62 years five months. The mean age of the female children is 16 years four months with a total age span of 11 years two months to 21 years three months. The mean BMI of the fathers is 28.73 (overweight), ranging from 24.15 (healthy) to 37.32 (obese). The mean BMI of the female children is 21.71, a healthy ratio, which ranges from 16.03 (underweight) to 26.95 (obese). The average age difference between fathers and their daughters is 33 years, six months, ranging from 27 years 11 months to 45 years 11 months.

The first thing of note for the results in this category, shown in Table 6.24, is the large number of results which are not significant. The results from the Whole Face, all of the single features and all of the combined feature facial portions are non-significant to quite a high degree. The only result close to the < 0.05 boundary is that of the combination area Eyes and Nose and Mouth. So it is only the three joined feature areas that show significance and of those, the Eyes-Nose-Mouth has the greatest AUC. The area of 0.667 gives an optimal sensitivity of 82.61% and specificity of 53.67% at a threshold of ≤2.131mm, shown in Figure 6.12. This result correctly predicts 19 of the 23 actual father-daughter relationship pairs while incorrectly identifying four pairs. It also rightly predicts 439 of the 818 unrelated pairs misclassifying 379 of them.

It is Eyes-Nose which has the greatest specificity when it comes to ruling in the actual positive matches. A specificity of 23.84% is given by a threshold set at ≤2.2174mm. So whilst ensuring that all true matches are correctly identified a total of 195 of the 818 are also correctly predicted meaning that 623 are not. Overall this reduces the pairs being considered as possible matches from 841 to 646 which is a reduction of 23%.

Due to the oddity of the significance of this relationship category, relevant comparisons between the types of facial features are not possible. All that can be said is that the chin appears to have negatively influenced the results of the joined features, while the mouth was a positive addition.

Table 6.25 contains the ROC results after the removal of the poorest scans and an improvement can be seen immediately in the number of results which are no longer non-significant. Only the Whole Face, Eyes and Eyes and Nose retain a non-significant result. Of the significant results the Eyes-Nose is the only portion not to increase the AUC, the largest increase is shown by Nose and Mouth combination and it is this portion that tops the ranking. The Rule In specificities either change little or increase dramatically, although the Nose and Mouth portion also showed a dramatic increase in this value, it is the Mouth which has the largest increase from 1.47% to 29.78% again topping the list.

Number of comparisons: 841
Matches: 23

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.579	***0.2047***	39.13	81.05	9.90
Eyes	0.609	***0.0997***	69.57	60.39	5.87
Nose	0.587	***0.1297***	69.57	51.34	0.86
Mouth	0.565	***0.2793***	69.57	47.80	1.47
Eyes-Nose	0.650	0.0023	65.22	64.55	23.84
Eyes-Nose-Mouth	0.667	0.0002	82.61	53.67	12.35
Eyes-Nose-Mouth-Chin	0.663	0.0006	82.61	51.83	18.70
Eyes and Nose	0.620	***0.0632***	52.17	75.55	4.40
Eyes and Mouth	0.622	***0.0735***	65.22	70.42	1.10
Nose and Mouth	0.586	***0.1225***	73.91	46.45	2.57
Eyes and Nose and Mouth	0.626	***0.0544***	52.17	80.68	6.36

Table 6.24 ROC results for fathers compared to female children.

Area	No of comparisons	Matches	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	621	17	0.593	***0.1758***	41.18	80.13	19.87
Eyes	700	19	0.584	***0.2847***	68.42	59.03	5.87
Nose	729	19	0.626	0.0323	78.95	49.01	12.54
Mouth	667	19	0.654	0.0011	84.21	45.99	29.78
Eyes-Nose	702	19	0.644	0.0082	63.16	64.13	22.84
Eyes-Nose-Mouth	616	17	0.685	0.0005	88.24	52.09	14.86
Eyes-Nose-Mouth-Chin	616	17	0.681	0.0017	76.47	60.60	22.70
Eyes and Nose	648	18	0.643	***0.0637***	61.11	73.97	3.17
Eyes and Mouth	560	15	0.673	0.0346	73.33	72.29	14.13
Nose and Mouth	567	16	0.696	0.0001	81.25	57.17	27.59
Eyes and Nose and Mouth	513	15	0.688	0.0211	66.67	80.12	9.04

Table 6.25 ROC results for fathers compared to female children after removal of the ten poorest quality scans.

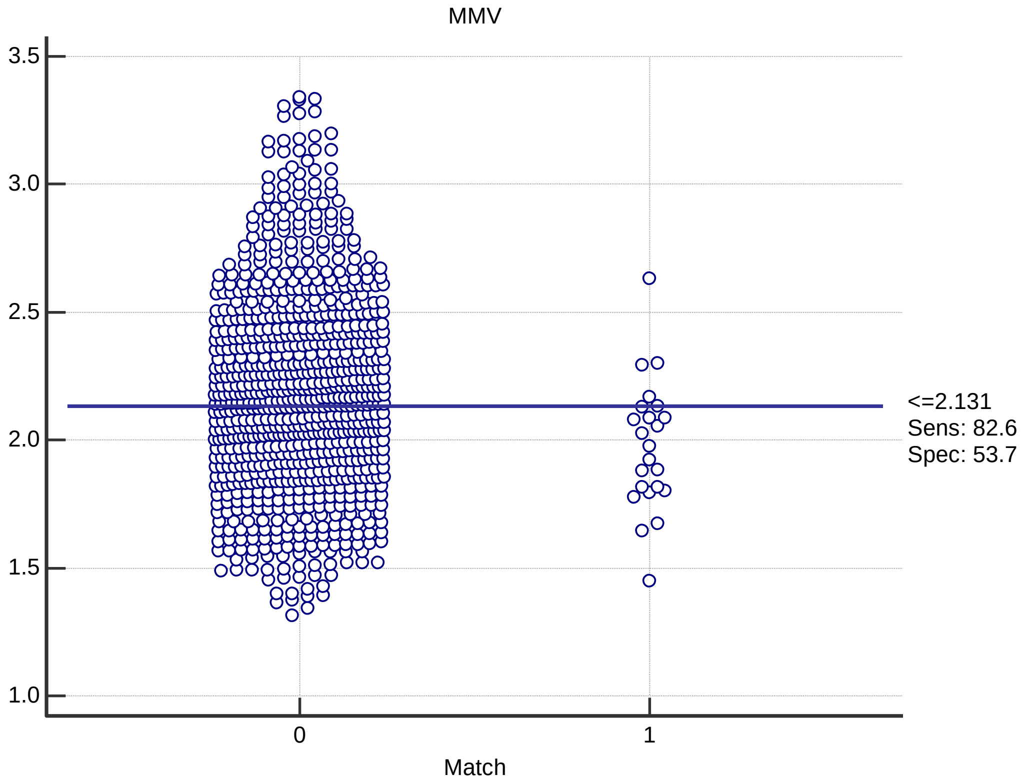


Figure 6.12 ROC Dot diagram for fathers to female children, Eyes-Nose-Mouth with optimal threshold.

6.3.3.2 Fathers to Male Children

The category of fathers to male children contains all possible comparisons between fathers and male children. In total there are 1,160 comparisons and 30 true matches indicating a father to son relationship. The mean age of the fathers is 49 years ten months ranging from 41 years one month to 62 years five months. The mean age of the male children is 15 years 10 months with a total age span of seven years one month to 24 years. The mean BMI of the fathers is 28.73 (overweight), ranging from 24.15 (healthy) to 37.32 (obese). The mean BMI of the male children is 21.16, a healthy ratio, which ranges from 14.67 (underweight) to 30.52 (obese). The average age difference between fathers and their son is 34 years three months, ranging from 25 years four months to 48 years seven months.

Only the Nose has a ROC result which is not significant in this relationship category of fathers to male children, shown in Table 6.26. The triple combination of Eyes and Nose and Mouth has the greatest AUC of 0.717 for which a threshold of ≤4.0528mm gives optimal sensitivity and specificity of 76.67% and 57.61% respectively in real terms this means that 23 out of the 30 actual father-son pairs are positively identified and 651 of the 1,130 unrelated pairs are correctly predicted as unrelated. This leaves seven actual father-son pairs and 479 unrelated pairs incorrectly classified as depicted in Figure 6.13.

For the one time in the study the same portion of the face has both the greatest AUC and the highest Rule In specificity. The triple combination of Eyes and Nose and Mouth when set a 100% sensitivity has a specificity of 25.04% and a threshold of ≤4.6823mm. This gives rise to the correct prediction of 283 of the 1,130 unrelated pairs whilst correctly predicting all of the 30 related pairs. Overall this enables a reduction in the total number of possible pairs from 1,160 to 877 a reduction of 24.3% as shown above the threshold line in Figure 6.14.

Both the Mouth and the Eyes have significant results and have a greater AUC than the Whole Face.

The joined features have similar values to the single features and all outperform the Whole Face, overall it is the largest four feature portions that have the greatest AUC of the joined features.

The combined features overall do have slightly higher AUC values than both the joined features and the Whole Face and once again it is the triple combination that provides the highest level of accuracy.

In comparison with the fathers to female children category, where the removal of the poorer scans reduced the number of non-significant results, for fathers to male children it increases them from one to six, as seen in Table 6.27. The Mouth, Nose and Mouth and all three joined feature portions are left as having significant results. Only the Eyes-Nose-Mouth-Chin was seen to increase its AUC and in doing so moved to the top position. The Rule In specificities also all decreased with the exception of Eyes-Nose-Mouth-Chin but this time the minimal increase was not enough to change its positon in the ranking leaving the combination of Nose and Mouth to top the percentages although at a lower value than before the removal of the poorer scans.

Number of comparisons: 1,160
Matches: 30

Area	AUC	Significance P-value	Optimal Sensitivity (%)	Optimal Specificity (%)	Rule In (%)
Whole Face	0.631	0.0083	83.33	38.58	13.72
Eyes	0.651	0.0020	60.00	70.00	12.39
Nose	0.563	***0.1914***	83.33	32.21	12.57
Mouth	0.687	0.0001	70.00	59.29	11.15
Eyes-Nose	0.684	0.0001	76.67	57.17	6.81
Eyes-Nose-Mouth	0.653	0.0025	66.67	61.42	5.84
Eyes-Nose-Mouth-Chin	0.685	0.0001	80.00	55.49	5.22
Eyes and Nose	0.663	0.0003	76.67	50.35	13.72
Eyes and Mouth	0.707	<0.0001	50.00	87.17	19.91
Nose and Mouth	0.686	<0.0001	73.33	60.35	23.19
Eyes and Nose and Mouth	0.717	<0.0001	76.67	57.61	25.04

Table 6.26 ROC results for fathers compared to male children.

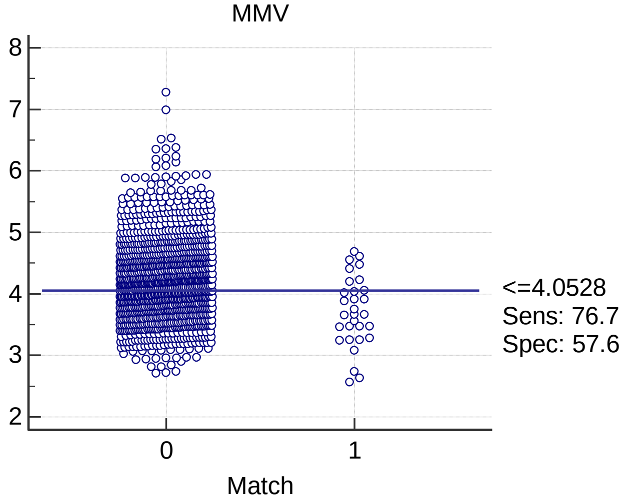


Figure 6.13 ROC Dot diagram for fathers to male children, Eyes and Nose and Mouth combination with optimal threshold.

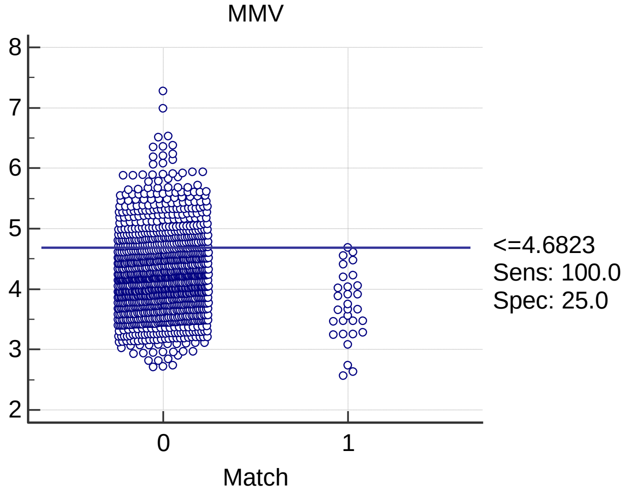


Figure 6.14 ROC Dot diagram for fathers to male children, Eyes and Nose and Mouth combination with Rule In threshold.

Area	**No of comparisons**	**Matches**	**AUC**	**Significance P-value**	**Optimal Sensitivity (%)**	**Optimal Specificity (%)**	**Rule In (%)**
Whole Face	874	23	0.591	***0.1032***	73.91	46.53	12.57
Eyes	900	23	0.602	***0.0939***	43.48	79.36	11.74
Nose	999	26	0.547	***0.3877***	84.62	29.60	10.59
Mouth	851	22	0.634	0.0279	40.91	80.82	9.77
Eyes-Nose	962	24	0.658	0.0065	75.00	57.14	5.86
Eyes-Nose-Mouth	836	22	0.653	0.0070	81.82	47.17	5.77
Eyes-Nose-Mouth-Chin	836	22	0.702	0.0002	81.82	58.97	6.63
Eyes and Nose	816	21	0.585	***0.1329***	90.48	29.18	9.31
Eyes and Mouth	660	18	0.617	***0.1113***	44.44	84.42	16.20
Nose and Mouth	714	20	0.668	0.0029	65.00	66.43	19.60
Eyes and Nose and Mouth	589	17	0.627	***0.0548***	76.47	50.00	20.98

Table 6.27 ROC results for fathers compared to male children after removal of the ten poorest quality scans

6.4 Review of the Areas of the Face

The results for each relationship category have been discussed in detail above and clearly show different results for each of the eleven relationship categories, however some patterns can be seen across the study as a whole. In seven of the eleven relationship categories the triple combination of ‘Eyes and Nose and Mouth produces the greatest AUC, with the Whole Face, Eyes-Nose-Mouth-Chin, Eyes-Nose-Mouth and Eyes-Nose each topping the list for one relationship category. The average AUC for each portion of the face across all relationships ranks them as follows:

1. Eyes and Nose and Mouth
2. Eyes and Mouth
3. Eyes and Nose
4. Eyes-Nose-Mouth-Chin
5. Eyes-Nose
6. Whole face
7. Eyes-Nose-Mouth
8. Eyes
9. Nose and Mouth
10. Nose
11. Mouth

As seen earlier the Mouth as a single feature performs poorly and can have a detrimental effect on other portions the face which include the mouth, this is especially true when comparing females. Taking this into account, overall the combinations performed better than the joined features which again performed better than the single features. The whole face was comparable with the joined features, superior to the single features but was outperformed by the combinations of single features.

The application of a Rule In scenario did produce different results in term of the best portions of the face. The combination Eyes and Mouth portion and the Eyes-Nose joined portion were each best for three relationships while the single Eyes portion ranked top for two results and the Whole Face and the combination portions Eyes and Nose, and Eyes and Nose and Mouth all proved to be superior for one specific relationship.

The removal of the ten poorest quality scans did improve the AUC in most cases but no clear pattern was observed in relation to the specific areas of the face which improved most or in the change in the ranking of the different portions of the face before and after.

6.4.1 The Effect of Age and BMI Difference.

For all of the relationships being considered there are likely to be factors, other than genetics, which influence the similarity or disparity seen between relatives. In order in address the impact of the two main factors, age and BMI, the absolute value of the difference in both of these between relations will be calculated and compared to the rank of MMV for each relationship type. Ranking, rather than the actual MMVs, are used as the MMVs are not proportional across the different areas of the face and the average rank across all areas of the face is required. The highest rank, one, is given to the relationship with the smallest MMV and the rest follow in size order. The results of these can be seen in Figures 6.15-6.36 which have lines of best fit plotted and the correlation coefficients for both age and BMI variation are in Table 6.28.

Generally, the age difference graphs show a positive correlation so the greater the age difference, the more likely the pairing will be ranked lower, owing to a greater disparity than those with a smaller age gap. There are three exceptions to this; sisters, mothers-daughters and fathers-daughters. The mother-daughter and father-daughter are negatively correlated, suggesting that the greater the age difference, the more similar the parent and child but the correlation coefficients are so low, -0.128 and -0.034 respectively, that the result is not very strong. The correlation between sisters is also negative but stronger than the parents to daughters relationships, with a correlation coefficient of -0.298. There are however only seven results for sister comparisons and so these are misleading. The overall age difference amongst the siblings has a much greater correlation than that between both parents and children. Brothers showed the greatest correlation with age difference, suggesting that the age gap between the brothers has a large impact on the similarity of the face.

The correlation coefficients of BMI difference are greater overall than those of age difference, although this is not the case for all relationships. The BMI difference is positively associated with the average rank/position in all relationships except mothers-sons. Mothers-sons has a correlation coefficient of -0.051 so, while it is negative, it demonstrates almost no correlation. Fathers-sons also has a correlation coefficient close to zero, 0.048, showing no correlation. As seen with age difference, the overall correlation of BMI difference was stronger between siblings than between parents and children although the fathers-daughters category does have a strong correlation. Sisters showed the strongest correlation of rank to BMI difference signifying that that BMI difference can have a negative effect on the calculated similarity between sisters.

Whilst correlation is shown between age and BMI difference and the average ranking of relatives, the plots are very scattered for all relationships showing no clear linear correlation exists. The specific correlations of MMV rank with age and BMI difference were also calculated for each portion of the face and in the most part they did not differ greatly from the result produced by the averaging of all facial portions. The facial portion which differed most between family members was the Mouth. In all of the sibling relationships, mothers-children relationships and mothers-daughters relationships the difference in BMI between family members had a negative correlation with the MMV ranking. The correlation with age difference also varied between the Mouth and the overall average for the same relationship categories, though not to the same extent as the BMI. Other portions of the face which included the mouth were affected in both cases but to a lesser degree. In three of the categories, fathers-children, fathers-sons and mothers-children the Whole Face had a stronger correlation to BMI than the average. Lastly the Nose had weaker correlations with age difference than the average in siblings, opposite sex siblings and with BMI in mothers-children, mothers-daughters, fathers-children and fathers-sons.

Correlation coefficients of rank		
	Age difference	**BMI difference**
Siblings	0.285	0.418
Opposite sex siblings	0.311	0.279
Same sex siblings	0.339	0.454
Sister	-0.298	0.549
Brothers	0.596	0.286
Mothers-Children	0.128	0.110
Mothers-Daughters	-0.128	0.245
Mothers-Sons	0.220	-0.051
Fathers-Children	0.131	0.208
Fathers-Daughters	-0.034	0.534
Fathers-Sons	0.251	0.048

Table 6.28 Correlation coefficients of average rank vs age and BMI difference.

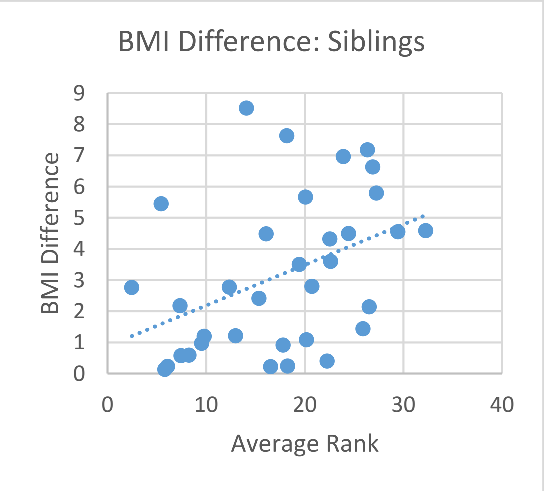


Figure 6.15 BMI: Siblings

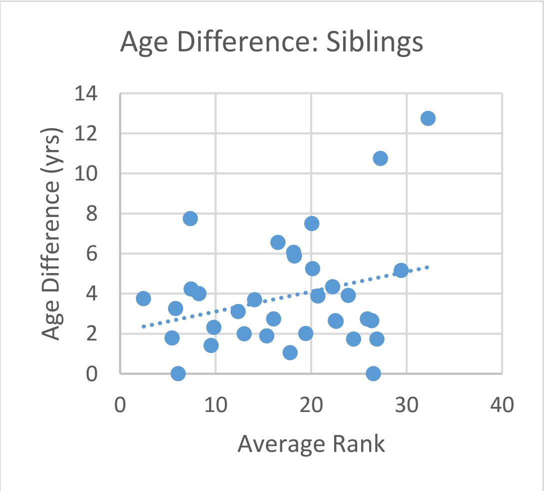


Figure 6.16 Age: Siblings

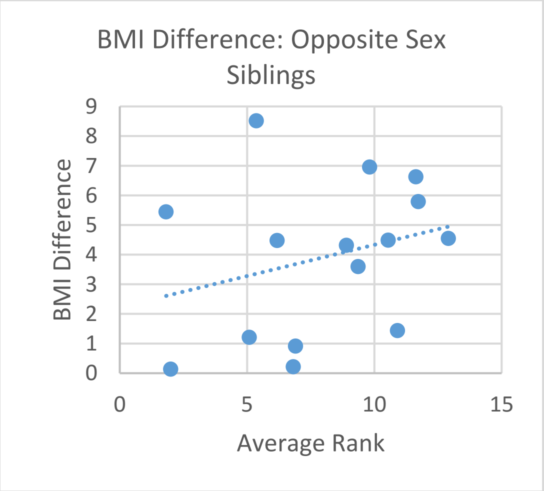


Figure 6.17 BMI: Opposite sex siblings

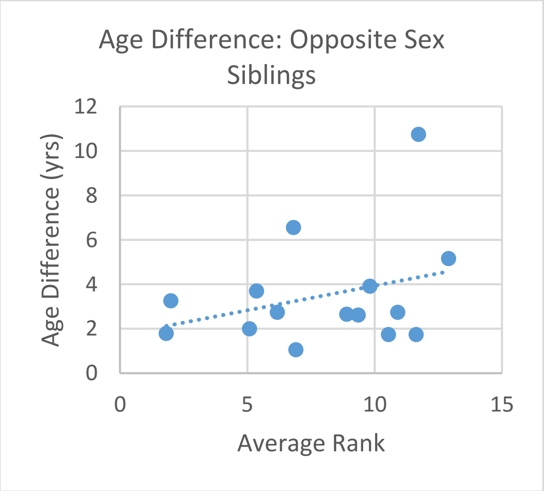


Figure 6.18 Age: Opposite sex siblings

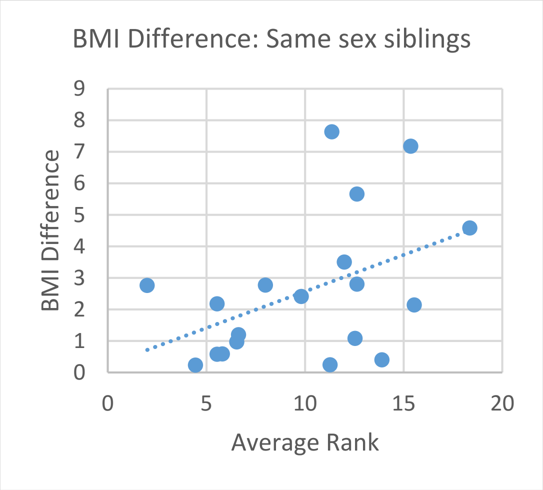


Figure 6.19 BMI: Same sex siblings

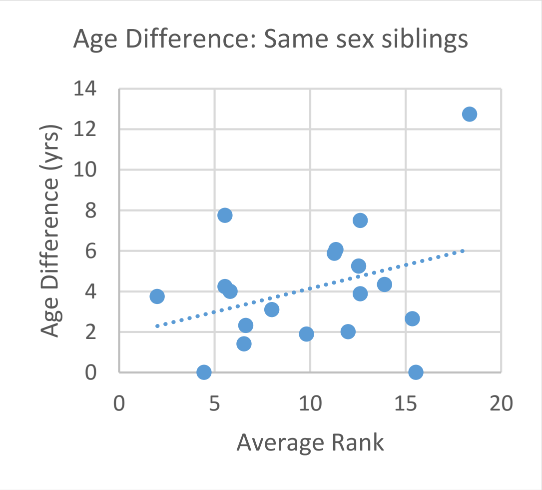


Figure 6.20 Age: Same sex siblings

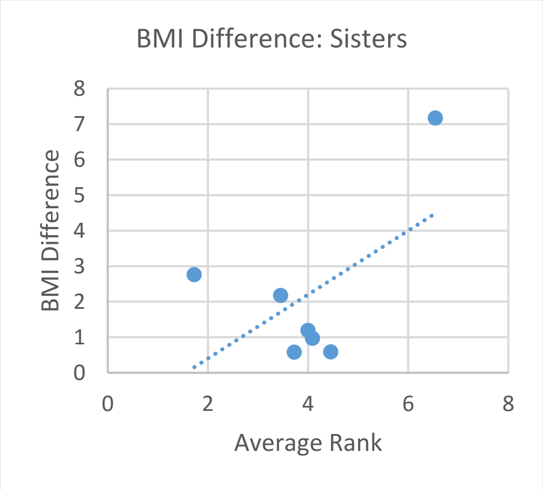


Figure 6.21 BMI: Sisters

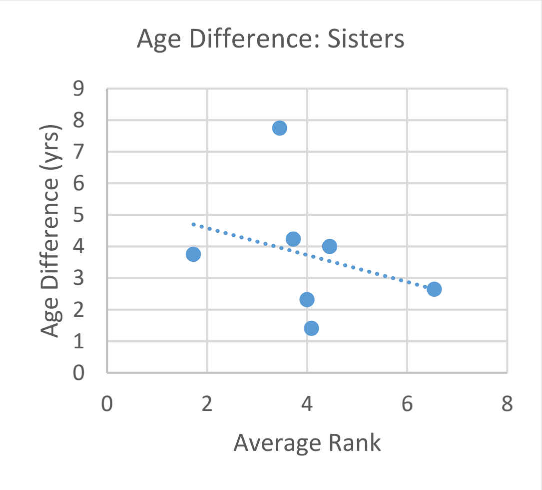


Figure 6.22 Age: Sisters

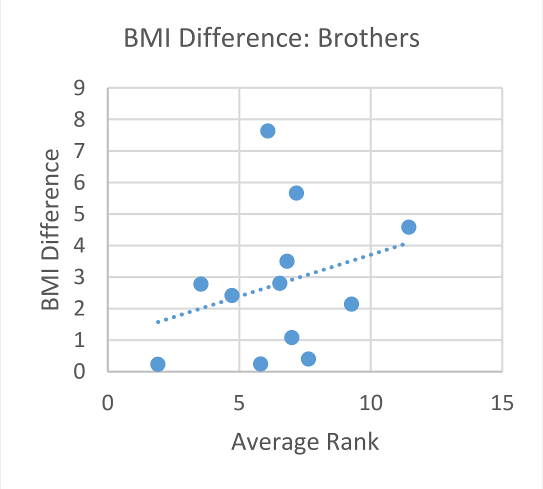


Figure 6.23 BMI: Brothers

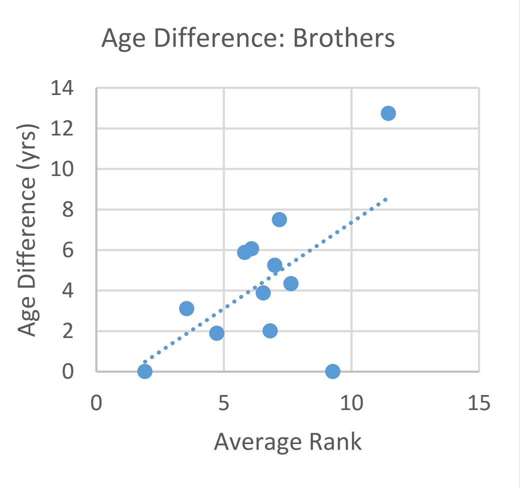
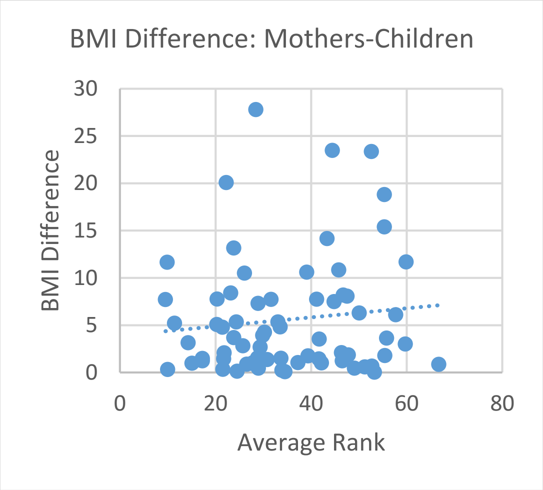


Figure 6.24 Age: Brothers



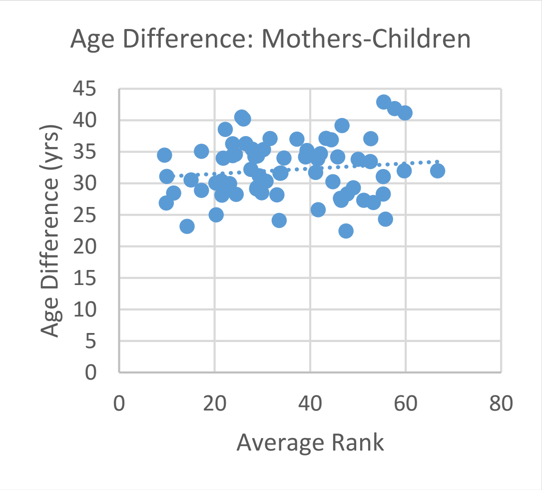
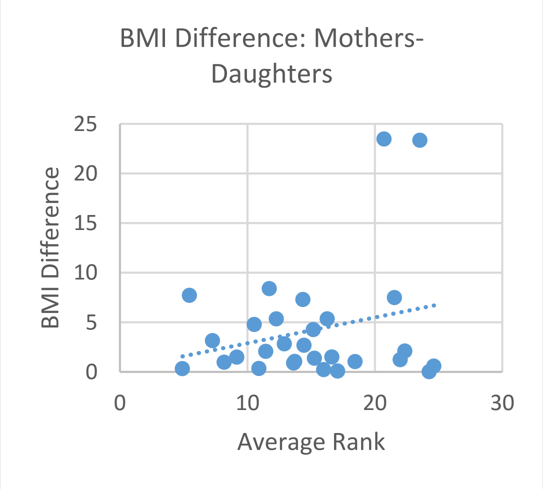


Figure 6.25 BMI: Mothers-Children

Figure 6.26 Age: Mothers-Children



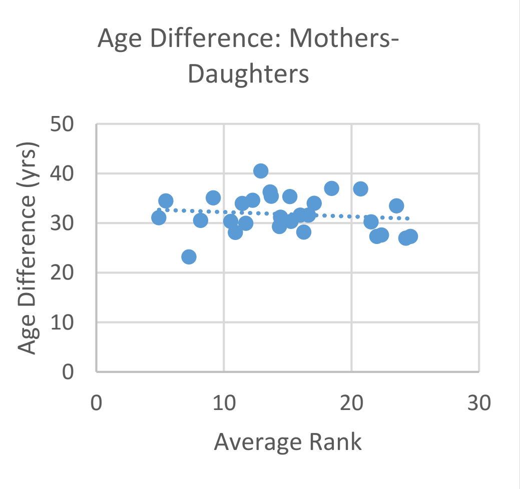
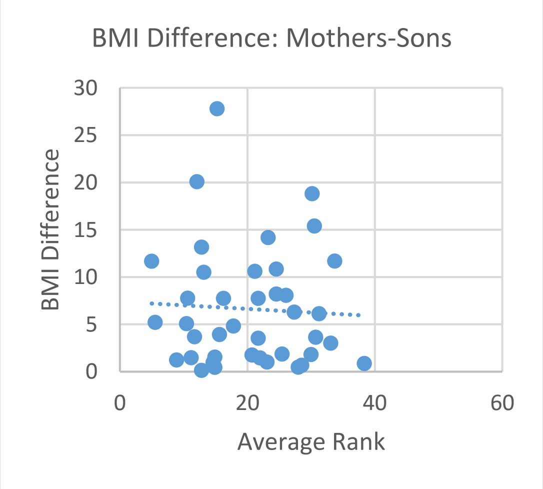


Figure 6.27 BMI: Mothers-Daughters

Figure 6.28 Age: Mothers-Daughters



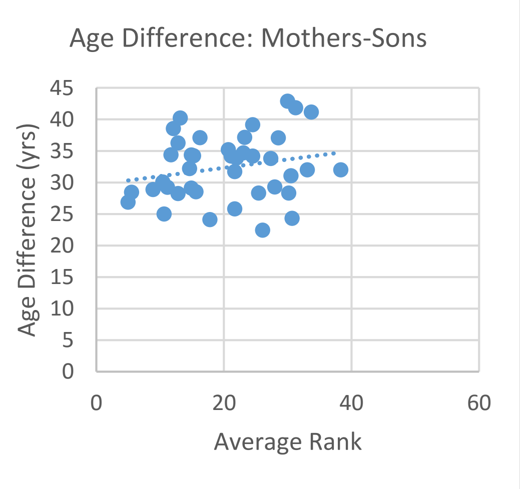


Figure 6.29 BMI: Mothers-Sons

Figure 6.30 Age: Mothers-Sons

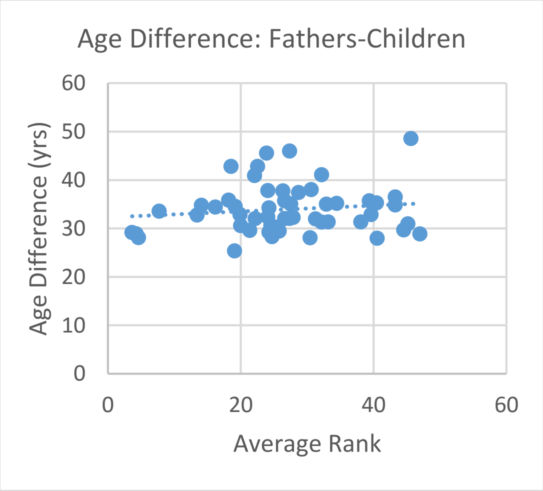

Age Difference: Fathers-Children

Figure 6.31 BMI: Fathers-Children

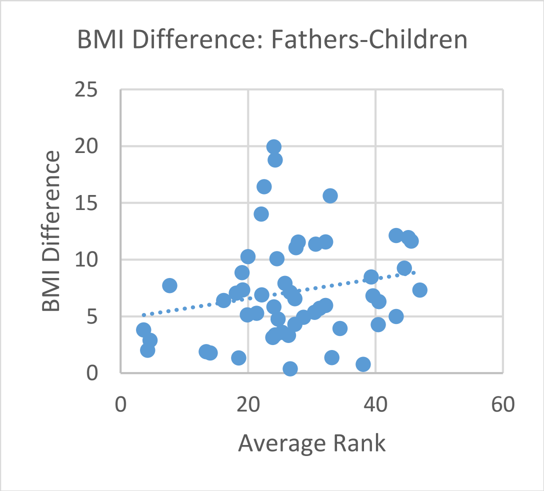

BMI Difference: Fathers-Children

Figure 6.32 Age: Fathers-Children

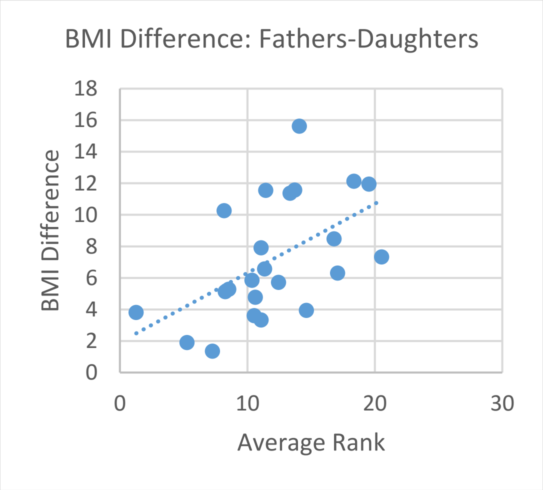

BMI Difference: Fathers-Daughters

Figure 6.33 BMI: Fathers-Daughters

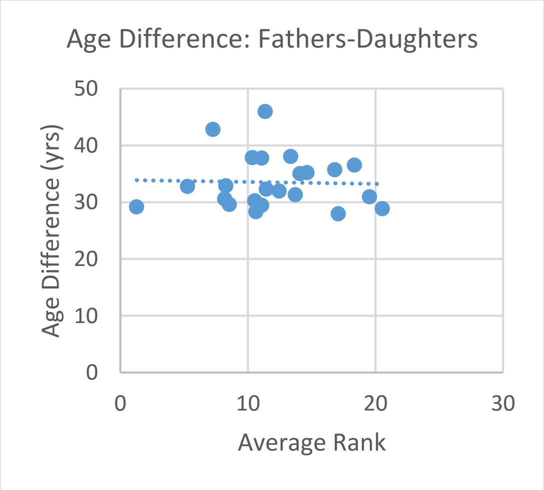

Age Difference: Fathers-Daughters

Figure 6.34 Age: Fathers-Daughters

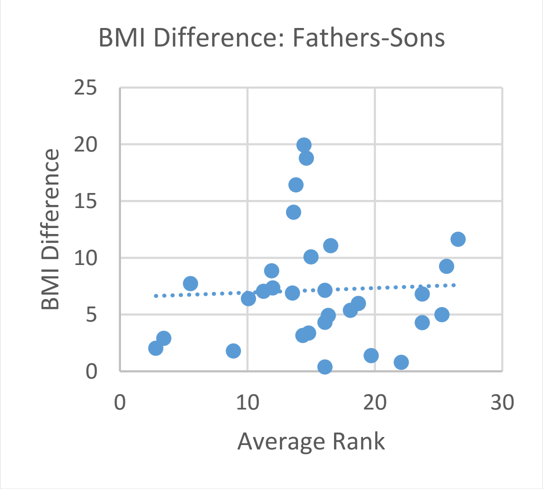

BMI Difference: Fathers-Sons

Figure 6.35 BMI: Fathers-Sons

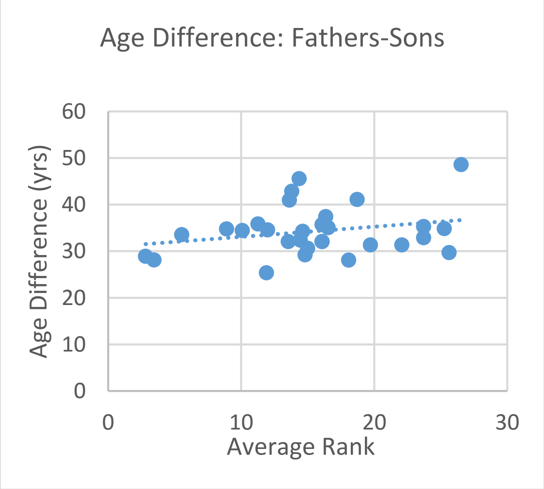

Age Difference: Fathers-Sons

Figure 6.36 Age: Fathers-Sons

Chapter 7: Discussion

The face is a complex area of the body, it never stops growing, exhibits sexual dimorphism, can be influenced by genetic and environmental factors and can be greatly changed by traumatic events and dental treatment (Jelenkovic et al. 2010, Naini & Moss 2004). Additionally, the skeletal and soft tissue components of the face can respond in different ways to all of the factors above (Baydaş et al. 2007). All of these influences make the recognition of a single individual from images taken, even at short intervals in time, a difficult process. Facial recognition technology has moved from human-based observation through two-dimensional linear and angular comparisons and on to three-dimensional landmarks and surface-based techniques.

The majority of studies conducted on the craniofacial area, whether for growth, heritability or facial recognition have been conducted in two-dimensions (Amini & Borzabadi-Farahani 2009, Ercan et al. 2007, Iblher et al. 2008, Jelenkovic et al. 2010). This is almost entirely due to the availability of technology and not because this was the preferred technique. For decades the best means of craniofacial assessment was the lateral cephalograph, this allowed for numerous measurements to be taken and compared either to a cross section of the population or in longitudinal studies of the same individuals. The main limitation of the lateral cephalograph is the lack of medio-lateral dimensions which consequentially have not been studied in the same amount of detail over recent years as have the anterio-posterior and superior-inferior dimensions. In the last ten to15 years the use of lateral cephalographs has diminished and been replaced with two-dimensional photographs, three-dimensional dimensional images and three-dimensional casts. There are two explanations for this change, firstly the advances in technology making three-dimensional surface data collection quick and relatively cheap and secondly, a better understanding of the potentially harmful effects of X-rays on the body resulting in a huge reduction in the use of X-rays for non-medical purposes. As described in the chapter on facial recognition two-dimensional photos have all sorts of associated issues like size, position and lighting variation making three-dimensions a better and less error prone choice as it eliminates these problems. The main difficulty with three-dimensional images is the complexity of their capture, whilst two-dimensional images are incredibly cheap, quick, available and can be taken from moving images such as CCTV, three-dimensional images require multiple cameras or a micro scribe (an apparatus for recording three-dimensional landmarks from subjects) and, at present, the cooperation of the subject. Even though the accuracy has increased the price of three-dimensional data acquisition has dropped dramatically over the last decade so that relatively accurate scanning devices can now be acquired for a couple of hundred pounds. The issue of subject cooperation is also improving with the speed of scans increasing so subjects are not required to stay still for more than a couple of seconds. So whilst the majority of current facial recognition systems operate in two-dimensions, looking to the future three-dimensional will inevitably take the place of two-dimensional as speed, cost and functionality improve. Taking everything into consideration this study chose to focus on three-dimensional image capture and analysis.

The purpose of the study was to assess the ability of algorithms, normally used in facial recognition software, at finding the relatives of a subject rather than the subject themselves. In order to conduct this research, three-dimensional representations of the faces of nuclear families were required. These needed to be processed and analysed using suitable mathematical algorithms and the results analysed. A study like this has never been conducted before so whilst comparison cannot be made for the results as a whole, elements of the study do relate to previous studies and can be compared in their own right.

Methods

The author had access to a structured white light surface scanner for the purposes of this study which performed well in most situations. The scanner is portable and has a capture time of approximately two –four seconds for each scan depending on the illumination of the subject. The resultant scan, a three-dimensional representation of the facial surface allowed for either landmark based or surface-based analysis. Landmarks can be placed manually or automatically on three-dimensional structures but automatic placement often incurs small errors which require manual adjustment (Gökberk et al. 2006, Lu et al. 2004). This is a time consuming task and often encounters intra and inter observer errors due to the uncertainty of landmarks and/or poor quality scans. Surface-based techniques do not require this initial stage and are thus less prone to observer error as the observer has no input. The main difference between landmark and surface-based methods is the amount of data used for the assessment of change or similarity. Landmarks can never fully portray the shape of the face, there will always be areas which will be not accounted for. Whilst the placement of fixed landmarks, those which are placed at specific recognisable points, is relatively accurate they tend to cover the midfacial area around the eyes nose and mouth and very limited data from the cheeks, forehead and chin are captured. Semi landmarks, placed at fixed distances or percentages between fixed landmarks, can cover areas with fewer definable features but still only convey the general shape of the face. The three dimensional surface-based studies of heritability (Djordjevic et al. 2013, Naini & Moss 2004) were able to determine the specific areas of the face which displayed the most heritability and found the lower cheeks in females and the forehead in males (Djordjevic et al. 2013) to be among them. Two dimensional studies and three-dimensional landmark based studies would have been unable to demonstrate this. Djordjevic et al. (2013) was also able to compare landmark based Procrustes superimposition and a surface-based Trimmed Iterative Closest Point algorithm (TrICP) and conclude that the surface-based approach was far better at assessing and visualising the heritability of the facial features. The present study concurs with the findings of Djordjevic et al. (2013) in that neither Procrustes superimposition nor Principal Component Analysis (PCA) were able to find any data components that were able to identify possible links between family members. The TrICP analysis was, however, able to find similarities between relatives. PCA has been used successfully in many facial recognition systems both for two-dimensional and three-dimensional analysis (Hesher et al. 2003, Tsalakanidou et al. 2003, Zhao et al. 2003) but has only shown success in heritability studies of same sex twins (Weinberg et al. 2013). Mohammadzade & Hatzinakos (2013) also saw an advantage in using the surface information when they reported that their facial recognition rates were better using ICP than PCA on landmarks.

The poor performance in the current study suggests that the variation between parents and children and singleton siblings of both sexes is too great for a simple landmark based analysis.

Variants of ICP have been used in many facial recognition studies and software with high recognition rate percentages accomplished. ICP has been chosen as it is the most widely used surface-based approach and although rigid in its comparison, the ability to select specific areas allows for feature based comparison in addition to using a holistic approach. Deformation models (Al-Osamini et al. 2009, Smeets et al. 2010) are possible and perform well for expression variation, however, unlike expression variation, where the general transformation between a neutral and a smiling face is known and so can be reversed through deformation, the ‘transformation’ between a child and parent is not known. ICP is also relatively simple to conduct and analyse, and therefore does not need to be conducted by mathematicians or engineers. It does not require images of the same resolution, so that data can be collected in a variety of conditions. Whilst a number of studies have looked into the algorithm’s ability to deal with facial expression and small degrees of ageing up to one year (Paone et al. 2014) none have assessed the ability of the algorithm to deal with extreme ageing. Some studies (Paone et al. 2014, Sun et al. 2010, Vijayan et al. 2011) did review the ability of facial recognition systems to correctly identify twins and in most cases the reported ability of the software was reduced when twins were introduced into the population. It is this reduction in the ability to correctly separate twins that suggests that the same algorithms could be used to identify relations, rather than tell them apart. The high recognition rates observed when comparing the same individual are expected to reduce when identifying relatives due to both the individual differences and age and sex variations between the various relationships. The expected reduction in recognition rates, when compared to other studies using the same algorithm, was observed in this study, meaning that while a specific genetic relative cannot be identified, the pool of possible relatives was able to be significantly reduced.

Results

Overall this study demonstrated that while it is not possible to determine a specific genetic relationship between two individuals using this methodology, it is possible to rule out a genetic link. This would enable a reduction of a pool of possible relations without having to use expensive and invasive tests such as DNA. The percentage of the population which is able to be ruled out differs greatly across the different types of relationship. The study also showed that the different areas of the face have different heritability patterns and that the best performing area differs for each relationship. The mouth however is seen to perform poorest overall suggesting that environmental influences on growth are very strong.

A large reduction in AUC and optimal specificity and sensitivity was encountered between the preliminary study and the full study (see Tables 5.6 and 6.5). There are a number of possible reasons for the differing results from the preliminary and the full study. Firstly, although randomly selected, the ten families considered in the preliminary study could have shown more genetic similarity than the rest of the study population. Secondly the scans themselves may have been of a higher quality than others. Thirdly, and most likely, is that due to the size of the study and the low number of participants, similarities are more likely to be observed. However, when the population size increases so does the range of facial shapes and the separation between the families becomes more blurred. Lu et al. (2006) had warned against the possibility of this occurring when they observed exactly the same issue in transitioning from a study of 18 participants to 200. There is a possibility that the accuracy of the current study would decrease further if the population size were to increase. Following the pattern that Lu et al. (2006) observed in their own study and in others, the reduction in accuracy is proportionally much higher between studies of below 100 participants and those over 100, than between studies of approximately 100 and 500 plus participants. So while a decrease would be expected should the study population increase, it would not be expected to be very large.

The results of the average MMVs whilst not a measure of heritability, generally concur with the levels of similarity seen between the different relationships in heritability studies. Both mothers and fathers show a tendency for female children to be more like them than the male children, which agrees with the results of AlKhudhairi & AlKofide (2010), Gelgör et al. (2006), Johannsdottir et al. (2005) and Saunders et al. (1980). Overall the MMVs for relationships between children were lower than those between adults and children, similar to the results seen by Nakata et al. (1973) and Saunders et al. (1980). This suggests that either heritability follows the same category pattern as the general similarity between unrelated individuals, or the actual matches follow the pattern and lower the overall averages as a result, or that the other heritability studies have presented results which include this bias. If the latter is true, then it may be that heritability levels in these other studies are in fact of a similar level across the different relationships. Although not directly comparable to the heritability values in the studies mentioned, the average AUC values for each relationship provide a better evaluation of heritability. Here the parents both demonstrate higher values for sons compared to daughters which counteracts the results of several authors (AlKhudhairi & AlKofide 2010, Gelgör et al. 2006, Johannsdottir et al. 2005, Saunders et al. 1980) although Nakata et al. (1973) did see this superiority but only in fathers. The results of the same sex sibling comparisons did outrank those of the parents but the categories which include opposite sexed sibling comparisons performed only slightly better than the fathers-daughters which disagrees with the work of Saunders et al. (1980) who not only saw higher levels of heritability in siblings but also reported the number of significant measurements to be greater for opposite sex twins than for same sex twins. Naini & Moss (2004) report differently and in this respect are in agreement with the current study stating that same sex dizygotic twins were more alike than those of opposite sex. The main source of disagreement amongst heritability studies is which parent displays the most likeness to their offspring. For both the average MMV and AUC this study reports that children resemble mothers more than fathers. The results of Johannsdottir et al. (2005) and Gelgör et al. (2006) are therefore in agreement with the current study whilst Hunter et al. (1970), Saunders et al. (1980), Nakata et al. (1973) and AlKhudhairi & AlKofide (2010) report the opposite. A large amount of the variation between this and other studies of heritability may simply be due to the nature of the data. All of the parental and singleton sibling studies were conducted in two dimensions and only three of the twins based studies used three-dimensional data. Put simply, until now a large proportion of the face has not been considered when assessing heritability. The studies which did use the whole face were focused more on the specific parts of the face which demonstrated heritability, rather than the comparison of the different levels of heritability between relations.

All previous heritability studies saw variation in the levels and significance of heritability in different parts or dimensions of the face and facial skeleton and this study was no different. The act of sectioning the face into features enabled each feature to be compared, both to the other features of the face and to the same feature in different relationship categories. The most notable result was the poor performance of the mouth especially when considering females. This ties in with the comparatively low similarity levels in the dental region compared to the rest of the craniofacial area reported in both parental and sibling heritability studies (Amini & Borzabadi-Farahani 2009, Gelgör et al. 2006, Hunter et al. 1970, Johannsdottir et al. 2005, Kitahara et al. 1996, Manfredi et al. 1997, Saunders et al. 1980). Of the three-dimensional facial surface studies Naini & Moss (2004) also indicate that the mouth does not display similar curvature between even monozygotic twins whereas Djordjevic et al. (2013) report that the lower lip in male twins did show similarities. Whilst Djordjevic et al.'s (2013) result does not entirely reflect those of this study, the AUC of the Mouth in same sex male children was greater than for the Eyes which corresponds with Djordjevic's findings. The mouth was only considered as a whole in this study, so it is possible that if the lower lip and underlying mandible does show strong heritability, hinted at by a number of studies (AlKhudhairi & AlKofide 2010, Amini & Borzabadi-Farahani 2009, Baydaş et al, 2007, Carels et al, 2001, Kitahara et al, 1996, Johannsdottir et al, 2005), then this could be masked by the upper lip and maxilla which are influenced to a greater degree by environmental factors. The reason for the low similarity levels of the mouth and dento-aveolar area may also have the same explanation as that reported by facial recognition papers. The mouth is a much more malleable and expressive feature than the eyes and nose and as such is difficult to capture in the same pose across multiple images, let alone between different individuals. To ensure continuity across a study population the same pose must be held during data collection. For the mouth this requires the positions of the upper and lower dentition and the state of the muscles, whether relaxed or taut, to remain fixed. Participants in this study were instructed to occlude the teeth and relax the facial muscles but other studies may not have followed the same procedures and thus may have resulted in different facial morphology being recorded.

In relation to facial recognitions systems and software, the recognition rates of this study are poor, the algorithm is however being used for an entirely different purpose. Rather than attempting to identify the same individual in different images, perhaps with expression or age variation, the algorithm is here attempting to identify a completely different person. A study like this has never been done before so the only comparisons that can be made are with the few studies which applied facial recognition systems to a population of twins (Paone et al. 2014, Sun et al. 2010, Vijayan et al. 2011). In all of these the systems reported that rates of recognition were reduced after the introduction of same sex twins with error rates of up to 20% (Paone et al. 2014). The addition of age variation of only one year, further increased the error rate for monozygotic twins to approximately 40% (Paone et al.2014). It can only be expected that when singleton siblings of different ages and sexes are also included that the recognition rates would drop yet further not to mention the extreme age variation introduced by the addition of parents.

Certain elements of the methodologies used to improve the ability of facial recognition systems generally and specifically in dealing with variations in expression were employed in this study. The removal of outlying and background data; the selection of the midfacial area; the exclusion of the ears, neck and hairline; the selection of features; the combination of different areas have all been employed before, mainly in an attempt to minimise the variation due to emotional expression (Chang et al. 2006, Mian et al. 2007, Smeets et al. 2010). As emotional changes tend to disfigure the mouth and cheeks more than the less malleable nose and eyes, it is usually data from the former which is removed. Whilst the mouth is shown to have had a detrimental effect in this study, possibly due to the malleability, the use of only single features is not seen to improve the result. The selection of joined features, thus not considering the edges of the face suggested by a few authors (Chang et al. 2006, Mian et al. 2007, Smeets et al. 2010) in order to reduce the scanning errors commonly found at the hairline and around the ears, did improve the accuracy. Further still, the combination of multiple parts of the face as employed by Smeets et al. (2010) saw the best results in both this study and their own study. For Smeets et al. (2010) the dominance of the combined approach was considered to be due to the reduction in overall expressive variance, as any extreme variation in one scan was diluted by the others. The superiority of the combined approach in the current study however is thought to be caused by the removal of the angles between features and the ‘addition’ effect. It is possible that children inherit similar features, such as the nose, from their parents but the position and angle at which they sit within the face differs. By considering them individually rather than as one inflexible surface these similarities can be better seen. The ‘addition’ effect then means that while only a small amount of similarity might be measured for each feature, if all the features exhibit similarity then the sum of all of them will create a clearer separation than from the values of dissimilar features.

Age and BMI

The genetic variation of the facial form between relatives would ideally be studied using scans taken of them at the same age. This would eliminate the variation caused by age difference and thus highlight the inherited features. A study of this kind, conducted in three-dimensional, would therefore take at least a generation to conduct and more if the age required is not that of an infant. Rather than wait a generation, growth studies, regression analysis and age difference correlation can all aid in identifying the amount and position of variation with respect to age. Overall a greater correlation of age difference and average MMV ranking was seen for the relatively small age gap between siblings rather than across the larger parent to child age difference. This observation strongly indicates that the developmental growth difference between siblings, even if only a few years apart, accounts for a much larger degree of variation than that which occurs in adulthood. This ties in with all growth studies which show that while soft tissue growth does not stop at adolescence, the rate of growth is much decreased (Iblher et al. 2008, Sforza et al 2010a, Sforza et al. 2010b, Zankl et al. 2002). The strongest correlation between age difference and MMV ranking was seen in brothers which, taking the age range of seven to 24 years into account, is understandable.

Developmental growth studies of the all the areas of the face show high rates of growth in males until approximately 18 years of age with a growth spurt at around 13-14 years old (Bishara et al. 1984, Ferrario et al. 1997, Ferrario et al. 2000, Oliver 1982, Sforza et al 2009, Sforza et al. 2010a, Sforza et al. 2010b, Sforza et al. 2013). As the age range of this study contains boys who have gone through the growth spurt and those who are yet to do so, the amount of variation will be high. This is not the case for the girls who range from 11-21 years and generally exhibit a growth spurt between 10-12 years. It is this difference in maturation levels which is likely to account for the greater correlation of age difference between boys and their parents than girls and their parents. As the girls have completed most of their developmental growth they are more likely to resemble their adult selves and thus their parents, while most of the boys still have a reasonable amount of growth to attain before reaching their adult proportions.

The mouth exhibits very different variation with respect to age than the other parts of the face particularly between mothers and children and more so in daughters than in sons (see Table 6.2 and Section 6.4.1). This corresponds to the greater degenerative change seen in the mouth proportionally to the other facial features. Ageing causes the mouth to change in size, proportion and position, the lips elongate and thin while the philtrum lengthens increasing the gap between the lips and the nose (Sforza et al. 2010b).

Whilst no conclusion can be made as the specific ages of the children was not correlated to the average ranking, the fact that a smaller age gap between parents and children did increase the likelihood of a low MMV does suggest that a child of higher years shows more similarity to the parents, as suggested by a number of authors (Gelgör et al. 2005, Johannsdottir et al. 2005, Kitahara et al. 1996, Saunders et al. 1980).

Body Mass Index (BMI) is not the only factor, other than age and genetics, to influence the morphology of the face but it can have a large impact. It is seen to correspond with the average ranking for all relationships except parents to their sons. It is the difference in BMI between relatives, rather than the BMI itself, that correlates with MMV ranking, indicating that similarities are more likely to be seen between relatives with a similar BMI, not a particular low or high BMI value. This suggests that relatives deposit fat on the face in the same way so even if the BMI is high, so long as the other relative also has a high BMI then the similarities are still noticeable. The correlations are stronger than those calculated for age suggesting that BMI difference has a greater influence than age proportionally. The average BMI is very similar for the daughters and sons though the range is greater for the sons which might explain the variation seen between the two categories. The fact that the whole face shows greater correlation than the average for all the cross generation categories, suggests that the portions of the face which are only found in the whole face area, the cheeks and forehead, exhibit greater variation caused by BMI than the rest of the face.

Limitations

Whilst every attempt has been made to conduct a robust study there are a number of limitations which could not be mitigated in the present study. Whilst the structured white light scanner is able to capture incredibly accurate three-dimensional representations of the human face, it was found that facial hair, make-up and movement all negatively affect the quality of scans. In the area of the eyebrows, holes in the scan surface often occur and scans of males with facial hair on the upper lip and lower facial area are often missing most of the data from the affected area. This can cause difficulty both in landmark positioning and surface comparison. Whilst the speed of the scanning was very short, two to four seconds depending on the lighting conditions, some participants were unable to stay still for this length of time. Each scan was reviewed quickly during the scanning process and those which exhibited obvious movement were repeated, but some were only noticed at a later stage when it was too late to rescan. As seven separate scans at different angles were merged to create the full scan of the face, any movement in the position of the face between each scan would cause difficulty in aligning each of the separate scans. Instruction on the positioning of the face did counteract this on most occasions but younger participants in particular struggled to comply with this request. The effect of the quality of the scans could be seen in the results following the removal of the ten scans of poorest quality. In general, an increase in AUC and Rule In ability was seen though it is possible that this was in part due to the reduced size of the study. In any future study, men with any amount of facial hair, the group which had the greatest detrimental effect on scans, would be excluded from the study though this would cause problems with the recruitment of volunteers, especially as the wearing of facial hair is both popular and widespread at the time of this study.

Recruiting participants for this study was a much more difficult task than expected. Due to the age of the child participants, ideally between 14-25 years, it was more unlikely that they would be routinely accompanied by their parents when out and about and many may not even live in the same city as their parents. This meant that specific groups with children of that age had to be targeted and specific scanning sessions arranged to encourage participation. This in turn limited the number of participants in the study which would ideally be increased in any future studies. Ideally, imposing a limit on BMI differences and age differences between family members would be helpful but this would also make it harder to apply this method in the real world as the resultant database would not represent the true population.

A possible limitation in using a volunteer sample is that the author only has a declaration that the parents and children in the study are in fact the biological relatives. The only solution to this would be to also conduct DNA testing which is very intrusive, expensive and time consuming.

A major limitation of the current TrICP algorithm is its computational complexity. It requires a high level of processing power to run it and even then, takes a long time to process. While the surfaces with fewer vertices, the single features, are faster to run than the whole face they are not close to the speed of landmark based comparisons. In total the 7,000 plus comparisons for each of the seven areas of the face took approximately 9,000 hours of processing. While a large proportion of that time is under automatic control, due to the setup of the current software, a large amount of manual input is still needed. The processing speed will improve with the improvement of processers and the algorithm may also be able to be improved. Once an initial database is set up, then each new scan of an individual will only need to be compared once to each other scan in the search for a relative, so most of the time input is focused on the initial stage required to calculate the threshold values.

With it being established that the combination of single features is the best method, then a more robust and less time consuming method of selecting and ‘cutting out’ each of the areas will need to be produced. At present there is room for error as they are manually cut out using the visual assessment of landmarks and features, this is both difficult and time consuming. The use of Gaussian curvature measurements to select specific landmarks and a template would enable the entire process to become automated and more repeatable.

Chapter 8: Conclusion

1. The results of this study show that it is not possible to determine a genetic relationship, parent-offspring or sibling-sibling, by analysis of facial morphology using the methodology outlined in this study.
2. It is possible to exclude the possibility of a genetic relationship by analysing facial morphology. In this study the best case scenario saw the reduction of the pool of possible relatives by 66.8%. (Using the combination of Eyes and Nose and Mouth for same sex male after removal of the poorest scan see Table 6.15)
3. The results of this study have increased the body of knowledge in the field of heritable facial morphology, and in particular:
    - The heritability shown between non-twin siblings
    - The heritability shown across entire family relationships, mother-father-daughter-son.
4. This study has shown that overall, sisters have the closest heritable links and are shown to be more similar to each other than other relationship types.
5. The results of this study show that BMI in itself is not an impediment when comparing heritability but that family members with widely varying BMI values are less likely to demonstrate a possible genetic link.
6. Differences in age, especially between siblings, were found to have an effect on heritable similarities with those closest in age showing greater similarities.
7. A larger study using the same methods would add validity to the results from this study. Ideally this would also include a longitudinal element so that scans could be taken at regular intervals, especially during adolescence. This would show how heritable traits increase or decrease over time.
8. In common with a few other studies this study has demonstrated the superiority of Trimmed Iterative Closest Point (TrICP) algorithms over Principal Component Analysis (PCA) as an analytical tool for facial morphology.
9. Until there are technological advances in both equipment and software, this type of three-dimensional facial imaging will not be practicable for use in identity checking or surveillance situations.

10. The methodology used in this study to measure and compare facial shapes can be usefully applied to other areas of investigation, for example the pair matching of commingled remains.
11. In combination with growth studies this methodology could also be used to create more realistic, age progressed, images of missing persons.

Bibliography

Abate, A.F., Ricciardi, S., & Sabatino, G. (2007). 3D face recognition in an ambient intelligence environment scenario. In K. Delac & M. Grgic (Eds.), *Face Recognition* (p 1-14). Vienna, Austria: I-Tech Education and Publishing.

Abrahams, P.H., Hutchings, R.T., & Marks, S.C. (1998). McMinn's colour atlas of human anatomy (4th ed.). Missouri, USA: CV Mosby.

Albert, A.M., Ricanek, K., & Patterson, E. (2007). A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, 172(1), 1-9.

AlKhudhairi, T.D., & AlKofide, E.A. (2010). Cephalometric craniofacial features in Saudi parents and their offspring. *The Angle Orthodontist*, 80(6), 1010-1017.

Al-Osaimi, F., Bennamoun, M., & Mian, A. (2009). An expression deformation approach to non-rigid 3D face recognition. *International Journal of Computer Vision*, 81, 302-316.

Amini, F., & Borzabadi-Farahani, A. (2009). Heritability of dental and skeletal cephalometric variables in monozygous and dizygous Iranian twins. *Orthodontic Waves*, 68(2), 72-79.

Amor, B.B., Ardabilian, M., & Chen, L. (2006). New experiments on ICP-based 3D face recognition and authentication. *IEEE Proceedings of 18th International Conference on Pattern Recognition,* 3, 1195-1199.

Baker, B.J., Dupras, T.L., & Tocheri, M.W. (2005). *The osteology of infants and children*. Texas: A&M University Press.

Bass, W.M. (1971). *Human osteology: a laboratory and field manual of the human skeleton* (5th ed.). Missouri: Missouri Archaeological Society.

Baydaş, B., Erdem, A., Yavuz, I., & Ceylan, I. (2007). Heritability of facial proportions and soft-tissue profile characteristics in Turkish Anatolian siblings. *American Journal of Orthodontics and Dentofacial Orthopedics*, 131(4), 504-509.

Belhumeur, P.N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 711-720.

Bennell, C., & Canter, D.V. (2002). Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. *Science & Justice*, 42(3), 153-164.

Berretti, S., Amor, B.B., Drira, H., Daoudi, S., Srivastava, A., del Bimbo, A., & Pala, P. (2013). Applications. In Daoudi, M., Srivastava, A., & Veltkamp, R. (Eds.), *3D Face Modelling, Analysis and Recognition* (p 149-202). London, UK: Wiley UK.

Besl, P.J., & McKay, N.D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 14(2), 239-256.

Bishara, S.E., Peterson L.C., & Bishara, E.C. (1984). Changes in facial dimensions and relationships between the ages of 5 and 25 years. *American Journal of Orthodontics,* 85(3), 238-252.

Blanz, V., & Vetter, T. (2003). Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 25(9), 1063-1074.

Boehnen, C., & Flynn, P. (2005). Accuracy of 3D scanning technologies in a face scanning scenario. *Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modelling*, 310-317.

Boehringer, S., van der Lijn, F., Liu, F., Günther, M., Sinigerova, S., Nowak. S., Ludwig, K.U., Herberz, R., Klein, S., Hofman, A., Uitterlinden, A.G., Niessen, W.J., Breteler, M.M.B., van der Lugt, A., Würtz, R.P., Nöthen, M.M., Horsthemke, B., Wieczorek, D., Mangold, E., & Kayser, M. (2011). Genetic determination of human facial morphology: Links between cleft-lips and normal variation. *European Journal of Human Genetics,* 19(11), 1192–1197.

Bowyer, K.W., Chang, C., & Flynn, P. (2006). A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Computer Vision and Image Understanding,* 101, 1-15.

Brewster, D. (1856). *The stereoscope; its history, theory, and construction, with its application to the fine and useful arts and to education,* London UK: John Murray.

Brigg, C.A., & Martakis, M. (1998). Craniofacial anatomy. In Clement, J.G. & Ranson, D.L. (Eds.), *Craniofacial Identification in Forensic Medicine* (p 37-48). London, UK: Arnold.

British Broadcasting Corporation News (2003). BBC NEWS | UK | England | Merseyside | Tourist mistaken for missing doctor. Retrieved 22 October 2014, from http://news.bbc.co.uk/1/hi/england/merseyside/3156819.stm

Bronstein, A. M., Bronstein, M. M., & Kimmel, R. (2005). Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1), 5-30.

Bruce, V. (2012). Familiar face recognition. In Wilkinson, C. & Ryan C. (eds.) *Craniofacial Identification* (p 1-10). Cambridge, UK: Cambridge University Press.

Buck, D.L., & Brown, C.M. (1987). A longitudinal study of nose growth from ages 6 to 18. *Annals of Plastic Surgery*, 18(4), 310-313.

Burstone, C.J. (1967). Lip posture and its significance in treatment planning. *American Journal of Orthodontics,* 53(4), 262-284.

Burton, A.M., & Jenkins, R. (2011). Unfamiliar face perception. In Calder, A., Rhodes, G., Johnson, M., & Haxby, J. (Eds.) *Oxford handbook of face perception* (p 287-306). Oxford, UK: Oxford University Press.

Burton, A.M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. *Psychological Science,* 10(3), 243-248.

Buschang, P.H., De La Cruz, R., Viazis, A.D., & Demirjian, A. (1993). Longitudinal shape changes of the nasal dorsum. *American Journal of Orthodontics and Dentofacial Orthopedics,* 104(6), 539-543.

Caisey, L., Gubanova, E., Camus, C., Lapatina, N., Smetnik, V., & Lévêque J.L. (2008). Influence of age and hormone replacement therapy on the functional properties of the lips. *Skin Research and Technology,* 14, 220-225.

Carels, C., Van Cauwenberghe, N., Savoye, I., Willems, G., Loos, R., Derom, C., & Vlietinck, R. (2001). A quantitative genetic study of cephalometric variables in twins. *Clinical orthodontics and research*, 4(3), 130-140.

Cartoux, J.Y., La Presse, J.T., & Richetin, M., (1989). Face authentication or recognition by profile extraction from range images. *Workshop on Interpretation of 3D Scenes*, 194–199

Cattaneo, C., Ritz-Timme, S., Gabriel, P., Gibelli, D., Giudici, E., Poppa, P., Nohrden, D., Assmann, S., Schmitt, R., & Grandi M. (2009). The difficult issue of age assessment on pedo-pornographic material. *Forensic Science International,* 183, e21-e24.

Chaconas, S. J. (1969). A statistical evaluation of nasal growth. *American Journal of Orthodontics*, 56(4), 403-414.

Chang, K.I., Bowyer, K.W., & Flynn, P.J. (2006). Multiple nose region matching for 3D face recognition under varying facial expression. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 28(10), 1695-1700.

Chen, Y., & Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing,* 10(3), 145-155.

Chetverikov, D., Svirko, D., Stepanov, D., & Krsek, P. (2002). The trimmed iterative closest point algorithm. *2002 IEEE Proceedings: 16th International Conference on Pattern Recognition,* 3, 545-548.

Cohen, B.J. (2005). *Memmler's The human body in health and disease* (10th ed.). Baltimore, USA: Lippincott, Williams & Wilkins.

Cook, J., Chandran, V., Sridharan, S., & Fookes, C. (2004). Face recognition from 3D Data using Iterative Closet Point Algorithm and Gaussian Mixture Models. *Proceedings 3D Data Processing Visualisation and Transmission*, Thessaloniki, Greece.

Cootes, T.F., Taylor, C.J., Cooper, D.H., & Graham, J. (1995). Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1), 38-59.

Crown Prosecution Service (2014) Crown Prosecution Service: Guidance on expert evidence (1st ed.). Retrieved from: http://www.cps.gov.uk

Curio, C., Bulthoff, H.H., & Giese, M.A. (Eds.). (2011). *Dynamic faces, insights from experiments and computation*. Boston, USA: Massachusetts Institute of Technology Press.

Curio, C., Giese, M.A., Breidt, M., Kleiner, M., & Bülthoff, H.H. (2011). Recognition of dynamic facial action probed by visual adaptation. In Curio, C., Bülthoff, H.H., & Giese, M.A., (Eds.). *Dynamic faces, insights from experiments and computation* (p 47-65). Boston, USA: Massachusetts Institute of Technology Press.

Damas, S., Wilkinson, C., Kahana, T., Veselovskaya, E., Abramov, A., Jankauskas, R., Jayaprakash, P.T., Ruiz, E., Navarro, F., Huete, M.I., Cunha, E., Cavalli, F., Clement, J., Lestón, P., Molinero, F., Briers, T., Viegas, F., Imaizumi, K., Humpire, D., & Ibáñez, O. (2015). Study on the performance of different craniofacial superimposition approaches (II): Best practices proposal. *Forensic Science International*. doi:10.1016/j.forsciint.2015.07.045

Davis, J.P, Valentine, T., & Wilkinson, C. (2012) Facial image comparison. In Wilkinson, C. & Ryan, C. (Eds.) *Craniofacial identification* (p 136-153). Cambridge, UK: Cambridge University Press.

Davis, J.P., & Valentine T. (2015). Human verification of identity from photographic images. In Valentine, T. & Davis, J.P. (Eds.). *Forensic facial identification, theory and practise of identification from eyewitnesses, composites and CCTV* (p 211-238). London, UK: Wiley Blackwell

Davis, J.P., Valentine, T., & Davis, R.E. (2010). Computer assisted photo-anthropometric analysis of full face and profile images. *Forensic Science International*, 200, 165-176.

De Jesus Mari, J., & Williams, P. (1985). A comparison of the validity of two psychiatric screening questionnaires (GHQ-12 and SRQ-20) in Brazil, using Relative Operating Characteristic (ROC) analysis. *Psychological Medicine*, 15(3), 651-659.

De Menezes, M., Rosati, R., Baga, I., Mapelli, A., & Sforza, C. (2011). Three-dimensional analysis of labial morphology: Effect of sex and age. *International Journal of Oral and Maxillofacial Surgery*, 40, 856-861.

Demayo, C., Torres, M., Sinco, A., & Bonachita-Sanguila, M. (2009). Geometric morphometric analyses of facial shape in twins. *The Internet Journal of Biological Anthropology,* 4(1).

Djordjevic, J., Jadallah, M., Zhurov, A. I., Toma, A. M., & Richmond, S. (2013). Three-dimensional analysis of facial shape and symmetry in twins using laser surface scanning. *Orthodontics and Craniofacial Research*, 16(3), 146-160.

Dodd, V. (2008). De Menezes police 'had no good photo' of real suspect. The Guardian. Retrieved 23 September 2014, from http://www.theguardian.com/uk/2008/sep/26/menezes.police

Duvdevani-Bar, S., Edelman, S., Howell, J., & Buxton, H. (1998). A similarity-based method for the generalization of face recognition over pose and expression. *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, 118-123.

Edmond, G., Davis, J.P., & Valentine T. (2015). Expert analysis: Facial image comparison In Valentine, T. & Davis, J.P. (Eds.). *Forensic facial identification, theory and practise of identification from eyewitnesses, composites and CCTV* (p 239-262). London, UK: Wiley Blackwell.

Eguchi, S., Townsend, G.C., Richards, L.C., Hughes, T., & Kasai, K. (2004). Genetic contribution to dental arch size variation in Australian twins. *Archives of Oral Biology,* 49(12), 1015-1024.

Ercan, I., Etoz, A., Guney, I., Ocakoglu, G., Turan-Ozdemir, S., Kan, I., & Kahveci, R. (2007). Statistical shape analysis of nose in Turkish young adults. *Journal of Craniofacial Surgery,* 18(1), 219-224.

Facial Identification Scientific Working Group (2012). *Guidelines for facial comparison methods version 1.0 2012.02.02.*

Facial Identification Scientific Working Group (2014). *Facial image comparison feature list for morphological analysis version 1.0 2014.08.15.*

Farkas, L.G., Katic, M.J., Forrest, C.R., & Litsas, L. (2001). Surface anatomy of the face in Down's syndrome: Linear and angular measurements in the craniofacial regions. *Journal of Craniofacial Surgery,* 12(4), 373-379.

Fawcett, T. (2006). Introduction to ROC analysis. *Pattern Recognition Letters,* 27, 861–874.

Fazekas, I.G., & Kósa, F. (1978). *Forensic fetal osteology*. Budapest, Hungary: Akadémiai Kiadó.

Federal Bureau of Investigation. (2015). *Facial recognition*. Retrieved from https://www.fbi.gov/about-us/cjis/fingerprints_biometrics/biometric-center-of-excellence/modalities/facial-recognition

Ferrario, V.F., Sforza, C., Poggio, C.E., & Schmitz, J.H. (1997). Three-dimensional study of growth and development of the nose. *The Cleft Palate-Craniofacial Journal,* 34(4), 309-317.

Ferrario, V.F., Sforza, C., Schmitz, J.H., Ciusa, V., & Colombo, A. (2000). Normal growth and development of the lips: A three-dimensional study from 6 years to adulthood using a geometric model. *Journal of Anatomy,* 196(3), 415-423.

Ferrario, V.F., Mian, F., Peretta, R., Rosati, R., & Sforza, C., (2007). Three-dimensional computerized anthropometry of the nose: landmark representation compared to surface analysis. *The Cleft palate-craniofacial journal*, *44*(3), 278-285.

Fledelius, H.C., & Stubgaard, M. (1986). Changes in eye position during growth and adult life. *Acta Ophthalmologica*, 64(5), 481-486.

Formby, W.A., Nanda, R.S., & Currier, G.F. (1994). Longitudinal changes in the adult facial profile. *American Journal of Orthodontics and Dentofacial Orthopedics*, 105(5), 464-476.

Franklin, D., & Cardini, A. (2007). Mandibular morphology as an indicator of human subadult age: Interlandmark approaches. *Journal of Forensic Sciences,* 52(5), 1015-1019.

Franklin, D., Cardini, A., O’Higgins, P., Oxnard, C.E., & Dadour, I. (2008). Mandibular morphology as an indicator of human subadult age: geometric morphometric approaches. *Forensic Science, Medicine, and Pathology,* 4(2), 91-99.

Garner, L.D. (1974). Soft-tissue changes concurrent with orthodontic tooth movement. *American Journal of Orthodontics*, 66(4), 367-377.

Gelgör, I.E., Karaman, A.I., & Zekiç, E. (2006). The use of parental data to evaluate soft tissues in an Anatolian Turkish population according to Holdaway soft tissue norms. *American Journal of Orthodontics and Dentofacial Orthopedics*, 129(3), 330.e1-330.e9.

Gilbert, J.G., & Segal, S. (1958). Growth of the nose and the septorhinoplastic problem in youth. *American Medical Association Archives of Otolaryngology*, 68(6), 673-682.

Gökberk, B., İrfanoğlu, M.O., & Akarun, L. (2006). 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, 24(8), 857-869.

Goldstein, M.S. (1939). Development of the bridge of the nose. *American Journal of Physical Anthropology*, 25(1), 101-117.

Greiner, M., Pfeiffer, D., & Smith, R.D. (2000). Principles and practical application of the Receiver Operating Characteristic analysis for diagnostic tests. *Preventive Veterinary Medicine*, 45, 23-41.

Greiner, M., Sohr, D., & Göbel, P. (1995). A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *Journal of Immunological Methods,* 185(1), 123-132.

Gribskov, M., & Robinson, N.L. (1996). Use of Receiver Operating Characteristic (ROC) analysis to evaluate sequence matching. *Computers and Chemistry,* 20(1)25-33.

Guo, B., Lam, K.M., Lin, K.H., & Siu, W.C. (2003). Human face recognition based on spatially weighted Hausdorff distance. *Pattern Recognition Letters*. 24(1), 499-507.

Hallinan, P.W. (1991). Recognizing human eyes. *Geometric Methods in Computer Vision,* 1570, 214-226.

Hammond, P., Hannes, F., Suttie, M., Devriendt, K., Vermeesch, J.R., Faravelli, F., Forzano, F., Parekh, S., Williams, S., McMullan, D., South, S.T., Carey, J.C., & Quarrell, O. (2012). Fine-grained facial phenotype-genotype analysis in Wolf-Hirschhorn syndrome. *European Journal of Human Genetics*, 20, 33-40.

Hancock, P.J.B. (2012). Unfamiliar face recognition. In Wilkinson, C. & Ryan, C. (Eds.) *Craniofacial Identification* (p 11-23). Cambridge, UK: Cambridge University Press.

Heidari, Z., Mahmoudzadeh-Sagheb, H., Khammar, T., & Khammar, M. (2009). Anthropometric measurements of the external nose in 18–25-year-old Sistani and Baluch aborigine women in the southeast of Iran. *Folia morphologica*, 68(2), 88-92.

Hesher, C., Srivastava, A., & Erlebacher, G. (2003). A novel technique for face recognition using range imaging. *Seventh International Symposium on Signal Processing and Its Applications,* 2, 201-204.

Hunter, W.S., Balbach, D.R., & Lamphiear, D.E. (1970). The heritability of attained growth in the human face. *American Journal of Orthodontics*, 58(2), 128-134.

Ibáñez, O., Vicente, R., Navega, D.S., Wilkinson, C., Jayaprakash, P.T., Huete, M.I., Briers, T., Hardiman, R., Navarro, F., Ruiz, E., Cavalli, F., Imaizumi, K., Jankauskas, R., Veselovskaya, E., Abramov, A., Lestón, P., Molinero, F., Cardoso, J., Çağdır, A.S., Humpire, D., Nakanishi, Y., Zeuner, A., Ross, A.H., Gaudio, D., & Damas, S. Study on the performance of different craniofacial superimposition approaches (I) *Forensic science International*. doi:10.1016/j.forsciint.2015.05.030

Iblher, N., Kloepper, L., Penna, V., Bartholomae, J.P., & Stark G.B. (2008). Changes in the aging upper lip - a photomorphometric and MRI-based study (on a quest to find the right rejuvenation approach). *Journal of Plastic, Reconstructive & Aesthetic Surgery,* 61, 1170-1176.

Jayaprakash, P.T., & Pritam, H.H. (2014). Workshop on Video Superimposition at MEPROCS. *International Conference on Craniofacial Superimposition*. Dundee, June 30th-July 4th.

Jelenkovic, A., Poveda, A., Susanne, C., & Rebato, E. (2010). Common genetic and environmental factors among craniofacial traits in Belgian nuclear families: Comparing skeletal and soft-tissue related phenotypes. *HOMO-Journal of Comparative Human Biology*, 61(3), 191-203.

Jing, X.Y., Wong, H.S., & Zhang, D. (2006). Face recognition based on 2D Fisherface approach. *Pattern Recognition*, 39(4), 707-710.

Johannsdottir, B., Thorarinsson, F., Thordarson, A., & Magnusson, T.E. (2005). Heritability of craniofacial characteristics between parents and offspring estimated from lateral cephalograms. *American Journal of Orthodontics and Dentofacial Orthopedics*, 127(2), 200-207.

Kaneko, S., Kondo, T., & Miyamoto, A. (2003) Robust matching of 3D contours using iterative closest point algorithm improved by M-estimation. *Pattern Recognition,* 36(9), 2041-2047.

Kapoutsis, C.A., Vavoulidis, C.P., & Pitas, I. (1999). Morphological iterative closest point algorithm. *IEEE Transactions on Image Processing*, 8(11), 1644-1646.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud. *Applied Cognitive Psychology,* 11(3), 211-222.

King, L., Harris, E.F., & Tolley, E.A. (1993). Heritability of cephalometric and occlusal variables as assessed from siblings with overt malocclusions. *American Journal of Orthodontics and Dentofacial Orthopedics*, 104(2), 121-131.

Kitahara, T., Ichinose, M., & Nakasima, A. (1996). Quantitative evaluation of correlation of skull morphology in families in an attempt to predict growth change. *The European Journal of Orthodontics,* 18(2), 181-191.

Koudelová, J., Dupej, J., Brůžek, J., Sedlak, P., & Velemínská, J. (2015). Modelling of facial growth in Czech children based on longitudinal data: Age progression from 12 to 15 years using 3D surface models. *Forensic Science International*, 248, 33-40.

Ksiezycki-Ostoya, B. K., McCollum, A.G., & Becker, P.J. (2009). Sagittal soft-tissue changes of the lower lip and chin associated with surgical maxillary impaction and consequent mandibular autorotation. *Seminars in Orthodontics* 15(3), 185-195.

Kunjur, J., Sabesan, T., & Ilankovan, V. (2006). Anthropometric analysis of eyebrows and eyelids: an inter-racial study. *British Journal of Oral and Maxillofacial Surgery*, 44(2), 89-93.

Lauweryns, I., Carels, C., & Vlietinck, R. (1993). The use of twins in dentofacial genetic research. *American Journal of Orthodontics and Dentofacial Orthopedics*, 103(1), 33-38.

Lévêque, J.L., & Goubanova, E. (2004). Influence of age on the lips and perioral skin. *Dermatology*, 208, 307–313.

Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., Wollstein, A., Lao, O., de Bruijne, M., Arfan Ikram, M., van der Lugt, A., Rivadeneira, F., Uitterlinden, A.G., Hofman, A., Niessen, W.J., Homuth, G., de Zubicaray, G., McMahon, K.L., Thompson, P.M., Daboul, A., Puls, R., Hegenscheid, K., Bevan, L., Pausova, Z., Medland, S.E., Montgomery, G.W., Wright, M.J., Wicking, C., Boehringer, S., Spector, T.D., Paus, T., Martin, N.G., Biffar, R., & Kayser, M. (2012). A genome-wide association study identifies five loci influencing facial morphology in Europeans. PLoS Genetics, 8(9), e1002932.

Lu, X., Colbry, D., & Jain, A.K. (2004). Three dimensional model based face recognition. *Proceedings of the International Conference of Pattern Recognition*, 362-366

Lu, X., Jain, A.K., & Colbry, D. (2006). Matching 2.5D face scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1), 31-43.

Lundström, A., & McWilliam, J.S. (1987). A comparison of vertical and horizontal cephalometric variables with regard to heritability. *The European Journal of Orthodontics*, 9(2), 104-108.

Lundström, A., & McWilliam, J.S. (1988). Comparison of some cephalometric distances and corresponding facial proportions with regard to heritability. *The European Journal of Orthodontics*, 10(1), 27-29.

Mamandras, A.H. (1984). Growth of lips in two dimensions: a serial cephalometric study. *American Journal of Orthodontics*, 86(1), 61-66.

Mamandras, A.H. (1988). Linear changes of the maxillary and mandibular lips. *American Journal of Orthodontics and Dentofacial Orthopedics,* 94(5), 405-410.

Manfredi, C., Martina, R., Grossi, G.B., & Giuliani, M. (1997). Heritability of 39 orthodontic cephalometric parameters on MZ, DZ twins and MN-paired singletons. *American Journal of Orthodontics and Dentofacial Orthopedics,* 111(1), 44-51.

Maurer, T., Guigonis, D., Maslov, I., Pesenti, B., Tsaregorodtsev, A., West, D., & Medioni, G. (2005). Performance of Geometrix Activated 3D Face Recognition Engine on the FRGC Data. *Proceedings of IEEE Conference Computer Vision and Pattern Recognition,* 154.

McCool, C., Sanchez-Riera J., & Marcel S. (2010). Feature distribution modelling techniques for 3D face verification. *Pattern Recognition Letters*, 31(11), 1324 -1330.

Medioni, G., & Waupotitsch, R. (2003). Face modeling and recognition in 3D. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures,* 232-233.

Meng, H.P., Goorhuis, J., Kapila, S., & Nanda, R.S. (1988). Growth changes in the nasal profile from 7 to 18 years of age. *American Journal of Orthodontics and Dentofacial Orthopedics*, 94(4), 317-326.

Meredith, H.V. (1958). A time series analysis of growth in nose height during childhood. *Child Development*, 19-34.

Metz, C.E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine,* 8(4), 283-298.

Mian, A., Bennamoun, M., & Owens, R. (2007). An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11), 1927-1943.

Milborrow, S., & Nicolls, F. (2008). Locating facial features with an extended active shape model. *Computer Vision– Proceeding of 10th European Conference on Computer Vision,* 4, 504-513.

Mohammadzade, H., & Hatzinakos, D. (2013) Iterative Closest Normal Point for 3D face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 35(2), 381-397.

Naini, F.B., & Moss, J.P. (2004). Three-dimensional assessment of the relative contribution of genetics and environment to various facial parameters with the twin method. *American Journal of Orthodontics and Dentofacial Orthopedics*, 126(6), 655-665.

Nakata, M., Yu, P.I., Davis, B., & Nance, W.E. (1973). The use of genetic data in the prediction of craniofacial dimensions. *American Journal of Orthodontics,* 63(5), 471-480.

Nanda, R.S., Meng, H., Kapila, S., & Goorhuis, J. (1990). Growth changes in the soft tissue facial profile. *The Angle Orthodontist,* 60(3), 177-190.

National Health Service (2015) Pregnant with twins. Retrieved 11 August 2015 from http://www.nhs.uk/conditions/pregnancy-and-baby/pages/what-causes-twins.aspx

O'Toole, A.J., & Phillips, P.J. (2015). Evaluating automatic face recognition systems with human benchmarks. In Valentine, T. & Davis, J.P. (Eds.). *Forensic facial identification, theory and practise of identification from eyewitnesses, composites and CCTV* (p 263-283). London, UK: Wiley Blackwell.

Oliver, B.M. (1982). The influence of lip thickness and strain on upper lip response to incisor retraction. *American Journal of Orthodontics,* 82(2), 141-149.

Paone, J.R., Flynn, P.J., Philips, P.J., Bowyer, K.W., Bruegge, R.W.V., Grother, P.J., Quinn, G.W., Pruitt, M.T., & Grant, J.M. (2014). Double trouble: Differentiating identical twins by face recognition. *IEEE Transactions on Information Forensics and Security,* 9(2), 285-295.

Papatheodorou, T., & Rueckert, D. (2004). Evaluation of automatic 4D face recognition using surface and texture registration. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition,* 321-326.

Papatheodorou, T., & Rueckert, D. (2007) 3D Face Recognition In: Delac, K. & Grgic, M. (Eds.). *Face recognition* (p 417-446). Vienna, Austria: I-Tech Education and Publishing.

Park, S., Kudlick, E.M., & Abrahamian, A. (1989). Vertical dimensional changes of the lips in the North American black patient after four first-premolar extractions. *American Journal of Orthodontics and Dentofacial Orthopedics,* 96(2), 152-160.

Pelc, K., Kornreich, C., Foisy, M.L., & Dan, B. (2006). Recognition of emotional facial expressions in Attention-Deficit Hyperactivity Disorder. *Pediatric Neurology,* 35(2), 93-97.

Penna, V., Stark, G.B., Eisenhardt, S.U., Bannasch, H., & Iblher, N. (2009). The aging lip: a comparative histological analysis of age-related changes in the upper lip complex. *Plastic Reconstructive Surgery,* 124, 624–628.

Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., & Worek, W. (2005). Overview of the face recognition grand challenge. *CVPR 2005. IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 1, 947-954.

Posen, J.M. (1967). A longitudinal study of the growth of the nose. *American Journal of Orthodontics*, 53(10), 746-756.

Prahl-Andersen, B., Ligthelm-Bakker, A.S.W.M.R., Wattel, E., & Nanda, R. (1995). Adolescent growth changes in soft tissue profile. *American Journal of Orthodontics and Dentofacial Orthopedics*, 107(5), 476-483.

Rankinen, T., Kim, S.Y., Perusse, L., Despres, J.P., & Bouchard, C. (1999). The prediction of abdominal visceral fat level from body composition and anthropometry: ROC analysis. *International Journal of Obesity*, 23(8), 801-809.

Rice, A., Phillips, P.J., Natu, Z., An, X., & O'Toole, A.J. (2013). Unaware person recognition from the body when face identification fails. *Psychological Science*, 24(11), 2235-2243.

Rousseeuw, P.J., & Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85(411), 633-639.

Rynn, C., & Wilkinson, C.M. (2006). Appraisal of traditional and recently proposed relationships between the hard and soft dimensions of the nose in profile. *American Journal of Physical Anthropology*, 130(3), 364-373.

Saunders, S.R., Popovich, F., & Thompson, G.W. (1980). A family study of craniofacial dimensions in the Burlington Growth Centre sample. *American Journal of Orthodontics*, 78(4), 394-403.

Savoye, I., Loos, R., Carels, C., Derom, C., & Vlietinck, R. (1998). A genetic study of anteroposterior and vertical facial proportions using model-fitting. *The Angle Orthodontist*, 68(5), 467-470.

Scheuer, L., & Black, S. (2000). *Developmental juvenile osteology*. Elsevier Academic Press. London, UK.

Scheuer, L., & Black, S. (2004). *The juvenile skeleton*. Elsevier Academic Press. London, UK.

Sforza, C., Elamin, F., Tommasi, D.G., Dolci, C., & Ferrario, V.F. (2013). Morphometry of the soft tissues of the orbital region in Northern Sudanese persons. *Forensic Science International*, 228(1), 180.e1-180.e11.

Sforza, C., Grandi, G., Binelli, M., Dolci, C., De Menezes, M., & Ferrario, V.F. (2010b). Age and sex related changes in three-dimensional lip morphology. *Forensic Science International*, 182.e1-182.e7.

Sforza, C., Grandi, G., Catti, F., Tommasi, D.G., Ugolini, A., & Ferrario, V.F. (2009). Age and sex related changes in the soft tissues of the orbital region. *Forensic Science International*, 185(1), 115.e1-115.e8.

Sforza, C., Grandi, G., De Menezes, M., Tartaglia, G.M., & Ferrario, V.F. (2010a). Age and sex related changes in the normal human external nose. *Forensic Science International*, 204(1), 205.e1-205.e9.

Simonian, S.J., Beidel, D.C., Turner, S.M., Berkes, J.L., & Long J.H. (2001). Recognition of facial affect by children and adolescents diagnosed with social phobia. *Child Psychiatry and Human Development,* 32(2), 137-145.

Singh, R.N. (1990). Changes in the soft tissue chin after orthodontic treatment. *American Journal of Orthodontics and Dentofacial Orthopedics,* 98(1), 41-46.

Smeets, D., Fabry, T., Hermans, J., Vandermeulen, D., & Suetens, P. (2010). Fusion of an isometric deformation modeling approach using spectral decomposition and a region-based approach using ICP for expression-invariant 3D face recognition. *IEEE 20th International Conference on Pattern Recognition,* 1172-1175.

Sparkes, M. (2014). Police trial facial recognition software that can ID suspects 'in seconds'. The Telegraph. 17 Jul 2014. Retrieved from: http://www.telegraph.co.uk/technology/news/10973185/Police-trial-facial-recognition-software-that-can-ID-suspects-in-seconds.html

Stephan, C.N., & Henneberg, M. (2003). Predicting mouth width from inter-canine width - a 75% rule. *Journal of Forensic Science*, 48, 725-727.

Stephan, C.N. & Murphy, S.J. (2008). Mouth width prediction in craniofacial identification: Cadaver tests of four recent methods, including two techniques for edentulous skulls. *Journal of Forensic Odonto-stomatology,* 26(1), 2-7.

Stephan, C.N., Henneberg, M., & Sampson, W. (2003). Predicting nose projection and pronasale position in facial approximation: a test of published methods and proposal of new guidelines. *American Journal of Physical Anthropology,* 122(3), 240-250.

Subtelny, J.D. (1961). The soft tissue profile, growth and treatment changes. *The Angle Orthodontist,* 31(2), 105-122.

Sun, Z., Paulino, A.A., Feng, J., Chai, Z., Tan, T., & Jain, A.K. (2010). A study of multibiometric traits of identical twins. *Proceedings of Biometric Technology for Human Identification* 7667:OT1-OT12.

Swets, J.A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative radiology,* 14(2), 109-121.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. *Science,* 240, 1285-1293.

Swets, J.A. (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*. London, UK: Psychology Press.

Tan, H., & Zhang, Y.J. (2006). A novel weighted Hausdorff distance for face localization. *Image and Vision Computing*, 24(7), 656-662.

Tanaka, J.W., Kay, J.B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition*, 5(4), 479-496.

Thayer, Z.M., & Dobson, S.D. (2010). Sexual dimorphism in chin shape: implications for adaptive hypotheses. *American Journal of Physical Anthropology*, 143(3), 417-425.

Tilotta, F.M., Glaunès, J.A., Richard, F.J., & Rozenholc, Y. (2010). A local technique based on vectorized surfaces for craniofacial reconstruction. *Forensic Science International*, 200(1), 50-59.

Tsalakanidou, F., Tzovaras, D., & Strintzis, M.G. (2003). Use of depth and colour eigenfaces for face recognition. *Pattern Recognition Letters,* 24(9), 427-1435.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 71-86.

Uzun, A., Akbas, H., Bilgic, S., Emirzeoglu, M., Bostancı, O., Sahin, B., & Bek, Y. (2006). The average values of the nasal anthropometric measurements in 108 young Turkish males. *Auris Nasus Larynx*, 33(1), 31-35.

van der Heijden, P., Korsten-Meijer, A.G., van der Laan, B.F., Wit, H.P., & Goorhuis-Brouwer, S.M. (2008). Nasal growth and maturation age in adolescents: a systematic review. *Archives of Otolaryngology – Head & Neck Surgery*, 134(12), 1288-1293.

Van Erkel, A.R., & Pattynama, P.M.T. (1998). Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. *European Journal of Radiology*, 27(2), 88-94.

Vig, P.S., & Cohen, A.M. (1979). Vertical growth of the lips: a serial cephalometric study. *American Journal of Orthodontics,* 75(4), 405-415.

Vijayan, V., Bowyer, K.W., Flynn, P.J., Huang, D., Chen, L., Hansen, M., Ocegueda O., Shah, S.K., & Kakadiaris, I.A. (2011). Twins 3D face recognition challenge. *Proceedings IEEE International Joint Conference on Biometrics,* 1–7.

Vivek, E.P., & Sudha, N. (2007). Robust Hausdorff distance measure for face recognition. *Pattern Recognition*, 40(2), 431-442.

Walder, C., Breidt, M., Bülthoff, H.H., Schölkopf, B., & Curio, C. (2011). Markerless tracking of dynamic 3D scans of faces. In Curio, C, Bülthoff, H.H. & Giese, M.A. (Eds.). *Dynamic faces, insights from experiments and computation* (p 255-276). Boston, USA: Massachusetts Institute of Technology Press.

Weinberg, S.M., Parsons, T.E., Marazita, M.L., & Maher, B.S. (2013). Heritability of face shape in twins: a preliminary study using 3D stereophotogrammetry and geometric morphometrics. *Dentistry 3000*, 1(1).

White, T.D., & Folkens, P.A. (2000) Human Osteology (2nd ed.) London, UK: Academic Press.

White, T.D., & Folkens, P.A. (2005). *The Human Bone Manual*. London, UK: Elsevier Academic Press.

Wilkinson, C., & Evans, R. (2009). Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Science and Justice*, 49(3), 191-196.

Wilkinson, C. (2004). *Forensic facial reconstruction*. Cambridge, UK: Cambridge University Press.

Yang, M. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using kernel methods. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 215-220.

Yu, C.C., Wan, S.Y., & Chang, C.Y. (2014). Commentary on "the labial aging process: A surface analysis-based three-dimensional evaluation" *Aesthetic Plastic Surgery*, 38(1), 242-243.

Yuille, A.L., Hallinan, P.W., & Cohen, D.S. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2), 99-111.

Zankl, A., Eberle, L., Molinari, L., & Schinzel, A. (2002). Growth charts for nose length, nasal protrusion, and philtrum length from birth to 97 years. *American Journal of Medical Genetics*, 111(4), 388-391.

Zhao, K., Xi, J., Yu, Y., & Chicharo, J.F. (2011). 3D face recognition based on a modified ICP method. *International Society for Optics and Photonics International Conference on Optical Instruments and Technology: Optoelectronic Imaging and Processing Technology*, 82000G1-82000G11.

Zhao, W., Chellappa, R., Phillips, J.P., & Rosenfeld, A. (2003). Face recognition: A literature survey. *ACM Computing Surveys*, 35(4), 399-458.

Zhu, G., Evans, D.M., Duffy, D.L., Montgomery, G.W., Medland, S.E., Gillespie, N.A., Ewen, K.R., Jewell, M., Liew, Y.W., Hayward, N.K., Sturm, R.A., Trent, J.M., & Martin, N.G. (2004). A genome scan for eye color in 502 twin families: most variation is due to a QTL on chromosome 15q. *Twin Research*, 7(02), 197-210.

Zhu, L.Y., Meng, T., Shi, B., & Deng, D.Z. (2008). Anthropometric study of the upper lip of 1500 healthy children in Chengdu, Western China. *British Journal of Oral and Maxillofacial Surgery*, 46, 554-560.

Appendix 1

Questionnaire

Thank you for agreeing to be scanned for my PhD project. Please can you answer the following questions to enable me to study a variety of variables. All details will only be used for this study and will not be passed onto anyone else. Personal details will only be read by me and no personal identification will be made in the publication of this study.

Name: ....

Date of Birth: ....

Date of scan: ....

Email .... (This is so I can inform you of the results of the study and provide you with the 3D scans of you and your family.)

Height .... (cm)

Weight .... (kg)

Ethnic origin ....(eg White British)

Parents/Grandparents Ethnic origin:

Mother		Father	
Grandmother	Grandfather	Grandmother	Grandfather

Have you suffered any injuries to your head, face or neck? (eg Broken nose) ....

If yes, what and when?

Have you had any surgical procedures performed to your head, face of neck? ....

If yes, what and when?

Do you have (or have had) any know diseases or illness? (eg arthritis).......................................

If yes, what, how long for and are you taking any medication or having treatment for it?

Do you have your wisdom teeth? Upper Left ☐ Upper Right ☐

Lower Left ☐ Lower Right ☐

Have you lost or had any teeth removed? ....

If yes, which one(s) and when?

Do you have (or have had) any food allergies or intolerances that would limit your diet in anyway or do you personally limit your diet in some way? (eg vegetarian) ....

If yes, what are they and how long have you been affected by them?

Declarations

Child's declaration:

I declare that the information above is correct to the best of my knowledge.

Mother's declaration:

I declare that the information above is correct to the best of my knowledge.

I certify that I am the biological mother of the child/children scanned.

I also testify that the father scanned is the true biological father of this/these child/children.

Father's declaration:

I declare that the information above is correct to the best of my knowledge.

I testify that to the best of my knowledge I am the biological father of the child/children scanned.

(Scan number(s) 00)

Signed ....   Date ....

Appendix 2:

10 Poorest Quality Scans

These scans were visually assessed to be of the poorest quality for each portion of the face.

Eyes	Nose	Mouth	Eyes - Nose	Eyes-Nose-Mouth	Eyes-Nose-Mouth-Chin	Whole face
F1F	F1C1	F1C2	F1F	F1F	F1F	F1F
F4C1	F5F	F1F	F5F	F7F	F7F	F6C3
F5M	F6C3	F2M	F6C3	F17F	F17F	F7F
F6C3	F6M	F3C1	F6M	F21C1	F21C1	F17F
F6F	F7F	F17F	F7F	F21C4	F21C4	F25C1
F7F	F17C2	F21C1	F21C4	F25C1	F25C1	F25F
F24C1	F24C1	F25F	F24C1	F25F	F25F	F26F
F26C1	F25C1	F26F	F25C1	F26F	F26F	F27F
F37C1	F25M	F27F	F28M	F27F	F27F	F31C2
F37F	F26M	F31F	F37C1	F31F	F31F	F37C1

Eyes + Nose
F1C1
F1F
F4C1
F5F
F5M
F6C3
F6F
F6M
F7F
F17C2
F24C1
F25C1
F25M
F26C1
F26M
F37C1
F37F

Nose + Mouth
F1C1
F1C2
F1F
F2M
F3C1
F5F
F6C3
F6M
F7F
F17C2
F17F
F21C1
F24C1
F25C1
F25F
F25M
F26F
F26M
F27F
F31F

Eyes + Mouth
F1C2
F1F
F2M
F3C1
F4C1
F5M
F6C3
F6F
F7F
F17F
F21C1
F24C1
F25F
F26C1
F26F
F27F
F31F
F37C1
F37F

Eyes + Nose + Mouth
F1C1
F1C2
F1F
F2M
F3C1
F4C1
F5F
F5M
F6C3
F6F
F6M
F7F
F17C2
F17F
F21C1
F24C1
F25C1
F25F
F25M
F26C1
F26F
F26M
F27F
F31F
F37C1
F37F

Appendix 3: Index for CD Appendix

The CD contains the ROC results, graphs, optimal and Rule In dot diagrams for each of the 121 variations of Relationship category and facial area. Each of the areas of the face are compiled in the same way as the Eyes illustrated below. A full index is also available on the CD.

```
+---Eyes
|   +---Full population
|   |   +---Graphs
|   |   |   +---Optimal dot diagrams
|   |   |   |       Full_Optimal_Eyes_Children Female only.emf
|   |   |   |       Full_Optimal_Eyes_Children Male only.emf
|   |   |   |       Full_Optimal_Eyes_Children Oppositesex.emf
|   |   |   |       Full_Optimal_Eyes_Children Same sex 2.emf
|   |   |   |       Full_Optimal_Eyes_Children Same sex.emf
|   |   |   |       Full_Optimal_Eyes_Children-Children.emf
|   |   |   |       Full_Optimal_Eyes_Fathers-Children.emf
|   |   |   |       Full_Optimal_Eyes_Fathers-Daughters.emf
|   |   |   |       Full_Optimal_Eyes_Fathers-Sons.emf
|   |   |   |       Full_Optimal_Eyes_Mothers-Children.emf
|   |   |   |       Full_Optimal_Eyes_Mothers-Daughters.emf
|   |   |   |       Full_Optimal_Eyes_Mothers-Sons.emf
|   |   |   |
|   |   |   +---ROC graphs
|   |   |   |       Full_ROC_Eyes_Children Female only.emf
|   |   |   |       Full_ROC_Eyes_Children Male only.emf
|   |   |   |       Full_ROC_Eyes_Children Opposite sex.emf
|   |   |   |       Full_ROC_Eyes_Children Same sex.emf
|   |   |   |       Full_ROC_Eyes_Children-Children.emf
|   |   |   |       Full_ROC_Eyes_Fathers-Children.emf
|   |   |   |       Full_ROC_Eyes_Fathers-Daughters.emf
|   |   |   |       Full_ROC_Eyes_Fathers-Sons.emf
|   |   |   |       Full_ROC_Eyes_Mothers-Children.emf
|   |   |   |       Full_ROC_Eyes_Mothers-Daughters.emf
|   |   |   |       Full_ROC_Eyes_Mothers-Sons.emf
|   |   |   |
|   |   |   \---Rule in dot diagrams
|   |   |           Full_Rule in_Eyes_Children Female only.emf
|   |   |           Full_Rule in_Eyes_Children Male only.emf
|   |   |           Full_Rule in_Eyes_Children Oppositesex.emf
|   |   |           Full_Rule in_Eyes_Children-Children.emf
|   |   |           Full_Rule in_Eyes_Fathers-Children.emf
|   |   |           Full_Rule in_Eyes_Fathers-Daughters.emf
|   |   |           Full_Rule in_Eyes_Fathers-Sons.emf
|   |   |           Full_Rule in_Eyes_Mothers-Children.emf
|   |   |           Full_Rule in_Eyes_Mothers-Daughters.emf
|   |   |           Full_Rule in_Eyes_Mothers-Sons.emf
|   |   |
|   |   \---Results
|   |           Full_Eyes_Children Female only.doc
|   |           Full_Eyes_Children Male only.doc
|   |           Full_Eyes_Children Opposite sex.doc
|   |           Full_Eyes_Children Same sex.doc
|   |           Full_Eyes_Children-Children.doc
|   |           Full_Eyes_Fathers-Children.doc
|   |           Full_Eyes_Fathers-Daughters.doc
|   |           Full_Eyes_Fathers-Sons.doc
|   |           Full_Eyes_Mothers-Children.doc
|   |           Full_Eyes_Mothers-Daughters.doc
|   |           Full_Eyes_Mothers-Sons.doc
```

```
|   |
|   \---Ten poorest removed
|       +---Graphs
|       |   +---Optimal dot diagrams
|       |   |
|       |   +---ROC graphs
|       |   |
|       |   \---Rule in dot diagrams
|       |
|       \---Results
|
+---Eyes and Mouth
|   +---Full population
|   \---Ten poorest removed
|
+---Eyes and Nose
|   +---Full population
|   \---Ten poorest removed
|
+---Eyes and Nose and Mouth
|   +---Full population
|   \---Ten poorest removed
|
+---Eyes-Nose
|   +---Full population
|   \---Ten poorest removed
|
+---Eyes-Nose-Mouth
|   +---Full population
|   \---Ten poorest removed
|
+---Eyes-Nose-Mouth-Chin
|   +---Full population
|   \---Ten poorest removed
|
+---Mouth
|   +---Full population
|   \---Ten poorest removed
|
+---Nose
|   +---Full population
|   \---Ten poorest removed
|
+---Nose and Mouth
|   +---Full population
|   \---Ten poorest removed
|
\---Whole face
    +---Full population
    \---Ten poorest removed
```